ÉCOLE DOCTORALE BIOLOGIE ET SANTÉ UNIVERSITÉ DE LILLE

Thèse d'Université

DÉTECTION AUTOMATISÉE DES HALLUCINATIONS AUDITIVES EN IRM FONCTIONNELLE ET PERSPECTIVES THÉRAPEUTIQUES DANS LA SCHIZOPHRÉNIE

Auteur

Thomas FOVET

Directeur

Pr. Renaud JARDRI

THÈSE SOUTENUE LE 15 DECEMBRE 2017
POUR L'OBTENTION DU GRADE DE DOCTEUR DE L'UNIVERSITÉ DE LILLE

DISCIPLINE : NEUROSCIENCES

JURY

Pr. Arnaud CACHIA, Université Sorbonne-Paris-Cité - *Rapporteur*

Dr. Emilie OLIE, CHU & Université de Montpellier - *Rapporteur*

Dr. Ingrid DE CHAZERON, CHU de Clermont-Ferrand - *Assesseur*

Pr. Pierre THOMAS, Université & CHU de Lille - *Assesseur*

Dr. Delphine PINS, CNRS & Université de Lille - *Assesseur*

Pr. Renaud JARDRI, Université & CHU de Lille - *Directeur de Thèse*

“There are two kinds of truth: the truth that lights the way and the truth that warms the heart. The first of these is science, and the second is art. Neither is independent of the other or more important than the other. Without art science would be as useless as a pair of high forceps in the hands of a plumber. Without science art would become a crude mess of folklore and emotional quackery. The truth of art keeps science from becoming inhuman, and the truth of science keeps art from becoming ridiculous.”

Raymond Chandler, Great Thought, February 19, 1938

TABLE DES MATIERES 2

LISTE DES FIGURES ET TABLEAUX 4

LISTE DES ABREVIATIONS 5

RESUME 6

MOTS-CLES 8

1 **INTRODUCTION 9**

1.1 HISTOIRE GENERALE DU CONCEPT 13
1.2 CONTEXTE CLINIQUE : LES HALLUCINATIONS DANS LA SCHIZOPHRENIE 14
1.3 PHYSIOPATHOLOGIE DES HALLUCINATIONS AUDITIVES DANS LA SCHIZOPHRENIE 20
1.3.1 PRINCIPAUX MODELES NEUROPSYCHOLOGIQUES 20
1.3.2 ÉTUDES D'IMAGERIE CEREBRALE 22
1.3.2.1 Imagerie structurale et hallucinations auditives 22
1.3.2.2 Imagerie fonctionnelle et hallucinations auditives 25
1.3.2.2.1 Principaux résultats des études "trait" 25
1.3.2.2.2 Principaux résultats des études "état" 26
1.3.2.3 Études de connectivité et hallucinations auditives 29
1.3.2.3.1 Connectivité fonctionnelle et connectivité effective 29
1.3.2.3.2 Connectivité anatomique (ou structurale) 32
1.3.2.4 Vers le développement de biomarqueurs spécifiques des hallucinations auditives ? 33
1.4 PRISE EN CHARGE DES HALLUCINATIONS 35
1.4.1 TRAITEMENTS MEDICAMENTEUX 35
1.4.2 TRAITEMENTS NON MEDICAMENTEUX : PSYCHOTHERAPIES 37
1.4.3 TRAITEMENTS NON MEDICAMENTEUX : LES NOUVELLES PISTES 38
1.4.3.1 Stimulation transcrânienne par courant continu 38
1.4.3.2 Stimulation magnétique transcrânienne et neuronavigation 39
1.4.3.3 Optimiser les stratégies thérapeutiques grâce à la capture hallucinatoire ? 40
1.5 OBJECTIFS DE LA THESE 41

2 DETECTION AUTOMATISEE DES HALLUCINATIONS AUDITIVES EN IRM FONCTIONNELLE 43

2.1 LA CAPTURE « *OFFLINE* » DE L'HALLUCINATION AUDITIVE 45
2.2 LA CAPTURE « *ONLINE* » DE L'HALLUCINATION AUDITIVE 69
2.2.1 GENERALITES SUR LES METHODES D'APPRENTISSAGE AUTOMATIQUE (*MACHINE-LEARNING*) 69
2.2.2 DEVELOPPEMENT D'UN CLASSIFICATEUR CAPABLE DE DETECTER LES PERIODES HALLUCINATOIRES : UTILISATION DU *LINEAR SUPPORT VECTOR MACHINE* 73
2.2.3 DEVELOPPEMENT D'UN CLASSIFICATEUR CAPABLE DE DETECTER LES PERIODES PRE-HALLUCINATOIRES ET COMPARAISON DU *LINEAR SUPPORT VECTOR MACHINE* AU *TV-ELASTIC-NET* 88

3 PERSPECTIVES THERAPEUTIQUES DANS LA SCHIZOPHRENIE 111

3.1 LES TECHNIQUES DE NEUROFEEDBACK 112
3.1.1 GENERALITES 112
3.1.2 INITIATIVES ACTUELLES : LA CREATION DU GROUPE NEXT 127
3.1.2.1 Recommandations de bonne pratique 127
3.1.2.2 Formation 128
3.1.2.3 Promotion de la recherche 128
3.1.2.4 Diffusion 128
3.1.3 POLEMIQUES ACTUELLES SUR L'EFFICACITE DU NEUROFEEDBACK 174
3.2 LE NEUROFEEDBACK GUIDE PAR IRMF : UTILISATION EN PSYCHIATRIE 179
3.3 LE NEUROFEEDBACK DANS LA PRISE EN CHARGE DES HALLUCINATIONS AUDITIVES 205
3.3.1 NEUROFEEDBACK GUIDE PAR EEG 205
3.3.2 NEUROFEEDBACK GUIDE PAR IRMF 206

4 DISCUSSION 231

4.1 QUELLE INTERFACE DE NEUROFEEDBACK ? 232
4.2 PROTOCOLE D'ENTRAINEMENT ET EVALUATION 235
4.3 UNE VERITABLE PSYCHOTHERAPIE GUIDEE PAR L'IMAGERIE ? 240
4.4 PERSPECTIVES 242

5 CONCLUSION 248

6 ANNEXES 250

BIBLIOGRAPHIE 254

LISTE DES FIGURES ET TABLEAUX

Figure 1. Illustration de l'approche dimensionnelle pour explorer les hallucinations auditives.

Tableau 1. Description des principaux types d'hallucinations.

Figure 2. Co-occurrence des modalités hallucinatoires dans une cohorte de 171 patients souffrant de schizophrénie et d'hallucinations selon au moins une modalité.

Tableau 2a. Principales échelles d'évaluation des hallucinations auditives

Tableau 2b. Principales échelles d'évaluation des hallucinations (toutes modalités sensorielles confondues)

Figure 3. Modifications cérébrales associées aux hallucinations auditives en imagerie structurale.

Figure 4. Activations per-hallucinatoires chez un sujet.

Tableau 3. Principaux résultats des méta-analyses incluant les études de capture de l'hallucination auditive.

Figure 5. Les différents types de connectivité.

Figure 6. Résumé du modèle quadripartite proposé par Lefebvre et collaborateurs.

Figure 6. Représentation du faisceau arqué.

Figure 7. Représentation du faisceau arqué.

Tableau 4. Exemples de biomarqueurs en imagerie dans le domaine des hallucinations auditives.

Figure 8. Evolution de la symptomatologie hallucinatoire chez 362 patients présentant un premier épisode psychotique, 1, 3, 6, 9 et 12 mois après la mise en place d'un traitement antipsychotique.

Figure 9. Présentation de la méthode permettant de labelliser les volumes d'une session IRMf de repos en périodes hallucinatoires ou non-hallucinatoires.

Figure 10. Principe de classification d'un linear Support Vector Machine (lSVM).

Figure 11. Evolution du nombre de publications sur le neurofeedback (recherche réalisée dans la base de données Medline avec le mot clé « neurofeedback »).

Figure 12. Principe du neurofeedback « décodé ».

Figure 13. Proposition d'une interface multi-classificateurs pour la pris en en charge des hallucinations auditives par neurofeedback guidé par IRMf.

Figure 14. Présentation du design d'un essai contrôlé randomisé visant à évaluer l'efficacité d'un protocole de neurofeedback guidé par IRMf dans le traitement des hallucinations auditives chez les patients souffrant de schizophrénie.

Figure 15. Résumé du parcours du participant à l'étude.

LISTE DES ABRÉVIATIONS

AFPBN : Association française de psychiatrie biologique et de neuropharmacologie

ARMS : *At Risk Mental State*

DTI : *Diffusion Tensor Imaging*

EEG : électroencéphalographie

lSVM : *linear Support Vector Machine*

IRM : imagerie par résonance magnétique

IRMf : imagerie par résonance magnétique fonctionnelle

MCI : *Mild Cognitive Impairment*

MEG : magnétoencéphalographie

NExT : *Neurofeedback Evaluation & Training*

rTMS : repetitive transcranial magnetic stimulation (rTMS)

TEP : tomographie par émission de positons

TDCS : *transcranial direct current stimulation*

RESUME

L'hallucination est une expérience subjective vécue en pleine conscience consistant en une perception impossible à distinguer d'une perception réelle, mais survenant en l'absence de tout stimulus en provenance de l'environnement externe. Les symptômes hallucinatoires, qui peuvent concerner toutes les modalités sensorielles, sont retrouvés dans divers troubles neurologiques et psychiatriques mais également chez certains sujets indemnes de toute pathologie. Dans le champ de la psychiatrie, la pathologie la plus fréquemment associée aux hallucinations reste la schizophrénie et la modalité auditive est la plus représentée, puisque 60 à 80% des patients souffrant de ce trouble sont concernés. Le retentissement fonctionnel des hallucinations auditives peut être important, altérant significativement la qualité de vie des patients.

Dans ce contexte, la prise en charge de ce type de symptômes s'avère un enjeu considérable pour les personnes souffrant de schizophrénie. Pourtant, les moyens thérapeutiques actuellement disponibles (traitements médicamenteux antipsychotiques notamment) ne permettent pas toujours une rémission complète de la symptomatologie hallucinatoire et l'on considère que 25 à 30% des hallucinations auditives sont « pharmaco-résistantes ». C'est à partir de ce constat que, ces dernières années, ont émergé, pour le traitement des hallucinations auditives, des techniques de neuromodulation comme la stimulation magnétique transcrânienne répétée ou la stimulation électrique transcrânienne par courant continu. Toutefois, les résultats de ces nouvelles thérapies sur les hallucinations auditives résistantes restent modérés et le développement de stratégies alternatives demeure un enjeu de recherche majeur.

Actuellement, les travaux en imagerie fonctionnelle permettent d'affiner les modèles physiopathologiques des hallucinations auditives, mais leur intérêt pourrait aller au-delà de la recherche fondamentale, avec possiblement des applications cliniques telles que l'assistance thérapeutique. Ce travail de thèse s'inscrit précisément dans le développement de l'imagerie cérébrale de « capture » des hallucinations auditives, c'est-à-dire l'identification des *patterns* d'activation fonctionnels associés à la survenue des hallucinations auditives.

La première partie de ce travail est consacrée à la détection automatisée des hallucinations auditives en IRM fonctionnelle. L'identification des périodes hallucinatoires survenues au cours d'une session d'IRM fonctionnelle est actuellement possible par une méthode de capture semi-automatisée validée. Celle-ci permet une labellisation des données acquises au cours d'une session de repos en périodes « hallucinatoires » et « non-hallucinatoires ». Toutefois, le caractère long et fastidieux de cette méthode limite largement son emploi. Nous avons donc souhaité montrer comment les stratégies d'apprentissage machine (*support vector machine ou SVM,* notamment) permettent l'automatisation de cette technique par le développement de classificateurs performants, généralisables et associés à un faible coût de calcul (indispensable en vue d'une utilisation en temps réel). Nous proposons également le développement d'algorithmes de reconnaissance de la période « pré-hallucinatoire », en mettant en évidence que ce type de classificateur présente aussi des performances largement significatives. Enfin, nous avons pu montrer que l'utilisation de stratégies d'apprentissage-machine alternatives au SVM (*e.g*, le *TV-Elastic-net*), obtient des performances significativement supérieures au SVM.

La deuxième partie de cette thèse propose une réflexion théorique sur les perspectives thérapeutiques offertes par le développement de ces stratégies de capture de l'hallucination auditive. Le neurofeedback est une méthode thérapeutique non-invasive consistant à mesurer l'activité d'une ou de plusieurs régions cérébrales chez un sujet et à lui présenter en temps réel l'enregistrement de cette activité. Nous présentons les avancées récentes de cette technique dans le champ de la médecine, mais rappelons également les polémiques qu'elle suscite de par le faible nombre d'essais contrôlés randomisés actuellement disponibles. A partir d'un travail de revue systématique de la littérature sur l'utilisation du neurofeedback guidé par IRM fonctionnelle pour le traitement des troubles psychiatriques, nous explorons les différentes stratégies envisageables pour mettre en place un protocole visant spécifiquement la prise en charge des hallucinations auditives.

Enfin, nous proposons à partir de cette réflexion théorique, un protocole de neurofeedback guidé par IRM fonctionnelle basé sur une interface multi-classificateurs permettant l'identification des différentes périodes de l'hallucination auditive (phase dite d'entrée dans l'hallucination, phase hallucinatoire, phase de sortie de l'hallucination, période sans hallucination) et nous présentons l'étude que nous souhaitons mettre en place afin de tester son efficacité dans un essai contrôlé randomisé.

ABSTRACT

Hallucination is a transient subjective experience perceived as real, but occurring in the absence of an appropriate stimulation coming from the external environment. Hallucinatory events, which can occur across every sensory modality, are observed in various neurological and psychiatric disorders but also among “non-clinical” populations. The most frequent disorder associated with hallucinations in the field of psychiatry is schizophrenia. Auditory-verbal experiences are particularly frequent, with a lifetime-prevalence of 60 to 80% in patients suffering from schizophrenia. Hallucinations may cause long-term disability and poorer quality of life.

In this context, the management of auditory-verbal hallucinations in patients with schizophrenia constitutes a major challenge. However, despite the increasing sophistication of biological and psychosocial research methods in the field, no significant therapeutic breakthrough has occurred in the last decade and a consensus exists that a significant proportion of patients with schizophrenia (i.e., around 25 %), exhibit drug-resistant auditory-verbal hallucinations. Non-pharmacological treatments, such as repetitive transcranial magnetic stimulation (rTMS) or transcranial direct current stimulation (tDCS) have been proposed as an option for addressing the unmet medical needs described above. However, these neuromodulation techniques show a moderate effect in alleviating drug-resistant auditory-verbal hallucinations and the development of innovative therapeutic strategies remains a major challenge.

In recent years, the number of brain imaging studies in the field of auditory-verbal hallucinations has grown substantially, leading to a better pathophysiological understanding of this subjective phenomenon. Recent progress in deciphering the neural underpinnings of AVHs has strengthened transdiagnostic neurocognitive models that characterize auditory-verbal hallucinations, but more specifically these findings built the bases for new therapeutic strategies. In this regards the development of auditory hallucinations “capture" brain-imaging studies (i.e. the identification of functional patterns associated with the occurrence of auditory hallucinations), was the main topic of this thesis.

The first part of this work is devoted to the automatized detection of auditory-verbal hallucinations using functional MRI (fMRI). The identification of hallucinatory periods occurring during a fMRI session is now possible using a semi-automatized procedure based on an independent component analysis applied to resting fMRI data combined with a post-fMRI interview (i.e. the patient is asked to report auditory-verbal hallucinations immediately after acquisition). This “two-steps method” allows for the identification of hallucination periods (ON) and non-hallucination ones (OFF). However, the time-consuming nature of this *a posteriori* labelling procedure considerably limits its use. In these regards, we show how machine-learning, especially support vector machine (SVM), allows the automation of hallucinations capture. We present new results of accurate and generalizable classifiers which could be used in real-time because of their low computational-cost. We also highlight that algorithms able to identify the "pre-hallucinatory" period exhibit significant performances. Finally, we propose the use of an alternative learning-machine strategy, based on TV-Elastic-net, which achieves slightly better performances and more interpretable discriminative maps than SVM.

The second part of this work presents theoretical considerations about the therapeutic perspectives offered by the development of fMRI “capture” methods for auditory-verbal hallucinations. Neurofeedback is a non-invasive technique enabling participants to achieve voluntary control over the neuronal activity of one or more brain regions. We give a description of the technique while recent advances in the field are discussed, particularly the debate about efficacy of neurofeedback due to the small number of randomized controlled trials. Based on a systematic review of the literature on the use of fMRI-based neurofeedback to treat psychiatric disorders, we explore the possible strategies for a protocol aiming to treat auditory-verbal hallucinations.

Finally, on the basis of this new theoretical framework, we propose an fMRI-neurofeedback protocol based on a multi-classifier interface allowing the identification of the different periods of auditory-verbal hallucination (“ignition of hallucination phase”, “hallucination phase”, “extinction of hallucination phase”, “period without hallucination”) and we describe a randomized controlled trial that will be implemented to test the efficiency of fMRI-based neurofeedback to relieve drug-resistant auditory hallucinations.

MOTS-CLES :
Hallucination, IRM fonctionnelle, imagerie de capture, apprentissage machine, temps-réel, neurofeedback.

KEY-WORDS:
Hallucinations, functional-MRI, «capture » brain imaging, machine-learning, real-time, neurofeedback.

1 . INTRODUCTION

“Quite recently I had a heart stroke, did I tell you that? It caused aphasia: a fluxion of the brain’s right side that yields hallucinations. Netley, I saw God. I knelt before him and he told me what to do.”

Alan Moore, From Hell

Classiquement, l'hallucination est définie comme une « perception sans objet à percevoir » (1). Il s'agit d'une expérience subjective vécue en pleine conscience consistant en une perception impossible à distinguer d'une perception réelle, mais survenant en l'absence de tout stimulus en provenance de l'environnement externe (2). L'hallucination doit être distinguée de l'illusion qui correspond à la perception déformée d'un objet réel.

Les hallucinations peuvent concerner l'ensemble des modalités sensorielles (vision, audition, odorat, goût, toucher, voir **Tableau 1**) (3). On peut les retrouver dans diverses pathologies comme les maladies neurologiques (*e.g.* maladie d'Alzheimer (4), maladie de Parkinson (5), démence à corps de Lewy (6)) ou les troubles psychiatriques (*e.g.* schizophrénie, trouble bipolaire (7), trouble de la personnalité borderline (8)), au cours de prises de substances psychoactives (drogues psychédéliques notamment) ou même parfois chez des sujets indemnes de toute pathologie (la fréquence des hallucinations auditives non-cliniques atteindrait 6% en population générale selon les résultats d'une méta-analyse récente (9)) (10–12).

Dans le champ de la psychiatrie, la pathologie la plus fréquemment associée aux hallucinations reste la schizophrénie. La modalité auditive est la plus représentée puisque 60 à 80% des patients souffrant de ce trouble sont concernés (12,13). La schizophrénie est une pathologie complexe dont la symptomatologie comprend plusieurs dimensions cliniques : les symptômes positifs (hallucinations, idées délirantes, etc.), les symptômes négatifs (repli, émoussement des affects, etc.), les symptômes de désorganisation (trouble du cours de la pensée, etc.) et les symptômes cognitifs (14). De fait, les études dans le champ de la schizophrénie privilégient de plus en plus une approche dimensionnelle permettant l'analyse d'une dimension symptomatique particulière, telle que celles proposées dans le *Research Domain Criteria* (*RDoC*) par le *National Institute of Mental Health* (15). Il s'agit, par cette méthode, de réduire l'hétérogénéité clinique de la maladie à l'étude et de travailler sur des dimensions plus élémentaires mais également de remettre en question l'approche catégorielle classique (voir **Figure 1**) pour privilégier une approche basée sur un continuum s'étendant du normal au pathologique (16). Dans ce contexte, la dimension « hallucinations », et plus particulièrement, celle des « hallucinations auditives » qui constituent la thématique centrale de cette thèse, apparait extrêmement intéressante (17).

Dans cette introduction, nous exposerons le contexte clinique et historique dans lequel ce travail de thèse s'inscrit, nous proposerons ensuite un état des lieux des connaissances actuelles sur la physiopathologie des hallucinations auditives dans la schizophrénie. Enfin, nous décrirons les principaux traitements disponibles pour la prise en charge de ce symptôme chez les patients souffrant de schizophrénie, avant d'envisager la manière dont les techniques d'imagerie pourraient permettre de développer des stratégies thérapeutiques innovantes de l'hallucination.

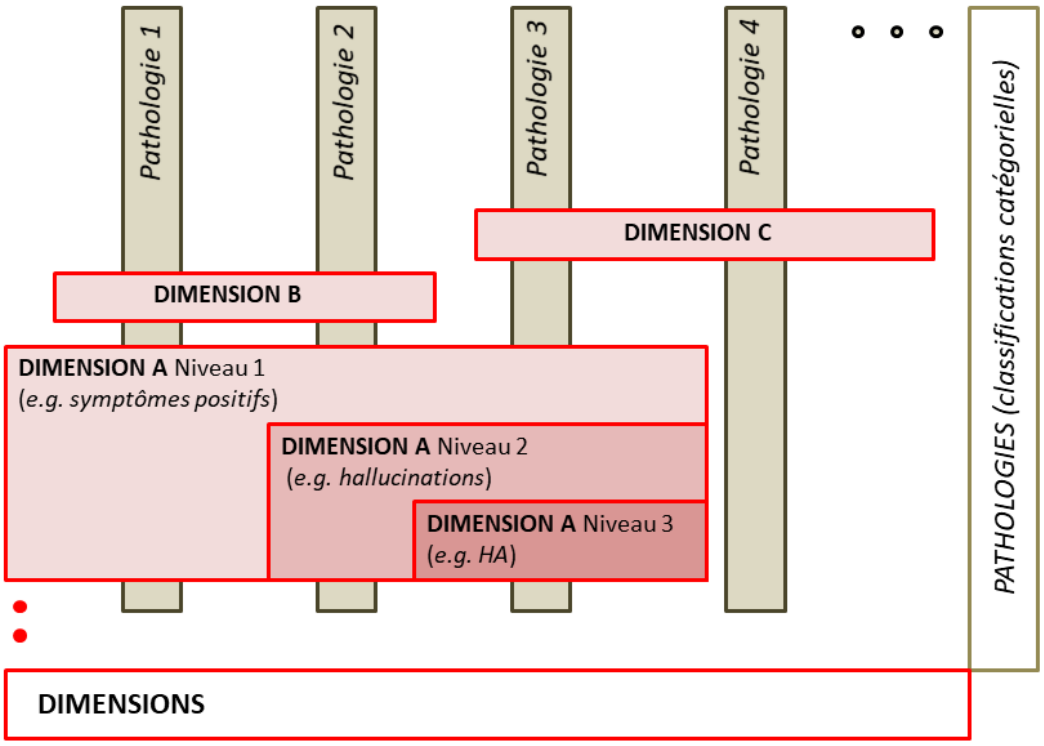


Figure 1. Illustration de l'approche dimensionnelle pour explorer les hallucinations auditives. *(HA : hallucinations auditives)*

TYPE D'HALLUCINATION	DESCRIPTION
Auditive	*Sons simples* : sonneries, mélodies, etc. *Sons complexes* : une ou plusieurs voix (hallucinations acoustico-verbales) qui s'adressent au patient à la deuxième ou troisième personne et qui peuvent converser entre elles.
Visuelle	*Images simples* : phosphènes, ombres, etc. *Images complexes* : scènes visuelles (par exemple : démons, Phoenix, animaux, etc.)
Gustative	Concerne le gout des aliments (sensation que la nourriture est pourrie ou empoisonnée car elle a « un drôle de gout »)
Olfactive	Perception de mauvaises odeurs, le plus souvent provenant du patient lui-même (par exemple odeurs de cadavres)
Tactile	Concerne la sensibilité externe (par exemple sensation de brûlure, de piqure, de filet sur la peau, de souffle, etc.)
Cénesthésique	Concerne la sensibilité interne : sensation d'être traversé (par un voile ou une balle), sensation de transformation corporelle, sensation de pourrissement des organes (comme dans le classique syndrome de Cotard), sensation de viol, etc.

Tableau 1. Description des principaux types d'hallucinations. Adapté de (3).

1.1 Histoire générale du concept

« Hallucinatio » signifie « erreur » ou « égarement » en latin. Ce terme serait lui-même dérivé du grec « *aluein »* qui signifie « errer, avoir l'esprit égaré » (18). Le terme est introduit en 1572 par Ludwig Lavater dans la langue anglaise et apparaitra un peu plus tard, au cours du XVII^{ème} siècle dans la langue française (19).

Dès l'Antiquité, les phénomènes hallucinatoires sont décrits (par exemple en Mésopotamie (20) ou en Grèce) Ils sont alors intimement liés aux croyances magiques et religieuses (possession, châtiment divin, etc.) mais les médecins grecs comme Hippocrate ou Galien évoquent déjà leur potentielle origine cérébrale.

Le Moyen-Age est souvent décrit comme une période d'obscurantisme au cours de laquelle l'héritage de la médecine grecque est oublié et les phénomènes hallucinatoires principalement considérés comme des manifestation du démon ou des anges (possession par le diable, sorcellerie, etc.) (19). Pourtant, durant cette période, les médecins arabo-musulmans, développent une approche psychosomatique des troubles mentaux qui sont alors considérés comme des maladies à part entière pour lesquelles se développent des prises en charge psychothérapeutiques ou pharmacologiques dans les Bîmâristâns (21).

Dans le monde occidental, ce n'est qu'à partir de la Renaissance que les théories sur le rôle des dysfonctionnements cérébraux dans la survenue des hallucinations deviennent prédominantes. Toutefois, c'est au cours de cette même période qu'aura lieu le « *Grand Renfermement* » décrit par Michel Foucault (22) : *« l'asile […] unit les fous aux vagabonds, aux pauvres, aux oisifs, à toutes sortes de dépravés* » (23). Celui-ci débutera avec la création de l'Hôpital Général en 1656 et il faudra attendre la fin du XVIII^{ème} siècle pour qu'une approche humaniste voie le jour avec des aliénistes comme Philippe Pinel et Jean-Baptiste Pussin.

Esquirol est le premier à introduire le concept d'hallucination en médecine. Celle-ci est alors définie comme une « perception sans objet » : « *Un homme qui a la conviction intime d'une sensation actuellement perçue, alors que nul objet extérieur propre à provoquer cette sensation n'est à portée de ses sens, est dans un état d'hallucinations* » (24). Cette définition sera très longtemps utilisée par des aliénistes comme J. Baillarger, J-P. Falret ou P. Guiraud (25). Baillarger, élève d'Esquirol, proposera la distinction entre hallucinations psychosensorielles (perception par les organes des sens) et

hallucinations psychiques (lorsqu'un sujet a l'impression que certaines de ses pensées lui sont étrangères ou imposées sans manifestations sensorielles ni objectivité spatiale) (26). Seglas quant à lui décrira les hallucinations psychomotrices verbales au cours desquelles les patients « *articulent avec leurs lèvres, leur langue, leur luette et tout leur appareil phonatoire, des propos qu'ils ne prennent pas à leur compte et qui, parfois, redoublent leurs hallucinations psychosensorielles*. » (Seglas cité dans (25)).

Plus tard, Henri Ey développe l'un des modèles les plus complet pour décrire les hallucinations survenant dans le cadre de la schizophrénie et propose en 1973 de compléter la définition d'Esquirol en précisant que l'hallucination est une « *perception sans objet à percevoir* » (1). Il souhaite en effet insister sur le processus hallucinatoire lui-même, basé sur une réalité subjective (la perception « vraie » d'une fausse réalité).

Enfin, dans les années 1980, Assad et Shapiro proposeront une autre définition du phénomène : « perception sans stimulus externe correspondant » (27). Cette définition, qui permet de sortir du débat « vraie » ou « fausse » perception, sera reprise dans les modèles neuropsychologiques qui seront développés ensuite.

1.2 Contexte clinique : les hallucinations dans la schizophrénie

Chez les patients souffrant de schizophrénie, les hallucinations peuvent survenir dans toutes les modalités sensorielles décrites précédemment. Toutefois, la prévalence de ce type de symptômes varie largement selon la modalité étudiée. On considère que les hallucinations auditives sont les plus fréquentes avec une prévalence vie entière chez ces patients, comprise entre 64 et 80%. Toutefois, longtemps considérées comme mineures, les hallucinations visuelles sont pourtant retrouvées chez plus d'un patient sur quatre (28). Enfin, selon une étude récente sur un échantillon de plus de 900 sujets souffrant de schizophrénie, la prévalence vie-entière pour les hallucinations olfactives serait comprise entre 6 et 10% et celle des hallucinations tactiles entre 9 et 19% (29).

Les hallucinations peuvent survenir dans plusieurs modalités chez un même patient : les hallucinations multimodales seraient retrouvées chez plus de la moitié des patients (30). L'association hallucinations visuelles / hallucinations auditives est la plus fréquente puisqu'environ un patient ayant présenté des hallucinations auditives sur trois a également présenté des hallucinations visuelles.

Cependant, toutes les combinaisons sont possibles (voir pour illustration les résultats de l'étude (29) présentés en **Figure 2**).

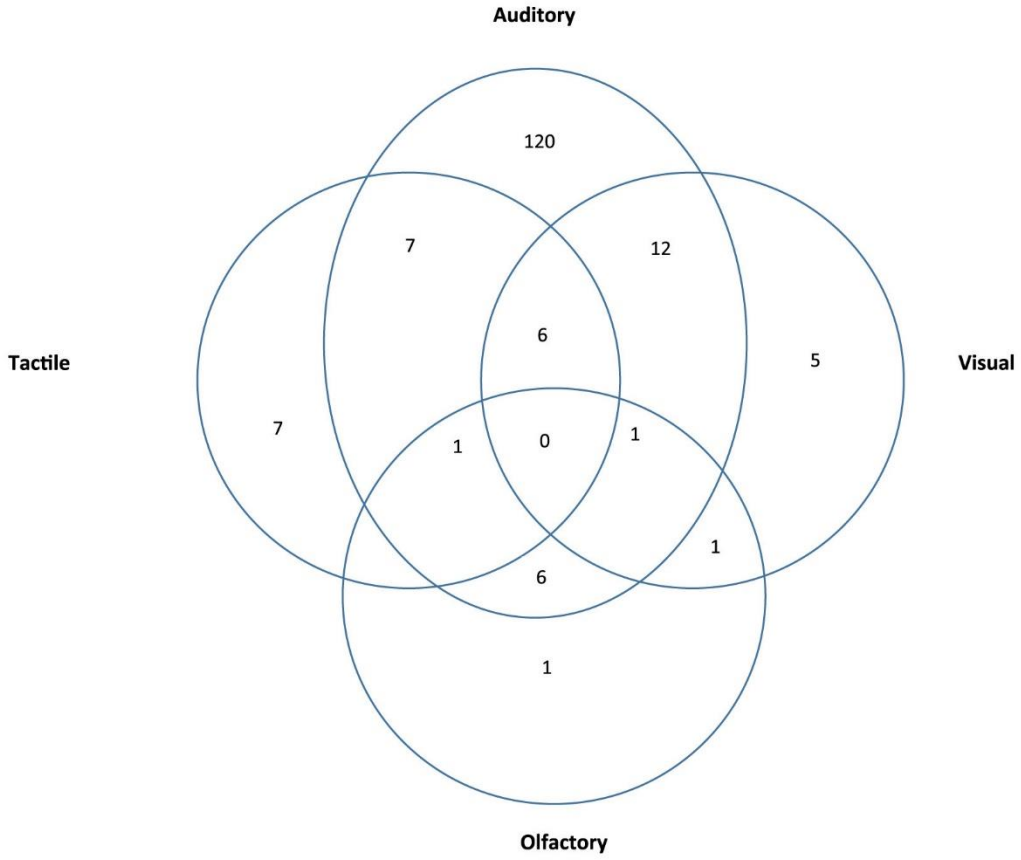


Figure 2. Co-occurrence des modalités hallucinatoires dans une cohorte de 171 patients souffrant de schizophrénie et d'hallucinations selon au moins une modalité. Tiré de (29).

Les associations ne figurant pas sur le diagramme sont les suivantes : visuel + Tactile, n=3 ; visuel + tactile + olfactive, n=1.

Les hallucinations auditives restent les plus fréquentes et les plus étudiées dans le champ de la schizophrénie. Il faut, au sein de cette catégorie, distinguer les hallucinations acoustico-verbales qui désignent les voix, des autres hallucinations auditives (sons divers, simples ou complexes). Tout confondu, ce type d'expériences survient, la plupart du temps, au cours d'épisodes de décompensation aigue. Cependant, les hallucinations auditives peuvent également être résistantes à un traitement médicamenteux bien conduit (dans 25 à 30% des cas environ) et donc persister chez les patients de manière chronique (31,32). Par ailleurs, le retentissement fonctionnel de ces symptômes peut être important (aboutissant par exemple à des tentatives de suicide (33), ou à des épisodes dépressifs caractérisés avec risque de passage à l'acte suicidaire (34,35)) et la qualité de vie des patients très affectée (36).

Les hallucinations auditives dans la schizophrénie apparaissent hétérogènes en ce qui concerne leur contenu et leurs caractéristiques phénoménologiques (37). Il peut s'agir de sons simples, d'onomatopées ou de mots uniques mais aussi de conversations complètes avec une prosodie bien définie. Dans ce cas, il peut y avoir une ou plusieurs voix, masculine ou féminine, familières ou non (12), qui commentent les actes de la personne, voire qui lui donnent des instructions et des ordres (c'est le classique « syndrome d'influence »). Les contenus sont très fréquemment à valence émotionnelle négative (38–40) et le degré de contrôle des patients sur leurs hallucinations est le plus souvent faible (39,41). Enfin, la localisation externe ou interne des voix n'est pas spécifique d'un diagnostic et n'est pas associée à un retentissement fonctionnel plus ou moins important (42) contrairement à ce qu'avait pu proposer G.G. de Clérambault sur la « spatialisation » des hallucinations auditives (d'abord intrapsychiques) au cours de l'évolution de la schizophrénie.

L'évaluation clinique des hallucinations auditives chez les patients souffrant de schizophrénie est un exercice délicat. En effet, en dehors des rares comportements observables comme les attitudes d'écoute (le sujet semble ailleurs, distrait ou préoccupé, parfois de façon très brusque), les réponses faites aux « voix » ou les véritables dialogues avec elles, l'exploration de ce type de symptômes repose essentiellement sur les éléments rapportés par le patient. Dans ce contexte, l'utilisation d'outils psychométriques standardisés apparait indispensable (43). La symptomatologie hallucinatoire est succinctement évaluée dans plusieurs échelles d'évaluation globale de la schizophrénie, notamment la *Scale for the Assessment of Positive Symptoms (SAPS)* dans laquelle on retrouve 7 items qui concernent les hallucinations ou la *Positive and Negative Syndrome Scale (PANSS)* avec l'item P3.

Mais des instruments d'évaluation spécifiques des hallucinations sont également disponibles. Le **Tableau 2a** regroupe des échelles évaluant spécifiquement les hallucinations auditives, parmi les plus utilisées, tandis que le **Tableau 2b** présente des outils permettant une évaluation générale de la symptomatologie hallucinatoire (quelle que soit la modalités sensorielle). En effet, certaines échelles, développées récemment, s'inscrivent dans l'approche dimensionnelle évoquée précédemment. C'est le cas de la *Psycho-sensory hallucinations scale (PSAS)*, une échelle multimodale basée sur une approche transdiagnostique visant à évaluer les hallucinations (toutes modalités) dans le cadre de la schizophrénie et de la maladie de Parkinson (44).

Des outils permettent aussi d'évaluer spécifiquement certains aspects associés à la symptomatologie hallucinatoire comme les stratégies de *coping* développées face aux hallucinations, (*i.e. Responses to Auditory Hallucinations Questionnaire, RAHQ*), ou les croyances associées aux hallucinations auditives (*i.e. Beliefs about Voices Questionnaire-Revised (BAVQ-R)*, *Voice Power Differential (VPD), Voice and You Scale (VAY)*). Le *Hallucinations change score (HCS)* permet quant à lui d'évaluer l'évolution des symptômes hallucinatoires au cours du temps. En lien avec les psychothérapies de type pleine conscience, certaines échelles d'acceptation sont disponibles comme la *Southampton Mindfulness of Voices Questionnaire (VAAS)*. Des outils évaluant la « propension » à présenter des hallucinations existent également comme la *Launay Slade Hallucination Scale (LSHS)* (45).

Enfin, des outils spécifiques ont également été développés pour l'évaluation des hallucinations précoces. C'est le cas de la *Multisensory HAllucinations Scale for Children (MHASC)* disponible sous la forme d'une application pour tablette (46).

Nom de l'échelle	Passation	Dimensions étudiées	Score
Psychotic Symptom Rating Scales – Auditory Hallucinations (PSYRATS-AH)	Entretien structuré.	Fréquence, durée, localisation (interne ou externe), intensité, conviction dans les croyance sur l'origine, quantification des contenus négatifs, intensité des contenus négatifs, quantification de la gène associée, intensité de la gène associée, perturbation dans le fonctionnement, contrôlabilité.	Chaque item est coté de 0 à 4. Privilégier une interprétation item par item plutôt que l'utilisation du score total.
Auditory Hallucinations Rating Scale (AHRS)	Entretien structuré.	Fréquence, intensité, sentiment de réalité, nombre de voix, contenu, saillance attentionnelle, gène associée.	Chaque item est coté de 0 à un score maximal variant de 5 à 9 selon l'item.
Hamilton Program for Schizophrenia Voices Questionnaire (HPSVQ)	Auto-questionnaire.	Fréquence, contenu négatif, intensité, durée, retentissement fonctionnel, souffrance associée, impact sur l'auto-évaluation, clarté, influence sur les actes, localisation, moment de la journée et situations associées aux hallucinations, localisation des voix.	Chaque item est coté de 1 à 5.
Characteristics of Auditory Hallucinations Questionnaire (CAHQ)	Auto-questionnaire.	Fréquence, intensité, contrôlabilité, clarté, ton, distractibilité et souffrance associée dans les 24 heures précédentes.	Chaque item est coté de 0 à 5.

Tableau 2a. Principales échelles d'évaluation des hallucinations auditives.

Nom de l'échelle	Passation	Dimensions étudiées	Score
Psycho-sensory hallucinations scale (PSAS)	Entretien structuré.	16 items évaluant l'ensemble des modalités sensorielles des hallucinations (auditive, visuelle, olfactive/gustative, cénesthésique) et d'un item spécifique « ange gardien » chez les patients souffrant de schizophrénie ou de maladie de Parkinson.	Chaque item est coté de 0 à 4.
Questionnaire for Psychotic Experiences (QPE)	Entretien structuré.	50 items (maximum) évaluant la présence, la sévérité et les caractéristiques des hallucinations, des illusions, et idées délirantes.	Cotation spécifique.
Multi-Modality Unusual Sensory Experiences Questionnaire (MUSEQ)	Auto-questionnaire.	43 items évaluant les expériences sensorielles inhabituelles dans 6 modalités : auditive, visuelle, gustative, olfactive, tactile/cénesthésique et sensation de présence.	Chaque item est coté de 1 à 5 (de 1 : expérience sensorielle inhabituelle subclinique à 5 : hallucination).

Tableau 2b. Principales échelles d'évaluation des hallucinations (toutes modalités sensorielles confondues).

1.3 Physiopathologie des hallucinations auditives dans la schizophrénie

Dans cette partie, nous abordons l'état actuel des connaissances en ce qui concerne les mécanismes cérébraux sous-tendant les hallucinations auditives dans la schizophrénie. La fréquente co-occurrence des hallucinations auditives et visuelles dans la schizophrénie suggère certaines spécificités de la physiopathologie des hallucinations auditives dans cette pathologie par rapport notamment aux pathologies neurodégénératives (*e.g*. maladie de Parkinson, démence à corps de Lewy) dans lesquelles les hallucinations surviennent plus fréquemment dans une seule modalité, de façon isolée (47). Seule la physiopathologie des hallucinations auditives dans la schizophrénie sera abordée ici.

1.3.1 Principaux modèles neuropsychologiques

D'un point de vue neurocognitif, les hallucinations sont habituellement considérées comme des perceptions erronées survenant en l'absence de stimuli externes. Elles sont liées à une intégration erronée de processus cognitifs et sensoriels qui peuvent influencer la perception consciente (48,49).

De nombreux modèles neurocognitifs ont été proposés pour expliquer la phénoménologie des hallucinations auditives qui, comme nous l'avons vu, apparait extrêmement hétérogène (50,51). Ces modèles s'appuient sur 3 principaux concepts : 1/ les hallucinations auditives sont générées par un phénomène interne attribué à une source externe, 2/ les hallucinations auditives sont associées à une sensation de perte de contrôle en ce qui concerne leur contenu, leur fréquence et leur survenue chez les patients, 3/ les hallucinations auditives sont associées à une charge émotionnelle le plus souvent négative.

L'attribution à une source externe de phénomènes auto-générés est une caractéristique qui a été notamment développée dans les travaux de Chris Frith (52). Ce modèle est basé sur l'hypothèse selon laquelle les mécanismes physiologiques permettant la reconnaissance de notre discours intérieur et de nos pensées comme propres sont perturbés chez les sujets souffrant de schizophrénie (défaut d'attribution de la source). Ce phénomène serait lié à un dysfonctionnement de la copie d'efférence (également dénommée décharge corollaire) qui est normalement associée à toute action (mouvement ou discours) et qui permet au sujet de prédire à la quasi-perfection la conséquence de ses propres actions (53,54). Lors de la survenue d'hallucinations auditives, le discours intérieur ne serait pas accompagné d'une décharge corollaire efficace. En l'absence de cette décharge corollaire, l'individu

développe un sentiment de ne pas être l'initiateur et le sujet de ses pensées et de ses actions (ici, le discours), phénomène également appelé « défaut d'agentivité ». Ce dernier pourrait aboutir à une attribution externe du discours et à l'émergence de l'hallucination. Ce modèle s'appuie notamment sur des études d'imagerie cérébrale ayant montré que le cortex auditif est anormalement hyperactivé lors de tâches de discours interne chez les patients souffrant de schizophrénie avec hallucinations auditives (54,55). Plusieurs travaux ont également pu montrer que ces mêmes patients présentent des difficultés à identifier les informations auto-générées (56–58). Toutefois, ce modèle reste discuté notamment en ce qui concerne la source de l'activation du cortex sensoriel auditif qui ne peut pas simplement être attribuée à l'absence de décharge corollaire (59). Certaines caractéristiques cliniques sont également difficiles à expliquer à partir de ce modèle (complexité des hallucinations auditives, voix d'une ou plusieurs personnes du sexe opposé, voix avec accent étranger, etc.) (60).

Un autre modèle, fondé sur le concept de « *reality testing* » (61) place le « source monitoring » au centre des mécanismes physiopathologiques impliqués dans les hallucinations. Il s'agit d'un processus de métacognition qui permet de former une représentation cohérente d'une expérience à partir des pensées et des croyances. L'hypothèse principale de ce modèle postule que les patients avec schizophrénie présentent un déficit dans la discrimination des évènements issus de l'environnement interne (l'imagination) et de ceux issus de l'environnement externe (la réalité). Ils présenteraient un biais spécifique d'attribution aux sources externes (62). Ces conceptions sont à rapprocher de celles de la théorie de détection du signal selon laquelle la détection d'un stimulus est basée sur deux principes : la sensibilité perceptuelle (efficience du système perceptuel) et un critère de décision subjectif (correspondant à la décision selon laquelle un évènement perçu est un stimulus). Ce critère serait perturbé chez les patients souffrant de schizophrénie et d'hallucinations auditives (63) avec une sensibilité accrue pour détecter des mots et des sons intégrés dans un bruit blanc par rapport aux patients souffrant de schizophrénie sans hallucinations (64).

Bien que ces modèles soient actuellement les plus développés (65,66), ils ne peuvent à eux seuls expliquer tous les aspects de la phénoménologie de l'hallucination auditive. Dans ce contexte, d'autres modèles ont été proposés. Par exemple, Allen et al. ont développé un modèle neuro-anatomique basé sur un réseau cérébral impliquant des régions impliquées dans les processus de perception mais également dans le langage ou la régulation émotionnelle et l'attention (67). Wilkinson et al. (68) ont adapté un modèle centré sur une prédiction statistique basée sur les expériences passées

du sujet (69) qui permet de minimiser l'erreur de prédiction (*i.e.* déviation des prédictions faites par le sujet). Cette erreur de prédiction est également au centre des modèles développés dans le champ des neurosciences computationnelles (70,71). Les prédictions en série s'appuyant principalement sur des réseaux de neurones hiérarchiques, permettent de créer une représentation de l'environnement extérieur constamment mise à jour pouvant avoir notamment un impact sur le système sensoriel, avec une activité accrue en cas d'évènement imprévu qui pourra être associée à la formation d'hallucinations et d'idées délirantes (72).

1.3.2 Études d'imagerie cérébrale

Les modèles neurocognitifs actuels s'inspirent largement des études en neuroimagerie qui peuvent être divisées en travaux d'imagerie cérébrale structurale, fonctionnelle et de connectivité.

1.3.2.1 Imagerie structurale et hallucinations auditives

Grâce aux progrès techniques en IRM et aux analyses de groupes, de nombreux travaux ont pu mettre en évidence l'existence de variations fines au niveau de la morphologie cérébrale chez les patients souffrant de schizophrénie et d'hallucinations auditives (voir **Figure 3**).

Ces études s'intéressent, pour la plupart, au volume de substance grise (*Grey Matter Volume*). On retrouve, en particulier, des variations subtiles du volume cérébral au niveau des régions impliquées dans le langage : aire de Broca et son homologue contro-latérale (73). Plusieurs travaux ont aussi mis en évidence un volume de substance grise altéré au niveau de la région temporale supérieure, chez les patients atteints de schizophrénie et souffrant d'hallucinations acoustico-verbales (72,74,75). En outre, la sévérité des hallucinations a été retrouvée corrélée avec la réduction de substance grise au niveau du gyrus temporal supérieur gauche (75), comprenant le gyrus de Heschl dont les modifications structurales ont pu être mises en évidence dans d'autres travaux (76). En dehors de ces anomalies, il a été montré que la surface du cortex insulaire était associée à la survenue d'hallucinations (77,78) (l'insula est une structure corticale fondamentale ayant des connections avec de nombreuses aires du cortex (79), on parle de « *rich-club hub* » pour désigner ces aires cérébrales). Des modifications d'autres *rich-club hubs*, comme le thalamus (80) et le cervelet (81), sont également associées aux hallucinations acoustico-verbales (73).

L'imagerie structurale ne se limite pas à la mesure quantitative de l'épaisseur et de la surface corticale. Il est désormais possible d'analyser finement la morphologie corticale, en particulier la forme et l'organisation des circonvolutions et sillons qui constituent un marqueur indirect du développement cérébral (82). En effet, le processus de plissement cortical, ou gyrification, commence à partir de la dixième semaine de vie fœtale et le cortex cérébral passe, au cours des deuxième et troisième trimestre de grossesse, d'une surface relativement lisse à une structure complexe plissée, qui s'avère ensuite relativement stable au cours de la vie (83). L'étude des variations de la morphologie corticale associées aux hallucinations auditives permet donc d'explorer l'impact des facteurs développementaux et la vulnérabilité précoce à ce symptôme (84).

Ces travaux ont pu mettre en évidence une diminution significative de la gyrification des régions corticales impliquées dans le langage (notamment sillons temporaux supérieurs, sillon frontal moyen gauche, aire de Broca) chez des patients souffrant de schizophrénie et présentant des hallucinations acoustico-verbales chroniques par rapport à des sujets témoins (85). Certains aspects phénoménologiques des hallucinations auditives ont également pu être associés à des modifications morphologiques dans les réseaux du langage. Par exemple, la perception par le patient d'une origine interne ou externe correspond à des déviations sulcales spécifiques au niveau de la jonction temporo-pariétale droite (86).

Ces résultats sont cohérents avec l'hypothèse neuro-développementale de la schizophrénie, selon laquelle cette dernière serait l'étape finale de processus neurodéveloppementaux anormaux ayant débutés plusieurs années avant le début de la maladie (87).

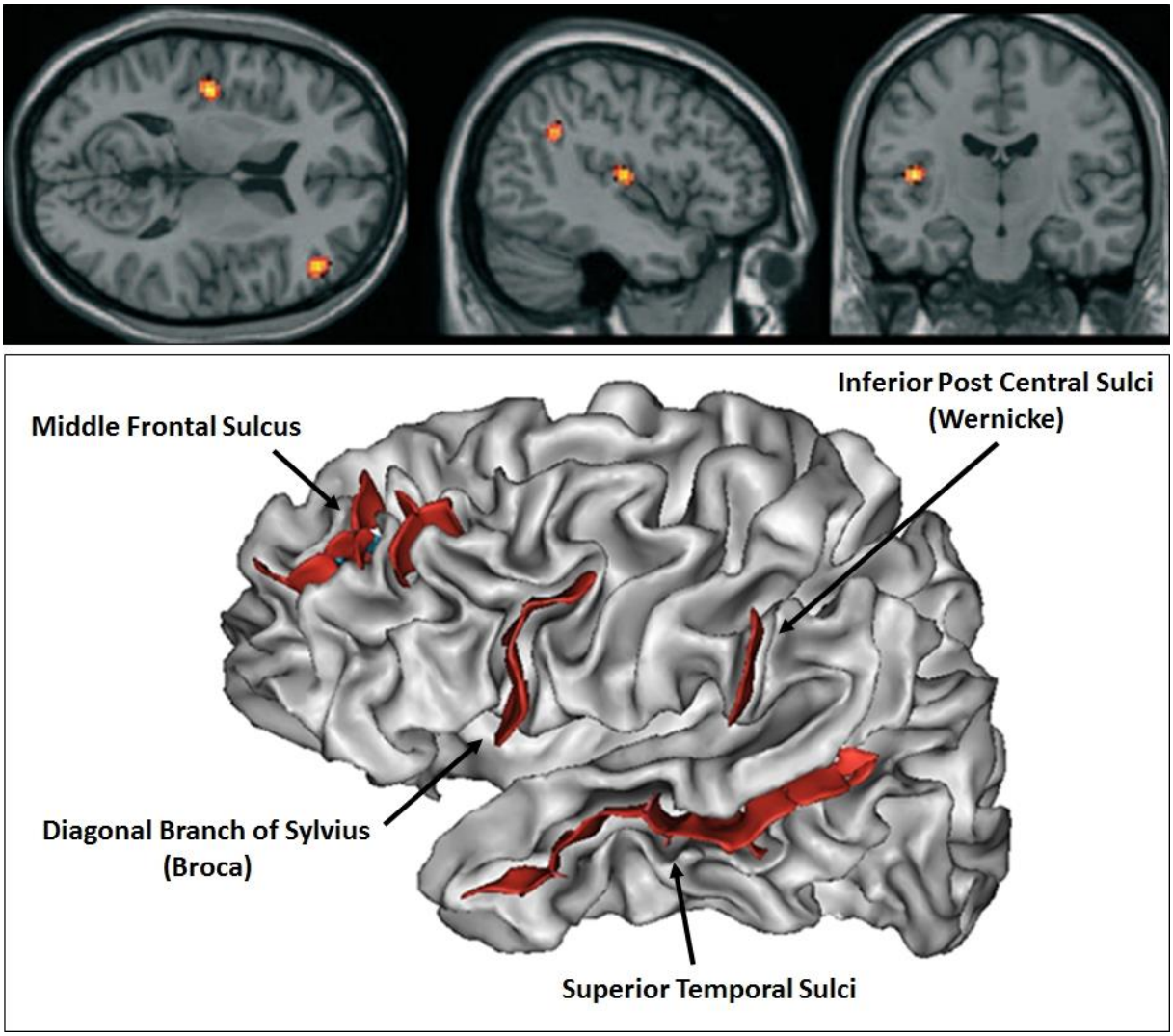


Figure 3. Modifications cérébrales associées aux hallucinations auditives en imagerie structurale.

(Cadrant supérieur) Corrélations entre la sévérité des hallucinations auditives et le volume de substance grise au niveau du cortex préfrontal droit (région homologue à droite de l'aire de Broca), du gyrus supramarginal inférieur gauche et du gyrus temporal transverse (gyrus de Heschl). Adapté de (88)*.*

(Cadrant inférieur) Modifications de la morphologie corticale associée aux hallucinations auditives (surface sulcale) spécifiques au niveau des aires du langage : sulcus temporal supérieur droit et gauche, sulcus frontal moyen gauche, sulcus post-central inférieur (aire de Wernicke) et branche diagonale de la fissure sylvienne gauche (aire de Broca). Adapté de (85)*.*

1.3.2.2 Imagerie fonctionnelle et hallucinations auditives

Parmi les études d'imagerie fonctionnelle, nous souhaitons d'emblée distinguer les études dites "traits" des études dites "état" (89). Les études "traits" visent à comparer l'activité cérébrale de sujets présentant des hallucinations à celle de sujets sans hallucination au cours de tâches cognitives, notamment celles impliquant le langage. Ces études permettent de comprendre le fonctionnement cérébral des sujets présentant des hallucinations, et ainsi, de mettre en évidence les bases neurales de la susceptibilité à halluciner. Souvent, les scores de sévérité de la symptomatologie hallucinatoire sont utilisés comme régresseurs durant l'analyse de l'activité cérébrale (90). Toutefois, ces études ne donnent pas accès à ce qui se passe au niveau fonctionnel dans le cerveau d'un sujet qui vit une hallucination. C'est justement l'objectif des études "état", au cours desquelles l'activité cérébrale en période hallucinatoire est comparée à celle des périodes sans hallucination chez les mêmes sujets.

1.3.2.2.1 Principaux résultats des études "trait"

Les études « trait » mesurent l'activité cérébrale au cours de taches spécifiques chez des patients avec et sans hallucinations auditives. Le plus souvent, un questionnaire *a posteriori* permet de s'assurer de l'absence d'hallucinations auditives au cours de la tache afin d'éviter toute confusion avec les études « état ».

Le principal résultat de ces études est la diminution d'activité fonctionnelle au niveau des régions temporales classiquement impliquées dans le traitement de la voix humaine et du langage intérieur chez les sujets présentant des hallucinations acoustico-verbales (91–94). La susceptibilité à halluciner semble être associée à l'utilisation des ressources de ces régions, responsable par exemple, de difficultés dans le traitement de l'information sensorielle normalement traitée par ces structures. En effet, cette information pourrait entrer en compétition avec le traitement des hallucinations acoustico-verbales (95).

Un second résultat d'intérêt est l'activation accrue du cortex cingulaire antérieur chez les patients présentant des hallucinations auditives. Cette zone a été montrée impliquée dans le processus d'attribution d'un stimulus à une origine interne ou externe à soi (96–98). Nous avons vu précédemment que l'hallucination pouvait être considérée comme une "non reconnaissance" de son propre discours intérieur par le patient que celui-ci attribue donc à une source externe. Ces modèles apparaissent aussi

cohérents avec les données structurales récentes mettant en évidence des atteintes spécifiques du cortex cingulaire antérieur chez les patients avec hallucinations auditives (98–101).

1.3.2.2.2 Principaux résultats des études "état"

Ces études, également appelées "étude de capture hallucinatoire", restent difficiles à mettre en place car elles impliquent que le sujet fasse l'expérience d'hallucinations auditives au cours de la session d'imagerie fonctionnelle. Toutefois, des résultats intéressants ont été obtenus à partir de données IRMf ou tomographie par émission de positons (TEP) de patients souffrant de schizophrénie, au cours d'épisode hallucinatoire. Cinq méta-analyses dont les résultats sont résumés dans le **Tableau 3** ont été publiées. Les résultats montrent que c'est l'activation d'un vaste réseau cérébral qui est à l'origine de l'expérience hallucinatoire (91,102–105) avec une augmentation de l'activité cérébrale au niveau des régions perceptuelles (cortex auditif primaire et associatif) (59,106–108), des zones de production et de perception verbale (aire de Broca et aire de Wernicke) (55,102,106,109,110) mais également le cortex hippocampique et para-hippocampique (111) (voir **Figure 4**). Ces dernières régions sont impliquées dans les phénomènes d'accès et de rappel en mémoire à long terme, ces résultats semblent donc compatibles avec l'hypothèse selon laquelle les hallucinations acoustico-verbales pourraient être des réactivations spontanées de traces mnésiques au niveau du cortex auditif associatif (112,113). Bien que cela soit encore débattu, il semble que le cortex auditif primaire ne soit pas indispensable à la survenue d'hallucinations acoustico-verbales même si plusieurs études ont pu mettre en évidence son activation au cours des épisodes (114,115). Il semblerait que cette structure puisse être d'avantage lié à certaines propriétés phénoménologiques, comme le sentiment de réalité de l'expérience sensorielle (116).

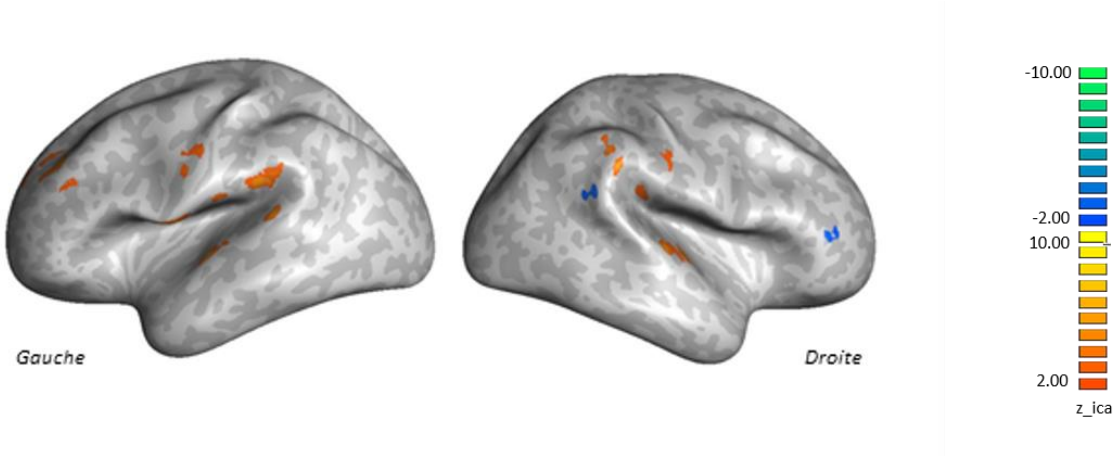


Figure 4. Activations per-hallucinatoires chez un sujet.

Une session IRMf de capture a été réalisée chez un patient souffrant d'hallucinations auditives fréquentes. La méthodologie décrite dans (117) *a été appliquée. Les régions identifiées ont été projetées sur des hémisphères « insufflés », afin de mieux discerner les activations pouvant se trouver dans les anfractuosités corticales.*

Références	Méthodologie	Principaux résultats
Jardri et al. **(102)**	Méta-analyse des études de capture incluant 10 études.	Activation concomitante de la survenue des HA dans un réseau bilatéral comprenant aire de Broca, insula antérieur, gyrus précentral, operculum, gyri temporaux supérieur et moyen, lobule pariétal inférieur, région hippocampique et para-hippocampique.
Kompus et al. **(103)**	Comparaison d'une méta-analyse incluant 11 études incluant sujets avec schizophrénie et sujets contrôles durant écoute de stimuli externes avec une méta-analyse incluant 12 études incluant des sujets présentant des HA au repos.	Activation augmentée en l'absence de stimulus auditif externe et activation diminuée en présence de stimulus externe dans le cortex auditif primaire gauche et le cortex préfrontal rostral droit chez les patients avec HA.
Kuhn et Gallinat. **(91)**	Méta-analyse de 10 études « état ».	Activation au niveau des gyri frontaux inférieurs, gyri postcentraux et operculum parietal gauche.
Van Lutterveld et al. **(118)**	Comparaison d'une méta-analyse incluant 10 études « état » dans lesquelles les HA étaient signalées par bouton avec une méta-analyse de 11 études détection stimuli auditifs signalés par bouton.	Activation spécifique des HA par rapport à la détection de stimuli auditifs au niveau des régions suivantes : claustrum, pulvinar, corps géniculé médial, pyramis, culmen, putamen, insula, gyrus parahippocampique, gyrus frontal médial, gyrus precentral, gyrus postcentral, gyrus temporal supérieur et gyrus frontal inférieur droit.
Zmigrod et al. **(104)**	Méta-analyse incluant 13 études « état »	Activation au niveau des gyri temporaux supérieurs, de l'insula, du gyrus frontal inférieur droit, du gyrus préfrontal droit

Tableau 3. Principaux résultats des méta-analyses incluant les études de capture de l'hallucination auditive.

HA : hallucinations auditives.

1.3.2.3 Études de connectivité et hallucinations auditives

L'étude de la connectivité entre différentes régions cérébrales tient aujourd'hui une place extrêmement importante dans la compréhension des hallucinations. On distingue plusieurs types de connectivité cérébrale : respectivement fonctionnelle, effective et anatomique (voir **Figure 5**) (119). La connectivité fonctionnelle se définit comme une corrélation d'activation entre deux régions cérébrales spatialement distantes, sans information sur le sens de cette interaction. La connectivité effective correspond quant à elle à l'influence directe d'une région cérébrale sur une autre et fournit donc des informations sur la causalité de relation entre deux zones cérébrales. Enfin, la connectivité anatomique correspond à l'étude des faisceaux de substance blanche reliant différentes régions cérébrales par une technique particulière d'IRM appelée Imagerie en Tenseur de Diffusion (*Diffusion Tensor Imaging : DTI*).

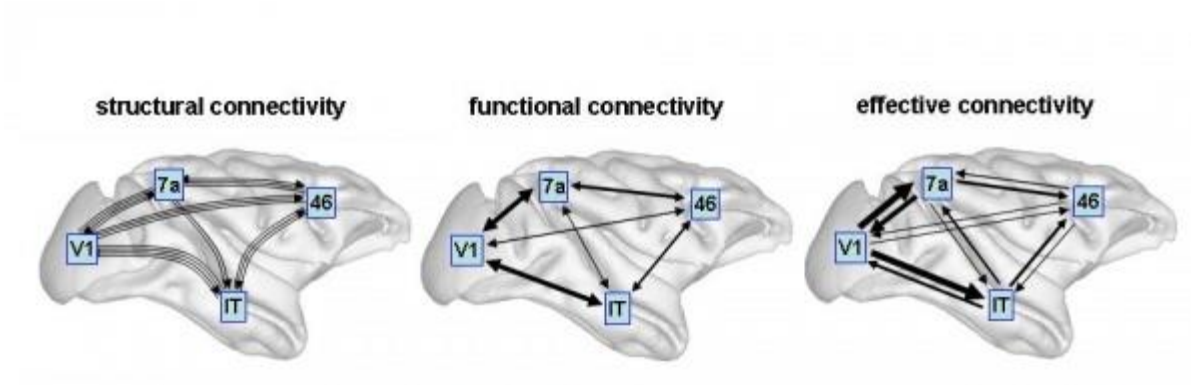


Figure 5. Les différents types de connectivité. Tiré de (120).

1.3.2.3.1 Connectivité fonctionnelle et connectivité effective

La connectivité fonctionnelle peut être évaluée par IRMf ou par TEP chez des sujets au repos ou réalisant des tâches spécifiques (tâches de langage notamment). Ces méthodes ont permis de valider l'existence d'une dysconnectivité dans la schizophrénie en général mais aussi propre aux hallucinations. Que ce soit au repos (121,122) ou lors de taches cognitives impliquant le langage (123,124), les travaux en connectivité fonctionnelle et effective ont ainsi montré qu'au-delà des anomalies structurales et fonctionnelles retrouvées chez les patients souffrant d'hallucinations acoustico-verbales, des anomalies de connectivité entre ces mêmes régions cérébrales jouent un rôle majeur dans l'émergence du symptôme. Par ailleurs, ces anomalies apparaissent différentes selon la modalité sensorielle de l'hallucination étudiée (125,126). Dans le cas des hallucinations auditives, ce

sont des anomalies de connectivité fronto-temporales gauches qui ont été mises en évidence au cours des études impliquant des tâches de langage (127).

Les études réalisées chez des patients au repos ont également obtenu des résultats intéressants qui ont permis de mettre en évidence le rôle de dysfonctionnement au niveau du réseau de mode par défaut (*Default Mode Network, DMN*) dans la survenue d'hallucinations auditives. Le DMN est un réseau de régions cérébrales activées lors des états d'éveil au repos (*i.e.* en l'absence de tache cognitive) et serait impliqué dans des processus tels que l'introspection, la référence à soi ou la mémoire épisodique. Il est composé du cortex préfrontal médial, du cortex cingulaire postérieur, du lobule pariétal inférieur et du lobe temporal inférieur (128). Le rôle du DMN dans la physiopathologie des troubles psychiatriques, et en particulier dans la schizophrénie, est actuellement largement étudié (129). Chez les patients souffrant d'hallucinations auditives, des anomalies de connectivité fonctionnelle de repos ont été mises en évidence notamment entre la jonction temporo-pariétal gauche, le cortex auditif et d'autres régions corticales et sous-corticales, même si plusieurs résultats apparaissent contradictoires et peu ont été répliqués (121). Des interactions spécifiques entre le DMN et d'autres réseaux de repos (*Resting-State Networks*) comme le réseau central exécutif (*Central Executive Network, CEN*); composé du cortex préfrontal dorsolatéral et le cortex pariétal postérieur (130) et le réseau de salience (*Salience Network, SN*), ont également été identifiées chez les sujets souffrant d'hallucinations auditives (131). Le SN est composé de la portion antérieure de l'insula et le cortex cingulaire. Il est impliqué dans la sélection des stimuli pertinents de l'environnement afin de permettre une orientation attentionnelle et une réponse comportementale adaptée, assurant la transition entre le DMN et le CEN. Dans un travail récent, Lefebvre et collaborateurs ont exploré la séquence chronologique des interactions entre DMN, CEN et SN et complexe hippocampique gauche (dont l'implication dans la survenue des hallucinations a pu être évoquée précédemment) au cours des différentes phases de l'hallucination. Un modèle de connectivité fonctionnelle par *Dynamic Causal Modelling* a permis de montrer que la période précédant la survenue d'une hallucination est caractérisée par une déstabilisation des connexions entre le SN, le CEN et le DMN qui rend le système plus vulnérable aux stimuli inappropriés provenant du complexe hippocampique. Au cours de la période ON, une information est donc transmise du complexe hippocampique gauche au SN qui induit en conséquence une diminution de l'activité du DMN. La phase d'extinction quant à elle était associée, dans ce travail, à un renforcement des connections du SN activant le CEN, ce qui aboutirait à la fin de la période

hallucinatoire (voir **Figure 6**). Ces résultats suggèrent que l'orientation volontaire de l'attention pourrait être impliquée dans la phase d'extinction des hallucinations (l'activation du CEN étant associée à des taches de focalisation de l'attention), ce qui apparait cohérent avec les stratégies favorisées dans les programmes de thérapie cognitive et comportementale des hallucinations (132).

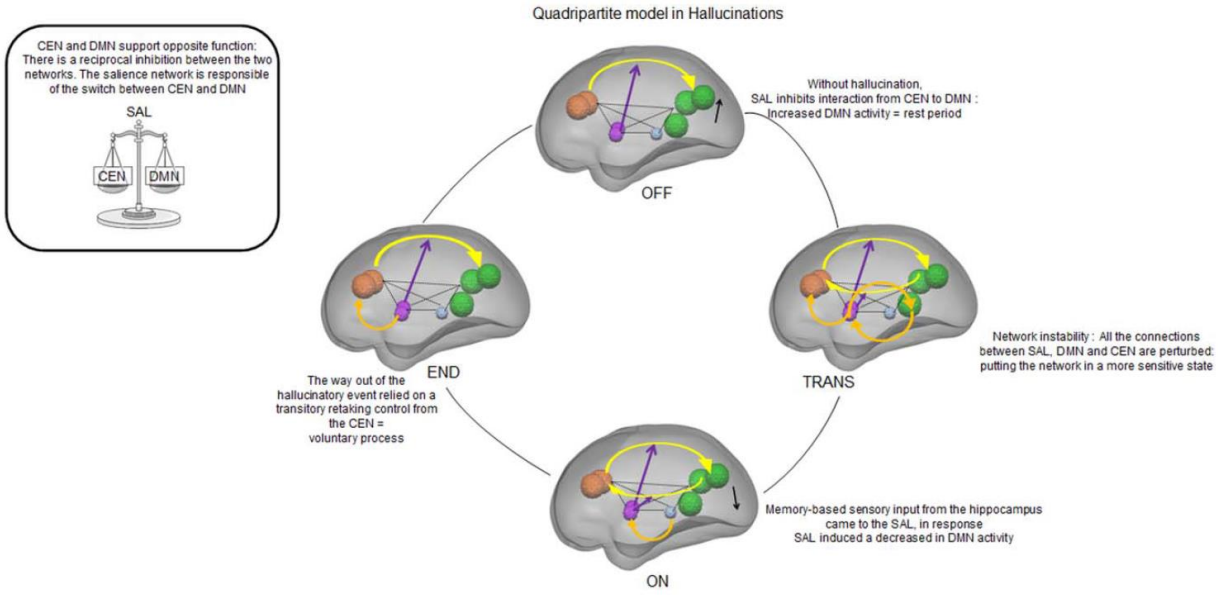


Figure 6. Résumé du modèle quadripartite proposé par Lefebvre et collaborateurs. Tiré de (133).

En orange = réseau central exécutif ; en violet = réseau de salience ; en vert = réseau de mode par défaut ; flèches jaunes = connections impactées par une modulation issue du réseau de salience (flèches violettes) ; flèches oranges = connections impactées impactée par la phase de la période halluciantoire.

1.3.2.3.2 Connectivité anatomique (ou structurale)

La connectivité anatomique (ou structurale) permet une évaluation indirecte de l'intégrité structurale et l'orientation des grands faisceaux de substance blanche grâce à la technique de DTI qui évalue la diffusion des molécules d'eau dans le tissu cérébral. Cette technique a permis de mettre en évidence une altération de la connectivité structurale chez les patients souffrant de schizophrénie, notamment au niveau des lobes frontaux et temporaux (122,134).

Des différences de connectivité, au niveau des aires du langage (dysconnectivité fronto-temporale gauche), ont également été mises en évidence en comparant (i) des patients souffrant de schizophrénie avec hallucinations, (ii) des patients souffrant de schizophrénie non-hallucinés et (iii) des sujets non schizophrènes (76,135,136). Ces différences apparaissent particulièrement importantes au niveau du faisceau arqué gauche (voir **Figure 7**), qui relie l'aire de Broca au niveau du lobe frontal à l'aire de Wernicke au niveau temporo-pariétal (137–140). Geoffroy et collaborateurs ont ainsi montré dans une méta-analyse incluant 5 études comparant des sujets souffrant de schizophrénie avec hallucinations auditives (n=106) et des sujets contrôles sans trouble psychiatrique (n=150), une fraction d'anisotropie diminuée (en faveur d'anomalies de connectivité anatomique) au niveau de cette structure chez les patients présentant des hallucinations auditives. De manière tout à fait intéressante, dans une autre étude, ces altérations apparaissent corrélées à la sévérité de la symptomatologie hallucinatoire (140). D'autres modifications de la substance blanche ont pu être mises en évidence chez les patients souffrant de schizophrénie avec hallucinations auditives notamment au niveau du faisceau arqué droit (135,141,142), suggérant une atteinte bilatérale de ce faisceau chez les patients avec schizophrénie souffrant d'hallucinations auditives (143). Enfin, le rôle d'anomalies au niveau du corps calleux dans la survenue d'hallucinations auditives est également évoqué par certains résultats suggérant une augmentation de la connectivité structurale de cette région (144).

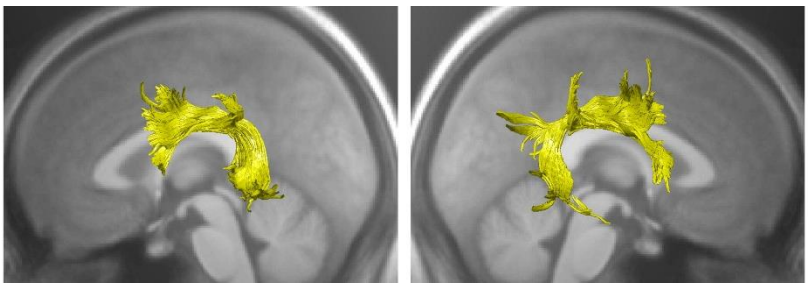


Figure 7. Représentation du faisceau arqué. Tiré de (141).

1.3.2.4 Vers le développement de biomarqueurs spécifiques des hallucinations auditives ?

L'ensemble des travaux décrits précédemment ont permis des avancées importantes dans la compréhension des mécanismes physiopathologiques sous-tendant l'hallucination auditive mais ils permettent également d'entrevoir le développement de marqueurs basés sur la neuroimagerie pour les hallucinations auditives. Les perspectives sont nombreuses : marqueurs de pronostic permettant de définir des niveaux de risque d'évolution pour chaque patient, marqueurs de diagnostic, marqueurs de prédiction de la réponse aux traitements médicamenteux, etc. (voir **Tableau 4** pour une illustration par des exemples). Mais des travaux restent nécessaires pour valider des biomarqueurs fiables, reproductibles, non-invasifs, simples à mettre en pratique et peu onéreux.

Etude	Type de biomarqueur	Nombre de sujets	Biomarqueur choisi	Objectif	Critère principal / Gold Standard	Validation (Performances, caractéristiques)	Faisabilité (non-invasif, simple à mettre en place, faible coût)	Utilisation actuelle en pratique clinique
Milev et al., 2003 (145)	*Pronostic*	123	VSG du lobe temporal	Corrélation entre VSG initiale au niveau du lobe temporal et persistance des HA à 5 ans.	Semaines par année avec HA évaluées comme au moins modérément sévères.	VSG au niveau du lobe temporal plus faible corrélé à la persistance des HA (r=0,24; p<0,01)	IRM structurale (non invasive)	Non
Kindler et al., 2013 (146)	*Marqueur d'efficacité du traitement*	30 (placebo: n=15; rTMS: n=15)	Flux sanguin cérébral au niveau du cortex auditif primaire de l'aire de Broca, du gyrus cingulaire évalué au cours d'une session d'IRMf au repos.	Corrélation entre amélioration Clinique et baisse du flux sanguin cérébral dans les régions étudiées.	Réponse thérapeutique (baisse du score de sévérité des HA)	Amélioration clinique significativement corrélée à la baisse du débit sanguin cérébral dans le cortex auditif primaire.	IRM fonctionnelle au repos (non invasive)	Non
Homan et al., 2012 (147)	*Prédiction de la réponse thérapeutique*	24	Débit sanguin cérébral au niveau du gyrus temporal supérieur évalué au cours d'une session IRMf de repos avant mise en place du traitement.	Corrélation entre le debit sanguine cerebral au niveau du gyrus temporal supérieur gauche et le réponse thérapeutique (distinction répondeurs / non répondeurs à la rTMS)	Réponse thérapeutique (baisse >50% des scores de sévérités des HA)	Débit sanguin cérébral au niveau du gyrus temporal supérieur gauche ASC = 0,96 (IC95% [0,87-1]) pour la prédiction répondeur / non répondeur	IRM fonctionnelle au repos (non invasive)	Non

Tableau 4. Exemples de biomarqueurs en imagerie dans le domaine des hallucinations auditives.

1.4 Prise en charge des hallucinations

De par le retentissement fonctionnel qu'elles génèrent, les hallucinations auditives nécessitent souvent une prise en charge dédiée. Plusieurs moyens thérapeutiques sont actuellement disponibles pour soulager les patients avec schizophrénie de leurs voix.

1.4.1 Traitements médicamenteux

Chez les patients souffrant de schizophrénie, les antipsychotiques sont largement prescrits pour traiter les symptômes positifs, dont les hallucinations auditives font partie. Aucun essai clinique ayant évalué spécifiquement l'efficacité des différentes molécules antipsychotiques sur la symptomatologie hallucinatoire n'est disponible. Toutefois, il est possible de dégager des résultats intéressants à partir de certaines études internationales comme l'étude *European First-Episode Schizophrenia Trial (EUFEST)* (148). Dans ce travail, 498 patients présentant un premier épisode psychotique ont été randomisés en 5 groupes et l'efficacité de 5 traitements antipsychotiques (halopéridol, olanzapine, amisulpride, quétiapine, ziprasidone) a été comparée après 12 mois de traitement. Aucune différence entre les groupes n'a été mise en évidence en ce qui concerne l'efficacité du traitement mais le taux d'interruption du traitement était plus important dans le groupe halopéridol et moins important dans le groupe olanzapine et le groupe amisulpride. Pour ce qui est de la symptomatologie hallucinatoire, une analyse incluant les 362 patients présentant un score supérieur à 3 à l'item P3 de la *Positive and Negative Syndrome Scale*, a montré que le score moyen passe de 4,4 à 2,5 après 4 semaines de traitement et il est de 1,5 après 6 mois de traitement (voir **Figure 8**) (149). Aucune différence significative d'efficacité selon la molécule n'a pu être mise en évidence.

Le principal mécanisme d'action des antipsychotiques est le blocage des récepteurs dopaminergiques. Leur effet sur les symptômes positifs serait lié à cette action sur les récepteurs dopaminergiques D2 au niveau de la voie méso-limbique, tandis qu'un certain nombre de leurs effets indésirables est lié au blocage des récepteurs dopaminergiques sur d'autres voies (voie nigro-striée pour les effets extra-pyramidaux, voie tubéro-infundibulaire pour les effets endocriniens, etc.). Classiquement, on distingue les antipsychotiques de première et de seconde génération, dont le profil d'efficacité reste proche, malgré des différences en termes de tolérance.

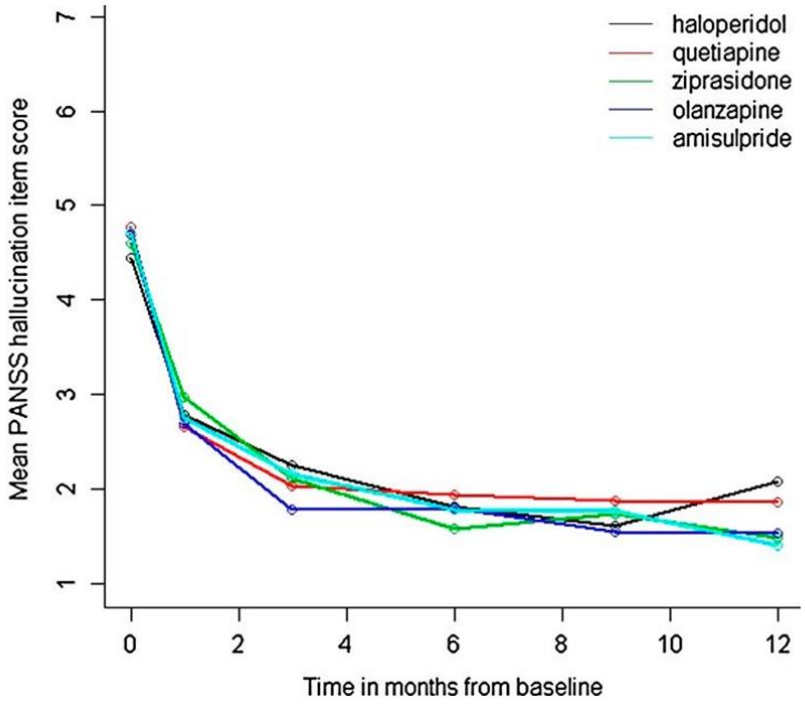


Figure 8. Evolution de la symptomatologie hallucinatoire chez 362 patients présentant un premier épisode psychotique, 1, 3, 6, 9 et 12 mois après la mise en place d'un traitement antipsychotique. Tiré de (149).

Plusieurs recommandations internationales suggèrent d'éviter la prescription de clozapine ou d'olanzapine en première intention (150) étant donné les effets secondaires (notamment le risque de syndrome métabolique). Pourtant, ces molécules seraient associées à une meilleure efficacité et à des taux de réhospitalisation plus faibles (151). En pratique, même si cela reste discuté, un changement de traitement antipsychotique doit être envisagé en cas de non-réponse après 2 à 4 semaines de traitement. Un traitement par clozapine doit être envisagé après échec de 2 traitements antipsychotiques bien menés (152).

La meilleure stratégie pour le traitement de maintien reste également discutée même si certains travaux semblent indiquer que la poursuite du traitement à la dose initiale efficace apparait associée à un taux de rechutes inférieur par rapport à la poursuite du traitement à une posologie plus faible (153). Les antipsychotiques d'action prologée peuvent également présenter un intérêt en terme de prévention de la rechute et d'amélioration du fonctionnement social (154).

Des travaux en imagerie fonctionnelle ont pu montrer que chez les patients traités par antipsychotiques, on retrouve une activité accrue au niveau du noyau caudé droit et une activité

diminuée au niveau du gyrus frontal médial, du cervelet et du thalamus droit (155) avec des variations selon la molécule (modification d'activité plus importante dans les zones sous-corticale pour les antipsychotiques de première génération ; modification d'activité préférentiellement dans les régions frontales pour les antipsychotiques de deuxième génération) par rapport aux patients non traités.

En imagerie structurale, l'utilisation des antipsychotiques au long cours serait associée à une diminution du volume de substance grise notamment au niveau du cortex temporal et pré-frontal, variable selon la molécule mais plus marqué avec des posologies importantes (156,157).

1.4.2 Traitements non médicamenteux : psychothérapies

Plusieurs techniques de psychothérapie ont montré une efficacité dans la prise en charge des hallucinations auditives notamment en combinaison avec la pharmacothérapie et les programmes de réhabilitation (158,159).

Au premier plan, on retrouve les thérapies cognitives et comportementales (TCC) qui ont pour objectif une meilleure adaptation à l'expérience des hallucinations auditives d'un point de vue cognitif, comportemental et affectif (160). L'intérêt des TCC dans la prise en charge des hallucinations auditives a été validé. On citera notamment une méta-analyse de Wykes qui rapporte une taille d'effet de 0,35 à 0,44 pour ce type de stratégie thérapeutique (149,161). Plusieurs programmes ont été développés au sein desquels l'identification des croyances sur les voix et leur remise en question sont les grands principes. L'identification et le renforcement de nouvelles stratégies de *coping* apparaissent fondamentaux dans cette approche (162–164). Ces programmes peuvent cibler spécifiquement la symptomatologie hallucinatoire ou s'intégrer dans des protocoles plus généraux de prise en charge de la schizophrénie. Cependant, les programmes de TCC ciblant un symptôme en particulier semblent plus efficaces (165). Il peut s'agir de thérapies individuelles (164) ou de groupe parfois centrées sur le renforcement de l'estime de soi (166,167).

Dans la lignée de ces programmes se sont également développés des thérapies d'acceptation et d'engagement, ainsi que des thérapies cognitives basées sur la pleine conscience (168).

Enfin, la thérapie intégrative centrée sur les hallucinations (*Hallucination-focused Integrated Treatment*, *HIT*) est un programme complet organisé en modules (dont certains s'adressent à des sous-

populations : enfants, sujet âgé, etc.) bénéficiant d'influences variées : TCC, thérapie familiale systémique ou encore techniques de remédiation cognitive (169,170).

A noter, des recommandations cliniques pour l'utilisation de ces TCC en association avec un traitement médicamenteux auprès de patients souffrant de schizophrénie ont été proposées par le *National Institute for Clinical Excellence*.

1.4.3 Traitements non médicamenteux : les nouvelles pistes

Malgré la disponibilité de ces moyens thérapeutiques, la fréquence des hallucinations résistantes à un traitement bien conduit est importante (32). Il apparait donc indispensable de développer de nouvelles pistes thérapeutiques. Au cours des dernières années, des techniques thérapeutiques de neuromodulation innovantes telle la stimulation magnétique transcranienne répétée (rTMS) ou la stimulation transcrânienne par courant continu (notée tDCS pour *transcranial direct current stimulation*) ont pu voir le jour dans le champ des pathologies psychiatriques.

1.4.3.1 Stimulation transcrânienne par courant continu

La tDCS est une technique de stimulation cérébrale non invasive et indolore qui implique la circulation d'un courant électrique de faible intensité (1-2 mA) entre 2 électrodes placées sur le scalp d'un sujet. Récemment, il a été montré qu'elle permettait de moduler l'excitabilité corticale et les comportements de manière transitoire et sécurisée, sans effet secondaire notable (171). La tDCS pourrait constituer un traitement de choix au rang des alternatives thérapeutiques non-pharmacologiques dans la prise en charge des pathologies psychiatriques pharmaco-résistantes, en particulier les hallucinations auditives dans la schizophrénie.

Les protocoles de tDCS dans la schizophrénie reposent sur le modèle de la dysconnectivité fronto-temporale (172) marquée par une hyperactivité de la région temporo-pariétale gauche pendant l'expérience hallucinatoire (173) et une hypoactivité des cortex préfrontaux (174) liée aux symptômes négatifs. Le montage ciblant les hallucinations auditives est un montage fronto-temporal gauche avec une stimulation inhibitrice (cathode) au niveau de la jonction temporo-pariétale gauche et une stimulation excitatrice concomitante (anode) au niveau du cortex préfrontal (CPF) gauche (175).

Le premier essai contrôlé randomisé en double aveugle chez 30 patients souffrant de schizophrénie a utilisé le montage fronto-temporal gauche et a rapporté une diminution significative de

31% des hallucinations auditives dans le groupe actif (d=1,58 ; p<0,001). Les effets observés sur les hallucinations auditives ont été maintenus au moins 3 mois après la fin des séances (176). Plusieurs études de cas et plusieurs études en ouvert ont par la suite corroboré ces résultats (177). Parmi ces études, certaines ont également observé une diminution significative des autres symptômes de la schizophrénie (symptômes négatifs) et une amélioration de l'insight (177). Par exemple, dans une étude ouverte chez 21 patients, Bose et collaborateurs ont montré que la réduction des hallucinations auditives après tDCS fronto-temporale gauche (32,7%) était significativement corrélée à l'amélioration de l'insight (178). Mondino et collaborateurs ont étudié l'effet de la tDCS fronto-temporale sur la reconnaissance de soi (un trouble cognitif lié aux hallucinations auditives) chez 28 patients et sur la connectivité fonctionnelle de la jonction temporo-pariétale gauche chez 23 patients dans deux essais contrôlés randomisés en double aveugle chez des sujets présentant des hallucinations auditives pharmaco-résistantes. Dans leur première étude, les patients reconnaissaient mieux le langage auto-généré après la tDCS active comparativement à la tDCS placebo (montage fronto-temporal gauche) ; l'amélioration des performances des patients dans la tâche de reconnaissance de soi était corrélée à la diminution de la fréquence des hallucinations auditives (179). Dans leur seconde étude, ces auteurs rapportent que la diminution des hallucinations auditives est corrélée à la diminution de la connectivité fonctionnelle entre la jonction temporo-pariétale gauche et l'insula gauche (180).

En ce qui concerne le maintien au long cours des effets cliniques de la tDCS, dans une étude de cas, Andrade et collaborateurs (181) ont rapporté un bénéfice clinique de séances d'entretien quotidiennes durant 3 ans (montage fronto-temporal gauche) chez un patient présentant des symptômes résistants à la clozapine.

En conclusion, malgré des résultats prometteurs, de nouvelles études sont attendues afin de confirmer l'efficacité clinique de cette technique et de déterminer les paramètres optimaux de stimulation (182).

1.4.3.2 Stimulation magnétique transcrânienne et neuronavigation

La rTMS est une technique permettant de modifier, de manière focale et non invasive, l'excitabilité corticale par l'utilisation d'un champ magnétique généré par une bobine. Le traitement par rTMS est actuellement en cours d'évaluation dans de nombreux troubles psychiatriques comme l'épisode dépressif caractérisé (183) ou les troubles obsessionnels compulsifs (184). L'utilisation de la

rTMS illustre comment, ces dernières années, une approche translationelle se développe, des études théoriques en neurosciences et en neuroimagerie aux applications cliniques.

Chez les patients présentant des hallucinations auditives, l'utilisation de la rTMS permet de réduire l'excitabilité corticale des régions retrouvées anarchiquement activées dans les hallucinations en imagerie cérébrale. La stimulation répétée à basse fréquence au niveau de la jonction temporo-pariétale gauche (site T3-P3) permet en effet de diminuer l'intensité et la fréquence des hallucinations auditives (185). La place de cet outil dans la stratégie de prise en charge des hallucinations auditives est actuellement bien définie (186) et son intérêt en traitement de seconde intention des hallucinations résistantes aux traitements antipsychotiques, bien démontré (187). Cependant, le repérage de la région cible comme actuellement proposé dans les études internationales s'avère peu précis et basé sur l'utilisation du système EEG 10-20 sur le scalp du sujet. Il semble que l'effet de la rTMS puisse être optimisé grâce à l'utilisation de données d'IRMf individuelles (188), on parle alors de neuronavigation. La réalisation d'une IRMf de capture destinée à guider le traitement par rTMS des hallucinations réfractaires est actuellement en cours d'évaluation (189). Les techniques d'imagerie pourraient donc avoir un intérêt dans l'optimisation des techniques thérapeutiques de neuromodulation, actuellement en plein essor.

1.4.3.3 Optimiser les stratégies thérapeutiques grâce à la capture hallucinatoire ?

Comme nous l'avons vu précédemment, les travaux en imagerie fonctionnelle permettent aujourd'hui d'affiner les modèles physiopathologiques des hallucinations auditives, mais leur intérêt ne se limite pas à la recherche fondamentale et les applications cliniques se développent rapidement, notamment dans l'assistance thérapeutique.

Dans cette optique, le développement de l'imagerie de « capture » des hallucinations auditives, c'est-à-dire de l'identification des *patterns* d'activation fonctionnels associés à la survenue des hallucinations auditives, est prometteur. Outre l'optimisation des protocoles de neuronavigation, il pourrait en effet permettre le développement de stratégies thérapeutiques intermédiaires entre neuromodulation et psychothérapie telles le neurofeedback guidé par IRM fonctionnelle.

1.5 Objectifs de la thèse

L'objectif principal de cette thèse est de développer et valider des techniques automatisées d'analyse du signal IRMf permettant de détecter la survenue des hallucinations auditives chez les patients souffrant de schizophrénie. Il s'agit plus particulièrement de développer des classificateurs performants capables de reconnaitre les *patterns* d'activation cérébrale associés aux hallucinations auditives.

L'objectif secondaire de ce travail est de proposer des techniques de thérapie guidée par l'image utilisant ces classificateurs en temps réel. Nous nous focaliserons notamment sur le neurofeedback guidé par IRMf.

2 . DETECTION AUTOMATISEE DES HALLUCINATIONS AUDITIVES EN IRM FONCTIONNELLE

“Artificial Intelligence is defined as the opposite of natural stupidity.”

Woody Allen

.

Comme nous l'avons décrit précédemment, les récents progrès en imagerie cérébrale se sont accompagnés de larges avancées dans la compréhension des mécanismes physiopathologiques impliqués dans les hallucinations auditives (voir 1.3.2. Etudes d'imagerie cérébrale). Parmi les travaux actuellement disponibles en imagerie cérébrale fonctionnelle, nous avons distingué études « traits » et études « état ». Ce sont ces dernières, également appelées « études de capture » et consistant en une mesure directe des activations cérébrales concomitantes à la survenue des hallucinations auditives, qui vont nous intéresser dans cette partie. Les études de capture sont particulièrement complexes à mettre en place car elles nécessitent que le patient présente des symptômes hallucinatoires mais également des périodes libres de tout symptôme (pouvant être considérées comme un état de base neurophysiologique), au cours d'une session d'imagerie cérébrale fonctionnelle (TEP, IRMf ou magnétoencéphalographie (MEG)). Outre la nature imprévisible et suggestive du phénomène hallucinatoire, la composante émotionnelle de ce symptôme peut également être à l'origine de difficultés puisque le sujet doit rester immobile au cours de l'intégralité de la session d'enregistrement. Ces conditions, toutes particulières, expliquent probablement les petits effectifs inclus dans ces études et qui en constituent la principale limite (173). Les méta-analyses apparaissent précieuses pour minimiser cette limite et ont permis d'identifier cinq grands pôles d'activation au cours des hallucinations auditives : au niveau du gyrus frontal inférieur (pars opercularis, gyrus précentral gauche, insula antérieure bilatérale, operculum frontal bilatéral), au niveau temporal (gyri temporaux moyen et supérieur gauches), au niveau du lobule pariétal inférieur gauche (gyrus supramarginal gauche), au niveau des régions hippocampales et para-hippocampales, au niveau du globus pallidus interne droit (173).

Les études de capture nécessitent également une méthodologie spécifique afin de déterminer les périodes au cours desquelles le patient présente des hallucinations auditives. Plusieurs techniques sont retrouvées dans la littérature. Certaines études proposent l'utilisation d'un bouton-réponse permettant au patient de signaler la survenue d'hallucinations au cours de la session d'imagerie fonctionnelle (111,114,190–192). Dans d'autres travaux, une technique d'acquisition discontinue est utilisée : le sujet est au repos au cours de la session mais on lui demande de signaler toutes les 30 ou 60 secondes (périodes définies aléatoirement) les symptômes ressentis au cours des dernières secondes (méthode dite de « *random sampling analysis* ») (105,193). Enfin, une procédure combinant analyse « *data-driven* » (*e.g.* analyse en composante indépendante) et entretien post-IRMf immédiat a

également été proposée (116). Cette dernière procédure, qui permet une labellisation des différents temps de la session d'IRMf en périodes avec hallucinations (dites périodes ON) et périodes sans hallucinations (dites périodes OFF) a récemment fait l'objet d'un travail de réplication et de validation qui sera présenté dans une première partie de la thèse (voir **Article 1**). Dans un deuxième temps, il sera question d'automatiser cette procédure longue et fastidieuse au moyen de stratégies de *machine-learning* (voir **Article 2**, **Article 3**, **Article 4**).

2.1 La capture « *offline* » de l'hallucination auditive

Comme nous venons de l'aborder, plusieurs techniques ont pu être utilisées dans les études de capture pour déterminer les périodes hallucinatoires chez les patients souffrant d'hallucinations auditives et présentant ce type de symptômes au cours d'une session d'IRMf. Notamment, la plupart de ces études sont basées sur une signalisation par le patient lui-même de la survenue des symptômes. Toutefois, un certain nombre de limites sont à signaler concernant cette méthodologie : impact de l'activité motrice (appui sur un bouton au cours de la session) sur l'acquisition du signal (194), reconnaissance parfois perturbée des symptômes chez les patients (195), non détection des patterns précoces d'entrée dans l'hallucination (133).

Dans ce contexte, l'**Article 1** propose la validation d'une méthode en deux étapes (dont la **Figure 9** propose une synthèse) basée sur une analyse en composante indépendante permettant de séparer les régions cérébrales co-activées pendant l'hallucination auditive sans modèle temporel prédéfini (196). Cette analyse est combinée à l'utilisation d'un questionnaire permettant de recueillir *a posteriori* les caractéristiques de la symptomatologie hallucinatoire présentée par le sujet durant la session d'enregistrement. Plusieurs taches permettant de s'assurer de la reproductibilité et de la fiabilité de cette « méthode en deux étapes », initialement développée dans des travaux précédents de notre équipe (197–199), sont présentées.

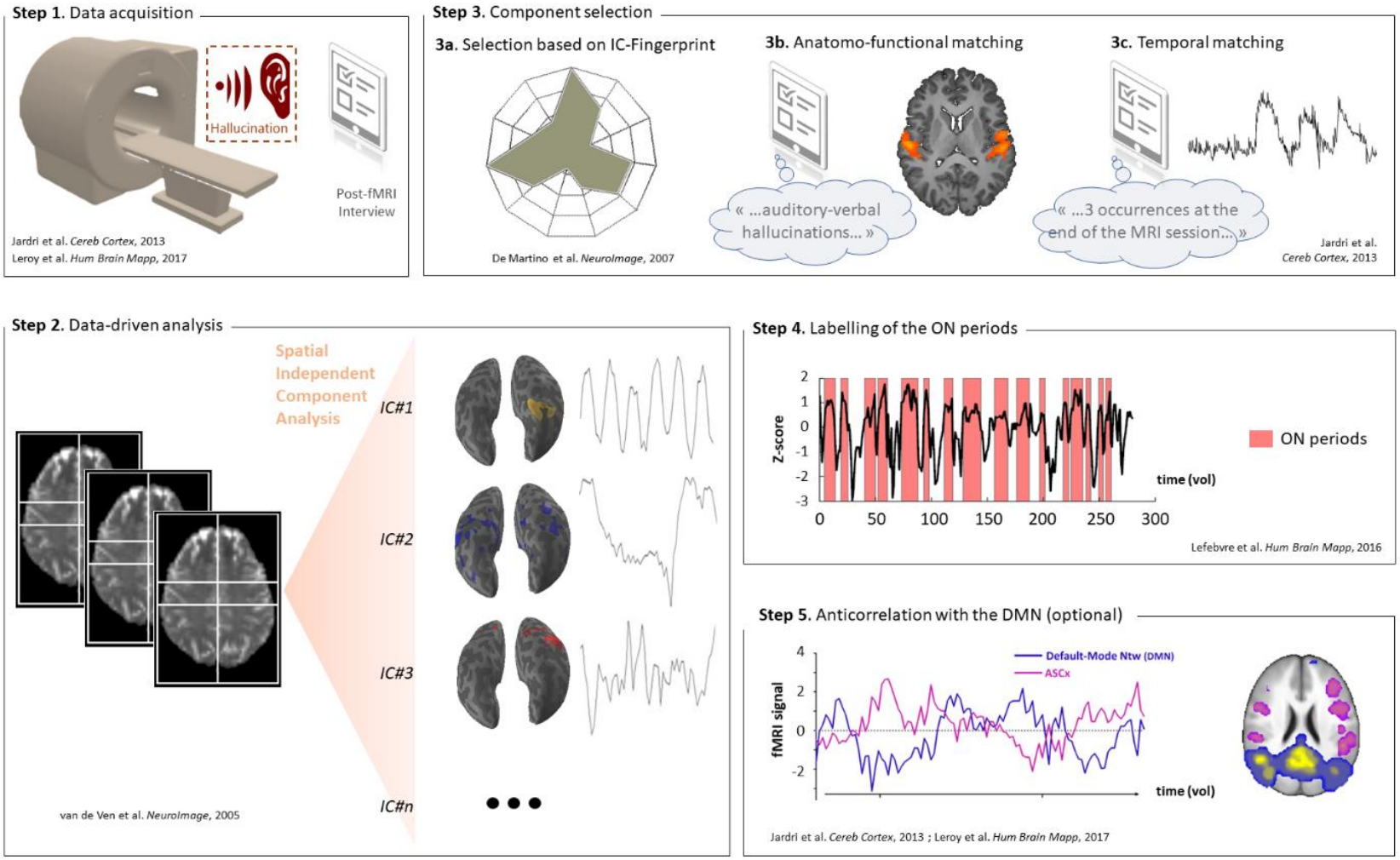


Figure 9. Présentation de la méthode permettant de labelliser les volumes d'une session IRMf de repos en périodes hallucinatoires ou non-hallucinatoires. Adapté de la méthode décrite dans (117).

ARTICLE 1

fMRI capture of auditory hallucinations: validation of the two-steps method

(Abbreviated title: fMRI capture of hallucinatory experiences)

Arnaud LEROY[1,2], Jack R. FOUCHER[3,4], Delphine PINS[1,2], Christine DELMAIRE[5],

Pierre THOMAS[1,2], Mathilde M. ROSER[4], Stéphanie LEFEBVRE[1,2], Ali AMAD[1,2]

Thomas FOVET[1,2], Nemat JAAFARI[6,7], Renaud JARDRI[1,2]

1. Univ. Lille, CNRS, UMR 9193 - SCALab - Sciences Cognitives et Sciences Affectives, F-59000 Lille, France

2. CHU Lille, Clinique de Psychiatrie, CURE, F-59000 Lille, France

3. Univ. Strasbourg, CNRS, UMR 7357 - ICube - Laboratoire des Sciences de l'Ingénieur, de l'Informatique et de l'Imagerie & Fédération de Médecine Translationnelle de Strasbourg (FMTS), F-67000 Strasbourg, France

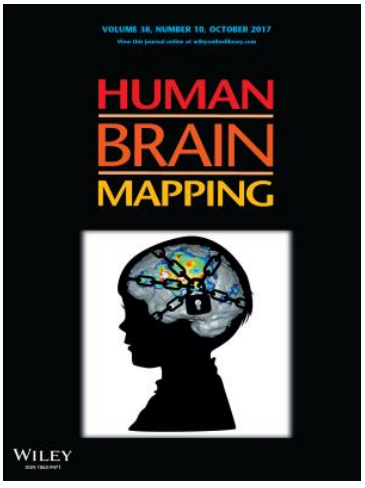
4. CHU Strasbourg, CEntre de neuroModulation Non Invasive de Strasbourg (CEMNIS), F-67000 Strasbourg, France

5. CHU Lille, Pôle Imagerie, Neuroradiologie, F-59000 Lille, France

6. Centre Hospitalier Henri Laborit, Unité de recherche clinique intersectorielle en psychiatrie à vocation régionale Pierre Deniker, F-86022 Poitiers, France

7. Univ. Poitiers & CHU Poitiers, INSERM, CIC-P 1402 & U-1084 Experimental and Clinical Neurosciences Laboratory, F-86022 Poitiers, France

Human Brain Mapping. 2017 ; 38(10) : 4966-4979.



Abstract

Our purpose was to validate a reliable method to capture brain activity concomitant with hallucinatory events, which constitute frequent and disabling experiences in schizophrenia. Capturing hallucinations using fMRI remains very challenging. We previously developed a method based on a two-steps strategy including (i) multivariate data-driven analysis of per-hallucinatory fMRI recording and (ii) selection of the components of interest based on a post-fMRI interview. However, two tests still need to be conducted to rule out critical pitfalls of conventional fMRI capture methods before this two-steps strategy can be adopted in hallucination research: replication of these findings on an independent sample and assessment of the reliability of the hallucination-related patterns at the subject level. To do so, we recruited a sample of 45 schizophrenia patients suffering from frequent hallucinations, 20 schizophrenia patients without hallucinations and 20 matched healthy volunteers; all participants underwent four different experiments. The main findings are (i) high accuracy in reporting unexpected sensory stimuli in an MRI setting; (ii) good detection concordance between hypothesis-driven and data-driven analysis methods (as used in the two-steps strategy) when controlled unexpected sensory stimuli are presented; (iii) good agreement of the two-steps method with the online button-press approach to capture hallucinatory events; (iv) high spatial consistency of hallucinatory-related networks detected using the two-steps method on two independent samples. By validating the two-steps method, we advance toward the possible transfer of such technology to new image-based therapies for hallucinations.

Key words: Independent Component Analysis; Interview; Schizophrenia; Hallucinations; fMRI; Reproducibility; Reliability.

Introduction

Validating reliable methods to explore the neural bases of consciousness is a crucial aim in neuroscience. This question has a strong impact on our attempts to correlate brain activation with a given behavioral experience. Here, we would like to illustrate how recent functional magnetic resonance imaging (fMRI) developments allow objective “capture” of the neural correlates of unpredictable and subjective mental events, such as hallucinations. Hallucinations are percepts in the absence of external stimuli (Ey, 1973). In schizophrenia, hallucinations are frequent and may cause long-term disability (Hor and Taylor, 2010). In adults, auditory-verbal hallucinations (AVHs) are most frequent (Andreasen and Flaum, 1991), although hallucinations may occur across every sensory modality (David et al., 2011; Llorca et al., 2016). Anatomical and functional disturbances in both primary and association sensory cortices have been proposed to account for AVHs (Allen et al., 2008; Jardri et al., 2011), but the detection of their occurrence while scanning a participant (hallucination capture methods) has long remained very challenging.

In a first subset of capture studies, AVH occurrences were signaled online by asking the participant to press a response button in the MRI scanner (Silbersweig et al., 1995; Dierks et al., 1999; Lennox et al., 2000; Sommer et al., 2008). The subsequent sequence of self-reports serves as a model for brain activity. Despite the cleverness of this method (later called the “button-press” method), several drawbacks were noted. First, the cerebral activations linked to motor readiness were shown to disturb the acquisition of resting state signals (Bazán et al., 2015). Second, the reliability of this method was questioned due to the poor insight and executive dysfunctions that may exist in patients with schizophrenia (Tan, 2009). Finally, activity related to AVHs may precede the button press (Diederen et al., 2010) and exhibit complex dynamics (Lefebvre et al., 2016).

A second line of capture studies utilized discontinuous acquisition methods (also called the “random-sampling” approach), in which many fMRI volumes were acquired at random intervals. Patients were asked for their sensory experiences immediately after each stop (Shergill et al., 2000, 2001). These two strategies (i.e., “button-press” and “random-sampling”) both relied on hypothesis-driven fMRI data analyses in that they were based on patient self-report of AVHs during scanning. This drawback made these approaches particularly vulnerable to a drop in performance in signaling hallucination occurrences.

A third line of studies used more data-driven approaches, such as spatial *independent component analysis* (ICA). Applied to fMRI, this statistical method allows the co-activated brain regions to be separated without a pre-defined temporal model of brain activity (Formisano et al., 2004). Even though the first studies combined ICA with online self-reports (van de Ven et al., 2005; Jardri et al., 2009), this method mainly paved the way to more simple designs for hallucinating patients, since they were only asked to report AVHs after acquisition, using a post-fMRI interview (Jardri et al., 2007, 2009). Data from this interview was also used to help select the most relevant components among those blindly generated by ICA, i.e., spatial functional patterns that best matched the hallucinations' time of occurrence and phenomenology. We named this approach the *two-steps method for hallucination fMRI capture* (the *"2S"* method), for which a proof-of-concept study has been published (Jardri et al., 2013).

Although promising, two tests still need to be conducted to rule out critical pitfalls of conventional fMRI capture methods before the *"2S"* strategy can be adopted in hallucinations research: (a) replication of these findings on an independent sample (reproducibility); and (b) assessment of the consistency of the AVH-related patterns at the subject level (reliability). In this paper, we addressed these issues by recruiting 85 participants in four different experiments. We successively studied the patients' ability to a posteriori report their sensory experiences (i), the concordance between the *"2S"* and the "button-press" methods on controlled stimuli (ii) and on hallucinations (iii), and finally, the consistency of the AVH-related neural networks identified using the *"2S"* procedure on independent samples (iv).

Materials and Methods

Population

We recruited 5 independent samples of participants who were free from any sensory deficit: 20 schizophrenia patients without hallucinations, 20 healthy subjects, and 3 samples of 5, 20 and 20 schizophrenia patients suffering from frequent AVHs. Patients were assigned to the "no-hallucination" group if they had not experienced hallucinations in the week prior to participation (task 1). They were assigned to the "AVH" group if the PANSS P3 item score was ≥ 3, with hallucination experiences frequent enough to occur during an MRI session (tasks 3 & 4). Please note that in task 3, five schizophrenia patients were selected for their good self-report of hallucinatory events (a necessary criterion for using the "button-press" approach) and that in task 4, two different subsets of twenty patients each were recruited to control for the possible influence of age and medication on replicability. The main characteristics of these samples are reported in **Table 1**. All of the patients enrolled in tasks 3 only had hallucinations in the auditory modality. For the patients in task 4, 88 % of these experiences occurred in the auditory modality, whereas 35, 12 and 12% were coenesthetic, visual and olfactory, respectively. All the participants were recruited at the University Hospital of Lille, except for those participating in task 3, which was performed at the University Hospital of Strasbourg.

Task	***1***	***2****	***3***	***4****	***4***
Sample	#1	#2	#3	#4	#5
Number of subjects	20	20	5	20	20
Population	Schizophrenia without hallucinations	Healthy subjects	Schizophrenia with hallucinations	First episode psychosis with hallucinations	Schizophrenia with hallucinations
Sex Ratio	17/3	15/5	3/2	17/3	14/6
Age (years)	39,5 +/- 10	12,9+/-1,6	34,4 +/-9,3	13,1+/-1,8	33,7+/-8,2
Dose of antipsychotic treatment (EqOZ)	20,9 +/- 12,7	0	21,2 +/-10,8	0	36,2+/- 17,3
PANSS-P	13,9+/-4,2	NA	22,8 +/-4,2	29,4+/-5,3	22+/- 4,4
PANSS-P3	1+/-0	1 +/- 0	5,2 +/- 0,4	5,1+/-1,3	5,3+/-0,9
Type of acquisition	MR-simulator	Single shot EPI	Single shot EPI	Single shot EPI	3D-PRESTO
Acquisition time	10 minutes	10 minutes	20 minutes	10 minutes	10 minutes
Sequence parameters					
Echo time (msec)	NA	70	43	70	30
Repetition time (msec)	NA	3000	3000	3000	1000
Voxel size (mm^3)	NA	4	4	4	3.3
Number of scans	NA	300	400	300	900
Acquisition per subject	1	1	4	1	1

Characteristics of the enrolled samples (mean ± sd). EqOZ: *equivalent olanzapine*; PANSS: *positive and negative syndrome scale*; PANSS-P: positive sub-score of the PANSS scale; PANSS-P3: P3 sub-score of the PANSS scale; NA: not available; EPI: *echo-planar imaging*; 3D-PRESTO: *PRinciples of Echo-Shifting with a Train of Observations*; *: Data from Jardri et al., 2013; There was no overlap between the two samples recruited in task 4.

Experimental Procedures

Task 1

Task 1 was designed to determine if schizophrenia patients could a posteriori report, with good precision, sensory experiences that occurred in a controlled experimental setting (i.e., using real auditory stimuli, with known characteristics in terms of time of onset, duration, amplitude, etc.). Task 1 was performed in an MRI simulator. We selected patients without AVHs for this first experiment to avoid any confusion in reporting task-related auditory stimuli vs. endogenous percepts (i.e., AVHs). Patients were asked to lie down at rest without falling asleep and were put in a dark environment. They were only asked to report auditory stimuli a posteriori. For 10 minutes, the sound of an EPI sequence was delivered without real MRI scan acquisition. In complement, a variable number of unexpected auditory stimuli were randomly presented through the headphones using E-Prime 1.3 (*Psychology Software Tools Inc., Pittsburgh, USA*) (normalized amplitude = 75 dB SPL). We used verbal material and selected 0 to 4 voices/participant (male voices, all unknown to the participants), as this is the mean number of AVHs usually reported during an fMRI session (Lefebvre et al., 2016). Stimulus presentations lasted from 6 to 30 seconds. Patients were interviewed immediately after the experiment about what they heard using a post-fMRI questionnaire (see the *Analysis* section). The number of voices heard and the moments of occurrence were reported. Voice detection performance was also measured.

Task 2

Task 2 was designed to evaluate the inter-method reliability of the *"2S"* method compared with detection of controlled stimuli using hypothesis-driven analysis. Task 2 was performed in an MRI scanner. Healthy participants were asked to lie down at rest without falling asleep while wearing MR-compatible headphones that transmitted audible stimuli and attenuated the ambient noise of the scanner. They were only asked to report auditory stimuli a posteriori. During the 10-minute fMRI session, a variable number (n) of words or sentences were presented through the headphones using E-Prime 1.3 (*Psychology Software Tools Inc., Pittsburgh, USA*) (normalized amplitude = 75 dB SPL). Compared to task 1, and because the purpose was no longer to test the quality of reporting of the patients, we chose to enhance power by increasing the total number of stimuli presented from [0-4] to [0-10]. Stimulus presentations lasted from 6 to 30 seconds. Patients were interviewed immediately after the experiment about what they heard, using a post-fMRI questionnaire (see the *Analysis* section). The number of heard stimuli and their moments of occurrence were reported. We then compared the *"2S"* method with a general linear model built using the exact time points of stimulus presentation.

Task 3

Task 3 was designed to evaluate the agreement between capture methods (i.e., between the *"2S"* and the “button-press” methods) in patients with a good self-report of their hallucinatory events. Task 3 was performed in an MRI scanner. The patients were asked to lie down at rest without falling asleep during acquisition. Each patient completed four different 20-minute fMRI sessions. During the first 3 sessions, the patients were instructed to signal the onset of their hallucinations with a response button (right hand) and to release it when the hallucinations stopped, i.e., they were explicitly asked to report hallucinations online, which referred to as the "button-press" condition. In the last session, the "2S" procedure was applied, and the patients were interviewed immediately after this last acquisition about what they heard, using a post-fMRI questionnaire (see the *Analysis* section). The number of AVHs and their moments of occurrence were reported.

Task 4

Task 4 was designed to test the reproducibility of the "2S" procedure. Task 4 was performed in an MRI scanner. Patients with AVH were asked to lie down at rest without falling asleep during acquisition. They were only asked to report hallucinations a posteriori. Each patient had a 10-minute fMRI session, and the “2S” procedure was applied to identify AVH periods during scanning. Two complementary analyses were conducted. First, we computed the spatial similarity between the AVH-related functional brain networks obtained at the subject level. Second, the between-sample consistency in hallucination detection between the current dataset and a previous independent sample (Jardri et al., 2013) as well as with coordinate-based meta-analytic findings from 10 different studies (Jardri et al., 2011) (Cf. **Table 1**) was evaluated for the hallucination-related network (association sensory cortices, ASC) (Jardri et al., 2013) and the default mode network (DMN), which is considered a standard, well-replicated and ubiquitous neural network.

Analyses

A posteriori voice detection performance

This analysis used the data collected in task 1. To normalize performance across subjects, sensitivity was recorded as 1 if all the voices were detected, and specificity was recorded as 1 if there were no additional recognized sounds. In all other cases, sensitivity and specificity were recorded as 0. We further generated random data for 20 mock participants and matched these data with those of the patients according to the number of voices presented. We generated random detection values using the RAND function (in *Matlab R2016a*). Each simulated recording was randomly divided into periods with and without voices, and a number of 0 or 1 was randomly assigned to each. Then, as for the patients, if the number was 1 for all of the periods with voices, a sensitivity of 1 was reported. If the number was 0 for all of the periods without voices, the specificity was 1. Accuracy was defined as (true detection + true no detection) / (true detection + false detection + true no detection + false no detection). The patient and simulated data were compared using a permutation test with an α level of 0.05 using R software for statistical computing v3.3. The patient and simulated data were compared using a two-sample

permutation test with 1000 iterations (Monte Carlo method) and an α level of 0.05, and these analyses were performed using the 'perm' package with on R software for statistical computing v3.3.

The two-steps hallucination fMRI capture procedure

This analysis was conducted on data collected in tasks 2, 3 and 4. Our capture method is divided into two consecutive steps (Jardri et al., 2013) (cf. Figure 1a, **Supplementary figure 1**). This method was developed to capture unpredictable events, such as hallucinations and unexpected stimuli presented to healthy participants. Step 1 is resting-state fMRI acquisition in participants with or without AVHs. Step 2 occurs immediately after MRI acquisition. Using a standardized post-fMRI interview, each participant is asked to report all the sensory experiences that occurred during scanning, including the sensory modality and number of events as well as their approximate times of occurrence (a translated version of the interview is available by request to the corresponding author).

Conventional preprocessing steps were conducted on anatomical and functional data (as detailed in our previous publications, (Jardri et al., 2013; Lefebvre et al., 2016). The functional data were preprocessed using a slice scan time correction, a 3D head motion correction, smoothing using a spatial Gaussian filter (full-width at half-maximum [FWHM] = 6.0 mm), a temporal high-pass filtering with 2 sin/cos, and linear trend removal. The anatomical data were subjected to an intensity inhomogeneity correction algorithm, resampled to a 0.5mm3 resolution, and normalized in Talairach's stereotactic space (Talairach J, Tournoux P, 1988). Data from the head tissue, subcortical structures, and cerebellum were then removed with the aim of advanced cortical segmentation processing. This segmentation was performed at the gray/white matter and the gray matter/cerebrospinal fluid boundaries. A boundary-based registration was finally used to align the functional/anatomical datasets.

Data obtained from step 1 are first blindly analyzed using cortex-based ICA analysis (Formisano et al., 2004). For each patient, cb-ICA (using the spatial decomposition algorithm "FastICA" (Hyvärinen and Oja, 2000)) is used to extract (20% of the total volume) independent components (ICs) from the rs-fMRI signal of the cortical voxels of the matrix, i.e., 30 ICs for task 2 and task 4 and 40 ICs for task 3. We referred to a fixed-point ICA algorithm, i.e., FastICA, which minimizes the mutual information of the components using a robust approximation of the negentropy as a contrast function, and a rapid, iterative (nonadaptive) algorithm for its maximization. The deflation approach was used to run FastICA, as previously described by Hyvärinen et al. (1999) and Formisano et al. (2004). The resulting ICs corresponded to 3D clusters of voxels with |Z|-normalized values greater than 2.5. Among these ICs, the most relevant are first selected using the IC-fingerprint method (De Martino et al., 2007; Jardri et al., 2013). Because ICA does not naturally order the resulting components according to their relevance, we referred to the IC-fingerprint method, which jointly uses 7 spatial and temporal signal properties for IC classification purposes (de Martino, 2007). These properties were measured post hoc for each IC to preserve the "data-driven" characteristic of the analysis. This step allowed us to discard noise-related ICs (e.g., EPI susceptibility, motion artefacts, high-frequency noise...), with the aim to only retaining the components related to a neurophysiological source, which were characterized by a high spatial and temporal structure (i.e., degree of clustering and one-lag serial auto-correlation, respectively) and by a

high entropy, coupled with a maximum power contribution in the low-frequency range (0.01 Hz - 0.1 Hz; see also Roquet et al. (2014)). This allows one to retain only the components related to a neurophysiological source (BOLD) for the next step. The surviving ICs are then compared to the post-fMRI interview data, in terms of the number, times of occurrence, and functional networks of interest (e.g., speech-related for AVHs, etc.). Data preprocessing, cortex-based ICA and IC-fingerprinting were performed using *Brain Voyager* v20.2.

Inter-method reliability in fMRI stimulus detection

This analysis used the data collected in task 2. Two parallel analyses were applied to the fMRI data in *Brain Voyager*: 1) the *"2S"* analysis, as described in the previous section and based on cb-ICA; 2) a general linear model (GLM) fitted to the experimental protocol generated for each participant using *E-Prime*. This GLM was based on controlled stimulus timing, as the time of stimulus presentation was known *a priori*. Because of the massive univariate nature of GLM analysis applied to fMRI data, the resulting statistical maps were thresholded using a false discovery rate approach ($q < 0.01$ (Genovese et al., 2002)). In addition to conventional correlation analysis, which we considered insufficient to confirm agreement of the results of the two analyses for the same dataset, we performed Deming regression (Cornbleet and Gochman, 1979) to account for observation errors on both the x- and y-axes (i.e., on the BOLD dynamics from GLM and cb-ICA, respectively).

Inter-method reliability in fMRI AVH capture

This analysis was based on data collected in task 3 and used *Matlab R2012b* with the *SPM8, statistical non-parametric mapping (SnPM) and FMRLab v2.3* toolboxes. For the "button-press" condition, we referred to the GLM approach described in the previous section. The brain activity expected to be related to AVHs was modeled by convolving the box-car time course of the button-press from the participants with the canonical hemodynamic response function (HRF), i.e., a two-gamma function using SPM standard parameters. This procedure was used to determine the BOLD-related component with the highest correlation coefficient between its temporal vector and the subject's signaling. As we previously reported that "button-press" components in the same subjects were highly reproducible, they were averaged for each patient (Foucher, 2013). For the "*2S"* condition, we referred to the ICA approach described in the "two-steps hallucination fMRI capture procedure" section.

Although a high spatial correlation coefficient can be considered a measure of inter-method reliability, here, we used Cohen's kappa coefficient, κ, to assess whether this agreement remains true at the voxel level. The *"button-press" spatial components* and *"2S" spatial components* were successively thresholded at z = 1.5, 2, 2.6, 3 and 3.6 to make binary maps of 0 = [no-AVH voxel], 1 = [AVH voxels] to measure the κ coefficient. Last, possible systematic differences between the spatial *"button-press" components* and *"2S" components* were assessed using a multi-subject pseudo-paired t-test design with SnPM. A permutation test was adopted due to the limited number of subjects in this task. Significance was set at pseudo-t > 2 with an extension $k > 1$ cm^3 (125 voxels) within the regions of interest, which were defined as regions that were positively active in either the signaling or resting condition, i.e., "button-press" or *"2S"* component.

Spatial consistencies in hallucination detection

This analysis used the data collected in task 4 (samples #4 & #5). After a first-level analysis (based on the *"2S"* capture method) was conducted, a secondary analysis was conducted by submitting individual ICs to a self-organizing group IC algorithm (*sog-ICA*, Esposito et al., 2005). An iterative cluster-size thresholding procedure based on Monte Carlo simulations (n = 1000) further corrected the resulting random-effects statistical maps, which were used to evaluate the regional stability of these AVH-related neural networks.

The between-subjects' spatial consistency was first tested using multidimensional similarity clustering (MDS) on sample #5 (See **table 1**). The MDS algorithm was applied on the sog-ICA decomposition of per-hallucinatory fMRI data, and the MDS linear projections were plotted in 2-dimensional space (Torgerson, 1952) in *Brain Voyager 20.2*. To help identify cluster plots of interest, the random-effects sog-ICA validation maps were visualized using the same color codes.

Between-samples spatial consistency was also tested using probabilistic mapping between sample #4 (Jardri et al., 2013) and #5 (replication sample). Note that these independent samples were obtained from different scanners using different sequences (single-shot EPI and 3D-PRESTO, respectively) and different magnetic fields strength (1.5T and 3T, respectively) (Cf. **Table 1**). At each spatial location, functional maps were generated to represent the relative number of subjects leading to significant activation patterns within the networks of interest for the initial sample (#4 (Jardri et al., 2013), n = 20), the replication sample (#5, n = 20), and coordinate-based meta-analytic findings (Jardri et al., 2011).

Linear regression analysis between default mode and AVH-related signal time courses

This analysis used the data collected in task 4. The AVH-related ICs were selected using samples #4 (Jardri et al., 2013) and #5 (task 4) according to the *"2S"* procedure. In parallel, we used the same data sets and selected ICs related to the DMN using a "goodness-of-fit" (GoF) procedure. For each participant, the IC with the highest GoF score (i.e., absolute correlation coefficient with a DM template taken from Laird et al., 2009 (Laird et al., 2009)) was assumed to be the DM component. To explore the dynamics of the AVH-related and DM-related networks, we normalized their fMRI signals to relative variations with respect to the mean value of the participants' individual time series (Deco et al., 2009). AVH-related and DM-related networks signal fluctuations were compared using the Pearson product moment correlation in samples #4 (Jardri et al., 2013) and #5, respectively.

Ethical issues/ study ID

All patients gave written informed consent. The study ID for task #3 is CPP03/45-PSY 2003/52S, while that for tasks #1, 2, and 4 is 2009-A00842-55. All reported experiments performed by the authors complied with the Helsinki declaration and its amendments.

Results

Are schizophrenia patients able to report with precision sensory experiences a posteriori (task 1)?

We tested the ability of schizophrenia patients to report off-line unexpected sensory events and their times of occurrence in the scanning context. The accuracy of a posteriori voices labeling in schizophrenia patients was measured at 95% ($^{95\%}$IC = 85.3-99.9), while random detection for 0 to 4 events would be 20% ($^{95\%}$IC = 9.61-36.14). This difference was highly significant (permutation testing, mean difference = 77.5, p = 0.002; Cf. **Figure 1b**). Sensitivity was 100% ($^{95\%}$IC = 80.0-100) and specificity was 95% ($^{95\%}$IC = 0.73-99.7), while random detection for 0 to 4 events would be 20% ($^{95\%}$IC = 6.61-44.3) for both.

Inter-method reliability in detecting controlled auditory stimuli (task 2)

A Deming regression analysis was used to account for observation errors on both the x- and y-axes (r = 0.57, with a significant test for slope $F_{2,98} = 2.1$; $p < 0.0001$), and confirmed the high degree of precision of the *"2S"* method even in the absence of online report, as shown by the good agreement with the GLM analysis based on controlled stimuli (Cf. **Figure 1c**).

Inter-method reliability in detecting online AVHs (task 3)

The average spatial correlation coefficient between the *"2S"* and "button-press" components was $r = 0.68 \pm 0.1$ (Cf. **Figure 2a**). The average Cohen's kappa coefficient was 0.50 ± 0.08 and was relatively consistent regardless of the z-score threshold. **Figure 2b** shows the plot of each individual κ according to the z-score threshold. The SnPM comparison between the *"2S"* and "button-press" spontaneous activity maps did not provide any evidence of a significant difference despite the use of a lenient threshold (Cf. **Figure 2c**).

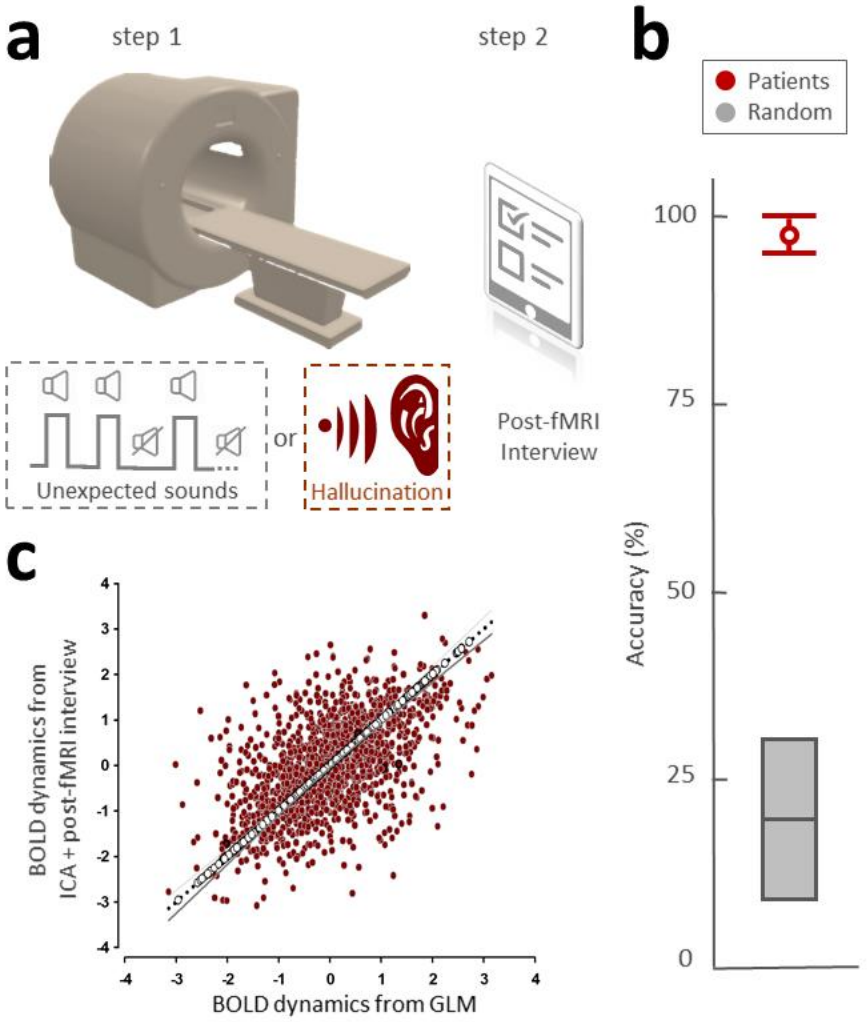


Figure 1. **The hallucination capture method**. ***a) Description of the two-steps procedure***. Step 1 occurs while participants are laying down in the MRI scanner. Two populations were tested with a variation on step 1: (i) healthy participants, who were exposed to unexpected voices during scanning (grey dotted square); and (ii) schizophrenia patients with frequent hallucinations, who were scanned without stimulus presentation because hallucinations constitute internally generated percepts (red dotted square). Step 2 occurs immediately after MRI acquisition. Using a standardized post-fMRI interview (see Methods), participants were asked to report sensory experiences that occurred during scanning as well as their precise time of occurrence. The collected data were then used to select the most appropriate components resulting from blind multivariate analysis of the fMRI signal (cortex-based independent component analysis or ICA). The results for healthy volunteers (n = 20) are presented in Figure 1c, while those for hallucinators (n = 20) are presented in Figures 3 and 4. ***b) Task 1: The ability of schizophrenia patients to report off-line the number of sensory events and their times of occurrence***. Twenty new patients without hallucinations were also tested using the "unexpected voices" procedure in an fMRI simulator. The accuracy of a posteriori sound labeling in the schizophrenia patients was plotted in red (mean 97.5%; $^{95\%}$IC = 85.3-99.9), while random detection for 0 to 4 events would be 20% ($^{95\%}$IC = 9.61-36.14, grey). Mean difference = 77.5, p = 0.002. ***c) Task 3: Validation of ICA + interview versus gold standard analysis***. Two parallel analyses (i.e., ICA + interview and general linear model analysis using the sound presentation protocol) were conducted on the same healthy volunteer dataset. Deming regression analysis was used to account for observation errors on both the x- and y-axes (r = 0.57, with a test for slope $F_{2,98} = 2.1$; p < 0.0001) and confirmed the high degree of precision of our capture method, even in the absence of online report.

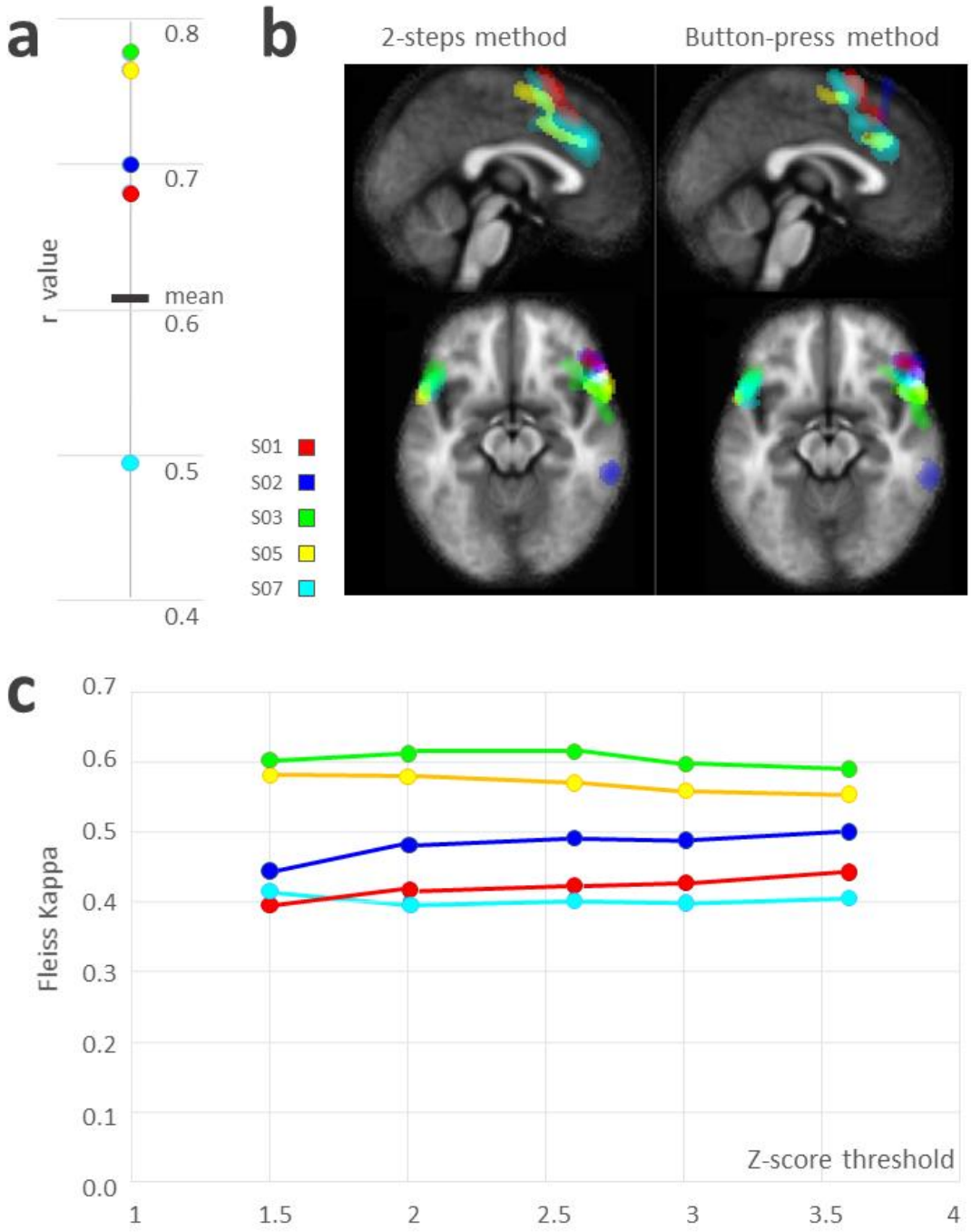


Figure 2. Task 2: Inter-method reliability in auditory-verbal hallucination (AVH)-related networks. Five schizophrenia patients with refractory hallucinations underwent four different 20-minute sessions of 400 single-shot EPI fMRI. During the 3 first sessions, they were instructed to signal the onset of an AVH with a response-button (right hand) and to release it when the AVH stopped. In the last session, the patients were instructed to lie down with their eyes closed without falling asleep. At the end of this session, they completed a post-fMRI interview to precisely report the times AVHs occurred during the scan. Components of interest were detected using the "button-press" method for the 3 first sessions and then averaged, while they were detected using the "*two-steps" (2S)* method for the last session. Each color represents one of the five patients. **a)** Correlation between the "*2S"* and the "button-press" methods for each participant; **b)** Components of interest chosen during AVH experiences for each participant using the *"2S"* and the "button-press" methods. **c)** Fleiss's kappa value, i.e., intersession concordance according to different statistical thresholds for SPM analysis.

Reproducibility of neural networks identified during AVHs (task 4)

After MDS projection, four main clusters were identified; these clusters represented the sensorimotor network (cluster 1), the AVH-related network (cluster 2), the salience network (cluster 3), and the visual rest network (cluster 4) (Cf. **Figure 3a**). Random-effects activation maps resulting from sog-ICA are presented in a glass brain (Cf. **Figure 3b**). The AVH-related network encompasses widespread cortical-subcortical areas, as listed in **Table 2**.

Identified Clusters	**Talairach and Tournoux Coordinates (x,y,z)**	**Number of voxels**
Right Cerebrum, Sub-lobar, Insula	41, 8, 5	10850
Left Cerebellum, Anterior Lobe, Culmen	-2, -59, -9	7104
Left Cerebrum, Sub-lobar, Insula,	-47, 8, 4	7069
Left Cerebrum, Parietal Lobe, Inferior Parietal Lobule	-55, -30, 36	6078
Right Cerebrum, Limbic Lobe, Cingulate Gyrus,	2, 23, 31	4152
Right Cerebrum, Frontal Lobe, Sub-Gyral	21, -7, 57	3079
Left Cerebrum, Frontal Lobe, Middle Frontal Gyrus	-28, 47, 22	1763
Left Cerebrum, Occipital Lobe, Lingual Gyrus	-19, -85, -4	1335
Right Cerebrum, Occipital Lobe, Middle Occipital Gyrus	34, -85, 9	1253
Right Cerebellum, Anterior Lobe, Culmen	46, -37, -28	1021
Right Cerebrum, Limbic Lobe, Uncus	27, 10, -25	943
Left Cerebrum, Frontal Lobe, Inferior Frontal Gyrus	-16, 21, -16	923
Left Cerebrum, Occipital Lobe, Cuneus	-16, -91, 14	740
Right Cerebrum, Frontal Lobe, Sub-Gyral	18, 28, -14	700
Left Cerebellum, Posterior Lobe, Cerebellar Tonsi	-36, -55, -39	653
Left Cerebrum, Occipital Lobe, Middle Occipital Gyrus	-36, -72, -7	574
Right Cerebellum, Posterior Lobe, Inferior Semi-Lunar Lobule	7, -76, -39	540
Right Cerebrum, Temporal Lobe, Transverse Temporal Gyrus	54, -18, 11	522
Left Cerebrum, Limbic Lobe, Uncus	-18, 5, -22	406

Regions involved in the hallucination-related network after the group-ICA decomposition of per-hallucinatory fMRI data. Data indicate x-y-z coordinates in stereotaxic space (TAL) of the weighted center for each identified cluster as well as the total number of voxels.

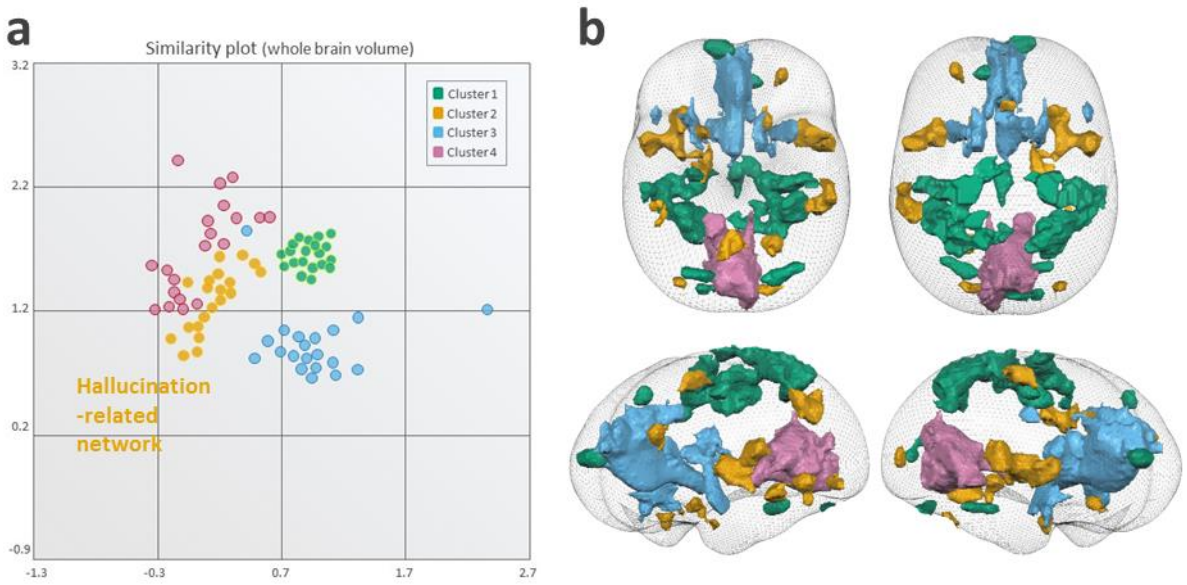


Figure 3. Task 4: Between-subjects' spatial consistency in auditory-verbal hallucination (AVH) detection (n = 20). **a)** *Cluster plots identified after multi-dimensional similarity clustering projection.* Each circle represents an individual IC taken from the 20 enrolled schizophrenia patients who experienced AVHs while scanning. Four clusters were identified and represented the sensorimotor network (cluster 1, green), the AVH-related network (cluster 2, orange), the salience network (cluster 3, blue), and the visual rest network (cluster 4, purple), **b)** *Random-effects activation maps* resulting from self-organizing group ICA presented in a glass brain, with colors assigned according to the cluster plot (shown in a). The AVH-related network is plotted in orange and encompasses the precentral gyrus, culmen, insula, inferior parietal lobule, cingulate gyrus, middle frontal gyrus, superior frontal gyrus, middle occipital gyrus, inferior frontal gyrus, cerebellar tonsil, fusiform gyrus, inferior semi-lunar lobule, transverse temporal gyrus and limbic lobe.

In a second step, we overlaid the results of the replication sample (2016) with those of the 2013 sample and with coordinate-based meta-analytic findings. At the group level, a negative correlation was identified between the BOLD signal of the AVH-related and DM-related networks, in both the 2013 sample ($r^2 = 0.38$, $p < 0.0001$; taken from Jardri et al. (2013)) and the current 2016 replication sample ($r^2 = 0.39$, $p < 0.0001$; Cf. **Figure 4a**). The spatial consistencies in the AVH-related and DM functional networks across these two independent samples of hallucinators (2013 and 2016) and with coordinate-based meta-analytic findings were also computed (Cf. **Figure 4b-c**). Important overlap was evident within the ASC and the DMN network.

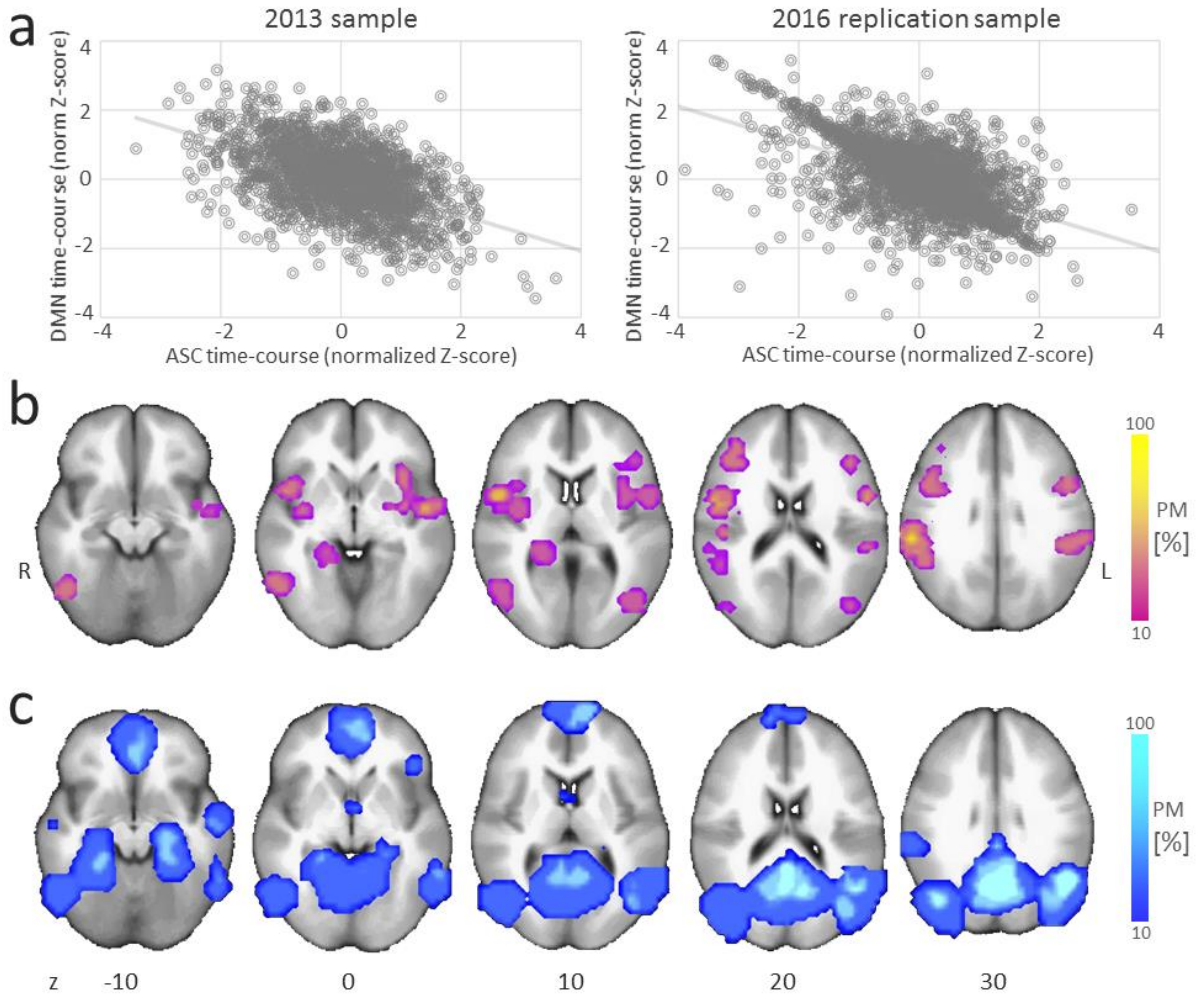


Figure 4. Task 4: Between-sample consistency in auditory-verbal hallucination (AVH) detection. a) At the group-level, a negative correlation was observed between the AVH-related network (within association cortices, ASC) and the default mode network (DMN) BOLD fluctuations, in both sample #4 (taken from Jardri et al., 2013) and the current 2016 replication sample (#5). **b, c)** Spatial consistency in AVH-related and DM-related functional networks across the two independent samples of hallucinators (#4 and #5). At each spatial location, functional maps represent the relative number of subjects leading to significant activation patterns within the networks of interest for the initial sample (#4, 2013, n = 20) and the replication sample (#5, 2016, n = 20) as well as for coordinate-based meta-analytic findings for hallucination capture **(b)** and for the anti-correlated DMN **(c)**. PM: probabilistic mapping; R/L indicate the right/left hemispheres.

Discussion

A major drawback in fMRI capture methods today remains the absence of a gold standard in detecting hallucinations during scanning. Because hallucinations are complex sensory experiences (David et al., 2011) that are often associated with negative affective states, reporting these symptoms online quickly becomes very challenging in the context of an MRI examination, especially for the most disabled patients. The subjects included in such studies are indeed specifically selected for their ability to report their symptoms online. Thus, to extend AVH capture to the field of clinical applications, developing a method applicable whatever the age and AVH severity appears critical.

Despite its limitations, fMRI capture of hallucinations based on online self-reports have received some validations in the literature (e.g., Sommer, 2008). Using a button-press approach, Diederen et al. (2013) notably confirmed the good reproducibility in brain activations obtained through fMRI capture of AVHs after two scans. Using a meta-analytical approach, the same research group also compared the brain activity measured during auditory stimulus detection with the activity concomitant to AVH (van Lutterveld, 2013). These authors were able to disentangle specific activation related to AVH from the spatial patterns associated with button-press signaling. Interestingly, previous works emphasized the pertinence of ICA-based approaches and their compatibility with button-press methods. In a study that combined ICA with online self-reports, van de Ven et al. (2005) demonstrated that a positive correlation exists between the average BOLD time-course obtained from the positive voxels of the component of interest and the button press reference model. Using a similar approach, Foucher (2013) showed the superiority of ICA over GLM for the analysis of the "button-press" method of hallucination capture. These encouraging findings pave the way for the assessment of inter-method reliability, external consistency and quality of sensory experiences reported by patients with hallucinations, as reported here.

Our purpose was thus to validate the *"2S"* method for fMRI hallucination capture, as initially introduced in a previous paper from our group (Jardri et al., 2013). The use of a post-fMRI interview proved capable of detecting a large range of modality-dependent experiences, without needing to put the participant in a dual-task situation (i.e., experiencing vivid hallucinations and at the same time pressing a response button). Several lines of support for the *"2S"* approach emerged from the present experiments. In a behavioral task, we first showed that schizophrenia patients were able to report controlled unexpected auditory stimuli with high accuracy in an fMRI environment. Using the same task while scanning healthy participants, we also demonstrated good concordance between a model-based analysis and the *"2S"* approach, which combined blind fMRI analysis with a post-fMRI interview. In a third experiment, we confirmed good agreement between the *"2S"* and online button-press approaches to capturing AVHs. Finally, the neural networks (e.g., AVH-related and DM-related networks) detected using the *"2S"* strategy in two independent samples were found to be highly comparable, supporting the good reproducibility of this method.

We showed in task 1 that despite the presence of an invalidating disorder, patients were fully able to *a posteriori* report the occurrence of unexpected voices presented during an MR simulation session. The reliability of the patients' report was very high despite very restrictive statistical analysis (if

a patient did not recognize one voice out of all of the voices presented, she/he was considered “not able to report”). This result constitutes the first level of validation for the post-fMRI interview in a population of schizophrenia patients. In the second task, we evaluated the reliability of the *“2S”* method compared with GLM analysis of controlled stimuli in healthy subjects and confirmed the high degree of precision of the *“2S”* method in a real fMRI setting, even without online report.

Based on the analysis of repeated scans in patients suffering from hallucinations, we further evidenced the stability of data obtained using the *“2S”* procedure and the conventional button-press approach in patients who were able to signal AVHs online. To date, the “button press” method is the most common accepted method, but it has important limitations, as previously listed (mainly due to motor readiness, executive dysfunction in schizophrenia, and the complex neural dynamics of AVHs). Consequently, we could only include 5 patients with good insight who were able to report their sensory experiences online (task 3). In contrast, we expect the *“2S”* method to be applicable to all patients with schizophrenia (we were able to recruit larger samples for tasks 2 and 4 for instance). The simplicity of the experimental setting of the *“2S”* method also constitutes an advantage over other capture methods, especially for patients who could have difficulties reporting hallucinations online, such as older participants or children (see, for example, Jardri et al. (2007)). Overall, tasks 1-3 confirmed the feasibility and reliability of the *“2S”* method despite the use of a post-fMRI interview. These results are important since a key strength, but also a potential limit of the *“2S”* method, specifically resides in the a posteriori nature of our interview. This question constitutes a hot but still unresolved topic in consciousness research.

Indeed, two types of methods have been proposed in experiments that test conscious access: (i) *report-based* paradigms and (ii) *no-report* paradigms (Tsuchiya et al., 2015), such as those based on eye-tracking methods. Crucially, no-re*port* paradigms could overestimate the occurrence of AVH-linked neural activation by including activation that occurs just before or after the activation directly related to hallucinations. These activations could be linked to post-perceptual processes (i.e., cognitive processes) or pre-perceptual processes (i.e., pre-neural correlates of AVHs) (Overgaard and Fazekas, 2016). Moreover, the occurrence of AVHs remains strongly subjective and patient dependent, even though we were able to demonstrate good reproducibility in the current study. Currently, we have no reason to prefer subjective variation linked to the operator in *no-report* paradigms to the individual variation observed in *report-based* paradigms. Furthermore, *report-based* paradigms could underestimate AVH occurrences because AVHs are linked to cognitive processes such as attention, working memory, decision making, and action planning. For example, reduced reporting was observed in the context of inattentional amnesia or experience without access (Tsuchiya et al., 2015). Here, our goal was to validate a method for reporting conscious experiences with good reliability and to correlate them with neural activations. By combining the use of a *report-based paradigm* (i.e., the interview) and a *no-report* paradigm (i.e., blind fMRI analysis), the *“2S”* fMRI capture method appears fully compatible with recent recommendations on conscious access paradigms to limit issues related to the unpredictable nature of the events of interest (Tsuchiya et al., 2016). We think it would be interesting in the near future to test

whether the fMRI-based approach described in this paper could be extended to other spontaneous phasic mental events, such as obsessions and tics.

We also studied the internal and external consistency of the results found using the *"2S"* procedure by testing the degree of overlap between (i) the cortical areas associated with AVH experiences as reported in the literature (Jardri et al., 2011) and (ii) the results obtained in two independent samples of hallucinators (in 2013 and 2016). The overlap was maximal within the association cortices (ASCs); these areas, including the insula and temporo-parietal junction, are known to play a core role in AVH experiences (Jardri et al., 2011). Interestingly, in these datasets, a similar degree of overlap was found for well-replicated intrinsic connectivity networks (Laird et al., 2009), such as the DMN. The negative correlation in BOLD fluctuations between the AVH-related and DM-related networks found in sample #4 (Jardri et al., 2013) was replicated in sample #5. By replicating previous findings, these results provide further support for the existence of anticorrelation between the DMN and sensory cortices during AVH experiences and of a central role of DMN dynamics in these phenomena (Alderson-Day et al., 2016; Lefebvre et al., 2016). This finding allows us to add the anticorrelation of DM-related/AVH-related time courses as a complementary selection criterion for the component of interest in the *"2S"* method (Lefebvre et al., 2016). These findings reinforce the consistency of the method as applied to fMRI capture of hallucinations. The overlap in the speech-related network and in the hippocampal complex was up to 90% and 65%, respectively, supporting the previously suggested core role of these areas in hallucinations (Allen et al., 2012; Amad et al., 2014). In contrast, other areas in this network may reflect the phenomenological content of the experiences, which is only shared by a minority of hallucinators (Ffytche et al., 1998; Jardri et al., 2013).

Importantly, the use of different samples of patients recruited from different centers as well as different scanners with various MR field strengths and fMRI sequences constitutes a strength of this paper. Although some of the patients came from the same center (CHU Lille), we avoided overlap between the samples involved in the different tasks. Heterogeneity in terms of age or symptom severity between the tasks further supports the reliability of the *"2S"* method in various populations, including patients who could have poor reporting ability, such as adolescents experiencing acute psychosis or adults suffering from severe chronic schizophrenia. This approach is further strengthened by the reference to multivariate statistics, such as ICA, which enables better control of false-positive rates compared to conventional massive univariate approaches (GLM) (Eklund et al., 2016) while also providing access to effect-size estimates (i.e., fMRI changes during hallucinations at the component level), a criterion recently recommended for good practice in fMRI research (Chen et al., in press).

From a methodological point of view, ICA presents several advantages in the context of fMRI capture of hallucinations. A first one relies on the use of multivariate statistics. Interestingly, the performance of such algorithms appears to substantially benefit from dimensionality reduction (Formisano et al, 2004) compared with more massive univariate methods. Indeed, we made the choice to perform a cb-ICA based on the idea that only 20% of the voxels lie within the cortex. Readers should be aware that hallucinations might involve complex cortical-subcortical interactions (e.g., Hoffman et al,

2011). However, our choice to restrict analyses to the part of the matrix containing cortex stays justified in the context of target definition for neuromodulation tools. Again, the consideration that hallucinations may result from neural dysconnectivity (e.g., Curcic-Blake et al, 2017) favors ICA over more conventional activation-based approaches, such as GLM. Here, the ICA decomposition of time-series provides a direct equivalent of functional connectivity components, which is more in line with the process we want to capture.

Finally, we believe that the *"2S"* method may have crucial therapeutic implications in the near future, notably, in optimizing strategies for repetitive transcranial magnetic stimulation (rTMS) for refractory hallucinations. Although this non-invasive brain stimulation method has shown moderate, but significant, efficacy in reducing the severity of hallucinations (Demeulemeester et al., 2012), it remains a source of debate (Slotema et al., 2011). Its moderate effect may result from inter-subject variation in the brain areas associated with AVH, since most rTMS protocols systematically target the left temporo-parietal junction. Identifying with high reliability the functional networks recruited during AVH in a given individual could pave the way for new subject-based neuronavigation strategies for rTMS treatment of hallucinations. A randomized controlled trial is currently running to test the superiority of such an fMRI-guided strategy over conventional rTMS in the treatment of drug-resistant hallucinations (ClinicalTrials.gov Identifier: NCT01373866).

Acknowledgements:

This study was supported by the GDR CNRS-3557 Institut de Recherche en Psychiatrie, as well as by a grant from the Programme Hospitalier de Recherche Clinique (PHRC-N MULTIMODHAL) to RJ.

References

Alderson-Day B, Diederen K, Fernyhough C, Ford JM, Horga G, Margulies DS, McCarthy-

Jones S, Northoff G, Shine JM, Turner J, van de Ven V, van Lutterveld R, Waters F, Jardri R (2016) Auditory Hallucinations and the Brain's Resting-State Networks: Findings and Methodological Observations. Schizophr Bull 42:1110–1123.

Allen P, Larøi F, McGuire PK, Aleman A (2008) The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. Neurosci Biobehav Rev 32:175–191.

Allen P, Modinos G, Hubl D, Shields G, Cachia A, Jardri R, Thomas P, Woodward T, Shotbolt P, Plaze M, Hoffman R (2012) Neuroimaging auditory hallucinations in schizophrenia: from neuroanatomy to neurochemistry and beyond. Schizophr Bull 38:695–703.

Amad A, Cachia A, Gorwood P, Pins D, Delmaire C, Rolland B, Mondino M, Thomas P, Jardri R (2014) The multimodal connectivity of the hippocampal complex in auditory and visual hallucinations. Mol Psychiatry 19:184–191.

Andreasen NC, Flaum M (1991) Schizophrenia: the characteristic symptoms. Schizophr Bull 17:27–49.

Bazán PR, Biazoli CE, Sato JR, Amaro E (2015) Motor Readiness Increases Brain Connectivity Between Default-Mode Network and Motor Cortex: Impact on Sampling Resting Periods from fMRI Event-Related Studies. Brain Connect 5:631–640.

Chen G, Taylor PA, Shin Y-W, Reynolds RC, Cox RW (in press) Untangling the relatedness among correlations, Part II: Inter-subject correlation group analysis through linear mixed-effects modeling. NeuroImage

Cornbleet PJ, Gochman N (1979) Incorrect least-squares regression coefficients in method-comparison analysis. Clin Chem 25:432–438.

Ćurčić-Blake B, Ford JM, Hubl D, Orlov ND, Sommer IE, Waters F, Allen P, Jardri R, Woodruff PW, David O, Mulert C, Woodward TS, Aleman A (2017): Interaction of language, auditory and memory brain networks in auditory verbal hallucinations. Prog Neurobiol 148:1–20.

David CN, Greenstein D, Clasen L, Gochman P, Miller R, Tossell JW, Mattai AA, Gogtay N, Rapoport JL (2011) Childhood onset schizophrenia: high rate of visual hallucinations. J Am Acad Child Adolesc Psychiatry 50:681–686.e3.

Deco G, Jirsa V, McIntosh AR, Sporns O, Kötter R (2009) Key role of coupling, delay, and noise in resting brain fluctuations. Proc Natl Acad Sci U S A 106:10302–10307.

De Martino F, Gentile F, Esposito F, Balsi M, Di Salle F, Goebel R, Formisano E (2007) Classification of fMRI independent components using IC-fingerprints and support vector machine classifiers. NeuroImage 34:177–194.

Demeulemeester M, Amad A, Bubrovszky M, Pins D, Thomas P, Jardri R (2012) What is the real effect of 1-Hz repetitive transcranial magnetic stimulation on hallucinations? Controlling for publication bias in neuromodulation trials. Biol Psychiatry 71:e15–e16.

Diederen KMJ, Charbonnier L, Neggers SFW, van Lutterveld R, Daalman K, Slotema CW, Kahn RS, Sommer IEC (2013) Reproducibility of brain activation during auditory verbal hallucinations. Schizophr Res 146:320–325.

Diederen KMJ, Neggers SFW, Daalman K, Blom JD, Goekoop R, Kahn RS, Sommer IEC (2010) Deactivation of the parahippocampal gyrus preceding auditory hallucinations in schizophrenia. Am J Psychiatry 167:427–435.

Dierks T, Linden DE, Jandl M, Formisano E, Goebel R, Lanfermann H, Singer W (1999) Activation of Heschl's gyrus during auditory hallucinations. Neuron 22:615–621.

Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci U S A 113:7900–7905.

Esposito F, Scarabino T, Hyvarinen A, Himberg J, Formisano E, Comani S, Tedeschi G, Goebel R, Seifritz E, Di Salle F (2005) Independent component analysis of fMRI group studies by self-organizing clustering. NeuroImage 25:193–205.

Ey H (1973) Traité des hallucinations : tome 1. Paris: Masson. Available at: http://bsf.spp.asso.fr/index.php?lvl=notice_display&id=79490.

Ffytche DH, Howard RJ, Brammer MJ, David A, Woodruff P, Williams S (1998) The anatomy of conscious vision: an fMRI study of visual hallucinations. Nat Neurosci 1:738–742.

Formisano E, Esposito F, Di Salle F, Goebel R (2004) Cortex-based independent component analysis of fMRI time series. Magn Reson Imaging 22:1493–1504.

Foucher JR (2013) Perspective in brain imaging and computer-assisted technologies for the treatment of hallucinations. In: The Neuroscience of Hallucinations. New York, NY: Springer New York. p 529–547.

Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15:870–878.

Hoffman RE, Fernandez T, Pittman B, Hampson M (2011): Elevated functional connectivity along a corticostriatal loop and the mechanism of auditory/verbal hallucinations in patients with schizophrenia. Biol Psychiatry 69:407–414.

Hor K, Taylor M (2010) Suicide and schizophrenia: a systematic review of rates and risk factors. J Psychopharmacol Oxf Engl 24:81–90.

Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. Neural Netw Off J Int Neural Netw Soc 13:411–430.

Jardri R, Pins D, Bubrovszky M, Lucas B, Lethuc V, Delmaire C, Vantyghem V, Despretz P, Thomas P (2009) Neural functional organization of hallucinations in schizophrenia: multisensory dissolution of pathological emergence in consciousness. Conscious Cogn 18:449–457.

Jardri R, Pins D, Delmaire C, Goeb J-L, Thomas P (2007) Activation of bilateral auditory cortex during verbal hallucinations in a child with schizophrenia. Mol Psychiatry 12:319.

Jardri R, Pouchet A, Pins D, Thomas P (2011) Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. Am J Psychiatry 168:73–81.

Jardri R, Thomas P, Delmaire C, Delion P, Pins D (2013) The neurodynamic organization of modality-dependent hallucinations. Cereb Cortex N Y N 1991 23:1108–1117.

Laird AR, Eickhoff SB, Li K, Robin DA, Glahn DC, Fox PT (2009) Investigating the functional heterogeneity of the default mode network using coordinate-based meta-analytic modeling. J Neurosci Off J Soc Neurosci 29:14496–14505.

Lefebvre S, Demeulemeester M, Leroy A, Delmaire C, Lopes R, Pins D, Thomas P, Jardri R (2016) Network dynamics during the different stages of hallucinations in schizophrenia. Hum Brain Mapp 37:2571–2586.

Lennox BR, Park SB, Medley I, Morris PG, Jones PB (2000) The functional anatomy of auditory hallucinations in schizophrenia. Psychiatry Res 100:13–20.

Llorca PM, Pereira B, Jardri R, Chereau-Boudet I, Brousse G, Misdrahi D, Fénelon G, Tronche A-M, Schwan R, Lançon C,

Marques A, Ulla M, Derost P, Debilly B, Durif F, de Chazeron I (2016) Hallucinations in schizophrenia and Parkinson's disease: an analysis of sensory modalities involved and the repercussion on patients. Sci Rep 6:38152.

Overgaard M, Fazekas P (2016) Can No-Report Paradigms Extract True Correlates of Consciousness? Trends Cogn Sci 20:241–242.

Roquet DR, Pham BT, Foucher JR (2014) Manual selection of spontaneous activity maps derived from independent component analysis: criteria and inter-rater reliability study. J Neurosci Methods 223:30-4.

Shergill SS, Brammer MJ, Williams SC, Murray RM, McGuire PK (2000) Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. Arch Gen Psychiatry 57:1033–1038.

Shergill SS, Cameron LA, Brammer MJ, Williams SC, Murray RM, McGuire PK (2001) Modality specific neural correlates of auditory and somatic hallucinations. J Neurol Neurosurg Psychiatry 71:688–690.

Silbersweig DA, Stern E, Frith C, Cahill C, Holmes A, Grootoonk S, Seaward J, McKenna P, Chua SE, Schnorr L (1995) A functional neuroanatomy of hallucinations in schizophrenia. Nature 378:176–179.

Slotema CW, Blom JD, de Weijer AD, Diederen KM, Goekoop R, Looijestijn J, Daalman K, Rijkaart A-M, Kahn RS, Hoek HW, Sommer IEC (2011) Can low-frequency repetitive transcranial magnetic stimulation really relieve medication-resistant auditory verbal hallucinations? Negative results from a large randomized controlled trial. Biol Psychiatry 69:450–456.

Sommer IEC, Diederen KMJ, Blom J-D, Willems A, Kushan L, Slotema K, Boks MPM, Daalman K, Hoek HW, Neggers SFW, Kahn RS (2008) Auditory verbal hallucinations predominantly activate the right inferior frontal area. Brain J Neurol 131:3169–3177.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain, Thieme, New York.

Tan B-L (2009) Profile of cognitive problems in schizophrenia and implications for vocational functioning. Aust Occup Ther J 56:220–228.

Torgerson WS (1952) Multidimensional scaling: I. Theory and method. Psychometrika 17:401–419.

Tsuchiya N, Frässle S, Wilke M, Lamme V (2016) No-Report and Report-Based Paradigms Jointly Unravel the NCC: Response to Overgaard and Fazekas. Trends Cogn Sci 20:242–243.

Tsuchiya N, Wilke M, Frässle S, Lamme VAF (2015) No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. Trends Cogn Sci 19:757–770.

van de Ven VG, Formisano E, Röder CH, Prvulovic D, Bittner RA, Dietz MG, Hubl D, Dierks T, Federspiel A, Esposito F, Di Salle F, Jansma B, Goebel R, Linden DEJ (2005) The spatiotemporal pattern of auditory cortical responses during verbal hallucinations. NeuroImage 27:644–655.

van Lutterveld R, Diederen KMJ, Koops S, Begemann MJH, Sommer IEC (2013a) The influence of stimulus detection on activation patterns during auditory hallucinations. Schizophr Res 145:27–32.

2.2 La capture « *online* » des hallucinations auditives

La « méthode en deux étapes » décrite dans la partie précédente permet une labellisation fiable et reproductible en périodes hallucinatoires (dites périodes ON) et non hallucinatoires (dites périodes OFF) de la série temporelle recueillie en IRMf chez un sujet présentant des hallucinations auditives au cours de l'acquisition. Cependant, ce travail de *labelling* s'avère extrêmement chronophage et ne peut en aucun cas être utilisé en temps réel. Aussi, dans cette seconde partie, nous nous sommes focalisés sur des méthodes d'automatisation de cette procédure à partir notamment de techniques de *machine-learning* appliquées aux données IRMf. L'objectif est de développer un classificateur performant capable de déterminer en temps réel le label ON ou OFF d'un volume fonctionnel donné chez un sujet donné.

2.2.1 Généralités sur les méthodes d'apprentissage automatique (*machine-learning*)

L'apprentissage automatique consiste à faire « apprendre » à un ordinateur des capacités qui n'auront pas été préalablement programmées. En imagerie cérébrale, il s'agit le plus souvent d'un apprentissage supervisé, c'est-à-dire qu'un algorithme sera entrainé à classer des sujets en deux catégories, en fonction d'un modèle issu de l'expertise du clinicien.

La première étape consiste donc à demander à l'expert de classer les sujets en deux groupes (par exemple : groupe de sujets malades et groupe de sujets non malades). Cette phase de classement de l'information (dite **phase de « *labelling* »**) par l'expert est ici utilisée pour constituer un modèle. Puis l'algorithme est implémenté sur ce modèle et entrainé par apprentissage automatique à correctement labéliser les données d'imagerie associées aux sujets (**phase d'entrainement**), en groupes (dans notre exemple il associera les données d'imagerie de chaque sujet soit au groupe « malade », soit au groupe « non malade »). Les capacités de l'algorithme à prédire (et donc correctement classer) le statut (*i.e.* « malade » ou « non-malade »), de nouveaux sujets non utilisés lors de la phase d'entrainement (groupe indépendant) sur la base de leur IRM, sont ensuite évaluées lors d'une **phase dite de test**.

Il existe différentes méthodes d'apprentissage automatique mais la plus utilisée actuellement en imagerie est le *linear Support Vector Machine* (lSVM) (200,201). Les données d'imagerie sont rapportées dans un espace à plusieurs dimensions et il s'agit alors de définir dans cet espace un plan permettant de séparer de manière optimale les deux groupes de sujets. Ce plan est appelé hyperplan, ou encore fonction de décision. Ainsi, dans notre exemple présenté en **Figure 10**, le logiciel analyse 2 voxels de l'ensemble des images IRM classées par l'expert pour déterminer un plan qui sépare de

manière optimale les deux groupes de sujets, malades (en bleu) ou non malades (en vert). En imagerie cérébrale, le nombre de voxels est bien sûr beaucoup plus important, il n'y a plus 2 dimensions mais *n* dimensions et le plan est alors remplacé par un hyperplan.

Comme nous l'avons évoqué, suite à la phase d'apprentissage, le modèle utilisera cet hyperplan pour déterminer à quel groupe sont associées les données de prochains sujets auxquels il n'a jamais été exposé (la phase de test). Dans notre exemple, le modèle sera ainsi capable de discriminer, sans l'intervention de l'expert, si les données d'un sujet sont associées au groupe malade ou non malade.

En ce qui concerne les études de validation de méthodes de *machine-learning* en clinique, elles nécessitent une phase d'apprentissage supervisée par l'expert, et une phase de validation du modèle, à partir d'un échantillon indépendant. C'est à partir de la capacité du modèle à classer correctement les sujets de l'échantillon de validation que se mesure la capacité de discrimination du modèle (*Sensibilité* et *Spécificité* du classificateur).

Le *machine-learning* a d'abord été utilisé dans des études explorant la capacité d'un modèle à établir un diagnostic positif, en comparant l'imagerie structurale en IRM de sujets malades à celle de sujets non malades au cours de la phase d'apprentissage. La méthode de lSVM permet une classification de qualité pour de nombreuses pathologies neurologiques et psychiatriques (202). Ainsi, les modèles en SVM ont montré une précision de la classification allant de 83 % à 100 % pour la maladie d'Alzheimer ; de 71 % à 100 % pour les *Mild Cognitive Impairment* (MCI) ; de 72 % à 97 % pour la maladie de Parkinson ; de 68 % à 86 % pour l'épisode dépressif caractérisé ; de 59 % à 85 % pour l'épisode dépressif caractérisé résistant au traitement médicamenteux ; de 79 % pour le trouble bipolaire ; de 68 % à 90 % pour les troubles du spectre autistique et de 81 % à 92 % pour la schizophrénie (pour une revue de la littérature sur l'utilisation du machine learning dans le diagnostic de la schizophrénie, voir (203)). Ces résultats indiquent une grande capacité de discrimination des modèles de lSVM entre sujets malades et sujets non malades. Cependant les études actuellement disponibles concernent le plus souvent des sujets à un stade relativement avancé de la pathologie, au cours duquel les anomalies IRM sont plus marquées qu'en début d'évolution du trouble comparativement à des participants non malades et pour lequel l'apport potentiel du *machine learning* à la clinique apparait de fait moins intéressant.

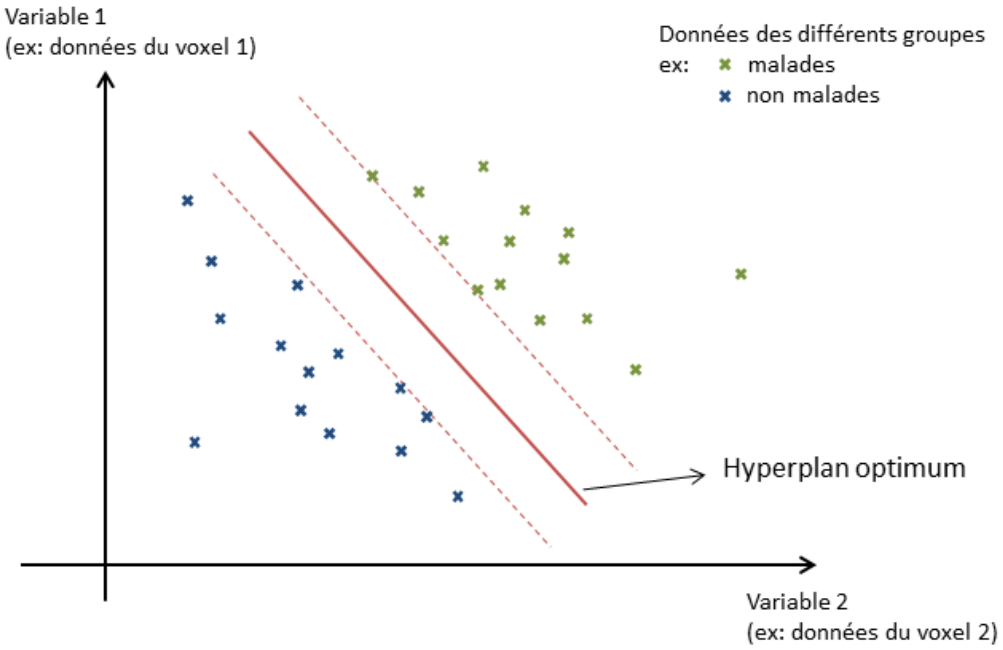


Figure 10. Principe de classification d'un *linear Support Vector Machine* (lSVM).

Le logiciel apprend à classer les sujets selon les groupes prédéterminés par l'expert en définissant un « hyperplan », représenté ici par la droite qui sépare de manière optimale les deux groupes de sujets.

Compte tenu de ces observations, plusieurs travaux ont évalué l'intérêt du *machine-learning* chez des personnes à risque de développer des pathologies neurologiques ou psychiatriques. En effet, chez des sujets présentant des prodromes ou dans des groupes de sujets à risque, le *machine-learning* pourrait apporter une prédiction individuelle du risque de transition vers la pathologie cible à partir d'images IRM anatomiques. Plusieurs résultats sont d'ores et déjà disponibles : la méthode de lSVM a démontré une précision de classification pour la transition vers la maladie d'Alzheimer chez les sujets sains à risque de 94,3% ; vers un MCI chez les sujets sains à risque de 81,8% à 91,4% ; vers la maladie d'Alzheimer chez des sujets MCI de 60% à 98,4% et vers les pathologies psychiatriques chez des sujets à risque de trouble psychiatrique (ARMS, pour *At-Risk Mental State*) de 82% à 92,3%. Des études en lSVM ont également montré des résultats intéressants pour la prédiction du décours évolutif de la schizophrénie une fois le diagnostic posé, avec une précision de la classification entre type épisodique ou continue de 67 à 70%. En ce qui concerne le premier épisode psychotique, une seule étude, à ce jour, a utilisé la lSVM pour prédire l'évolution clinique (à 6 ans), avec une sensibilité de 71% et une spécificité variant de 61% à 68% (204).

Plus récemment, le ISVM a été utilisé en prédiction de la réponse thérapeutique. Ces modèles utilisent généralement, pour la phase d'apprentissage, les données d'imagerie recueillies avant la mise en place d'un traitement, puis le suivi des sujets permet de constituer les groupes de sujets répondeurs ou non répondeurs. Les études lSVM ont montré une précision de la classification pour la réponse au traitement dans l'épisode dépressif caractérisé allant de 65,22% à 88,9%. Concernant la schizophrénie, si de nombreuses études ont exploré les altérations cérébrales en fonction du pronostic et les marqueurs d'imagerie permettant de distinguer les patients selon l'absence de rémission, le *machine learning* et les techniques de lSVM n'ont pas encore été utilisés pour prédire l'absence de rémission dans cette pathologie (205).

Malgré les résultats très encourageants du *machine learning* en psychiatrie et en neurologie, l'impact de ces techniques sur la pratique clinique psychiatrique a été, jusqu'à présent, très limité. En effet, les modifications anatomiques et de connectivité dans les pathologies psychiatriques sont généralement discrètes dans les stades précoces de la maladie, y compris dans les pathologies où des modifications morphologiques cérébrales sont les plus marquées, comme la schizophrénie (206,207). Les capacités de discrimination du *machine-learning* restent malheureusement encore insuffisantes par rapport aux capacités diagnostiques des psychiatres et son utilisation à visée diagnostique, pronostique ou thérapeutique est donc limitée par ce manque de précision. Cependant les performances du *machine-learning* pourraient être grandement améliorées dans les années futures par l'utilisation de plusieurs modalités d'imagerie dans le modèle (IRM structurale, IRM fonctionnelle, DTI, EEG etc.) ou par l'adjonction de facteurs biologiques autres aux données d'imagerie (génétique, expression génique, dosages sanguins de facteurs associés aux pathologies etc.). Ainsi, si des données suggèrent que les approches d'imagerie multimodales augmenteraient la sensibilité et la spécificité des modèles de machine learning (208), il semble licite d'espérer intégrer à l'avenir ces modèles dans la pratique clinique. C'est l'objectif du projet *PSYMAC* mené dans la région Hauts de France : construire un algorithme de prédiction du devenir à un an des patients expérimentant un premier épisode psychotique, et ainsi déterminer le risque pour ces patients de développer une schizophrénie, ou un trouble bipolaire, ou à voir leur état stabilisé à la normale, ceci à partir de données IRM mais également de données cliniques et génotypiques supplémentaires.

2.2.2 Développement d'un classificateur capable de détecter les périodes hallucinatoires : utilisation du *linear Support Vector Machine*

Le principe du *machine learning* ne se limite pas à une utilisation sur des données d'imagerie anatomique et il est tout à fait possible d'appliquer les stratégies de lSVM à des données fonctionnelles. Ceci permet de développer des classificateurs capables de labelliser chaque volume temporel d'une session d'IRMf selon deux classes préalablement établies (et pour lesquelles le classificateur a été entrainé), *i.e.*, les périodes ON (hallucinatoires) et les périodes OFF (non-hallucinatoires).

Ce type de classificateur peut être développé (phase d'entrainement et phase de test) avec des données fonctionnelles issues d'un seul et même sujet, on parle alors de classification « intra-sujet ». Toutefois, cela implique le développement de classificateurs spécifiques à chaque sujet, ce qui s'avère extrêmement chronophage, nécessitant une session d'enregistrement spécifique.

Dans ce contexte, des approches « inter-sujets » permettant de développer des algorithmes avec des données fonctionnelles issues de plusieurs sujets ont été proposées. L'objectif est le développement de classificateurs dont l'utilisation peut être généralisée à une population, chez des sujets indépendants (on parle de « *subject-independant classifier* »). C'est la question du développement possible de ce type de classificateur pour la détection des hallucinations auditives qui est au centre de ce travail de thèse.

Ainsi, l'**Article 2** présente le développement d'un classificateur (« inter-sujets ») capable de détecter les périodes hallucinatoires grâce au lSVM, l'étape préalable de *labelling* ayant été réalisé grâce à la technique décrite dans l'**Article 1**.

ARTICLE 2

Multivariate auditory-verbal hallucinations detection in schizophrenia

Thomas FOVET[1,2]*, Pierre YGER[3], Renaud LOPES[4], Pierre THOMAS[1,2], Philippe DOMENECH[5,6], Renaud JARDRI[1,2]

1. Univ. Lille, CNRS UMR 9193, Laboratoire de Sciences Cognitives et Sciences Affectives (SCALab-PsyCHIC), F-59000 Lille, France

2. CHU Lille, Pôle de Psychiatrie, Unité CURE, F-59000 Lille, France

3. Institut de la Vision, INSERM UMRS 968, UPMC UM 80, CNRS UMR 7210, Paris

4. Univ. Lille, INSERM, CHU Lille, U1171 - Degenerative and Vascular Cognitive Disorders, F-59000 Lille, France

5. AP-HP, Hôpital Henri-Mondor, DHU Pepsy, Neurochirurgie / Université Paris Est-Créteil, France

6. Behavior, Emotion & Basal Ganglia (BEBG) team, INSERM U1127- CNRS U 7225, Institut du Cerveau et de la Moelle épinière, Paris, France

En préparation

ABSTRACT

Auditory-verbal hallucinations (AVH) can be defined as auditory percepts in the absence of corresponding external stimuli. AVH are frequent experiences in schizophrenia (60 to 80% of patients) which may cause long-term disability. Recent functional Magnetic Resonance Imaging (fMRI) developments allowed for the objective “capture” of AVH’ occurrences while scanning a participant. Our team developed a semi-automatized procedure combining data-driven analysis of resting-state fMRI data with a post-fMRI interview (*i.e.* the patient is asked to report hallucinations’ occurrences and main clinical features of these experiences after acquisition). This “*two-steps method*” allows for the identification and distinction of fMRI periods with AVH (ON) from periods without (OFF). However, this detection scheme, notably the *a posteriori* labelling procedure, stays very time-consuming. Multicentric and multi-subject generalization would clearly benefit from an automatization of this fMRI capture method using machine-learning. *Multi-Voxel Pattern Analysis applied to fMRI data,* notably *linear Support Vector Machine* (lSVM), a supervised classification algorithm, is gaining recognition in accurately discriminating between complex cognitive states. Here, we present a validated fully-automated reliable procedure to detect AVH’ occurrences when applying lSVM to a per-AVH fMRI dataset. We demonstrated good between-subjects cross-validity, especially because contributing voxels are localized in a restraint set of brain regions. Adapting this method for real-time decoding will pave the way for innovative brain-based treatment of AVH, such as fMRI-neurofeedback.

KEY-WORDS

hallucination; automated; schizophrenia; generalization; between-subject; linear support vector machine; capture fMRI

INTRODUCTION

Auditory-verbal hallucinations (AVH), defined as the experience of hearing voices or sounds in the absence of appropriate external stimuli, are core symptoms of schizophrenia. Approximately 60 to 80% of patients with schizophrenia exhibit AVH (1), an experience also commonly associated with depressive disorders (2,3), poor quality of life (4) and increased risk of suicide (5). Moreover, the need for therapeutic innovation is high since, for 25 to 30% of patients with AVH, only a partial remission can be reached with antipsychotic drugs (6).

In this context, a better understanding of AVH' pathophysiology may constitute a precious way for the development of new treatments able to relieve patients from their voices. Especially, over the last decade, neuroimaging techniques provided significant advances in the knowing of AVH' neural underpinnings (7,8). On the one hand, functional brain imaging provided information about the neural bases of the susceptibility to hallucinate in studies measuring brain activity during specific tasks in patients who hallucinate and those who don't (i.e. "trait" studies) (9,10) They revealed altered functional activity in the temporal lobes of patients with AVH (probably emerging from a competition between AVH and normal external speech for processing sites within the temporal cortex), but also a decrease in the functional activity of the rostral dorsal anterior cingulate cortex, a structure known to be involved in the allocation of an internal or external origin for a given stimulus (11). On the other hand, functional brain imaging allowed for directly measuring brain activation associated with the occurrence of AVHs in "state" (or "capture" studies). Recent meta-analyses of such studies showed increased activity within a complex and distributed network including temporal, parietal, frontal and subcortical regions during AVH (12,13). Particularly, speech production and comprehension areas (i.e. Broca's and Wernicke's areas) have been shown to be involved, but in addition to this network, brain areas involved in contextual memory (i.e. hippocampal complex) seem to play a role (13–15).

Interestingly, these findings have strengthened transdiagnostic neurocognitive models that characterize AVH, but more specifically they built the bases for new therapeutic strategies. For example, rTMS protocols for AVH focus on reducing excitability of the left temporoparietal junction, judged to be a key-region in AVH' pathophysiology (13,16). In addition to optimizing neuronavigation protocols (e.g. for rTMS target location), the development of "capture" neuroimaging of AVH could enable the development of intermediate therapeutic strategies between neuromodulation and psychotherapy such as fMRI-guided neurofeedback (fMRI-NF), by providing reliable tools to detect in real-time the occurrence of AVH.

The current paper is an effort toward validating a reliable method for fMRI detection of AVH. The ideal detection method should be: (i) fully automated; (ii) accurate (i.e. with an area under the Receiver operating characteristic (ROC) curve close to 1); (iii) characterized by high inter-subject generalization properties and (iv) with a low-calculation-cost to be easily implemented in a closed-brain computer interface loop.

The detection of the AVH' occurrence while scanning a participant has long remained very challenging. The "button-press" approach (i.e. when the participant is asked to press a response button to signal AVH during the fMRI session; e.g. (17–20)) and the "random-sampling" method (i.e. discontinuous acquisition method in which the participant is asked for his sensory experiences after each stop; e.g. (21,22)) were criticized because of their vulnerability to a drop in patients' performance in signalling AVH occurrences in the MRI scanner (23). Recently, our team developed a more data-driven approach based on *independent component analysis* (ICA), called the *two-steps method for hallucination fMRI capture* (the "*2S"* method; (24)). This semi-automated procedure, combining a post-fMRI interview during which the patient is asked to report AVH, and ICA applied to per-hallucinatory fMRI data, has been validated as reliable and replicable in its ability to capture brain activity concomitant with AVH (23). Despite its validity, this *a posteriori* labelling procedure stays very time-consuming with multiple post-processing steps and analyses, and its use for real-time capture appears almost inconceivable. Here, we would like to present a new automated, accurate and generalizable AVH-capture method, based on *linear Support Vector Machine* (lSVM), a supervised classification method able to accurately discriminate between complex cognitive states (25).

RESULTS

Between-subjects lSVM classification performances for AVH periods (ON) versus (OFF and REST) volumes

Twenty-three patients suffering from schizophrenia (DSM-IV-TR criteria) and exhibiting very frequent and resistant AVH (more than 10 episodes per hour) were enrolled in this study (Cf. **Table 1**). After clinical evaluation, each participant performed a single MRI session encompassing an anatomical run, a 14-minutes resting state-fMRI run and the post-session interview. **Figure 1** describes how the "2S" method allows for the labeling of AVH periods (ON periods) and non-AVH ones (OFF periods). These ON/OFF periods were used to train a lSVM classifier to predict the activation patterns concomitant to AVH.

	Duration of SCZ	PANSS-P	P3-item	PANSS-Tot	CPZeq (mg)
Valid	23	23	23	23	23
Missing	0	0	0	0	0
Mean	17.83	21.04	4.783	81.17	353.6
Std. Deviation	10.45	4.666	0.8505	13.44	273.6
Minimum	5.000	13.00	4.000	59.00	100.0
Maximum	46.00	30.00	7.000	110.0	1200

Table 1: Clinical characteristic of the recruited sample (n=23).

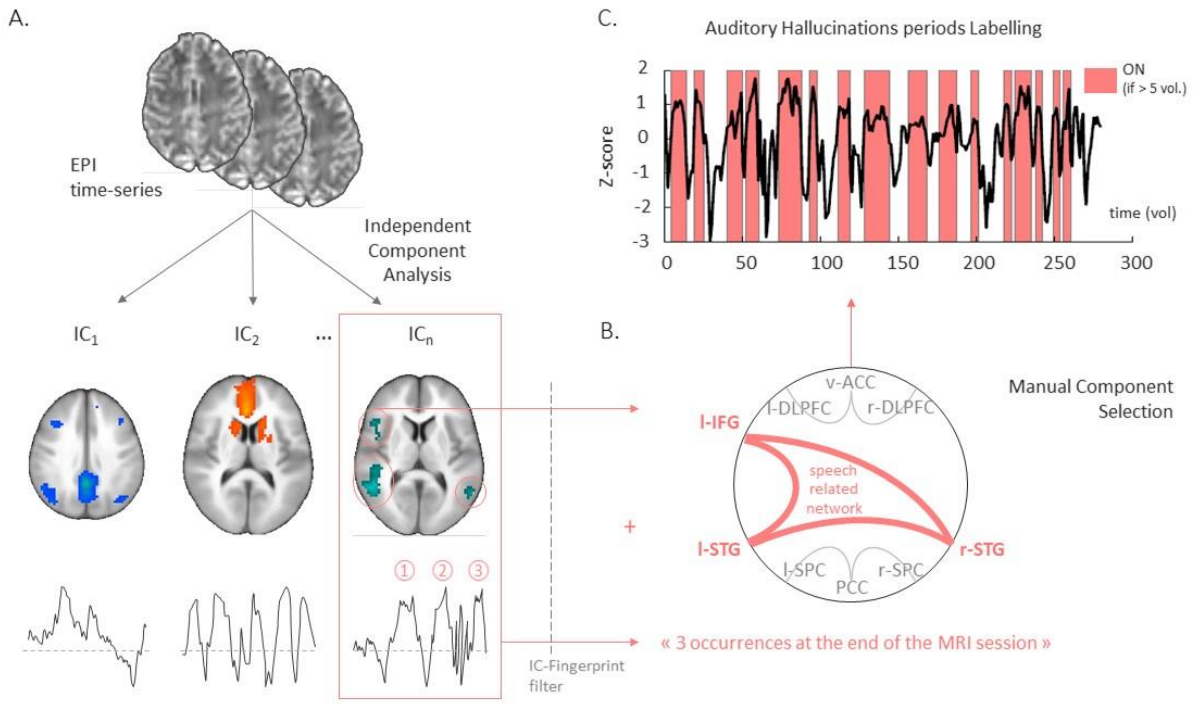


Figure 1: Description of the *two-steps method for hallucination fMRI capture* used for labelling the whole set of functional volumes for each subject.

The patients' states at different acquisition times were assigned to one of the following three categories: (i) "ON" (i.e. periods characterized by the occurrence of auditory-verbal hallucinations); (ii) "OFF" (i.e. periods characterized by the absence of auditory-verbal hallucinations); (iii) REST periods (i.e. periods that the method is unable to classify as "ON" or "OFF") (for a validation study of this method see (Leroy et al. 2017)).

A. Spatial *independent component analysis* (ICA) is applied to resting fMRI time series of patients with frequent auditory-verbal hallucinations (Formisano et al. 2004).

B. The most relevant component among those blindly generated by ICA is selected based on (i) the typical BOLD "fingerprint" of the independent component (i.e. the characterization of a component with respect to several spatial and temporal features making it possible to automatically classify it as related to BOLD responses, motion artefacts, etc) (De Martino et al. 2007); (ii) the time-course that best matches the hallucinations' timing and duration (information collected during a post-fMRI session interview) (iii) the spatial pattern that best matches the auditory-verbal nature of hallucinations (involvement of speech related network).

C. The time-course of the selected independent component is finally Z-transformed. Periods of increased BOLD signal (Z-scores > 0) that were maintained for at least 4 consecutive volumes (i.e., 12 seconds) were considered eligible for the ON period. The OFF periods were defined as a decrease in the BOLD signal (Z-score < 0) during at least 3 consecutive volumes and separated from ON periods by at least two volumes. The REST periods were the periods not labelled ON or OFF.

As learning with hundreds of samples using high-dimensional data is associated with a significant risk of overfitting leading to poor performances on independent subjects, we opted for a *region of interest* (ROI) approach to reduce the "curse of dimensionality". An ANOVA-based feature selection (n= 100 voxels) was thus conducted before running the lSVM.

To discriminate between ON periods (i.e., periods with AVH) *versus* OFF and REST periods, we chose the lSVM algorithm, a supervised method able to linearly separate features using the *hyperplane* (or *decision boundary*) that provides the largest margin between two conditions (26). We first used a leave-one subject-out cross validation procedure on the training set to assess the accuracy of the classifier.

Classification results achieved an above chance level decoding performance (see **Figure 2, A)**. We obtained an AUC of 0,79 (see **Figure 2, B**). To identify voxels driving the prediction, we extracted the discriminative weight map, i.e., the spatial patterns that best discriminate between the ON states and the (OFF or REST) states (see **Figure 2, D**). We also tested for the effect of varying the number of subjects included in the classifier's training. We showed that the classifier reached a peak of performance for 15 participants or more (see **Figure 2, C**).

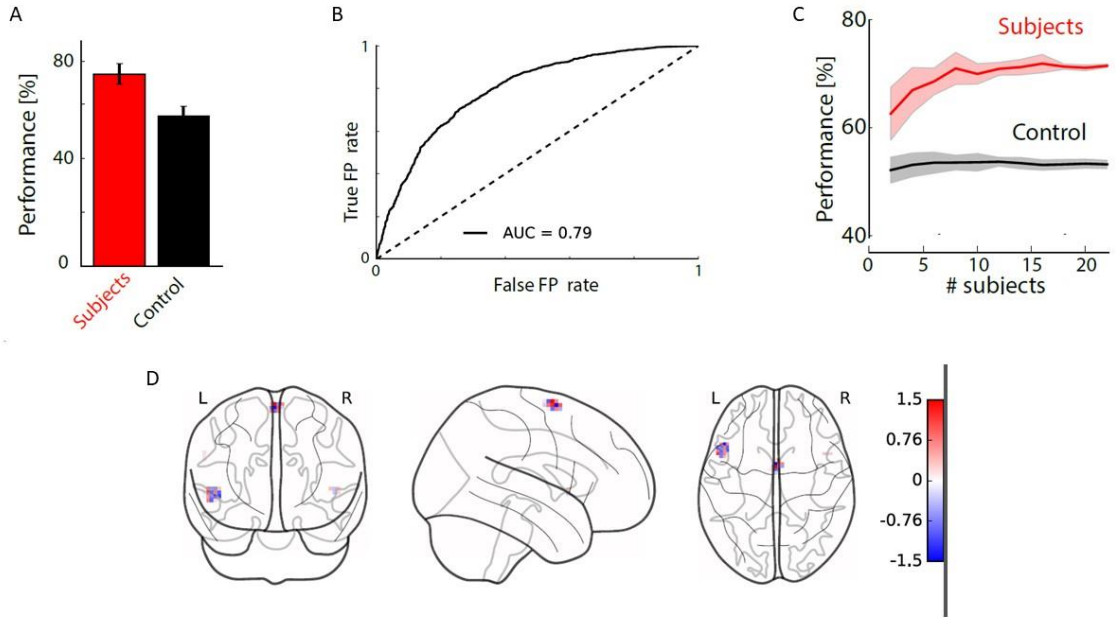


Figure 2: Decoding performances of a classifier able to identify ON periods *versus (*OFF and REST) periods.

A. Accuracy of the classifier compared with control (the level of chance).

B. Receiver operating characteristic (ROC) curve of the classifier.

C. Evolution of the accuracy of the classifier when varying the number of subjects included for training.

D. The weight map of the classifier reveals the spatial patterns that best discriminate the two cognitive states: (i) positive weights indicate a contribution toward predicting ON periods; (ii) negative weight indicate a contribution towards predicting the OFF and REST periods.

Classifier's improvement for ON versus OFF decoding.

Because our labelling procedure allows for the rigorous identification of OFF periods (see **Figure 1**), we also used our lSVM procedure to test the performances of a classifier trained to detect ON *and* OFF periods. This choice may be justified by the indeterminate nature of the REST periods, which contain volumes not characterized as pure ON or OFF periods. By excluding the REST volumes, performances of the classifier significantly improved compared to the first classifier with an AUC of 0,92 (see **Figure 3, A and B**).

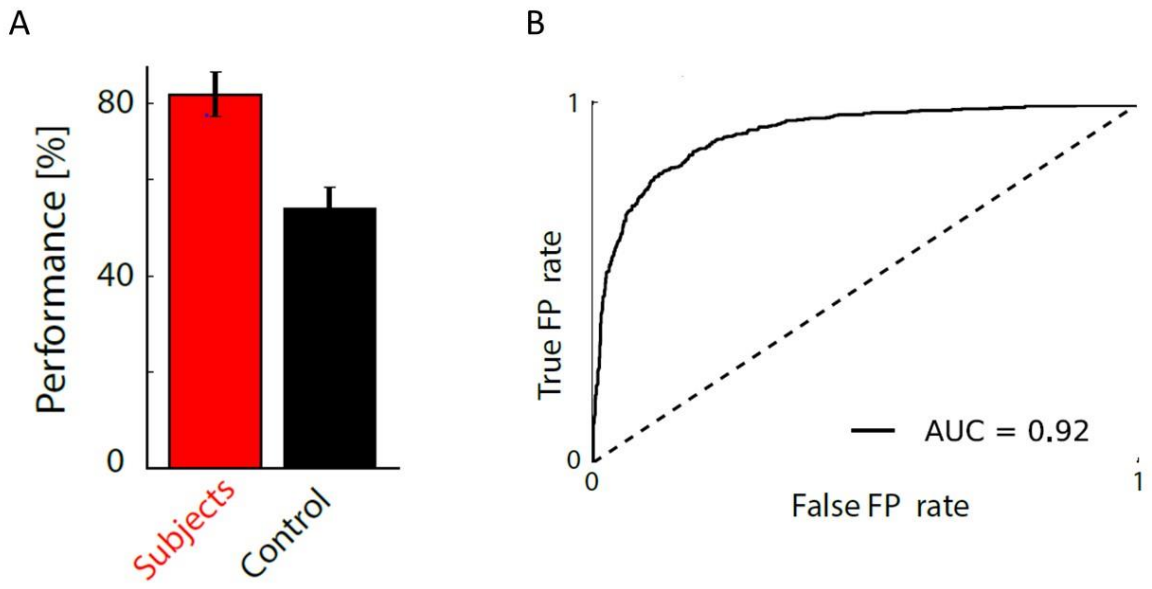


Figure 3: Decoding performances of a classifier able to identify ON *versus* OFF periods.
A. Accuracy of the classifier compared with control (chance level).
B. Receiver operating characteristic (ROC) curve of the classifier.

Specificity of the AVH spatial patterns

It has been shown that the period preceding the occurrence of the AVH (i.e., the few seconds corresponding to the brain transition from a resting-state to a full AVH state) has specific features. Particularly, a deactivation of the left parahippocampus cortex during this "pre-AVH" stage has been shown in two studies (27,28) and other studies identified specific "pre-AVH" activation patterns (29,30). Even if this has been less investigated, the same observation was done for the "post-AVH" period which could be characterized by specific changes in resting-state networks' dynamics (31).

In this regards, we tested the lSVM performances of the ON *versus* OFF classifier when sliding from ON to the 3 preceding volumes and from the ON to the 3 following volumes (see **Figure 4**). Interestingly, we showed that during these transition periods, the lSVM scores (i.e. result of the lSVM discriminant function) rapidly drop as the volumes are far from the ON period. This finding supports the idea of specific AVH patterns captured by the lSVM classifier.

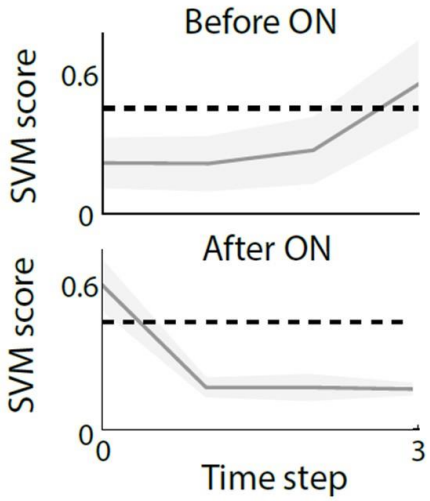


Figure 4: Evolution of the lSVM scores (i.e. result of the lSVM discriminant function) when sliding towards the 3 volumes preceding the ON period (upper panel), or when sliding towards the 3 volumes following the ON period (lower panel).

DISCUSSION

In this study, we developed a reliable and fully-automatized method for the detection of AVHs on a per-AVH fMRI dataset. The choice of a lSVM classifier was justified because of the good generalization performances of this technique when exposed to new inputs, noteworthy in the verbal domain (32). By extracting information about an individual's subjective mental state, in our case AVH, we demonstrated that spontaneous and iterative representational contents can be decoded from fMRI activity. Indeed, we showed that our lSVM algorithm was accurate in detecting AVH but furthermore exhibited good inter-subject generalization properties. We were notably able to detect AVH at the subject-level using prior detections in the same individual (see **Supplementary Figure 1**) but also with a between-subject approach (see **Figure 2**), using prior detections in other participants. Finally, despite the use of a ROI-strategy to discard non-discriminative voxels, AUC was shown extremely high, even if a limited number of voxels were used for lSVM (n= 100 voxels), which is very exciting for later real-time applications for which low-calculation cost is needed such as fMRI-NF.

fMRI-NF is an emerging method allowing participants to achieve control of their own brain activity using the real-time feedback of the activity (measured indirectly based on the BOLD signal) of a particular brain region or network (33). Producing durable changes in brain activity, neurofeedback have been described by some authors, as equivalent to an "endogenous" (i.e. without any physical intervention) rTMS (34). Bearing this in mind, fMRI-NF could constitute a very promising non-invasive treatment for drug-resistant AVH (35–37). A major issue in the development of fMRI-NF protocols for AVH lies on the complexity of the neurophysiological "signature" of AVH. Interestingly, we showed that the left inferior frontal gyrus (Broca's area) and, to a lesser degree, its right homologue appears as the most important regions to drive the prediction of AVH (see discriminative weight map). Even if this is in

line with the recent neuroimaging studies about AVH (Sommer et al. 2008), this also can be surprising with regards to the labelling technique we used (see **Figure 1**) focusing on temporo-parietal speech-related networks. One could hypothesize that the activity of these more posterior networks should be more variable, depending on the phenomenology of the symptom, contrary to the frontal regions involved in AVH.

In this paper, we also showed that during transition periods (from OFF to ON), the lSVM scores rapidly drop as the volumes are far from the ON period. Crucially, recent works indicated that AVH cannot be reduced to a simple ON/OFF process in a single region but that the chronological sequence of resting-state networks dynamics (i.e. from ignition to full perception and finally to extinction) must be taken into account (31). Our results are in line with these findings, showing that this transition period is probably characterized by specific patterns.

From a methodological point of view, we chose linear classifiers rather than non-linear ones, since this last option appear more vulnerable to overfitting (38), despite more complex class boundary definition (39), linear classifiers have been shown to perform better than non-linear classifiers on fMRI data-sets (40). In addition to their stability, linear classification findings are easier to interpret which appears crucial in a clinical perspective. Among those algorithms, the lSVM has the advantage to not assume multivariate normality of the data and thus searches for the decision boundary that maximize the margins between the two patterns to distinguish (41).

We also used a ROI-approach to reduce the curse of dimensionality. Some authors proposed to reduce the number of voxels analyzed using the "searchlight method" (42). However, a first disadvantage of this method is the strong assumption that discriminative information is located in small brain regions. In the context of the neural bases of AVH, we have arguments from the literature that support the involvement of a distributed network during AVH experiences. The "search-light" method furthermore needs to correct for multiple comparisons for each iteration of the search, which reduces statistical power (43). Here, we combined signals from multiple regions of interest (44), based on an ANOVA (45).

A step further will be to extend our detection scheme to mind-reading (43), i.e. model the representational content of hallucinations. Given the multidimensional nature of hallucinatory experiences (46), multivariate pattern analysis (MVPA) might be particularly well suited to decode its neural bases. Besides, recent works indicated specificities in the neural “signature” of hallucinatory symptoms according to fine-grain clinical and semiotical dimensions (e.g. for AVH, verbal *versus* musical hallucinations or inner *versus* outer space hallucinations (47,48) that MVPA could be able to detect.

CONCLUSION

We developed an accurate, generalizable, subject-independent classifier able to detect the AVH occurrence with a low-computational cost, paving the way for innovative treatment based on the fMRI real-time decoding of brain activity such as fMRI-neurofeedback.

MATERIALS AND METHODS

Participants

The population is composed of 23 patients with schizophrenia (DSM-IV-TR) suffering from very frequent AVH (i.e., more than 10 episodes/hour) and resistant AVHs as evaluated with item P3 of the Positive and Negative Syndrome Scale (PANSS). The exclusion criteria included the presence of an Axis-II diagnosis secondary Axis-I diagnosis, neurological or sensory disorder, and a history of drug abuse, which was based on a clinical interview and urine tests that were administered at admission. Clinical characteristics of the recruited subjects are summarized in **Table 1**

Participants were asked to lie in the scanner in a state of wakeful rest with eyes closed. The subjects experienced an average of 4 hallucinatory episodes per scan.

Data acquisition

After clinical evaluation, each patient participated in a single MRI acquisition session. The acquisition included: an anatomical run and a 14 minutes resting-state fMRI run. This resting-state run was followed by a post-fMRI interview (see Jardri et al. 2013). This interview explored the phenomenology of AVHs occurring during the resting-state run, i.e. relative moment(s) of AVH occurrence, duration of AVH experiences, and frequency of AVHs during scanning. If AVHs did not occur during the resting-state run, a new session was proposed and only resting-state runs during which patients described at least one AVH experience were included in the study.

Imaging parameters

Participants underwent a 10-minutes anatomical T1-weighted 3D multishot turbo-echo scan (1,5-T Philips Achieva; 150 transverse slices, field of view = 256 mm^2, voxel size = 1 mm^3). A set of 280 blood oxygen level-dependent (BOLD) fMRI volumes were acquired (single-shot sensitivity-encoded echo-planar imaging sequence, 30 transverse slices, field of view = 240 mm^2, voxel size = 4 mm3, repetition time = 3000 ms, echo time = 70 ms, total acquisition time = 14 minutes). All participants wore headphones and earplugs to attenuate scanner noise and were asked to lie in the scanner in a state of wakeful rest with eyes closed.

Ethics

The study was approved by the local ethical committee (CPP Nord-Ouest France IV), and written informed consent was obtained for each participant enrolled in the study.

fMRI preprocessing

The anatomical and functional data were pre-processed using BrainVoyager software (BVQX v3.6, Maastricht).

The pre-processing of the functional data consisted in a slice time correction, a 3D motion correction for head movements using a rigid body algorithm, smoothing using a spatial Gaussian filter (full-width at half-maximum [FWHM] = 6,0 mm), a temporal high-pass filtering with 2 sin/cos. The anatomical data were submitted to an intensity inhomogeneity correction algorithm. Coregistration using gradient-driven affine transformations with 9 alignment parameters (3 translations, 3 rotations and 3 scales) between functional runs and 3D-T1 weighted scans of each patient were performed automatically, and then manually corrected. All anatomical and functional volumes were spatially normalized to the MNI space.

Data labelling

For each patient, cortex-based ICA was performed to extract 30 components from the resting-state fMRI signal of the cortical voxels. The ICA was performed using the “FastICA” algorithm as implemented in the BrainVoyager software. To identify AVH periods, we used the method described in **Figure 1**.

Data analysis

The scikit-learn library [24] was then used to implement the lSVM classifiers using Python (Pedregosa et al. 2011).

Computation of samples. All volumes are processed with a common brain mask, computed across all volumes and all subjects. Within this mask, to reduce the dimensionality of the data, we used a classical univariate feature selection based on F-test, namely ANOVA. The number of voxels kept for the analysis is roughly 5% of the total number of voxels.

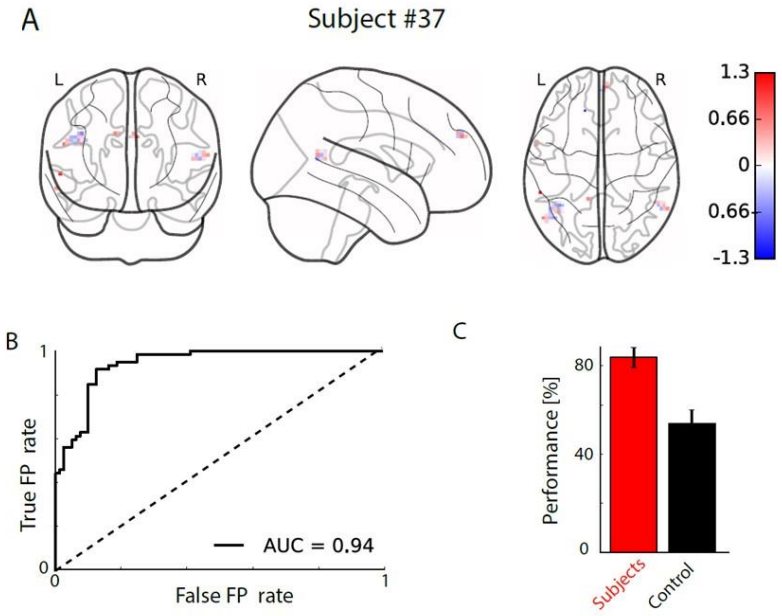
Supervised analysis with lSVM. With labels (ON, OFF, REST) provided by the external labelling (see Figure 1), we trained a linear Support Vector Machine to distinguish between ON and OFF+REST.

Performance metric and cross-validation

Cross validation. We used a leave-one subject-out cross validation procedure.

Result significance. AUC were calculated for each trained classifier.

Discriminative maps. To identify voxels driving the prediction, we extracted the discriminative weight map, i.e., the spatial patterns that best discriminate between the 2 states.



Supplementary Figure 1: Individual results of the decoding performances of the ON vs OFF + REST linear support vector machine classifier.

A. The weight map of the classifier reveals the spatial patterns that best discriminate AVH from non-AVH periods: (i) positive weights indicate a contribution toward predicting ON periods; (ii) negative weight indicate a contribution towards predicting the OFF and REST periods.

B. Receiver operating characteristic (ROC) curve of the classifier.

C. Accuracy of the classifier compared with control (chance level).

DECLARATION OF INTERESTS

This study was supported by the *Agence Nationale de la Recherche* (ANR-16-CE37-0015 awarded to RJ).

REFERENCES

Birchwood, Max, Zaffer Iqbal, et Rachel Upthegrove. 2005. « Psychological Pathways to Depression in Schizophrenia: Studies in Acute Psychosis, Post Psychotic Depression and Auditory Hallucinations ». *European Archives of Psychiatry and Clinical Neuroscience* 255 (3): 202-12. doi:10.1007/s00406-005-0588-4.

Cantor-Rivera, Diego, Ali R. Khan, Maged Goubran, Seyed M. Mirsattari, et Terry M. Peters. 2015. « Detection of Temporal Lobe Epilepsy Using Support Vector Machines in Multi-Parametric Quantitative MR Imaging ». *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society* 41 (avril): 14-28. doi:10.1016/j.compmedimag.2014.07.002.

Cortes, Corinna, et Vladimir Vapnik. 1995. « Support-Vector Networks ». *Machine Learning* 20 (3): 273-97. doi:10.1023/A:1022627411411.

Cutting, J, et F Dunne. 1989. « Subjective Experience of Schizophrenia ». *Schizophrenia Bulletin* 15 (2): 217-31.

De Martino, Federico, Francesco Gentile, Fabrizio Esposito, Marco Balsi, Francesco Di Salle, Rainer Goebel, et Elia Formisano. 2007. « Classification of FMRI Independent Components Using IC-Fingerprints and Support Vector Machine Classifiers ». *NeuroImage* 34 (1): 177-94. doi:10.1016/j.neuroimage.2006.08.041.

Demeulemeester, Morgane, Ali Amad, Maxime Bubrovszky, Delphine Pins, Pierre Thomas, et Renaud Jardri. 2012. « What Is the Real Effect of 1-Hz Repetitive Transcranial Magnetic Stimulation on Hallucinations? Controlling for Publication Bias in Neuromodulation Trials ». *Biological Psychiatry* 71 (6): e15-16. doi:10.1016/j.biopsych.2011.10.010.

Diederen, Kelly M. J., Sebastiaan F. W. Neggers, Kirstin Daalman, Jan Dirk Blom, Rutger Goekoop, René S. Kahn, et Iris E. C. Sommer. 2010. « Deactivation of the Parahippocampal Gyrus Preceding Auditory Hallucinations in Schizophrenia ». *The American Journal of Psychiatry* 167 (4): 427-35. doi:10.1176/appi.ajp.2009.09040456.

Dierks, T., D. E. Linden, M. Jandl, E. Formisano, R. Goebel, H. Lanfermann, et W. Singer. 1999. « Activation of Heschl's Gyrus during Auditory Hallucinations ». *Neuron* 22 (3): 615-21.

Dyck, Miriam S., Krystyna A. Mathiak, Susanne Bergert, Pegah Sarkheil, Yury Koush, Eliza M. Alawi, Mikhail Zvyagintsev, Arnim J. Gaebler, Sukhi S. Shergill, et Klaus Mathiak. 2016. « Targeting Treatment-Resistant Auditory Verbal Hallucinations in Schizophrenia with FMRI-Based Neurofeedback - Exploring Different Cases of Schizophrenia ». *Frontiers in Psychiatry* 7: 37. doi:10.3389/fpsyt.2016.00037.

Formisano, Elia, Fabrizio Esposito, Francesco Di Salle, et Rainer Goebel. 2004. « Cortex-Based Independent Component Analysis of FMRI Time Series ». *Magnetic Resonance Imaging* 22 (10): 1493-1504. doi:10.1016/j.mri.2004.10.020.

Fovet, Thomas, Renaud Jardri, et David Linden. 2015. « Current Issues in the Use of FMRI-Based Neurofeedback to Relieve Psychiatric Symptoms ». *Current Pharmaceutical Design* 21 (23): 3384-94.

Fovet, Thomas, Natasza Orlov, Miriam Dyck, Paul Allen, Klaus Mathiak, et Renaud Jardri. 2016. « Translating Neurocognitive Models of Auditory-Verbal Hallucinations into Therapy: Using Real-Time FMRI-Neurofeedback to Treat Voices ». *Frontiers in Psychiatry* 7: 103. doi:10.3389/fpsyt.2016.00103.

Harkavy-Friedman, Jill M., David Kimhy, Elizabeth A. Nelson, David F. Venarde, Dolores Malaspina, et J. John Mann. 2003. « Suicide Attempts in Schizophrenia: The Role of Command Auditory Hallucinations for Suicide ». *The Journal of Clinical Psychiatry* 64 (8): 871-74.

Haxby, James V., Andrew C. Connolly, et J. Swaroop Guntupalli. 2014. « Decoding Neural Representational Spaces Using Multivariate Pattern Analysis ». *Annual Review of Neuroscience* 37: 435-56. doi:10.1146/annurev-neuro-062012-170325.

Hoffman, Ralph E., Adam W. Anderson, Maxine Varanko, John C. Gore, et Michelle Hampson. 2008. « Time Course of Regional Brain Activation Associated with Onset of Auditory/Verbal Hallucinations ». *The British Journal of Psychiatry: The Journal of Mental Science* 193 (5): 424-25. doi:10.1192/bjp.bp.107.040501.

Jardri, Renaud, Alexandre Pouchet, Delphine Pins, et Pierre Thomas. 2011. « Cortical Activations during Auditory Verbal Hallucinations in Schizophrenia: A Coordinate-Based Meta-Analysis ». *The American Journal of Psychiatry* 168 (1): 73-81. doi:10.1176/appi.ajp.2010.09101522.

Jardri, Renaud, Pierre Thomas, Christine Delmaire, Pierre Delion, et Delphine Pins. 2013. « The Neurodynamic Organization of Modality-Dependent Hallucinations ». *Cerebral Cortex (New York, N.Y.: 1991)* 23 (5): 1108-17. doi:10.1093/cercor/bhs082.

Larøi, Frank, Iris E. Sommer, Jan Dirk Blom, Charles Fernyhough, Dominic H. Ffytche, Kenneth Hugdahl, Louise C. Johns, et al. 2012. « The Characteristic Features of Auditory Verbal Hallucinations in Clinical and Nonclinical Groups: State-of-the-Art Overview and Future Directions ». *Schizophrenia Bulletin* 38 (4): 724-33. doi:10.1093/schbul/sbs061.

Lefaucheur, Jean-Pascal, Nathalie André-Obadia, Andrea Antal, Samar S. Ayache, Chris Baeken, David H. Benninger, Roberto M. Cantello, et al. 2014. « Evidence-Based Guidelines on the Therapeutic Use of Repetitive Transcranial Magnetic Stimulation (RTMS) ». *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology* 125 (11): 2150-2206. doi:10.1016/j.clinph.2014.05.021.

Lefebvre, Stéphanie, Morgane Demeulemeester, Arnaud Leroy, Christine Delmaire, Renaud Lopes, Delphine Pins, Pierre Thomas, et Renaud Jardri. 2016. « Network Dynamics during the Different Stages of Hallucinations in Schizophrenia ». *Human Brain Mapping* 37 (7): 2571-86. doi:10.1002/hbm.23197.

Lennox, B. R., S. B. Park, P. B. Jones, P. G. Morris, et G. Park. 1999. « Spatial and Temporal Mapping of Neural Activity Associated with Auditory Hallucinations ». *Lancet (London, England)* 353 (9153): 644.

Lennox, B. R., S. B. Park, I. Medley, P. G. Morris, et P. B. Jones. 2000. « The Functional Anatomy of Auditory Hallucinations in Schizophrenia ». *Psychiatry Research* 100 (1): 13-20.

Leroy, Arnaud, Jack R. Foucher, Delphine Pins, Christine Delmaire, Pierre Thomas, Mathilde M. Roser, Stéphanie Lefebvre, et al. 2017. « FMRI Capture of Auditory Hallucinations: Validation of the Two-Steps Method ». *Human Brain Mapping*, juin. doi:10.1002/hbm.23707.

McCarthy-Jones, Simon. 2012. « Taking Back the Brain: Could Neurofeedback Training Be Effective for Relieving Distressing Auditory Verbal Hallucinations in Patients with Schizophrenia? » *Schizophrenia Bulletin* 38 (4): 678-82. doi:10.1093/schbul/sbs006.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. « Scikit-learn: Machine Learning in Python ». *Journal of Machine Learning Research* 12 (Oct): 2825-30.

Rocca, Paola, Filomena Castagna, Tullia Mongini, Cristiana Montemagni, et Filippo Bogetto. 2010. « Relative Contributions of Psychotic Symptoms and Insight to Quality of Life in Stable Schizophrenia ». *Psychiatry Research* 177 (1-2): 71-76. doi:10.1016/j.psychres.2009.01.030.

Ros, Tomas, Moniek A. M. Munneke, Diane Ruge, John H. Gruzelier, et John C. Rothwell. 2010. « Endogenous Control of Waking Brain Rhythms Induces Neuroplasticity in Humans ». *The European Journal of Neuroscience* 31 (4): 770-78. doi:10.1111/j.1460-9568.2010.07100.x.

Shergill, S. S., M. J. Brammer, S. C. Williams, R. M. Murray, et P. K. McGuire. 2000. « Mapping Auditory Hallucinations in Schizophrenia Using Functional Magnetic Resonance Imaging ». *Archives of General Psychiatry* 57 (11): 1033-38.

Shergill, S. S., L. A. Cameron, M. J. Brammer, S. C. Williams, R. M. Murray, et P. K. McGuire. 2001. « Modality Specific Neural Correlates of Auditory and Somatic Hallucinations ». *Journal of Neurology, Neurosurgery, and Psychiatry* 71 (5): 688-90.

Shergill, S. S., R. M. Murray, et P. K. McGuire. 1998. « Auditory Hallucinations: A Review of Psychological Treatments ». *Schizophrenia Research* 32 (3): 137-50.

Shergill, Sukhwinder S., Mick J. Brammer, Edson Amaro, Steve C. R. Williams, Robin M. Murray, et Phillip K. McGuire. 2004. « Temporal Course of Auditory Hallucinations ». *The British Journal of Psychiatry: The Journal of Mental Science* 185 (décembre): 516-17. doi:10.1192/bjp.185.6.516.

Silbersweig, D. A., E. Stern, C. Frith, C. Cahill, A. Holmes, S. Grootoonk, J. Seaward, P. McKenna, S. E. Chua, et L. Schnorr. 1995. « A Functional Neuroanatomy of Hallucinations in Schizophrenia ». *Nature* 378 (6553): 176-79. doi:10.1038/378176a0.

Slotema, Christina W., Jan D. Blom, Remko van Lutterveld, Hans W. Hoek, et Iris E. C. Sommer. 2014. « Review of the Efficacy of Transcranial Magnetic Stimulation for Auditory Verbal Hallucinations ». *Biological Psychiatry* 76 (2): 101-10. doi:10.1016/j.biopsych.2013.09.038.

Sommer, Iris E. C., Kelly M. J. Diederen, Jan-Dirk Blom, Anne Willems, Leila Kushan, Karin Slotema, Marco P. M. Boks, et al. 2008. « Auditory Verbal Hallucinations Predominantly Activate the Right Inferior Frontal Area ». *Brain: A Journal of Neurology* 131 (Pt 12): 3169-77. doi:10.1093/brain/awn251.

Upthegrove, R., M. R. Broome, K. Caldwell, J. Ives, F. Oyebode, et S. J. Wood. 2016. « Understanding Auditory Verbal Hallucinations: A Systematic Review of Current Evidence ». *Acta Psychiatrica Scandinavica* 133 (5): 352-67. doi:10.1111/acps.12531.

Upthegrove, Rachel, M. Birchwood, K. Ross, K. Brunett, R. McCollum, et L. Jones. 2010. « The Evolution of Depression and Suicidality in First Episode Psychosis ». *Acta Psychiatrica Scandinavica* 122 (3): 211-18. doi:10.1111/j.1600-0447.2009.01506.x.

Zmigrod, Leor, Jane R. Garrison, Joseph Carr, et Jon S. Simons. 2016. « The Neural Mechanisms of Hallucinations: A Quantitative Meta-Analysis of Neuroimaging Studies ». *Neuroscience and Biobehavioral Reviews* 69 (octobre): 113-23. doi:10.1016/j.neubiorev.2016.05.037.

2.2.3 Développement d'un classificateur capable de détecter les périodes pré-hallucinatoires et comparaison du *linear Support Vector Machine* au *TV-Elastic-net*

Dans l'**Article 2**, nous avons démontré qu'il était possible de développer un classificateur à même de détecter les périodes hallucinatoires (ON vs OFF ou même ON vs REST). Toutefois, dans une perspective thérapeutique de neurofeedback, il est crucial de pouvoir détecter le plus précocement possible le phénomène hallucinatoire. Dans le travail suivant, nous nous sommes donc intéressés à la période « pré-hallucinatoire » (*i.e.* les quelques secondes qui précèdent la survenue d'une hallucination). Cette période présente des caractéristiques particulières en terme d'activation cérébrale. Il a notamment pu être montré qu'une désactivation du cortex parahippocamique gauche est fréquente (209,210). La même désactivation du cortex parahippocampique est retrouvée dans les secondes qui précèdent les processus de rappel conscients chez les sujets sains (211), il pourrait donc s'agir d'un « *trigger* » inadéquat d'activation des aires du langage retrouvées au cours de la période hallucinatoire. Cet état de transition est également accompagné d'une déactivation de plusieurs autres régions cérébrales : le gyrus temporal supérieur gauche, le gyrus frontal moyen gauche, le gyrus frontal inférieur droit, l'insula droite et l'hémisphère gauche du cervelet (209). Le cortex cingulaire antérieur pourrait également être impliqué (210). Deux autres études, montrent une spécificité des *patterns* d'activation au cours de cette période (212,213). Lennox et collaborateurs (212) montrent une activité au niveau du gyrus temporal moyen droit tandis que Shergill et collaborateurs (213) mettent en évidence une activité au niveau du gyrus frontal inférieur gauche et du gyrus temporal moyen droit au cours de la transition vers l'hallucination. Enfin, cette période a été étudiée par connectivité effective dans le travail de Lefebvre et collaborateurs (214) présenté précédemment (voir 1.3.2.3.1. Connectivité fonctionnelle et connectivité effective).

Dans ce contexte, nous nous sommes focalisés, dans l'**Article 3**, sur le développement d'un classificateur permettant d'identifier spécifiquement la période « pré-hallucinatoire ». Dans ce même travail, et dans un objectif d'optimisation des performances de ce classificateur, nous avons comparé les performances du lSVM et du TV-Elastic-net, un algorithme de *machine learning* alternatif dont les caractéristiques et la validation ont fait l'objet de l'**Article 4** (le manuscrit de cet article, plus technique, n'a pas été intégré dans la thèse, mais il est disponible en ligne).

L'intérêt principal du TV-Elastic-net est qu'il permet d'obtenir des cartes discriminatives beaucoup plus faciles à interpréter par rapport à la technique de lSVM (qui fournit des cartes quasiment impossibles à interpréter). En effet, cette technique se fonde sur l'utilisation de trois types de contrainte de régularisation spatiale (l_1, l_2 *and Total Variation*) basées sur le principe de parcimonie mais aussi tirant profit des caractéristiques spécifiques du signal d'IRMf cérébrale (données en 3 dimensions) et des connaissances actuelles sur la structure et la connectivité du cerveau (215).

ARTICLE 3

Prediction of activation patterns preceding hallucinations in patients with schizophrenia using machine learning with structured sparsity

Amicie de PIERREFEU[1], Thomas FOVET [2,3], Fouad HADJ-SELEM [5], Tommy LÖFSTEDT [6], Philippe CIUCIU [1], Stephanie LEFEBVRE [2,3], Pierre THOMAS [2,3], Renaud LOPES [7], Renaud JARDRI [2,3*], Edouard DUCHESNAY [1]

1. NeuroSpin, CEA, Paris-Saclay, Gif-sur-Yvette, France

2. Univ. Lille, CNRS UMR 9193, Laboratoire de Sciences Cognitives et Sciences Affectives (SCALab-PsyCHIC), F-59000 Lille, France

3. CHU Lille, Pôle de Psychiatrie, Unité CURE, F-59000 Lille, France

4. CHU Lille, Imaging Dpt., Neuroradiology unit, Lille F-59000, France

5. Energy Transition Institute: VeDeCoM, France

6. Department of Radiation Sciences, Umeå University, Umeå, Sweden

7. Univ. Lille, INSERM, CHU Lille, U1171 - Degenerative and Vascular Cognitive Disorders, F-59000 Lille, France

Soumis

Abstract

Despite significant progress in the field, the detection of fMRI signal changes during hallucinatory events remains difficult and time consuming. In the first part, this paper proposes a machine-learning algorithm to automatically identify resting-state fMRI periods that precede hallucinations versus periods that do not. When applied to whole-brain fMRI data, state-of-the-art classification methods, such as support vector machines (SVM), yield dense solutions that are difficult to interpret without arbitrary thresholding. We propose to extend existing sparse classification methods by taking the spatial structure of brain images into account with structured sparsity using the total variation penalty. Based on this approach, we obtained reliable classifying performances associated with interpretable predictive patterns composed of two clearly identifiable clusters in speech-related brain regions. The variation in transition-to-hallucination functional patterns from one patient to another but also from one occurrence to the next (e.g., also depending on the sensory modalities involved) appears to be the major difficulty when developing effective classifiers. Consequently, in the second part, this paper aims at characterizing the variability within the pre-hallucination patterns across subjects and occurrences using an extension of principal component analysis with spatial constraints. The principal components (PCs) and the associated basis patterns shed light on the intrinsic structures of the variability present in the dataset. Such results are promising in the scope of innovative fMRI-based therapy for drug-resistant hallucinations, such as fMRI-based neurofeedback.

Keywords

Hallucinations, Schizophrenia, Real-time fMRI, Machine learning, Resting-state networks

Highlights

- Capturing hallucination-related fMRI signals is typically difficult and time consuming.
- Structured Machine Learning enables reliable real-time automatic fMRI decoding of hallucinations.
- Predictive patterns for the transition periods to hallucinations are clinically relevant.
- The functional pattern variability of transition-to-hallucination constrains classifiers' performance.

1. Introduction

Hallucinations are defined as abnormal perceptions in the absence of causative stimuli. These experiences, especially auditory hallucinations, constitute fundamental features of psychosis (64-80 % of lifetime prevalence among schizophrenia-diagnosed patients) and can lead to functional disability and a low quality of life (McCarthy-Jones et al. 2017).

Over the past years, auditory hallucinations have been studied in-depth using brain imaging methods, such as functional and structural magnetic resonance imaging (fMRI and sMRI), to decipher its underlying neural mechanisms. Numerous brain changes in patients suffering from auditory hallucinations have been extensively covered in a wide range of studies (e.g., Allen et al. 2008; Jardri et al. 2011; Bohlken et al. 2017; Sommer et al. 2008). Beyond location, the functional dynamics of the neural networks involved in auditory hallucinations have also been studied.

To address this important question, an increasing number of studies have focused on so-called Intrinsic Connectivity Networks (ICN) and their potential role in the onset of hallucinations (Alderson-Day et al. 2016; Northoff and Qin 2011). ICNs typically reveal interactions among brain regions when the subject is not engaged in any particular task. Frequently reported networks include the Default Mode Network (DMN), the Control Executive Network (CEN), the Salience Network (SAL) and the Sensorimotor Network (SMN) (Alderson-Day et al. 2016; Damoiseaux et al. 2006). Numerous studies assert that fluctuations in those ICNs are associated with the onset of hallucination periods. For instance, the emergence of hallucinations correlates with a disengagement of the DMN (Jardri et al. 2013). More recently, stochastic effective connectivity analyses revealed complex interactions among hallucination-related networks, DMN, SAL and CEN during the ignition, active phase, and extinction of hallucinatory experiences (Lefebvre et al. 2016).

Despite significant progress in the field, “capturing” the neural correlates of subjective mental events (such as hallucinations) remains a time-consuming task with multiple post-processing steps and analyses. However, recent progress in machine learning now paves the way for real-time automatic fMRI decoding of hallucination-related patterns. Such developments may have crucial impacts in implementing innovative fMRI-based therapy for drug-resistant hallucinations, such as fMRI-based neurofeedback (Arns et al. 2017; Fovet et al. 2015). During fMRI-based neurofeedback, brain activity is measured and fed back in real-time to the subject to help her/him progressively achieve voluntary control over her/his own neural activity. Precisely defining strong a priori strategies to choose the appropriate target brain area/network(s) for fMRI-based protocols appears critical. Interestingly, considering the rapid technical developments of fMRI techniques and the availability of high-performance computing, the pattern classification approach now appears as one of the potential strategies for fMRI-based neurofeedback sessions.

In this context, the feasibility of fMRI-based neurofeedback relies on robust and reliable classifying performances and the ability to detect hallucinations sufficiently early to allow the patients the necessary time to modulate their own cerebral activity (Fovet et al. 2016). Rather than detecting hallucinatory events per se, we aim to help patients become aware of the imminence of this experience based on

online detection of fMRI signal changes in key networks for the ignition of hallucinations. In this study, we thus specifically focus on the period preceding the occurrence of the hallucination, i.e., the few seconds corresponding to the brain transition from a resting-state to a full hallucinatory state. Interestingly, previous fMRI studies noted the existence of specific fMRI changes prior to hallucinations (Lennox et al. 1999; Hoffman et al. 2008; Diederen et al. 2010; Lefebvre et al. 2016)

Among the current machine-learning approaches available for fMRI analysis, Multi-voxel Pattern Analysis (MVPA), a supervised classification method, is gaining recognition for the potential to accurately discriminate between complex cognitive states (Fovet et al. 2016; Haxby et al. 2014). MVPA seeks to identify significantly reproducible spatial activity patterns differentiated according to mental states. Extending these methods to the prediction of the phenomena of transition towards hallucinations should provide better insight into the mechanisms of these subjective experiences. Thus, leveraging real-time pattern decoding capabilities and applying them in the case of hallucinations could lay the foundations for potential solutions for affected individuals.

Variation in transition-to-hallucination functional patterns from one patient to another (e.g., due to phenomenological differences) and from one occurrence to the next (e.g., depending on the modalities involved) appears to be the major potential shortcoming in developing an effective classifier. Indeed, such disparities may inexorably lead to a decrease in the decoding performances. Therefore, characterizing the variability within the pre-hallucination patterns across subjects and occurrences is highly desired. Principal Component Analysis (PCA) is one such unsupervised method that has been successfully applied to analyse the variability of a given dataset. The principal components (PCs) and the associated basis patterns shed light on the intrinsic structures of the variability present in the dataset. This unsupervised approach is complementary to the supervised approach described above, as it can help with interpreting the classification performances.

Here, we apply both supervised and unsupervised machine-learning methods to an fMRI dataset collected during hallucinatory episodes. The goal of this paper is two-fold: i) To predict the activation patterns preceding hallucinations using a supervised analysis, ii) To uncover the variability in activation patterns during the emergence of hallucinations using unsupervised analysis. The goals of these two analyses appear completely complementary in the context of future fMRI-based clinical and therapeutic applications.

2. Methods

2.1 Participants and experimental paradigms

The population is composed of 37 patients with schizophrenia (DSM-IV-TR criteria, average age = 35.23 years, 10 females/27 males) suffering from very frequent multimodal hallucinations (i.e., more than 10 episodes/hour). Participants were enrolled through the FR2SM network (Fédération Régionale de Recherche en Santé Mentale), grouping all the private/public institutions for mental health in the Hauts-de-France region (62% of the participants were hospitalized at the time recruited, 38% received outpatient care). This sample presents a partial overlap with previous works from our team (Lefebvre et al. 2016; Leroy et al. 2017). Clinical characteristics of the recruited subjects are summarized in **Table 1**.

fMRI was acquired at rest. Participants were asked to lie in the scanner in a state of wakeful rest with eyes closed. The subjects experienced an average of 4 hallucinatory episodes per scan. The patients' states at different acquisition times were labelled using a semi-automatic difficult procedure, as described in (Jardri et al. 2013; Lefebvre et al. 2016; Leroy et al. 2017) and assigned to one of the following four categories: transition towards hallucinations (trans), on-going hallucinations (on), no hallucinations (off) and end of hallucinations (end). This labeling task is a non-straightforward two-steps strategy: The first step is a data-driven analysis of the fMRI signal using an ICA in the spatial domain. The second step is the selection of the ICA components associated with possible sensory experiences that occurred while scanning. This pipeline is said to be semi-automatic since it combined: (a) an automatic denoising part, based on classifiers described in de Martino et al, 2007, and (b) a manual and time-consuming part, with the use of the immediate post-fMRI interview conducted with the patient, in which sensory modalities, number of episodes, phenomenological features of the experiences were specified.

The study was approved by the local ethical committee (CPP Nord-Ouest France IV), and written informed consent was obtained for each participant enrolled in the study.

Age (mean)	35.8 ± (9.8) years
Sex	10 F / 27 M
CGI (mean ± sd)	4.2 ± (1.6)
Dose of anti-psychotic treatment (EqOZ) (mg/d)	42.5 ± (22,4)
PANSS (mean ± sd)	82.4 ± (20.3)
AHRS (mean ± sd)	26 ± (7)
Average number of hallucination episodes per patient	4
Number of patients experiencing hallucinations (by modality) during the fMRI session	
Auditory	32
Tactile	7
Visual	6
Gustatory	2
Olfactory	2

Table 1. Clinical characteristics of the recruited samples. *CGI = Clinical Global Impressions Scale; EqOZ = Equivalent Olanzapine; PANSS = Positive and Negative Syndrome Scale; AHRS = Auditory Hallucination Rating Scale.*

2.2 Imaging parameters

The participants underwent an 11-minute anatomical T_1 weighted 3D multishot turbo-field-echo scan (3 T Philips Achieva X-series, with an 8-elements SENSE head coil). The field-of-view was 256 mm^2 with a voxel resolution of 1 mm in all directions. Participants also underwent a blood oxygen level-dependent (BOLD) fMRI session. The parameters of the 3D-PRESTO SENSE sequence were field-of-view 206 x 206 x 153 mm^3, TE = 9.6 ms, TR = 19.25 ms, EPI-factor = 15, flip angle = 9°, dynamic scan time = 1000 ms. Because of the multishot nature of the PRESTO sequence, the TR is not equivalent to the scan duration. Each fMRI session consisted of 900 volumes collected for a total acquisition time of 15 min.

2.3 fMRI Preprocessing

The anatomical and functional data were pre-processed using SPM12 (WELLCOME, Department of Imaging Neuroscience, London, UK) running on MATLAB R2016a (MathWorks, Inc., Natick, Massachusetts, USA). To control motion-induced artefacts, point-to-point head motion was estimated for each subject (Van Dijk et al. 2012). Excessive head motion (cumulative translation or rotation >3 mm or 3°) was applied as an exclusion criterion. Applying this filter, one patient was excluded from the analysis. Signal preprocessing consisted of motion correction (realignment of fMRI volumes) and voxelwise linear detrending. Given that we excluded subjects where motion was too important, we estimate that noise had a contained and therefore tolerable impact on the remaining subjects. Moreover, concerning the low frequency trends in the fMRI signal, we believe that these slow signal intensity drifts do not create excessive artefacts over the signal given that we are dealing with very short periods of transition. Hence, applying a linear detrending is likely sufficient.

Then, we performed a coregistration of the individual anatomical T1 images to the functional images and spatial normalization to the Montreal Neurological Institute (MNI) space using DARTEL based on segmented T1 scans. We did not perform any spatial smoothing step in the preprocessing pipeline. A brain mask was extracted automatically and used to restrict voxels considered in the subsequent steps to 67,665 voxels.

2.4 Computation of samples

Prior to training classifiers, the first step involved computing samples from the fMRI signal. The intention was to convert the fMRI signal into vectors of features reflecting the pattern of activity across voxels at a point in time. We opted against creating the samples directly from the fMRI signal. Instead, we created the samples by estimating the activity within each voxel using a linear model. The design of such a model was a crucial part of the learning process. We used a General Linear Model (GLM) to estimate the activity within each voxel. From each set of consecutive images under a pre-hallucination state (“trans” periods) or “off” state, we created one sample. On average, each “trans” or “off” state lasted for 8 consecutive EPI volumes, which appears sufficient to estimate activity. Based on the GLM, we regressed the fMRI signal time course on a linear ramp function for each set of scans. (See **Figure 1. A** as an example of regression within one voxel). This choice is based on the hypothesis that activation in some regions presents a ramp-like increase during the time preceding the onset of hallucinations. A sigmoid activation in some regions prior to the occurrence of an hallucination is potentially more realistic

than a ramp-like activation. However, in order to fit a sigmoid function to a set of points, two parameters need to be estimated. Given the fact we only had a limited set of 8 consecutive pre-hallucination EPI volumes, fitting a sigmoid would have meant leaving only 6 degrees of liberty. Given the arguments above and our wish to reach the highest possible level of robustness, we thus chose to use a ramp model in these the conditions.

The Figure 1. A represents the evolution of signal intensity in one single voxel, in the 8 consecutive volumes of a pre-hallucination period of a subject. In this specific voxel, the signal presents a ramp-like increase during the pre-hallucination period.

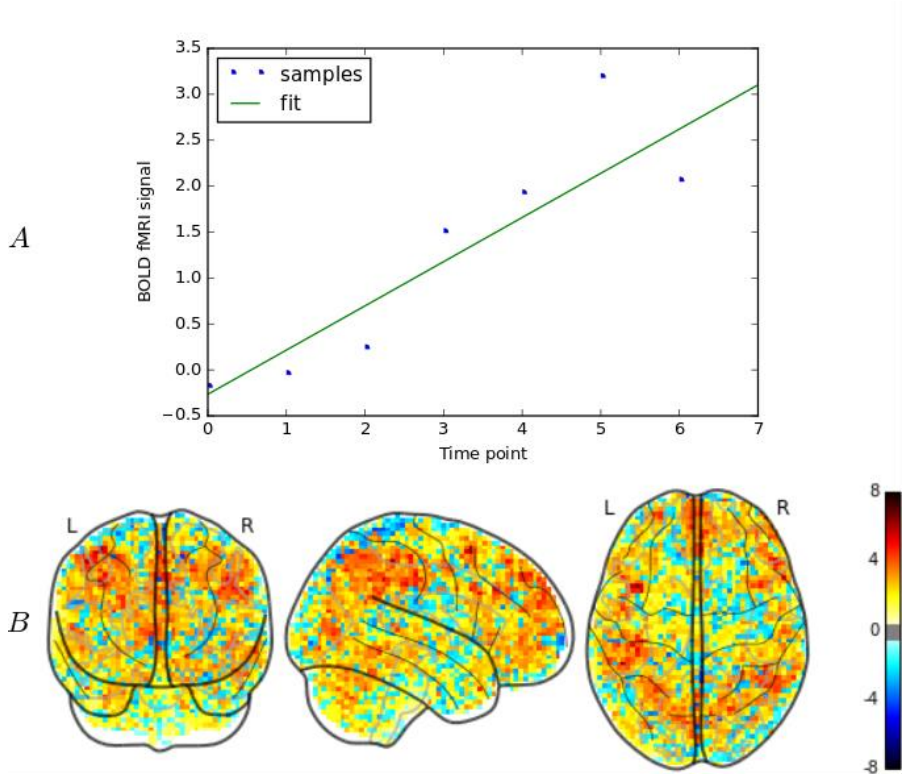


Figure 1: *A. Regression of the fMRI signal time course of a voxel on a linear ramp function (fit is represented in green). B: Sample created from one set of consecutive pre-hallucinations scans. The features are the T-statistic values associated with the coefficients of the regression in each voxel.*

2.5 Supervised analysis

All analyses were performed in Python using the scikit-learn toolbox (Pedregosa et al. 2011) and the pylearn-parsimony package (https://github.com/neurospin/pylearn-parsimony). Given the slow, partially manual and interview-intensive nature of the labelling pipeline (see Jardri et al. 2013), we constructed in parallel an algorithm to detect a transition-to-hallucination state in a real-time, automated fashion exclusively relying on imaging data. We focused the analysis on the transition towards a hallucination state (trans) with the intention of distinguishing it from the resting-state activity (off).

2.5.1 Classifiers

Learning with hundreds of samples (376) using high-dimensional data ($7x10^5$ voxels) is associated with a high risk of overfitting on the training subjects, leading to poor performances on independent subjects. Such issues of replicability can be addressed using state-of-the-art regularized learning algorithms, such as linear support vector machine (SVM), and based on a ridge (l_2) penalty or elastic net (Zou & Hastie 2005), and the combination of the sparsity inducing lasso (l_1) and the ridge penalties.

In this study, we compared two different linear classifiers for binary classification. First, we used a regular Linear SVM, based on a ℓ2 (Ridge) penalty on the coefficients vectors. The role of the Ridge penalty is to shrink the coefficients toward zero to control the variance of fitted coefficients. However, the SVM classifier cannot select significant variables and rather, tend to produces dense patterns of predictors that are difficult to interpret without arbitrary thresholding. In the context of predictive signature discovery, it is crucial to understand the brain activation patterns that underpin the prediction. We therefore seek an approach that selects a reduced number of predictive regions. Feature selection methods, such as recursive feature elimination (RFE) (Guyon et al. 2002), have been used to select a reduced set of predictors (De Martino et al. 2008). However, since they are prone to local minima, those ad hoc heuristics tend to be replaced by sparse models based on convex minimization problems that simultaneously optimize the prediction performances while performing the feature selection.

Another solution to obtain a limited number of predictors is the use of ℓ1-regularized classifiers (Lasso), that produce sparse patterns of predictors by enforcing many voxels to have zero-weights. The combination of Ridge and Lasso penalties in ElasticNet (Friedman et al, 2010) promotes sparse models while still maintaining the regularization properties of the ℓ2 penalty. However, despite the fact that Lasso or ElasticNet classifiers have often been advocated as leading to more interpretable models, they generally lead to scattered and unstable weight patterns (Grosenick et al. 2013; Dubois et al. 2014a).

Therefore, we propose to utilize the benefit of the known structure of brain fMRI images to force the solution to adhere to biological priors, producing more plausible interpretable solutions. Indeed, MRI data is naturally encoded on a 3-dimensional grid. Some voxels are neighbours, whereas others are not. Therefore, structured sparsity can be obtained by combining l_1, l_2 and Total Variation (TV) penalties. Such combination of penalties will enforce the spatial smoothness of the solution while segmenting predictive regions from the background.

Consequently, as a second classifier, we used logistic regression with 3 types of regularization penalties: l_1, l_2 and TV (Dubois et al. 2014), which was denoted TV-Elastic-net (TV-Enet). The l_1 and l_2 penalties served the purpose of addressing overfitting induced from the MRI data's high intrinsic dimensionality. The TV penalty regularizes the solution while taking advantage of the spatial 3D structure. Together, these penalties enable the generation of a coherent, parsimonious and interpretable weight map. Moreover, these penalties provide a segmentation of the predictive weight map into spatially contiguous parcels with constant values, which is a highly desirable characteristic in the scope of predictive signature discovery.

2.5.2 Performance metric, cross-validation and model selection

Performance was evaluated through a double cross-validation pipeline. The double cross-validation process consists of two nested cross-validation loops. In the outer (external) loop of double cross-validation, we employed a Leave-One-Subject-Out pipeline where all subjects except one are referred as the training data. The remaining subject was used as the test data. The test sets were exclusively used for model assessment, whereas the train sets were used in the inner 5-fold cross-validation loop for model fitting and model selection. Classifier performances were assessed by computing the balanced accuracy, sensitivity and specificity with which test samples were classified. Sensitivity is defined as the ability to identify the transition towards hallucination state (trans), whereas specificity evaluates the ability to identify the resting-state activity (off). The balanced accuracy score is defined as the average of the sensitivity and specificity. We also implemented the receiver operating characteristic (ROC) curve for each classifier, from which the area under the curve (AUC) was computed.

2.5.3 Result significance

To measure the significance of the prediction scores for both classifiers, we used an exact binomial test while leveraging a paired two-samples t-test to compare the decoding performances of the two classifiers.

2.5.4 Predictive pattern

To analyse the brain regions that drive the prediction, we refitted the model on all samples of the dataset and extracted the associated discriminative weight map. This weight map revealed the spatial patterns that best discriminate the two cognitive states (trans and off). The weights revealed the relative contribution of each voxel to the decision function. Positive weights indicated positive contribution towards predicting *trans* state*,* whereas negative weight signalled a positive contribution towards predicting the *off state.*

2.6 Unsupervised Analysis

2.6.1 Decomposition method

Subsequently, in addition to supervised analysis, we conducted an extensive analysis of the data using unsupervised machine learning. The goal was to characterize the variability within pre-hallucination scans. PCA can extract the significant mode of variation from high-dimensional data. However, its interpretability remains limited. Indeed, the components produced by PCA are often noisy and exhibit no visually meaningful patterns. Nonetheless, our ultimate goal was to understand the variability in the form of intelligible patterns. In this context, we used SPCA-TV (Sparse Principal Component Analysis-Total Variation), which is an extension of the regular PCA with l_1, l_2, and TV penalties on the PCA loadings, promoting the formation of structured sparse components that are relevant in a neuroscientific scope (de Pierrefeu et al. 2016). We hypothesized that the principal components extracted with SPCA-TV could uncover major trends of variability within pre-hallucination samples. Thus, the principal components might reveal the existence of subgroups of hallucinations, notably according to the sensory modality involved (e.g., vision, audition, etc.).

From the 376 samples, we retained the 210 elements corresponding to pre-hallucinations samples. We applied SPCA-TV to these 210 samples and interpreted the resulting principal components.

Additionally, we computed the explained variance of each component yielded by SPCA-TV and investigated whether those components are really capturing a signature of the cognitive process involved in the onset of hallucinations. To do so, we projected each activation maps, “off” and “trans” samples, in the principal components basis and use the subsequent associated scores to decode the mental state of subjects. We used an SVM using the same cross-validation pipeline described in the supervised analysis method section.

3. Results

3.1 Supervised analysis

3.1.1 Classification performances

Classification results are presented in **Table 2**. Classification of *resting state (*i.e*., non-hallucination) patterns* (off) versus *transition towards hallucinations patterns* (trans) achieved an above chance level decoding performances with both methods. Using the SVM classifier, we obtained an AUC of 0.73 and a balanced accuracy of 0.73 with a specificity of 0.78 and a sensitivity of 0.67. When using the TV-Enet classifier, we obtained an AUC of 0.79 and a balanced accuracy of 0.74 with a specificity of 0.76 and a sensitivity of 0.71.

Classifier	AUC	Acc	Spe	Sen
SVM	0.73	0.73	0.78	0.67
TV-Enet	0.79*	0.74	0.76	0.71

Table 2 *- The performance of the classifiers. Prediction accuracies: Sensitivity (Sen, recall rate of trans samples), Specificity (Spe, recall rate of off samples) and Balanced accuracy (Acc): (Sen+Spe)/2; AUC indicates area under the curve. We tested whether the scores obtained with SVM are significantly different from scores obtained with TV-Enet. Significance notations: * = p<0.01.*

TV-Enet yields significantly increased AUC and balanced accuracy compared with the SVM (T = 30.27, p = 3.3x10-27 and T = 11.58, p = 1.06x10-13, respectively).

Since the 37 patients included in this study are suffering of multimodal hallucinations (see **Table 1**), we also evaluated the performance of prediction of the TV-Enet model on two subsamples; On one hand, the 35 subjects suffering from auditory hallucinations among other modalities, and on the other hand, the 5 subjects without any auditory hallucinations (**Table 3**).

For the cohort of patients experiencing auditory hallucinations, we obtained an AUC of 0.80 and a balanced accuracy of 0.75, with a specificity of 0.76 and a sensitivity of 0.73. For the cohort of patients that are not experiencing auditory hallucination, we obtained decreased prediction performances: an AUC of 0.75, a balanced accuracy of 0.63, with a specificity of 0.74 and a sensitivity of 0.55.

Presence of Auditory Hallucinations	AUC	Acc	Spe	Sen
Yes	0.80	0.75	0.76	0.73
No	0.75	0.63	0.74	0.55

Table 3 *- Prediction performances of TV-Enet on the subgroup of patients experiencing auditory hallucinations among other modalities (top row) and on the subgroup of patients that are not experiencing auditory hallucinations. (bottom row)*

3.1.2 Predictive weight maps

When using the regular SVM classifier, the relevance of the obtained discriminative weight maps is limited (**Figure 2. A**). The whole brain seems to contribute to the prediction. It is clinically challenging to interpret the weight map. The TV-Enet classifier yields a more coherent weight map with two defined stable predictive clusters (**Figure 2. B**). The details of these two clusters are described in **Table 4**.

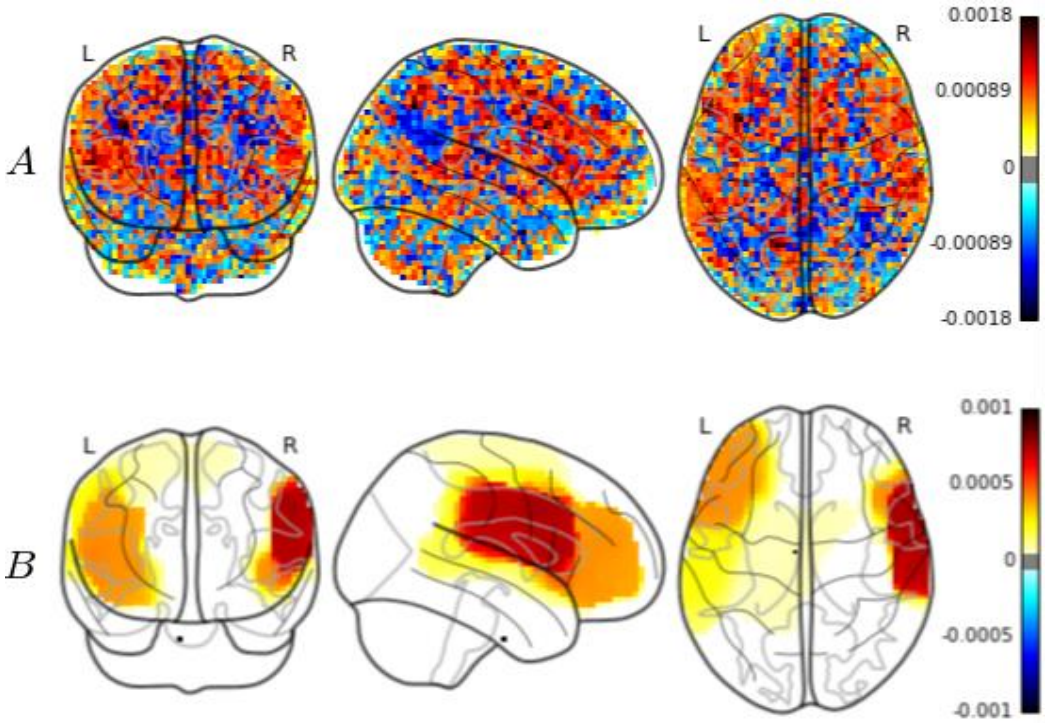


Figure 2*: A: Linear support vector machine (SVM) and B: TV-Enet predictive weight maps.*

3.2 Unsupervised analysis

3.2.1 Relevance of components

The first component explains 2.5% of the variance. The second component explains 1.4% of the variance. The third component explains 0.09% of the variance The fourth component explains 0.05% of the variance. The prediction of mental state based on the scores associated with each component yields a significant decoding performance: the classifier was able to distinguish “trans” samples from “off” samples, with an AUC of 65%, a recall mean of 65%, a sensitivity of 68% and a specificity of 64%.

3.2.2 Component weight maps

The components extracted with the SPCA-TV method are of great interest from a clinical point of view (see **Figure 3**). They reveal structured interpretable patterns of variability within the different pre-hallucinations periods from our sample. Details regarding the clusters present in each principal component are provided in **Table 5**.

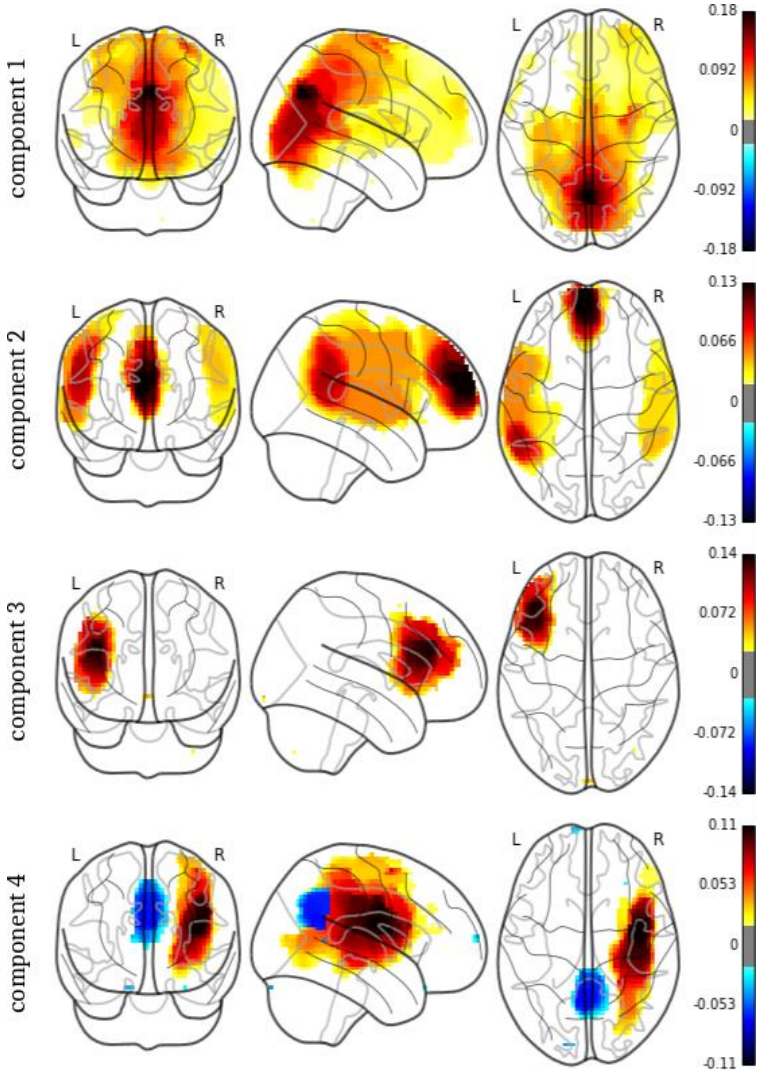


***Figure 3**: SPCA-TV Principal Components Note that the sign is arbitrary.*

Clusters	Center in MNI coordinates (x,y,z)	Cluster size (voxels)	Cluster mean	Cortical Regions involved	Laterality
1	(53,0,15)	3,541	$4.1e^{-4}$	Precentral Gyrus, Postcentral Gyrus, Inferior Frontal Gyrus, Central Opercular Cortex, Anterior and Posterior Supramarginal Gyrus, Insular Cortex, Frontal Pole, Middle Frontal Gyrus, Planum Temporale, Temporal Pole, Superior Temporal Gyrus	Right
2	(-36,0,28)	10,134	$2.0e^{-4}$	Precentral Gyrus, Frontal Pole, Postcentral Gyrus, Middle Frontal Gyrus, Superior Frontal Gyrus, Insular Cortex, Frontal Orbital Cortex, Central Opercular Cortex, Inferior Frontal Gyrus	Left

Table 4 *- Supervised analysis: The clusters in the discriminative weight map.*

PC	Clusters	Center in MNI coordinates (x,y,z)	Cluster size (voxels)	Cluster mean	Cortical Regions involved	Laterality
1	1	(8,-28,27)	22,002	0.05	Precuneus Cortex, Posterior Cingulate Gyrus, Precentral Gyrus, Postcentral Gyrus, Superior Frontal Gyrus, Frontal Pole, Lingual Gyrus	Right and Left
2	1	(-52,-25,28)	4,249	0.05	Postcentral Gyrus, Precentral Gyrus, Anterior and Posterior Supramarginal Gyrus, Angular Gyrus, Middle Frontal Gyrus, Superior Temporal Gyrus, Middle Temporal Gyrus	Left
2	2	(56,-18,25)	2,716	0.03	Postcentral Gyrus, Precentral Gyrus, Anterior and Posterior Supramarginal Gyrus, Angular Gyrus, Superior Temporal Gyrus,	Right
2	3	(-1,48, 25)	1,988	0.07	Frontal Pole, Paracingulate Gyrus, Anterior Cingulate Gyrus	Right and Left
3	1	(-41,25,1)	1,857	0.08	Middle Frontal Gyrus, Frontal Pole, Inferior Frontal Gyrus, Frontal Operculum Cortex, Insular Cortex	Left
4	1	(37,-23,26)	5,022	0.05	Precentral Gyrus, Postcentral Gyrus, Middle Frontal Gyrus, Insular Cortex, Superior Parietal Lobule, Angular Gyrus, Posterior Supramarginal Gyrus	Right
4	2	(1,-52,30)	1,173	0.04	Precuneus Cortex, Posterior Cingulate Gyrus	Right and Left

Table 5 *- Unsupervised analysis: The clusters in the weight maps associated with the first four PCs.*

4. Discussion

Here, we wanted to automatize the detection of specific functional patterns preceding hallucination occurrences in participants scanned at rest. First using supervised analysis, we found evidence of prediction scores with a reliable level of significance. Our prediction of the emergence of hallucinations appears accurate and yields highly interpretable associated weight maps. Second, using unsupervised analysis, we characterized the variability of pre-hallucinations patterns across occurrences and subjects in the form of intelligible components.

4.1 Supervised analysis

4.1.1 Decoding Performances

The present findings indicate that the two classification algorithms were able to significantly detect pre-hallucination patterns in brain activity at rest. Crucially, spatial regularization (TV) combined with the elastic net penalty significantly improved the prediction performances (increased AUC) and provided more balanced specificity and sensitivity. Indeed, traditional SVM naturally tends to allocate the “off” response, which subsequently leads to a good specificity but a reduced detection rate (sensitivity) of patterns preceding the occurrence of hallucinations.

The studied cohort contains patients who suffer from complex multimodal hallucinations. Thus, hallucinations captured during acquisition can be very heterogenous across subjects but also across occurrences. When evaluating the classifier’s performance on non-auditory hallucinations only, we obtained degraded prediction scores as opposed to the ones obtained on patients experiencing auditory hallucination among others modalities. This finding is to be expected since the learning of the model is conducted on 37 subjects, of which 32 exhibit auditory experiences. Therefore, our predictive model seems to be more specific to the prediction of auditory hallucinations than any other modalities. Considering the above, the inter-subject decoding performance achieved should be considered reasonably satisfactory.

Furthermore, a comparison to the seminal procedure used for labelling the scans (Jardri et al. 2013) places our result in perspective. Compared with that procedure that required the incorporation of information from post-fMRI interviews with patients into the labelling process, the proposed machine learning-based method is fully automatic, relying exclusively on the imaging data. Moreover, the learned model could be applied in real-time during data acquisition.

Despite the challenge which constitutes gathering so many subjects in a fMRI hallucinations capture dataset (n = 37 subjects), we expect that increasing the sample-size may improve performances. We believe that our prediction model can still gain additional useful information from more data. Even if this is difficult to define a clear cut-line for clinical applications, an accuracy of 80% could be considered as acceptable to be use in the scope of fMRI-based therapy for drug-resistant hallucinations, such as fMRI-based neurofeedback. 80% stays an arbitrary threshold here, but it is considered as satisfactory since detecting ⅘ hallucinations in a clinical setting is already promising.

4.1.2 Predictive weight map interpretation

The predictive maps obtained with the SVM method are dense and difficult to interpret without arbitrary thresholding. Even though the prediction performance is relatively good, a physician will never draw a conclusion from such a black-box model as presented in **Figure 2. A** in a clinical setting. Understanding the brain activation patterns that drive the prediction is crucial. In addition, the predictive map obtained with TV-Enet is considerably more interpretable given that it provides a smooth map composed of two clearly identifiable regions. Interestingly, these regions, especially speech-related brain regions, were previously shown to be involved in hallucinations (Ćurčić-Blake et al. 2017).

First, the two large stable predictive fronto-temporal clusters appear consistent with what we currently know on the networks involved in auditory hallucinations. Indeed, numerous studies have highlighted abnormal resting-state functional connectivity among some temporo-parietal, frontal and subcortical regions in patients with auditory hallucinations (Alderson-Day et al. 2015; Allen et al. 2008). Otherwise, patients experiencing auditory hallucinations while in the MRI scanner (in so-called fMRI "capture" studies) demonstrated significantly increased activation in the Broca's area, insula, inferior frontal gyrus, left middle and superior temporal gyrus, left inferior parietal lobule and left hippocampal region (Jardri et al. 2011). Second, the right cluster identified in our study also emphasizes the role of the right-sided homologues of the classical speech-related areas (i.e., right inferior frontal gyrus, right superior temporal and supramarginal gyrus) in auditory hallucinations as previously described in the literature. It has been hypothesized that activity in these regions, especially the insula and the right homologue of the Broca's area, is associated with the occurrence of auditory hallucinations (Jardri et al. 2011; Sommer et al. 2008), whereas language production in a natural context predominantly activates left-lateralized frontal and temporal language areas. The role of right-sided speech-related areas in the pathophysiology of auditory hallucinations was also mentioned by Mondino et al. 2016. By neuromodulating a speech-related fronto-parietal network, these authors demonstrated that a reduction in resting-state functional connectivity between the left temporo-parietal junction and right inferior frontal areas could be measured, and this reduction was associated with a significant reduction in the severity of the hallucinations.

The higher rate of auditory hallucinations in this sample may account for the speech-related regions identified in the predictive map. This explains the fact that these regions are crucial in the prediction process of pre-hallucinations patterns. Given the fact that 32 of the 37 patients suffered from auditory hallucinations, among other modalities, it is not surprising that such regions previously associated with auditory-verbal hallucinations are identified as highly predictive. Reversely, since the number of patients suffering from hallucinations in other modalities (visual, tactile and olfactory) is limited, their weights in the classifier appear minimal as opposed to auditory hallucinations predictive weights. Consequently, this explains the degraded prediction performances obtained on non-auditory hallucinations, presented in **Table 3**.

Classification algorithms may ideally benefit from a modality-specific training on more restrictive datasets of patients hallucinating in just one sensory modality. However, even if this could be easily performed for voice-hearing, this appears quite challenging for other modalities. Taken together, these results confirm that adding a penalty to account for the spatial structure of the brain seems relevant in

fMRI captures given that it significantly improves the classifier performance and results in clinically interpretable weight maps.

Here, we demonstrated that supervised classification methods can accurately predict the imminence of a hallucinatory episode. Thus, leveraging real-time pattern decoding capabilities and applying them in the case of hallucinations could lay the foundation for alternative solutions for affected patients in the near future, such as fMRI-based neurofeedback.

4.2 Unsupervised analysis

4.2.1 Relevance of weight maps

The total amount of explained variance is surprisingly low. Indeed, the activation maps of resting state fMRI data preceding hallucinations are very noisy and only a minor part of its variability can be captured.

However, when predicting the mental state of subjects based on SPCA-TV scores, the decoding accuracy is significant. Naturally, the performance is decreased compared to the performance obtained in the supervised part of this paper, which is expected since we are losing some information from the compression of the 67,655 features into 4 scores. However, the fact that we can still significantly distinguish pre-hallucination samples from resting state samples using those 4 component scores reveals that they make sense and are specifically related to hallucinations. Consequently, although the explained variance is low due to the resting state nature of the data, the components are relevant and capture the cognitive processes involved in the onset of hallucinations.

4.2.2 Weight map interpretation

The variability of pre-hallucination patterns across occurrences and subjects are represented in the form of intelligible components.

The first PC mainly includes weights in the precuneus cortex and the posterior cingulate cortex. The posterior cingulate cortex, which is part of the DMN, is associated with auditory hallucinations (Rotraska-Jagiela et al., 2010), We believe that this component may have captured the visual pathways typically involved in the occurrence of visual hallucinations.

The second PC is composed of one activation cluster in the paracingulate gyrus and the anterior cingulate gyrus and two symmetric bilateral activation clusters in the temporal cortex. This fronto-temporal component appears compatible with the dysconnectivity hypothesis of hallucinations and may have captured processes at the roots of auditory hallucinations. Interestingly, some processes involved in the occurrence of hallucinations, such as the monitoring of inner speech processes and error detection, are classical functions of the anterior cingulate cortex included in this component (Allen et al. 2008; Mechelli et al. 2007). This second PC yield regions classically involved in inhibition (paracingulate gyrus, anterior cingulate gyrus) (Allen et al. 2008; Mechelli et al. 2007). The severity of auditory hallucinations has been found inversely related to the strength of the functional connectivity between the temporal-parietal junction, the Anterior Cingulate Cortex (ACC) and the amygdala (Vercammen et al. 2010). This ACC dysconnectivity was supposed to drive the external misattribution observed during

auditory hallucinations (Allen et al. 2007; Mechelli et al. 2007), and might explain global inhibition impairments in hallucinations' pathophysiology (Jardri et al. 2016), that may account for this feature beyond the schizophrenia-spectrum, as for instance in LSD-induced hallucinations (Schmidt et al. 2017).

The third PC reveals a cluster in the frontal gyrus and the anterior insula. These regions are important for speech production, encompassing the well-known Broca's area (Small & Hickok 2016), and are involved in auditory hallucinations (Jardri et al. 2011; Sommer et al. 2008).

Finally, the fourth PC includes two clusters of opposing signs. On the right hemisphere, there is a large activation cluster that involves the temporo-parietal junction and a deactivation cluster that involves the precuneus cortex and the posterior cingulate gyrus. Interestingly, this PC reveals an activation of brain regions involved in auditory hallucination-related processes and in self-other distinction, such as the right temporo-parietal junction (Jardri et al. 2011; Decety & Lamm 2007; Plaze et al. 2015), together with a deactivation of key nodes from the DMN, including the posterior cingulate cortex, medial prefrontal cortex, medial temporal cortex and lateral parietal cortex (Buckner et al. 2008). Our results appear fully compatible with recent fMRI-capture findings demonstrating that aberrant activations of speech-related areas concomitant to hallucinatory experiences follow complex interactions between ICNs, such as the DMN and the CEN (Lefebvre et al. 2016). A disengagement of the DMN during goal-directed behaviours has been seminally evidenced in the resting-state literature (Raichle et al. 2001; Lefebvre et al. 2016; Fox et al. 2005), and similar mechanisms might be involved in hallucinatory occurrences (Jardri et al. 2013; Leroy et al., 2017). Such fluctuations of the ICNs are thus thought to be highly involved in the transition from a resting state to an active hallucinatory state.

4.3 Perspectives

In the present study, we chose to train a classifier to specifically detect periods preceding the occurrence of hallucinations (i.e., "trans" periods). As mentioned earlier, several studies demonstrated that this period is potentially associated with specific brain activations. Diederen et al. 2010 demonstrated reduced activity in the left parahippocampal gyrus, the left superior temporal gyrus (STG), the medial frontal gyrus and the right inferior frontal gyrus (IFG) prior to auditory hallucinations. A study by Hoffman & Hampson 2011 also revealed increased activation in the right posterior temporal area compared with its right homologue in the same period. The specific patterns observed in the "trans" period probably correspond to the triggering mechanisms of auditory hallucinations, which may have a component in memory (Ćurčić-Blake et al. 2017) and constitute a very interesting target for neurofeedback therapies. Real-time recognition of the "trans" period using the TV-Enet classifier could enable the delivery of visual information (i.e., visual feedback) about the imminent onset of hallucinations to the participant during an fMRI-NF session. Such a procedure could help the subject learn effective coping strategies to prevent the occurrence of hallucinations. Similarly, recent effective connectivity findings revealed that the extinction of auditory hallucinations ("end" periods) was associated with a takeover of the fronto-parietal CEN (Hoffman & Hampson 2011; Lefebvre et al. 2016). This finding suggests that terminating auditory hallucinations is a voluntary process that could benefit from and be reinforced by fMRI-NF learning. We believe that such fMRI-NF based on the TV-Enet classifier could reduce the associated distress based on improvement in the feelings of control and self-efficacy.

One of the major limits of such fMRI-based therapies remains the accessibility and cost of the equipment. It appears fundamental to develop less complex devices as potential second-line treatments for hallucinations, such as near-infrared spectroscopy (NIRS). From this technological transfer perspective, the discriminative maps obtained using the TV-Enet classifier also appear advantageous given that the identified clusters are cortical regions with activity that is easily measured with NIRS.

5. Conclusion

Because hallucinations were frequently multimodal in the sample of patients recruited for this study, we could expect more disparities in the functional patterns associated with their complex hallucinations and the transition towards this state compared with pure auditory experiences. In this context, the significant inter-subject decoding performances obtained appear satisfactory and are promising for future fMRI-based therapy for drug-resistant hallucinations.

Acknowledgements: This study was supported by the *Programme Hospitalier de Recherche Cliniqu* (PHRC-N2011 19-02 MULTIMODHAL awarded to RJ) and the *Agence Nationale de la Recherche* (ANR-16-CE37-0015 INTRUDE awarded to RJ).

Conflicts of interest: none

References

Alderson-Day, Ben, Kelly Diederen, Charles Fernyhough, Judith M. Ford, Guillermo Horga, Daniel S. Margulies, Simon McCarthy-Jones, et al. 2016. “Auditory Hallucinations and the Brain’s Resting-State Networks: Findings and Methodological Observations.” Schizophrenia Bulletin 42 (5): 1110–23.

Alderson-Day, Ben, Simon McCarthy-Jones, and Charles Fernyhough. 2015. “Hearing Voices in the Resting Brain: A Review of Intrinsic Functional Connectivity Research on Auditory Verbal Hallucinations.” Neuroscience and Biobehavioral Reviews 55 (August): 78–87.

Allen, Paul, Edson Amaro, Cynthia H. Y. Fu, Steven C. R. Williams, Michael J. Brammer, Louise C. Johns, and Philip K. McGuire. 2007. “Neural Correlates of the Misattribution of Speech in Schizophrenia.” The British Journal of Psychiatry: The Journal of Mental Science 190 (February): 162–69.

Allen, Paul, Frank Larøi, Philip K. McGuire, and Andrè Aleman. 2008. “The Hallucinating Brain: A Review of Structural and Functional Neuroimaging Studies of Hallucinations.” Neuroscience and Biobehavioral Reviews 32 (1): 175–91.

Arns, M., J-M Batail, S. Bioulac, M. Congedo, C. Daudet, D. Drapier, T. Fovet, et al. 2017. “Neurofeedback: One of Today’s Techniques in Psychiatry?” L’Encéphale 43 (2): 135–45.

Bohlken, M. M., K. Hugdahl, and I. E. C. Sommer. 2017. “Auditory Verbal Hallucinations: Neuroimaging and Treatment.” Psychological Medicine 47 (2): 199–208.

Buckner, Randy L., Jessica R. Andrews-Hanna, and Daniel L. Schacter. 2008. “The Brain’s Default Network.” Annals of the New York Academy of Sciences 1124 (1): 1–38.

Ćurčić-Blake, Branislava, Judith M. Ford, Daniela Hubl, Natasza D. Orlov, Iris E. Sommer, Flavie Waters, Paul Allen, et al. 2017. “Interaction of Language, Auditory and Memory Brain Networks in Auditory Verbal Hallucinations.” Progress in Neurobiology 148 (January): 1–20.

Damoiseaux, J. S., S A R, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, and C. F. Beckmann. 2006. “Consistent Resting-State Networks across Healthy Subjects.” Proceedings of the National Academy of Sciences 103 (37): 13848–53.

Decety, Jean, and Claus Lamm. 2007. “The Role of the Right Temporoparietal Junction in Social Interaction: How Low-Level Computational Processes Contribute to Meta-Cognition.” The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry 13 (6): 580–93.

De Martino, Federico, Giancarlo Valente, Noël Staeren, John Ashburner, Rainer Goebel, and Elia Formisano. 2008. “Combining Multivariate Voxel Selection and Support Vector Machines for Mapping and Classification of fMRI Spatial Patterns.” NeuroImage 43 (1): 44–58.

Diederen, Kelly M. J., Sebastiaan F. W. Neggers, Kirstin Daalman, Jan Dirk Blom, Rutger Goekoop, René S. Kahn, and Iris E. C. Sommer. 2010. “Deactivation of the Parahippocampal Gyrus Preceding Auditory Hallucinations in Schizophrenia.” The American Journal of Psychiatry 167 (4): 427–35.

Dubois, Mathieu, Fouad Hadj-Selem, Tommy Lofstedt, Matthieu Perrot, Clara Fischer, Vincent Frouin, and Edouard Duchesnay. 2014a. “Predictive Support Recovery with TV-Elastic Net Penalty and Logistic Regression: An Application to Structural MRI.” In 2014 International Workshop on Pattern Recognition in Neuroimaging. doi:10.1109/prni.2014.6858517.

———. 2014b. “Predictive Support Recovery with TV-Elastic Net Penalty and Logistic Regression: An Application to Structural MRI.” In 2014 International Workshop on Pattern Recognition in Neuroimaging. doi:10.1109/prni.2014.6858517.

Fovet, Thomas, Renaud Jardri, and David Linden. 2015. “Current Issues in the Use of fMRI-Based Neurofeedback to Relieve Psychiatric Symptoms.” Current Pharmaceutical Design 21 (23): 3384–94.

Fovet, Thomas, Natasza Orlov, Miriam Dyck, Paul Allen, Klaus Mathiak, and Renaud Jardri. 2016. “Translating Neurocognitive Models of Auditory-Verbal Hallucinations into Therapy: Using Real-Time fMRI-Neurofeedback to Treat Voices.” Frontiers in Psychiatry / Frontiers Research Foundation 7 (June): 103.

Fox, M. D., A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle. 2005. “From The Cover: The Human Brain Is Intrinsically Organized into Dynamic, Anticorrelated Functional Networks.” Proceedings of the National Academy of Sciences 102 (27): 9673–78.

Grosenick, Logan, Brad Klingenberg, Kiefer Katovich, Brian Knutson, and Jonathan E. Taylor. 2013. “Interpretable Whole-Brain Prediction Analysis with GraphNet.” NeuroImage 72 (May): 304–21.

Guyon, Isabelle, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. 2002. “Gene Selection for Cancer Classification Using Support Vector Machines.” Machine Learning 46 (1-3). Kluwer Academic Publishers: 389–422.

Haxby, James V., Andrew C. Connolly, and J. Swaroop Guntupalli. 2014. “Decoding Neural Representational Spaces Using Multivariate Pattern Analysis.” Annual Review of Neuroscience 37 (June): 435–56.

Hoffman, Ralph E., Adam W. Anderson, Maxine Varanko, John C. Gore, and Michelle Hampson. 2008. “Time Course of Regional Brain Activation Associated with Onset of Auditory/verbal Hallucinations.” The British Journal of Psychiatry: The Journal of Mental Science 193 (5): 424–25.

Hoffman, Ralph E., and Michelle Hampson. 2011. “Functional Connectivity Studies of Patients with Auditory Verbal Hallucinations.” Frontiers in Human Neuroscience 6 (December): 6.

Jardri, Renaud, Kenneth Hugdahl, Matthew Hughes, Jérôme Brunelin, Flavie Waters, Ben Alderson-Day, Dave Smailes, et al. 2016. “Are Hallucinations Due to an Imbalance Between Excitatory and Inhibitory Influences on the Brain?” Schizophrenia Bulletin 42 (5): 1124–34.

Jardri, Renaud, Alexandre Pouchet, Delphine Pins, and Pierre Thomas. 2011. “Cortical Activations during Auditory Verbal Hallucinations in Schizophrenia: A Coordinate-Based Meta-Analysis.” The American Journal of Psychiatry 168 (1): 73–81.

Jardri, Renaud, Pierre Thomas, Christine Delmaire, Pierre Delion, and Delphine Pins. 2013. “The Neurodynamic Organization of Modality-Dependent Hallucinations.” Cerebral Cortex 23 (5): 1108–17.

Lefebvre, Stéphanie, Morgane Demeulemeester, Arnaud Leroy, Christine Delmaire, Renaud Lopes, Delphine Pins, Pierre Thomas, and Renaud Jardri. 2016. “Network Dynamics during the Different Stages of Hallucinations in Schizophrenia.” Human Brain Mapping 37 (7): 2571–86.

Lennox, B. R., S. B. Park, P. B. Jones, P. G. Morris, and G. Park. 1999. “Spatial and Temporal Mapping of Neural Activity Associated with Auditory Hallucinations.” The Lancet 353 (9153): 644.

Leroy, Arnaud, Jack R. Foucher, Delphine Pins, Christine Delmaire, Pierre Thomas, Mathilde M. Roser, Stéphanie Lefebvre, et al. 2017. “fMRI Capture of Auditory Hallucinations: Validation of the Two-Steps Method.” Human Brain Mapping, June. doi:10.1002/hbm.23707.

McCarthy-Jones, Simon, David Smailes, Aiden Corvin, Michael Gill, Derek W. Morris, Timothy G. Dinan, Kieran C. Murphy, et al. 2017. “Occurrence and Co-Occurrence of Hallucinations by Modality in Schizophrenia-Spectrum Disorders.” Psychiatry Research 252 (June): 154–60.

Mechelli, Andrea, Paul Allen, Edson Amaro Jr, Cynthia H. Y. Fu, Steven C. R. Williams, Michael J. Brammer, Louise C. Johns, and Philip K. McGuire. 2007. “Misattribution of Speech and Impaired Connectivity in Patients with Auditory Verbal Hallucinations.” Human Brain Mapping 28 (11): 1213–22.

Mondino, Marine, Emmanuel Poulet, Marie-Françoise Suaud-Chagny, and Jerome Brunelin. 2016. “Anodal tDCS Targeting the Left Temporo-Parietal Junction Disrupts Verbal Reality-Monitoring.” Neuropsychologia 89 (August): 478–84.

Northoff, Georg, and Pengmin Qin. 2011. “How Can the Brain’s Resting State Activity Generate Hallucinations? A ‘Resting State Hypothesis’ of Auditory Verbal Hallucinations.” Schizophrenia Research 127 (1-3): 202–14.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. “Scikit-Learn: Machine Learning in Python.” Journal of Machine Learning Research: JMLR 12 (November). JMLR.org: 2825–30.

Pierrefeu, Amicie de, Tommy Löfstedt, Fouad Hadj-Selem, Mathieu Dubois, Philippe Ciuciu, Vincent Frouin, and Edouard Duchesnay. 2016. “Structured Sparse Principal Components Analysis with the TV-Elastic Net Penalty.” arXiv [stat.ML]. arXiv.http://arxiv.org/abs/1609.01423.

Plaze, Marion, Jean-François Mangin, Marie-Laure Paillère-Martinot, Eric Artiges, Jean-Pierre Olié, Marie-Odile Krebs, Raphaël Gaillard, Jean-Luc Martinot, and Arnaud Cachia. 2015. “‘Who Is Talking to Me?’ — Self–other Attribution of Auditory Hallucinations and Sulcation of the Right Temporoparietal Junction.” Schizophrenia Research 169 (1-3): 95–100.

Raichle, M. E., A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman. 2001. “A Default Mode of Brain Function.” Proceedings of the National Academy of Sciences 98 (2): 676–82.

Schmidt, A., F. Müller, C. Lenz, P. C. Dolder, Y. Schmid, D. Zanchi, U. E. Lang, M. E. Liechti, and S. Borgwardt. 2017. “Acute LSD Effects on Response Inhibition Neural Networks.” Psychological Medicine, October, 1–13.

Small, Steven L., and Gregory Hickok. 2016. “The Neurobiology of Language.” In Neurobiology of Language, 3–9.

Sommer, Iris E. C., Kelly M. J. Diederen, Jan-Dirk Blom, Anne Willems, Leila Kushan, Karin Slotema, Marco P. M. Boks, et al. 2008. “Auditory Verbal Hallucinations Predominantly Activate the Right Inferior Frontal Area.” Brain: A Journal of Neurology 131 (Pt 12): 3169–77.

Van Dijk, Koene R. A., Mert R. Sabuncu, and Randy L. Buckner. 2012. “The Influence of Head Motion on Intrinsic Functional Connectivity MRI.” NeuroImage 59 (1): 431–38.

Vercammen, Ans, Henderikus Knegtering, Johann A. den Boer, Edith J. Liemburg, and André Aleman. 2010. “Auditory Hallucinations in Schizophrenia Are Associated with Reduced Functional Connectivity of the Temporo-Parietal Area.” Biological Psychiatry 67 (10): 912–18.

ARTICLE 4

Structured Sparse Principal Components Analysis with the TV-Elastic Net Penalty

de Pierrefeu A[1], Lofstedt T[2], Hadj-Selem F[3], Dubois M[1], Jardri R[4], Fovet T[4], Ciuciu P[1], Frouin V[1], Duchesnay E[1].

1. NeuroSpin, CEA, Paris-Saclay, Gif-sur-Yvette - France
2. Department of Radiation Sciences, Ume°a University, Ume°a - Sweden.
3. Energy Transition Institute: VeDeCoM - France.
4. Univ Lille, CNRS UMR 9193, SCALab, CHU Lille, Pˆole de Psychiatrie (unit CURE), Lille, France

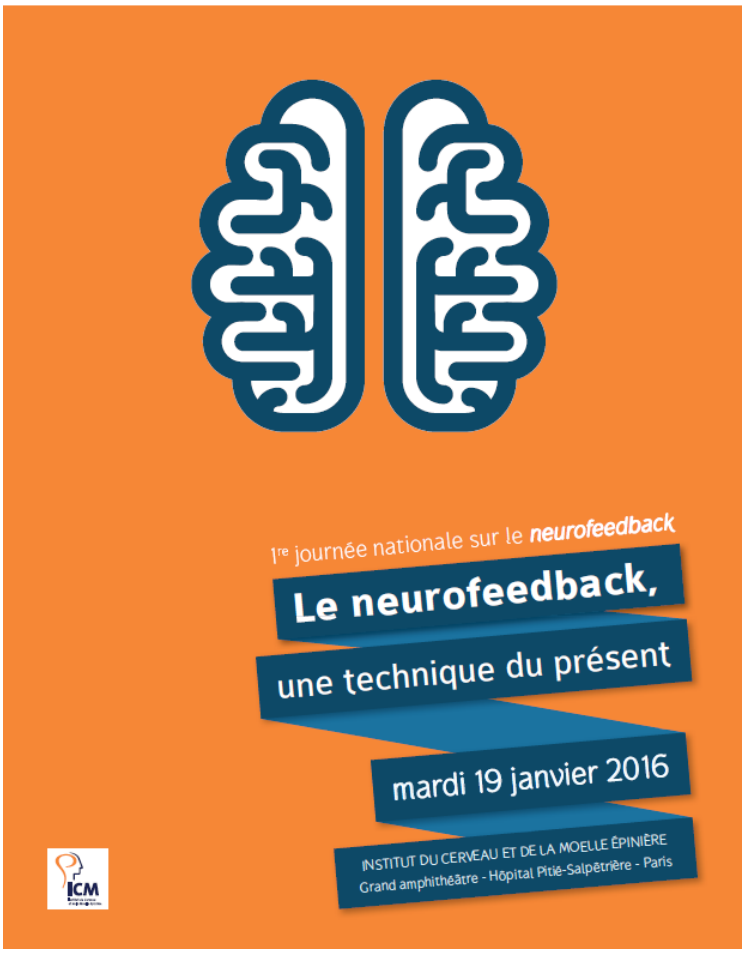
Disponible ici :
https://www.researchgate.net/publication/307862593_Structured_Sparse_Principal_Components_Analysis_with_the_TV-Elastic_Net_Penalty

IEEE Transactions on Medical Imaging. Online.

3 . PERSPECTIVES THERAPEUTIQUES DANS LA SCHIZOPHRENIE

“To succeed, planning alone is insufficient. One must improvise as well.”

Isaac Asimov, Foundation



!

3.1 Les techniques de neurofeedback

3.1.1 Généralités

Le neurofeedback est une méthode thérapeutique non-invasive consistant à mesurer l'activité d'une ou de plusieurs régions cérébrales chez un sujet et à lui présenter en temps réel l'enregistrement de cette activité. Grâce à ce dispositif, les participants peuvent apprendre à contrôler l'activité neuronale d'une ou de plusieurs région(s) cérébrale(s) déterminée(s), en se basant sur le feedback en temps réel des mesures d'activité de la cible retenue.

Le principe du neurofeedback fut démontré pour la première fois chez l'être humain en 1962, par James Kamiya à Chicago (entraînement volontaire des ondes alpha au niveau du cortex occipital). Cette première expérience fut suivie de nombreux autres travaux dans les années 1960-70, comme les études de Barry Sterman (216) pour le traitement de l'épilepsie ou celles de Joel Lubar pour la prise en charge des symptômes d'hyperactivité et les difficultés de concentration chez l'enfant (217).

Malheureusement, le grand enthousiasme suscité par ces travaux pionniers n'aboutit qu'à une diffusion prématurée et incontrôlée de la technique. Dans ce contexte, plusieurs entreprises commercialisèrent rapidement des « kits » de neurofeedback basés sur des fondements scientifiques peu rigoureux et alors même que les connaissances sur la méthode demeuraient minimes. C'est probablement en partie à cause de ce phénomène que les techniques de neurofeedback acquirent, à partir de cette période, une réputation de méthodes issues d'un charlatanisme *new-age*, qui demeure encore très présente actuellement au sein de la communauté scientifique. Aussi, à la fin des années 1970, les travaux de recherche sur le neurofeedback furent-ils limités.

Toutefois, dans les années 1980-90, de nouveaux protocoles voient le jour notamment les protocoles basés sur les potentiels corticaux lents (218,219). Mais ce n'est qu'avec le développement des techniques d'IRM fonctionnelle en temps réel au début des années 2000 qu'un véritable renouveau pour le concept de neurofeedback va émerger (220). En effet, la haute résolution spatiale de cette technique permet un accès aux structures cérébrales profondes du cerveau, pour lesquelles des dysfonctionnements ont pu être identifiés dans plusieurs troubles psychiatriques (221). La **Figure 11** montre l'évolution du nombre de publications dans le domaine depuis les années 1960.

Malgré ce nouvel intérêt, les mécanismes qui sous-tendent l'auto-régulation cérébrale au cours d'un entrainement par neurofeedback restent actuellement très peu connus et le développement de recommandations de bonne pratique extrêmement difficile (222). Dans l'**Article 5**, nous exposons les principales utilisations thérapeutiques du neurofeedback (guidé par électroencéphalographie (EEG) ou par IRMf) qui ont fait l'objet de publications scientifiques.

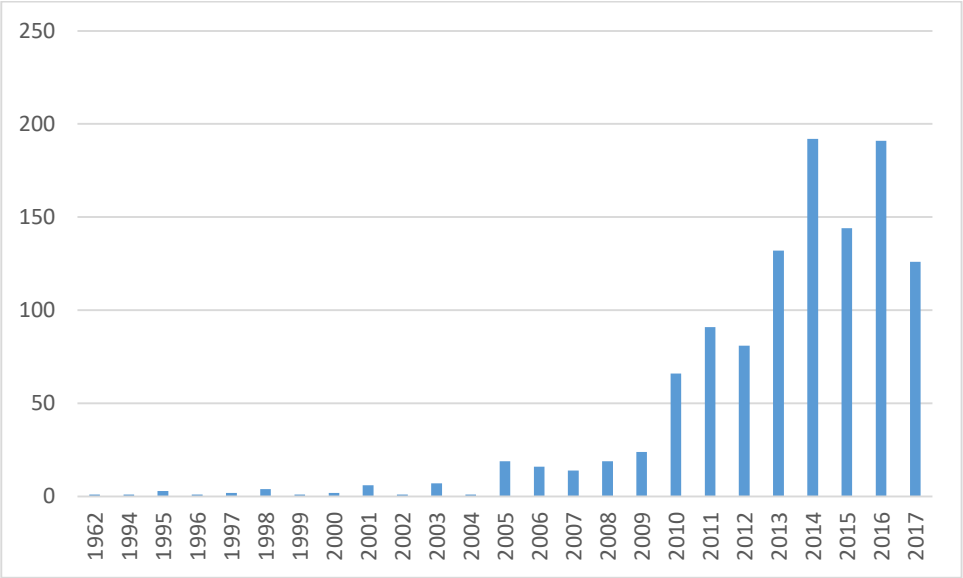


Figure 11. Evolution du nombre de publications sur le neurofeedback (recherche réalisée dans la base de données *Medline* avec le mot clé « *neurofeedback* »).

ARTICLE 5

Le neurofeedback en psychiatrie :

Les outils d'imagerie cérébrale et de neurophysiologie au service de la thérapeutique

Thomas FOVET[1], Renaud JARDRI[1], Jean-Arthur MICOULAUD-FRANCHI[2]

1. Univ Lille, CNRS, UMR-9193, SCA-Lab, équipe psyCHIC et CHU Lille, Pôle de psychiatrie (CURE), Hôpital Fontan, F-59000 Lille, France

2. Université de Bordeaux, CNRS, USR-3413 SANPSY et Service d'Explorations Fonctionnelles du Système Nerveux, Clinique du Sommeil, CHU de Bordeaux, 33076 Bordeaux, France

L'Information Psychiatrique. 2016 ; 92 : 285-93.

Introduction

Il existe actuellement un regain d'intérêt clinique pour le neurofeedback guidé par électroencéphalographie (EEG) en psychiatrie (1,2). Ce nouvel élan coïncide avec le développement des techniques de neurofeedback guidées par imagerie par résonnance magnétique fonctionnelle (IRMf) qui ont pu voir le jour grâce à l'avènement de l'IRMf en temps réel dans les années 2000 (3,4).
Dans cet article, nous décrirons en premier lieu le principe général du neurofeedback. Nous présenterons ensuite les résultats actuellement disponibles pour cette technique lorsqu'elle est guidée par EEG ou par IRMf pour la prise en charge des pathologies psychiatriques. Enfin, nous discuterons les liens qu'entretient le neurofeedback avec le champ de la remédiation cognitive et des psychothérapies, et nous évoquerons quelques perspectives enthousiasmantes de l'utilisation de cette technique dans la discipline.

Principes généraux

Qu'est-ce que le biofeedback ?

L'ouvrage de référence d'Anne et Antoine Rémond, *Biofeedback principes et applications* (5), propose la définition du biofeedback suivante : « *Groupe de procédés thérapeutiques qui utilise une instrumentation électronique ou électromécanique. Cette dernière permet de mesurer avec précision, traiter et représenter, sous forme analogique ou numérique, une information aux propriétés renforcées, sur l'activité neuro-musculaire ou l'activité autonome (normale ou anormale) des individus au moyen de signaux sonores ou optiques. Ses objectifs - d'autant mieux atteints qu'ils sont effectués sous l'égide d'un professionnel compétent dans le domaine du biofeedback - sont d'aider les individus à développer une meilleure conscience et un contrôle volontaire plus intense de leurs processus physiologiques, processus pratiquement inconscients (c'est-à-dire peu indépendants a priori, ou indépendants d'un contrôle volontaire), ceci en contrôlant d'abord le signal externe, puis finalement en utilisant des moyens psychophysiologiques internes* » (5).
Un paramètre mesurant une fonction physiologique est donc « traité » ou objectivé par une interface afin de fournir au sujet une information continue et en temps réel (« bio-feedback »), le plus souvent sous forme visuelle ou auditive (mais toutes les modalités sensorielles sont théoriquement envisageables). Les modifications de la variable ainsi traitée, réalisées dans la direction désirée sont renforcés positivement. Ce procédé permet au sujet de contrôler une activité biologique avec pour objectif de réduire l'intensité de symptômes cibles et donc d'obtenir un effet thérapeutique (6,7).

Qu'est-ce que le neurofeedback ?

Dans le cas du neurofeedback, qui peut être considéré comme un sous-type particulier de biofeedback, la variable traitée puis modulée par le sujet, est une activité cérébrale. Le neurofeedback est donc une méthode non-invasive qui consiste à mesurer l'activité d'une ou de plusieurs régions cérébrales chez un sujet et à lui présenter en temps réel l'enregistrement de cette activité (8). Grâce à ce dispositif, les participants peuvent apprendre à contrôler l'activité neuronale d'une ou de plusieurs région(s) cérébrale(s) déterminée(s), en se basant sur le feedback en temps réel des mesures d'activité de la cible retenue (9). Cet apprentissage peut ou non s'accompagner de modifications comportementales chez le sujet.

L'activité cérébrale peut être mesurée au moyen de différentes techniques. Les plus fréquemment utilisées sont l'EEG et l'IRMf mais d'autres méthodes comme la spectroscopie (*Near infrared spectroscopy* : NIRS) ou la magnétoencéphalographie (MEG) peuvent également être utilisées. Le principe général du neurofeedback est présenté en **Figure 1**.

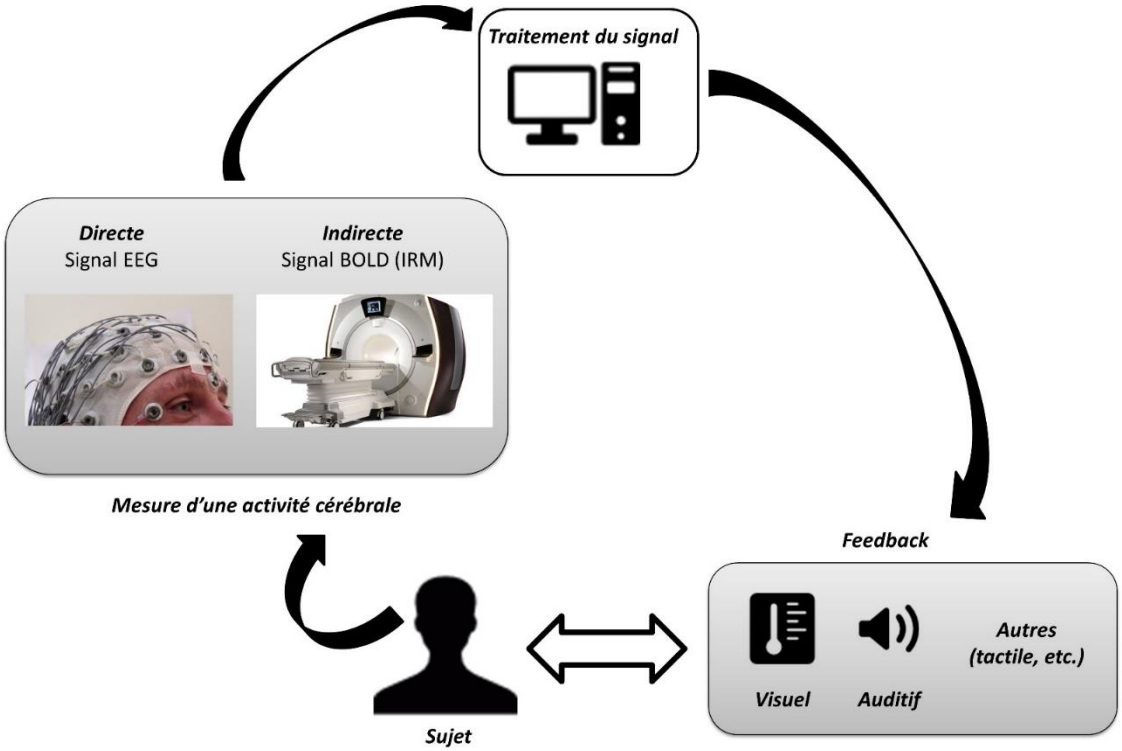


Figure 1 : principe du neurofeedback.
Une activité cérébrale est mesurée, soit directement grâce à l'EEG par exemple, soit indirectement via le signal BOLD (Blood oxygenation level dependent) qui est un signal hémodynamique que l'on considère corrélé à l'activité cérébrale. Le signal mesuré est traité par une interface informatique puis présenté au sujet via un feedback (le plus fréquemment visuel sous la forme d'une jauge dont le niveau varie selon l'intensité de l'activité mesurée). Le sujet peut alors moduler son activité cérébrale selon le feedback qui lui est renvoyé.

Les différentes techniques de neurofeedback : avantages et inconvénients

Bien que le principe général décrit plus haut s'applique quelle que soit la méthode utilisée pour mesurer l'activité cérébrale, chaque technique présente des avantages et des inconvénients. Par exemple, la haute résolution temporelle de l'EEG est souvent mise en avant alors que pour l'IRMf, c'est la haute résolution spatiale qui présente un intérêt majeur. Plusieurs caractéristiques sont extrêmement importantes à connaître pour la mise en place des protocoles de neurofeedback. Par exemple, le délai hémodynamique de plusieurs secondes avec les techniques d'IRMf implique un délai de 4 à 6 secondes avant l'affichage du feedback. Les patients doivent être informés de ce délai avant les séances de neurofeedback. Le **Tableau 1** reprend les avantages et inconvénients de l'EEG et de l'IRMf pour les protocoles de neurofeedback. Ces deux techniques sont à l'heure actuelle, les mieux étudiées. Les travaux en MEG et en NIRS restent assez rares mais devraient se développer dans les années futures (10). L'intérêt de la NIRS est, notamment, son faible coût par rapport à l'IRMf. Enfin, des interfaces combinant plusieurs techniques (EEG et IRMf par exemple) commencent à voir le jour (11).

	EEG	fMRI
Signal utilisé	Activité électrique des cellules pyramidales perpendiculaires au scalp	Contraste du signal BOLD (signal hémodynamique corrélé à l'activité neuronale)
Résolution temporelle	Millisecondes	Secondes
Résolution spatiale	Centimètres / Régions superficielles	Millimètres / Régions profondes
Délai du feedback	< 50 millisecondes	4-6 secondes (délai hémodynamique)
Portabilité	Oui	Non
Coût	Modéré	Elevé

Tableau 1 : caractéristiques de la technique de neurofeedback selon la technique employée pour recueillir l'activité cérébrale (IRMf ou EEG). *Adapté de* (12).

Neurofeedback guidé par EEG (EEG-neurofeedback) et troubles psychiatriques

L'EEG a été la première technique de mesure de l'activité cérébrale ayant fait l'objet d'une utilisation de type neurofeedback. Actuellement, on considère qu'il existe trois cibles neurophysiologiques principales (voir **Figure 2**), selon la pathologie psychiatrique ciblée, dans les protocoles d'EEG-neurofeedback :

- la cible est une augmentation de la puissance spectrale dans la bande béta associée à une diminution de la puissance spectrale dans la bande thêta dans le trouble du déficit de l'attention avec ou sans hyperactivité (TDA/H) ; l'objectif est d'augmenter l'éveil (« arousal ») qui apparait fréquemment diminué dans le TDA/H.
- la cible est une inversion du rapport droite/gauche de la puissance spectrale dans la bande alpha au niveau frontal afin de corriger la dysrégulation de la balance émotionnelle retrouvée dans l'épisode dépressif caractérisé (EDC).
- la cible est une augmentation de la puissance spectrale dans la bande alpha (voir dans la bande thêta) afin d'induire un état de relaxation et de diminuer l'hyperéveil corticale dans les troubles anxieux et les troubles addictifs,.

Les études qui seront présentées ici dans le cadre de l'utilisation du neurofeedback guidé par EEG dans les pathologies psychiatriques sont les travaux présentant une qualité méthodologique satisfaisante (étude contrôlée et randomisée, en ouvert ou en aveugle) et une cible neurophysiologique EEG identifiable comme décrit ci-dessus.

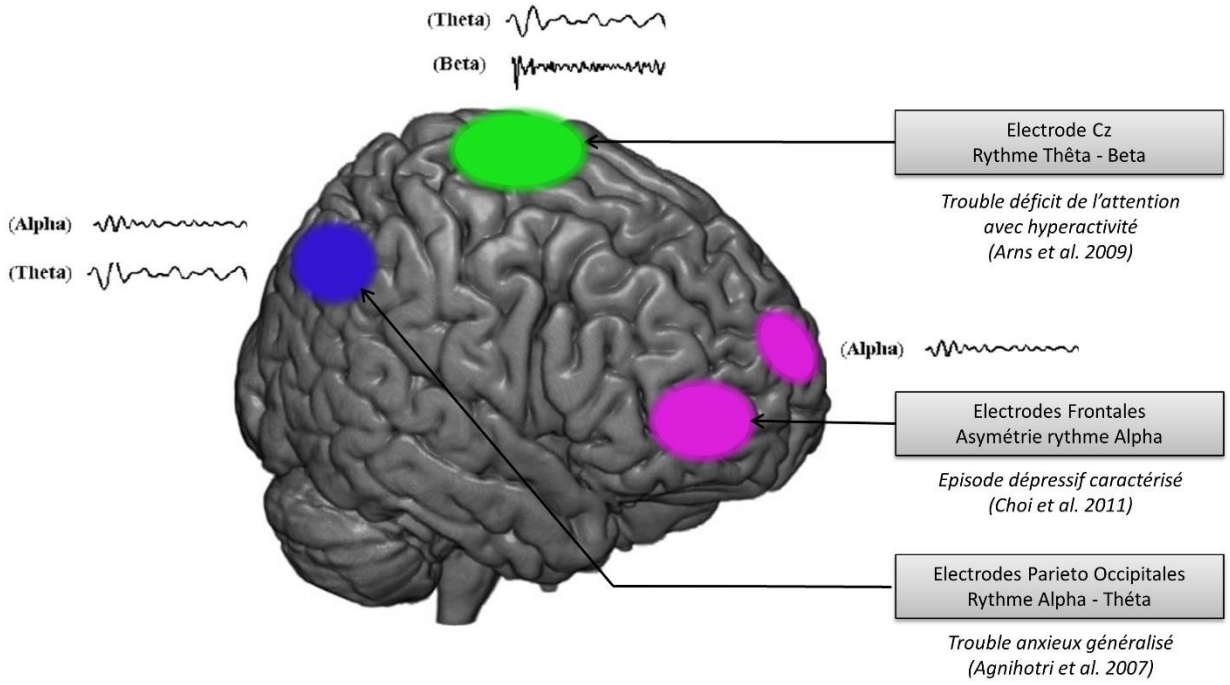


Figure 2 : **utilisation du neurofeedback guidé par EEG en psychiatrie** (13–15).

Trouble du déficit de l'attention avec ou sans hyperactivité

Trois méta-analyses ont exploré l'efficacité thérapeutique de l'EEG-neurofeedback dans le TDA/H (13,16,17). Les résultats montrent qu'une taille d'effet significativement supérieure au groupe témoin est retrouvée dans les études randomisées en ouvert (16). L'effet thérapeutique apparait plus important pour la dimension inattention que pour la dimension hyperactivité et l'amélioration pour cette dimension est proportionnelle au nombre de séances de neurofeedback réalisées (13). Par ailleurs, la taille d'effet reste significative uniquement pour la dimension inattention dans les études randomisées en aveugle (17). L'effet thérapeutique mis en évidence se maintiendrait dans le temps, au moins 6 mois après la fin de la cure (18).

Enfin, les travaux qui se sont intéressés à l'EEG-neurofeedback pour la prise en charge des troubles du spectre autistique (TSA), semblent indiquer que le neurofeedback n'aurait pas une efficacité sur les signes spécifiques du TSA, mais une efficacité sur les signes de TDA/H présents en comorbidité chez environ 40 à 50 % des sujets (19).

Episode dépressif caractérisé

Un seul essai randomisé en ouvert a exploré l'efficacité de l'EEG-neurofeedback pour la prise en charge de l'EDC (14). Cette étude était en faveur d'une efficacité de l'EEG-neurofeedback sur les évaluations cliniques par questionnaire de sévérité de l'EDC (*Beck Depression Inventory II, BDI II*), sur les pensées automatiques négatives et positives (*Automatic Thought Questionnaire-Positive, ATQ-P et Automatic Thought Questionnaire-Negative ATQ-N*) et sur les évaluations neuropsychologiques des fonctions exécutives (14). Sur 12 sujets dans le groupe avec neurofeedback, la moitié (n=6) ont été répondeurs (décroissance de 20 % des scores de sévérité sur 75 % des mesures par questionnaires). Il n'y avait aucun répondeur dans le groupe contrôle. Une étude pilote récente chez 9 sujets a également apporté des résultats en faveur d'une spécificité d'effet neurophysiologique de cette thérapeutique (20).

Troubles anxieux et trouble addictifs

Deux essais randomisés en ouvert ont exploré l'efficacité du neurofeedback dans le trouble anxieux généralisé (15,21) et un essai randomisé a été mené chez des patients souffrant de trouble obsessionnel et compulsif (22). L'ensemble de ces études apporte des résultats en faveur d'une efficacité de l'EEG-neurofeedback. La place de cette technique dans la prise en charge des troubles anxieux reste toutefois à déterminer. L'efficacité de l'EEG-neurofeedback devra notamment être comparée aux bénéfices établis des thérapies cognitivo-comportementales (TCC). Le neurofeedback pourrait s'imposer comme un outil complémentaire s'intégrant au sein de l'arsenal thérapeutiques des TCC (23). Enfin, comme pour les TSA, l'effet de l'EEG-neurofeedback dans les troubles addictifs pourrait être non spécifique, médié par l'intermédiaire d'un effet sur les comorbidités anxieuses ou TDA/H (24,25).

Neurofeedback guidé par IRMf (IRMf-neurofeedback) et troubles psychiatriques

L'IRMf est une technique non-invasive, de haute résolution spatiale, permettant d'enregistrer une réponse hémodynamique (signal BOLD : *Blood Oxygenation Level Dependant*), reflet indirect de l'activité neuronale (26). Aujourd'hui, il est possible d'avoir un accès immédiat aux résultats expérimentaux par l'analyse des données dès leur acquisition : c'est l'IRMf en temps réel (27). Ces avancées technologiques rendent possible la mise en place de protocoles de neurofeedback en IRMf (28,29). L'intérêt majeur est de pouvoir cibler toutes les régions cérébrales (avec, contrairement à l'EEG-neurofeedback, la possibilité de cibler des régions cérébrales profondes). De nombreuses études ont pu mettre en évidence qu'il est tout à fait possible d'apprendre à un sujet sain à réguler l'activité de certaines régions cérébrales comme : l'amygdale, le cortex cingulaire antérieur ou le cortex insulaire par exemple (pour un travail de synthèse, voir (30)).

Dans certains cas, cet apprentissage est associé à des modifications comportementales chez les participants (par exemple, la réponse à des stimuli aversifs après IRMf-neurofeedback ciblé sur le cortex insulaire antérieur (31)).

Dans la continuité de ces travaux, plusieurs études récentes utilisant le IRMf-neurofeedback ont obtenu des résultats prometteurs dans la prise en charge de symptômes subjectifs (32). Des résultats préliminaires extrêmement encourageants sont d'ores et déjà disponibles sur l'utilisation de cette technique dans la prise en charge, par exemple, des acouphènes (33) ou de la douleur (34).

Les progrès dans le domaine de l'imagerie cérébrale (notamment l'apport de l'IRMf) permettent actuellement une meilleure compréhension de la physiopathologie des troubles mentaux (35). En identifiant des cibles potentielles pour le IRMf-neurofeedback, ces avancées permettent d'envisager, au moins au niveau théorique, l'utilisation de cette technique à visée thérapeutique dans ces pathologies. Plusieurs résultats expérimentaux ont été publiés ces dernières années (voir **Figure 3**). Toutefois, contrairement aux neurofeedback guidé par EEG, les études explorant le neurofeedback guidé par IRMf contrôlées et randomisées apparaissent excessivement rares.

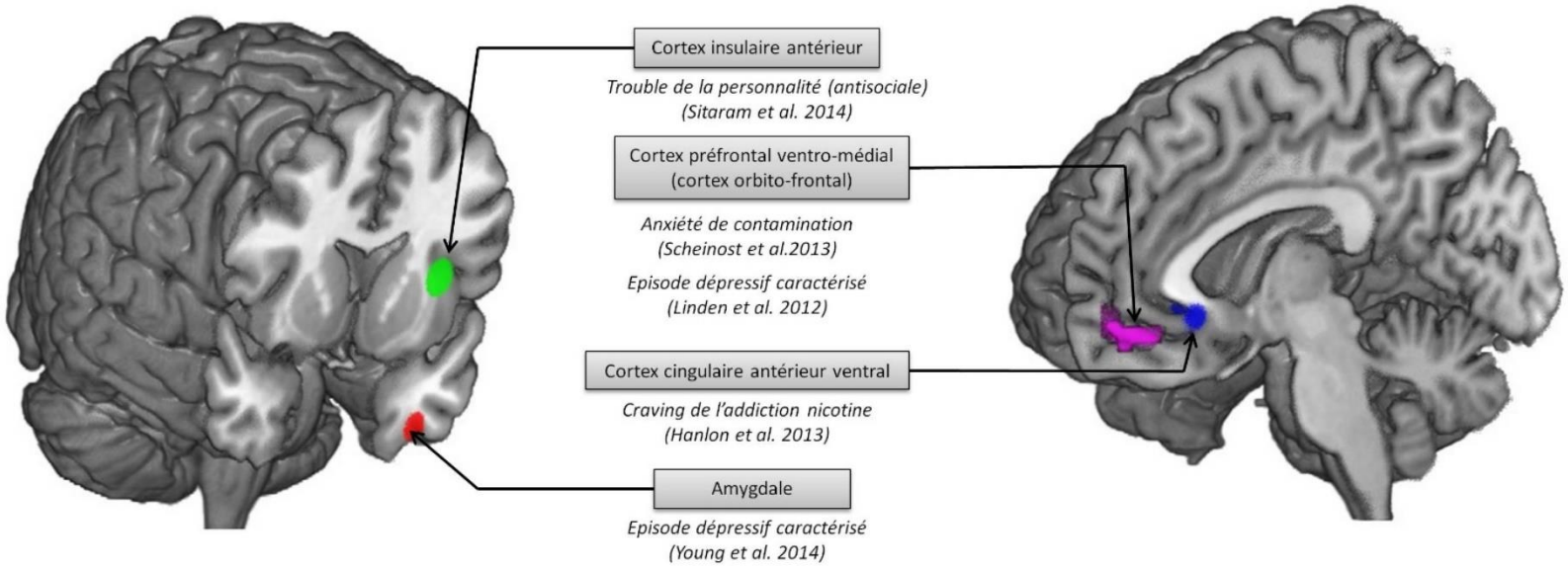


Figure 3 : **utilisation du neurofeedback guidé par IRMf en psychiatrie** (36–40).

Episode dépressif caractérisé

Deux études utilisant l'IRMf-neurofeedback chez des patients souffrant EDC sont actuellement disponibles (38,40) et plusieurs essais sont actuellement en cours (4). Linden et collaborateurs ont mis en évidence, dans une étude pilote, une amélioration clinique statistiquement significative chez 8 patients après un protocole de IRMf-neurofeedback. Ce protocole comprenait 4 séances au cours desquelles les patients devaient augmenter l'activité dans des régions cérébrales préalablement identifiées comme impliquées dans la présentation de stimuli à valence émotionnelle positive (notamment la région préfrontale ventro-médiale) (38). Une autre étude a permis de montrer que les patients souffrant d'EDC sont capables d'apprendre à réguler l'activité de l'amygdale cérébrale gauche au cours de la remémoration de souvenirs autobiographiques positifs et que ceci est associé à une diminution aux scores cliniques d'anxiété (40). Toutefois, ces résultats restent préliminaires et devront être validés grâce à des études contrôlées randomisées de plus grande ampleur.

Addictions

Des résultats prometteurs ont été publiés dans le domaine des addictions. Hanlon et collaborateurs ont mis en évidence qu'un protocole d'IRMf-neurofeedback visant à apprendre à décroitre l'activité du cortex cingulaire antérieur ventral, permettait de diminuer le *craving* dans la dépendance à la nicotine (39).

Trouble de la personnalité

Dans une étude pilote, Sitaram et collaborateurs ont proposé d'envisager l'IRMf-neurofeedback comme une piste thérapeutique pour les patients souffrant de trouble de la personnalité type antisocial. Malheureusement, dans ce travail, un seul patient a réalisé les 12 sessions d'entrainement prévues dans le protocole (qui consistait à augmenter l'activité dans le cortex cingulaire antérieur). Toutefois, chez ce participant, l'entrainement était suivi d'une modulation des réponses émotionnelles à des stimuli aversifs. Malgré de nombreuses limites, cette étude ouvre la voie à l'utilisation de l'IRMf-neurofeedback dans le cadre des troubles de la personnalité.

Trouble obsessionnel compulsif

L'IRMf-neurofeedback pourrait également constituer une approche thérapeutique prometteuse dans le trouble obsessionnel compulsif (TOC). Des résultats encourageants ont été publiés, mettant en évidence qu'un protocole entrainant les sujets à réduire l'activité d'une partie du cortex orbito-frontal (une région impliquée dans la physiopathologie du trouble obsessionnel compulsif (41)) permettait de réduire l'intensité clinique de l'anxiété de contamination, symptôme fréquent dans le TOC (37).

Discussion : neurofeedback et thérapeutique en psychiatrie

La place du neurofeedback dans le champ des moyens thérapeutiques en psychiatrique reste encore à définir. En effet, le neurofeedback se situe à l'interface de plusieurs disciplines : l'imagerie médicale et la neurophysiologie, les sciences cognitives, et les sciences psychologiques.

La prise en charge d'un patient par EEG-neurofeedback nécessite en général 25 à 50 séances de 45 à 60 minutes (avec une fréquence d'une à trois séances par semaine). Le nombre élevé de séances permet : l'apprentissage de stratégies efficaces permettant une diminution des symptômes ciblés, le transfert de ces stratégies dans la vie quotidienne, et l'obtention d'un effet sur la neuroplasticité cérébrale permettant le maintien de l'efficacité dans le temps. Ce nombre important de séances implique la nécessité d'une intégration cohérente du neurofeedback au sein des stratégies de prise en charge actuelles des pathologies psychiatriques (psychopharmacologie, électrothérapie, psychoéducation, thérapies cognitivo-comportementales, remédiation cognitive, self empowerment).

Le champ de la remédiation cognitive, par l'ancrage du neurofeedback dans les sciences cognitives, et le champ de l'*empowerment*, par l'ancrage du neurofeedback dans les conceptions de la psychologie de la santé (42) nous paraissent deux domaines structurants pour l'integration du neurofeedback au sein des thérapeutiques psychiatriques.

Remediation cognitive, apprentissage et neurofeedback

La pratique du neurofeedback peut s'inscrire dans le champ plus global de la remédiation cognitive (43). A ce titre il conduit à un apprentissage chez le sujet qui utilise la technique (44,45). Cependant, les théories classiques du comportementalisme comme le conditionnement opérant ne suffisent pas pour décrire le processus d'apprentissage impliqué dans le neurofeedback. Celui-ci s'avère un effet très complexe et, pour l'heure, peu connu. Il impliquerait à la fois des processus implicites (ou « automatiques ») et explicites (ou « contrôlés »), ces deux types de processus pouvant survenir de manière séquentielle ou en parallèle. Un modèle d'apprentissage pour le neurofeedback reste donc à construire.

L'objectif d'un protocole de neurofeedback est de conduire au développement de compétences nouvelles chez le sujet souffrant d'un trouble psychiatrique (2). Cette technique se différencie cependant des méthodes de remédiation cognitive « classique » puisqu'elle ne restreint pas la stratégie cognitive à développer en fournissant au sujet une consigne déterminée (par exemple identifier une cible parmi des distracteurs dans une tâche de détection de cible). La stratégie cognitive développée au cours du neurofeedback est celle pour laquelle le renforcement positif (en lien avec les caractéristiques neurophysiologiques ciblées par le protocole utilisé) est le plus important. Cette spécificité permet à la fois de renforcer le sentiment d'auto-efficacité du sujet, mais également le développement de stratégies cognitives d'autant plus efficaces qu'elles sont spécifiques au sujet.

Pour autant, la place d'un thérapeute formé est essentielle pour expliquer la technique, renforcer l'apprentissage, maintenir la motivation au cours des séances et permettre de suivre l'évolution des performances que ce soit au cours d'une séance donnée, ou au fils des séances successives. Le patient sera ensuite amené à transférer dans la vie quotidienne les compétences acquises.

Ces réflexions autour de l'apprentissage dans le neurofeedback sont essentielles pour des études d'efficacité de qualité. En effet, comme le notait en 1997 Rémond : « *Dans beaucoup d'essais d'utilisation du biofeedback qui ne se sont pas montrés satisfaisants, les auteurs de ces essais ont-ils pu ne pas se demander si un apprentissage convenable avait été effectué ? Devant tout nouveau patient, on doit en effet se poser la question suivante : la variable physiologique apparemment en cause est-elle sensible au biofeedback et, si oui, sa modification s'est-elle effectivement produite dans la population que l'on étudie ? Douter de l'efficacité d'un traitement par biofeedback d'une variable physiologique lorsque celui-ci est effectué sans essai préalable pour modifier cette variable, revient à douter de l'efficacité d'un médicament dans une maladie, lorsqu'en fait, il n'a pas été absorbé par le patient.* » (5). Ainsi, les premières études ayant testé l'efficacité du neurofeedback dans le TDA/H avaient mis l'accent sur la qualité du protocole et de l'effet d'apprentissage au cours des séances, aux dépens de la construction méthodologique du protocole lui-même (absence de randomisation et de groupe témoin). Paradoxalement, les études plus récentes se sont focalisées sur une méthodologie rigoureuse (avec groupes témoins et évaluations en aveugle) mais bien souvent au détriment de la qualité de l'apprentissage dans le groupe actif.

Il apparait donc essentiel que les futures études dans le domaine puissent allier la qualité méthodologique des études randomisées contrôlées en aveugle et la qualité des séances de neurofeedback autour du concept d'apprentissage.

Psychologie de la santé, empowerment et neurofeedback

L'empowerment est une notion récente, issue des conceptions de la psychologie de la santé (46) qui désigne l'augmentation de la capacité d'agir d'une personne souffrant de pathologie(s) par le biais du développement de son autonomie. Ces dernières années, ce concept a été largement développé dans le champ de la santé mentale. Parce qu'elles permettent aux patients de renforcer leur sentiment d'auto-efficacité (47) mais pourraient également les amener à retrouver un sentiment de contrôle sur des symptômes souvent envahissants et stigmatisant, il nous apparait important d'éclairer les techniques de neurofeedback à la lumière de ce concept. Ainsi, le neurofeedback pourrait, dans les années futures, dépasser les conceptions classiques d'apprentissage et de remédiation cognitive pour faire émerger une véritable *psychothérapie guidée par l'imagerie cérébrale*. Les stratégies cognitives développées au cours des protocoles de neurofeedback pourraient être reprises, améliorées et généralisées au cours de séances de psychothérapie « classique ».

Au sein même du cadre théorique du neurofeedback, ce concept s'avère intéressant, tout particulièrement en ce qui concerne la problématique du choix de la cible. En effet, il n'a, pour l'heure, pas pu être déterminé si la meilleure stratégie pour les protocoles de neurofeedback est d'apprendre au sujet à diminuer l'activité des circuits impliqués dans la physiopathologie du trouble ciblé ou de lui permettre d'augmenter l'activité des régions permettant de développer des mécanismes de compensation et de « coping ». La stratégie optimale pourrait varier d'un sujet à un autre et selon la pathologie, ce que devront déterminer les futurs travaux.

Conclusion

Rémond constatait en 1997, que le développement du biofeedback en France reste « *un vaste domaine à cultiver* » et soulignait « *l'intérêt et les avantages que pourrait apporter le biofeedback aux jeunes médecins, aux spécialistes de nombreuses disciplines, et en particulier, aux jeunes spécialistes de neurophysiologie clinique et à leurs techniciens* » (5). Ce constat semble plus que jamais d'actualité particulièrement dans le champ psychiatrique où le neurofeedback pourrait trouver une place de choix à l'interface de la thérapeutique et de la neurophysiologie clinique. Des initiatives ont récemment vu le jour dans cette direction puisque s'est tenue récemment la 1ère journée nationale sur le neurofeedback qui s'est donnée pour objectif de structurer la pratique clinique et scientifique du neurofeedback en France dans les années à venir (48).

REFERENCES

1. Arns M, Heinrich H, Strehl U. Evaluation of neurofeedback in ADHD: the long and winding road. Biol Psychol. janv 2014;95:108-15.
2. Micoulaud-Franchi J-A, McGonigal A, Lopez R, Daudet C, Kotwas I, Bartolomei F. Electroencephalographic neurofeedback: Level of evidence in mental and brain disorders and suggestions for good clinical practice. Neurophysiol Clin Clin Neurophysiol. déc 2015;45(6):423-33.
3. Micoulaud-Franchi J-A, Fakra E, Cermolacce M, Vion-Dury J. [Towards a new approach of neurophysiology in clinical psychiatry: functional magnetic resonance imaging neurofeedback applied to emotional dysfunctions]. Neurophysiol Clin Clin Neurophysiol. avr 2012;42(3):79-94.
4. Fovet T, Jardri R, Linden D. Current Issues in the Use of fMRI-Based Neurofeedback to Relieve Psychiatric Symptoms. Curr Pharm Des. 2015;21(23):3384-94.
5. Rémond A, Rémond A. Biofeedback: principes et applications. Masson; 1994. 242 p.
6. Evans P. Biofeedback for bowel disorders. Indep Nurse [Internet]. 1 juin 2007 [cité 10 févr 2016];2007(6). Disponible sur: http://www.magonlinelibrary.com/doi/abs/10.12968/indn.2007.1.6.74962
7. Kropotov JD. Quantitative EEG, Event-Related Potentials and Neurotherapy. Academic Press; 2010. 601 p.
8. deCharms RC. Applications of real-time fMRI. Nat Rev Neurosci. sept 2008;9(9):720-9.
9. Sulzer J, Haller S, Scharnowski F, Weiskopf N, Birbaumer N, Blefari ML, et al. Real-time fMRI neurofeedback: progress and challenges. NeuroImage. 1 août 2013;76:386-99.
10. Naseer N, Hong K-S. fNIRS-based brain-computer interfaces: a review. Front Hum Neurosci. 2015;9:3.
11. Simoes M, Lima J, Direito B, Castelhano J, Ferreira C, Carvalho P, et al. Feature analysis for correlation studies of simultaneous EEG-fMRI data: A proof of concept for neurofeedback approaches. Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf. août 2015;2015:4065-8.
12. Thibault RT, Lifshitz M, Raz A. The self-regulating brain and neurofeedback: Experimental science and clinical promise. Cortex J Devoted Study Nerv Syst Behav. 18 nov 2015;74:247-61.
13. Arns M, de Ridder S, Strehl U, Breteler M, Coenen A. Efficacy of neurofeedback treatment in ADHD: the effects on inattention, impulsivity and hyperactivity: a meta-analysis. Clin EEG Neurosci. juill 2009;40(3):180-9.
14. Choi SW, Chi SE, Chung SY, Kim JW, Ahn CY, Kim HT. Is alpha wave neurofeedback effective with randomized clinical trials in depression? A pilot study. Neuropsychobiology. 2011;63(1):43-51.
15. Agnihotri H, Paul M, Sandhu JS. Biofeedback Approach in The Treatment of Generalized Anxiety Disorder. Iran J Psychiatry. 1 août 2007;2(3):90-5.
16. Sonuga-Barke EJS, Brandeis D, Cortese S, Daley D, Ferrin M, Holtmann M, et al. Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. Am J Psychiatry. mars 2013;170(3):275-89.
17. Micoulaud-Franchi J-A, Geoffroy PA, Fond G, Lopez R, Bioulac S, Philip P. EEG neurofeedback treatments in children with ADHD: an updated meta-analysis of randomized controlled trials. Front Hum Neurosci. 2014;8:906.
18. Leins U, Goth G, Hinterberger T, Klinger C, Rumpf N, Strehl U. Neurofeedback for children with ADHD: a comparison of SCP and Theta/Beta protocols. Appl Psychophysiol Biofeedback. juin 2007;32(2):73-88.
19. Holtmann M, Steiner S, Hohmann S, Poustka L, Banaschewski T, Bölte S. Neurofeedback in autism spectrum disorders. Dev Med Child Neurol. nov 2011;53(11):986-93.
20. Peeters F, Oehlen M, Ronner J, van Os J, Lousberg R. Neurofeedback as a treatment for major depressive disorder--a pilot study. PloS One. 2014;9(3):e91837.

21. Rice KM, Blanchard EB, Purcell M. Biofeedback treatments of generalized anxiety disorder: preliminary results. Biofeedback Self-Regul. juin 1993;18(2):93-105.
22. Kopřivová J, Congedo M, Raszka M, Praško J, Brunovský M, Horáček J. Prediction of treatment response and the effect of independent component neurofeedback in obsessive-compulsive disorder: a randomized, sham-controlled, double-blind study. Neuropsychobiology. 2013;67(4):210-23.
23. Servant D, Logier R, Mouster Y, Goudemand M. [Heart rate variability. Applications in psychiatry]. L'Encéphale. oct 2009;35(5):423-8.
24. Peniston EG, Kulkosky PG. Alpha-theta brain wave neurofeedback for Vietnam veterans with combat related post traumatic stress disorder. Medical Psychotherapy. 1991;4, 1-14.
25. Scott WC, Kaiser D, Othmer S, Sideroff SI. Effects of an EEG biofeedback protocol on a mixed substance abusing population. Am J Drug Alcohol Abuse. 2005;31(3):455-69.
26. Boynton GM, Engel SA, Glover GH, Heeger DJ. Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci Off J Soc Neurosci. 1 juill 1996;16(13):4207-21.
27. LaConte SM. Decoding fMRI brain states in real-time. NeuroImage. 15 mai 2011;56(2):440-54.
28. Hinds O, Ghosh S, Thompson TW, Yoo JJ, Whitfield-Gabrieli S, Triantafyllou C, et al. Computing moment-to-moment BOLD activation for real-time neurofeedback. NeuroImage. 1 janv 2011;54(1):361-8.
29. Weiskopf N. Real-time fMRI and its application to neurofeedback. NeuroImage. 15 août 2012;62(2):682-92.
30. Ruiz S, Buyukturkoglu K, Rana M, Birbaumer N, Sitaram R. Real-time fMRI brain computer interfaces: self-regulation of single brain regions to networks. Biol Psychol. janv 2014;95:4-20.
31. Caria A, Sitaram R, Veit R, Begliomini C, Birbaumer N. Volitional control of anterior insula activity modulates the response to aversive stimuli. A real-time functional magnetic resonance imaging study. Biol Psychiatry. 1 sept 2010;68(5):425-32.
32. Sulzer J, Haller S, Scharnowski F, Weiskopf N, Birbaumer N, Blefari ML, et al. Real-time fMRI neurofeedback: progress and challenges. NeuroImage. 1 août 2013;76:386-99.
33. Haller S, Birbaumer N, Veit R. Real-time fMRI feedback training may improve chronic tinnitus. Eur Radiol. mars 2010;20(3):696-703.
34. deCharms RC, Maeda F, Glover GH, Ludlow D, Pauly JM, Soneji D, et al. Control over brain activation and pain learned by using real-time functional MRI. Proc Natl Acad Sci U S A. 20 déc 2005;102(51):18626-31.
35. Amad A, Fovet T, Geoffroy PA. Keep calm: Psychiatric disorders are organic! The power of words in medicine. Aust N Z J Psychiatry. janv 2016;50(1):100-1.
36. Sitaram R, Caria A, Veit R, Gaber T, Ruiz S, Birbaumer N. Volitional control of the anterior insula in criminal psychopaths using real-time fMRI neurofeedback: a pilot study. Front Behav Neurosci. 2014;8:344.
37. Scheinost D, Stoica T, Saksa J, Papademetris X, Constable RT, Pittenger C, et al. Orbitofrontal cortex neurofeedback produces lasting changes in contamination anxiety and resting-state connectivity. Transl Psychiatry. 2013;3:e250.
38. Linden DEJ, Habes I, Johnston SJ, Linden S, Tatineni R, Subramanian L, et al. Real-time self-regulation of emotion networks in patients with depression. PloS One. 2012;7(6):e38115.
39. Hanlon CA, Hartwell KJ, Canterberry M, Li X, Owens M, Lematty T, et al. Reduction of cue-induced craving through realtime neurofeedback in nicotine users: the role of region of interest selection and multiple visits. Psychiatry Res. 30 juill 2013;213(1):79-81.
40. Young KD, Zotev V, Phillips R, Misaki M, Yuan H, Drevets WC, et al. Real-time FMRI neurofeedback training of amygdala activity in patients with major depressive disorder. PloS One. 2014;9(2):e88785.
41. Mataix-Cols D, Wooderson S, Lawrence N, Brammer MJ, Speckens A, Phillips ML. Distinct neural correlates of washing, checking, and hoarding symptom dimensions in obsessive-compulsive disorder. Arch Gen Psychiatry. juin 2004;61(6):564-76.
42. Micoulaud Franchi JA, Pallanca O. Neurofeedback. In: Neurophysiologie clinique en psychiatrie. Elsevier Masson. Paris; p. 185-212.
43. Micoulaud-Franchi, J-A., Daudet, C. Neurofeedback par électroencéphalographie en psychiatrie : remédiation neurophysiologique. EMC - Psychiatrie. 2015;1-7.
44. Strehl U. What learning theories can teach us in designing neurofeedback treatments. Front Hum Neurosci [Internet]. 6 nov 2014 [cité 18 févr 2016];8. Disponible sur: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4222234/
45. Gevensleben H, Moll GH, Rothenberger A, Heinrich H. Neurofeedback in attention-deficit/hyperactivity disorder – different models, different ways of application. Front Hum Neurosci [Internet]. 21 oct 2014 [cité 18 févr 2016];8. Disponible sur: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4204462/
46. Micoulaud-Franchi J-A, Lancon C. L'inaccessible presque touché. Connaissance minimale en psychologie de la santé à l'usage de l'étudiant en médecine. Ann Méd-Psychol Rev Psychiatr. mai 2015;173(4):377-83.
47. Linden DEJ. Neurofeedback and networks of depression. Dialogues Clin Neurosci. mars 2014;16(1):103-12.
48. NE&T. 1ère journée nationale sur le neurofeedback [Internet]. Disponible sur: https://www.youtube.com/watch?v=FvZRv995xtI&list=PLvZCt4kFF1NqlMag6NRtLv6RE5bfDR21N

3.1.2 Initiatives actuelles : la création du groupe NExT

Comme nous l'avons présenté précédemment, depuis le début des années 2000 le neurofeedback bénéficie d'un regain d'intérêt notamment dans le champ de la psychiatrie. Malheureusement, l'absence de réglementation et de guide de bonne pratique dans le domaine a également entrainé une augmentation du nombre d'applications commerciales prétendument « innovantes » n'ayant montré aucune efficacité (223). Ceci conduit essentiellement à ajouter de la confusion au débat sur l'efficacité du neurofeedback, avec un risque considérable de rejeter une technique potentiellement efficace du fait d'utilisations non optimales. Dans ce contexte, le développement de guides de bonne pratique clinique et d'une formation adéquate dans le domaine du neurofeedback, ainsi que de la mise en place d'études d'évaluation sur ses mécanismes d'action apparaissent indispensables (224).

L'objectif de la section *Neurofeedback Evaluation & Training (NExT)* de l'Association Française de psychiatrie biologique et neuropharmacologie (AFPBN) est de promouvoir l'évaluation et l'utilisation rigoureuse des techniques de neurofeedback en psychiatrie. Cette section a été créée en 2015 par Jean-Marie BATAIL, Stéphanie BIOULAC, Christophe DAUDET, Dominique DRAPIER, Thomas FOVET, Renaud JARDRI et Jean-Arthur MICOULAUD FRANCHI (ordre alphabétique) ; elle est constituée de membres actifs qui publient dans le domaine du neurofeedback en psychiatrie.

Ses missions s'articulent autour de 3 grands axes : la diffusion de recommandations de bonne pratique, la formation et la promotion de la recherche dans le domaine.

3.1.2.1 Recommandations de bonne pratique

Les objectifs s'inscrivent dans la poursuite des travaux réalisés par les membres fondateurs :

- réaliser des gradations de niveau de preuve concernant l'utilisation du neurofeedback en psychiatrie,
- réaliser des accords d'experts concernant les niveaux de preuves si cela est nécessaire,
- réaliser des guides de bonne pratique technique sur l'utilisation clinique du neurofeedback.

3.1.2.2 Formation

Les objectifs sont de :

- former les praticiens du neurofeedback à l'utilisation des outils de neurophysiologie clinique et à la mise en place de séances de remédiation neurophysiologique fondées sur les paradigmes de l'apprentissage,
- mettre en place des modules de formation.

3.1.2.3 Promotion de la recherche

Les objectifs sont de :

- regrouper les professionnels, praticiens et chercheurs en neurofeedback afin de développer et de promouvoir la recherche dans le domaine de la psychiatrie, notamment par la mise en place de projets de recherches cliniques de type PHRC inter-régional,
- structurer un réseau autour du neurofeedback en psychiatrie par :
    - Un lien particulier avec le club *STEP* (*Stimulation Transcrânienne en Psychiatrie*) de l'AFPBN afin d'intégrer le neurofeedback dans le champ plus vaste des techniques de neuromodulation en psychiatrie,
    - un lien avec le projet Interface Cerveau Ordinateur (*Brain Computer Interface ; BCI*) financé par l'ITMO Neurosciences, Sciences Cognitives, Neurologie, Psychiatrie et le Groupe National de R&D sur les interfaces cerveau-machine (dirigé par Francois Cabestaing), afin d'intégrer le neurofeedback dans le champ plus vaste des BCI.

3.1.2.4 Diffusion

La diffusion des connaissances et des pratiques passe par :

- La réalisation de journées nationales sur le neurofeedback ;
- La réalisation de journées régionales sur le neurofeedback ;
- La rédaction d'articles de synthèse.

La première journée nationale sur le neurofeedback s'est tenue le 19 janvier 2016 à l'Institut du Cerveau et de la Moelle Epinière, en partenariat avec l'AFPBN. Cet évènement a rassemblé 220 personnes. La deuxième journée nationale sur le neurofeedback s'est tenue le 25 janvier 2017 à l'ESPCI Paristech. Les vidéos des interventions proposées au cours de ces journées sont disponibles sur la chaine *Youtube* de la section NExT : https://www.youtube.com/channel/UC1ZWzvYuSZQC4b_qPRgBpPg

Une journée régionale s'est également tenue à Bordeaux le 7 juillet 2016 sous la forme d'un *workshop*. Un article de synthèse issu de la première journée nationale sur le neurofeedback a été publié dans le journal *L'Encéphale* (voir **Article 6**) accompagné d'un éditorial signé par Martijn Arns (224). Le contenu de la deuxième journée a également fait l'objet d'une publication scientifique (voir **Article 7**) visant à proposer une réflexion sur la conception des protocoles de neurofeedback à partir d'une synthèse des connaissances actuelles sur la méthode.

Enfin, avant la 3ème édition de la journée nationale sur le neurofeedback prévue en mai 2018, la section NExT sera représentée au Congrès de L'Encéphale 2018 avec un symposium intitulé « *Le neurofeedback en psychiatrie : nouvelles perspectives* » présidé par le Professeur Fakra et le Professeur Drapier avec des interventions centrées sur le niveau de preuve de l'efficacité thérapeutique du neurofeedback en psychiatrie (Jean-Marie Batail), la mise en place de recommandations françaises de bonne pratique pour l'utilisation du neurofeedback en psychiatrie (Jean-Arthur Micoulaud Franchi) et les perspectives de recherche (Thomas Fovet).

ARTICLE 6

Neurofeedback: one of today's techniques in psychiatry?

Auteurs par ordre alphabétique : Martijn ARNS[1], Jean-Marie BATAIL[2], Stéphanie BIOULAC[3], Marco CONGEDO[4], Christophe DAUDET[5], Dominique DRAPIER[2], Thomas FOVET[6], Renaud JARDRI[6], Michel Le VAN QUYEN[7], Fabien LOTTE[8], David MEHLER[9], Jean-Arthur MICOULAUD FRANCHI[3]*, Diane PURPER-OUAKIL[10], François VIALATTE[11], The NExT group[12]

1. Research Institute Brainclinics Nijmegen, Netherlands ; Department of Experimental Psychology, Utrecht University Utrecht, Netherlands.
2. Academic Psychiatry Department, Centre Hospitalier Guillaume Régnier, Rennes, France; EA 4712 Behavior and Basal Ganglia, CHU Rennes, Rennes 1 University, France.
3. Univ. Bordeaux, SANPSY, USR 3413, F-33000 Bordeaux France, CNRS, SANPSY, USR 3413, F-Bordeaux, France.
4. GIPSA-lab, CNRS and Grenoble Alpes University, Grenoble, France.
5. Univ. Bordeaux, F-33000 Bordeaux France
6. Lille, CNRS, UMR-9193, SCA-Lab, équipe psyCHIC et CHU Lille, Pôle de psychiatrie (CURE), Hôpital Fontan, F-59000 Lille, France.
7. Institut du Cerveau et de la Moelle Epinière, UMRS 1127, CNRS UMR 7225, Hôpital de la Pitié-Salpêtrière, Paris, France, Université Pierre et Marie Curie - Paris 6, Paris, France
8. Inria Bordeaux Sud-Ouest, Project-Team Potioc / LaBRI – Talence, France.
9. Cardiff University Brain Imaging Centre (CUBRIC), School of Psychology, Cardiff University, United Kingdom; Department of Psychiatry and Psychotherapy, University of Muenster, Muenster, Germany
10. CHU Montpellier-Saint Eloi, Médecine psychologique pour enfants et adolescents (MPEA1), 80 avenue Augustin Fliche, 34295 Montpellier Cedex.
11. Laboratoire Plasticité du Cerveau, UMR 8249, ESPCI Paris Tech, PSL Research University, 10 rue Vauquelin 75005 Paris, France.
12. The NExT (Neurofeedback Evaluation & Training) group of the AFPBN

L'Encéphale. 2017 ; 43(2) : 135-145.

ABSTRACT

Objectives

Neurofeedback is a technique that aims to teach a subject to regulate a brain parameter measured by a technical interface to modulate his/her related brain and cognitive activities. However, the use of neurofeedback as a therapeutic tool for psychiatric disorders remains controversial. The aim of this review is to summarize and to comment the level of evidence of electroencephalogram (EEG) neurofeedback and real-time functional magnetic resonance imaging (fMRI) neurofeedback for therapeutic application in psychiatry.

Method

Literature on neurofeedback and mental disorders but also on Brain Computer Interfaces (BCI) used in the field of neurocognitive science has been considered by the group of expert of the NExT (Neurofeedback Evaluation & Training) section of the French Association of Biological Psychiatry and Neuropsychopharmacology (AFPBN).

Results

Results show a potential efficacy of EEG-neurofeedback in the treatment of attentional-deficit/hyperactivity disorder (ADHD) in children, even if this is still debated. For other mental disorders, there is too limited research to warrant the use of EEG-neurofeedback in clinical practice. Regarding fMRI-neurofeedback, the level of evidence remains too weak, for now, to justify clinical use. The literature review highlights various unclear points, such as indications (psychiatric disorders, pathophysiologic rationale), protocols (brain signals targeted, learning characteristics), and techniques (EEG, fMRI, signal processing).

Conclusion

The field of neurofeedback involves psychiatrists, neurophysiologists and researchers in the field of brain-computer-interfaces. Future studies should determine the criteria for optimizing neurofeedback sessions. A better understanding of the learning processes underpinning neurofeedback could be a key element to develop the use of this technique in clinical practice.

Keywords

Neurofeedback; EEG; real-time fMRI; psychiatric disorder

Introduction

Neurofeedback can be considered as a biofeedback technique (*i.e.* a technique which consists in measuring a physiological activity using a technical interface to extract a parameter of interest; this parameter is then presented in real-time to the participant, typically *via* visual or auditory feedback [1]; the goal is to teach the subject to modify the parameter). When the physiological activity is a brain activity, biofeedback is called neurofeedback. Thus, neurofeedback allows the subject to voluntary modulate his/her related brain and cognitive activities [1, 2] (see **Figure 1).**

The first observation of neurofeedback, was based on the classical conditioning principles applied to the electroencephalogram (EEG). Classical conditioning involves learning new behaviors through the process of association. Neurofeedback originates from the 1930s based on the work of Gustave Durup and Alfred Fessard, who were two emblematic figures of psychophysiology and neurophysiology in France. They observed that brain activity (alpha blocking response) could be modified according to the classical conditioning principles (*i.e.* to develop an association between an EEG activity (alpha blocking response), a behavior and cognitive response, and a signal of feedback [3]. In 1941, Jasper & Shagass published the first systematic study that investigated classical conditioning of EEG [4]. Subsequent studies in the 1960s confirmed that alpha blocking could indeed be conditioned and related to some specific cognitive activities of the trained subject [5].

After a serious decline during the 1980s and 1990s, mainly due to the poor reliability of methods used for recording brain activity, the technique gained ground again in the early 2000s with a renewed interest both in scientific and societal terms [6]. Thanks to the principle on which it is based and to the fertile dynamic nature of ongoing research in a range of clinical, therapeutic and fundamental topics, neurofeedback can be considered a technology of today [6, 7]. However, despite great interest in neurofeedback research [8-10], significant controversy exists, particularly in psychiatry and neurology [7, 11]. With regard to the efficacy of neurofeedback in brain disorders, opinions within the scientific community appear to be rather sharply divided [7, 9, 12] comprising an optimistic group who consider neurofeedback to be effective and a skeptical group who do neither assign scientific or therapeutic value to neurofeedback training. This article aims to review the evidence of EEG neurofeedback (EEG NF) and real-time functional magnetic resonance imaging neurofeedback (fMRI NF) in psychiatric disorders. The advantages and pitfalls for each of both neurofeedback techniques are discussed, and new perspectives are highlighted. Lastly, research on the learning process through the link between neurofeedback and brain computer interfaces (BCIs) is discussed.

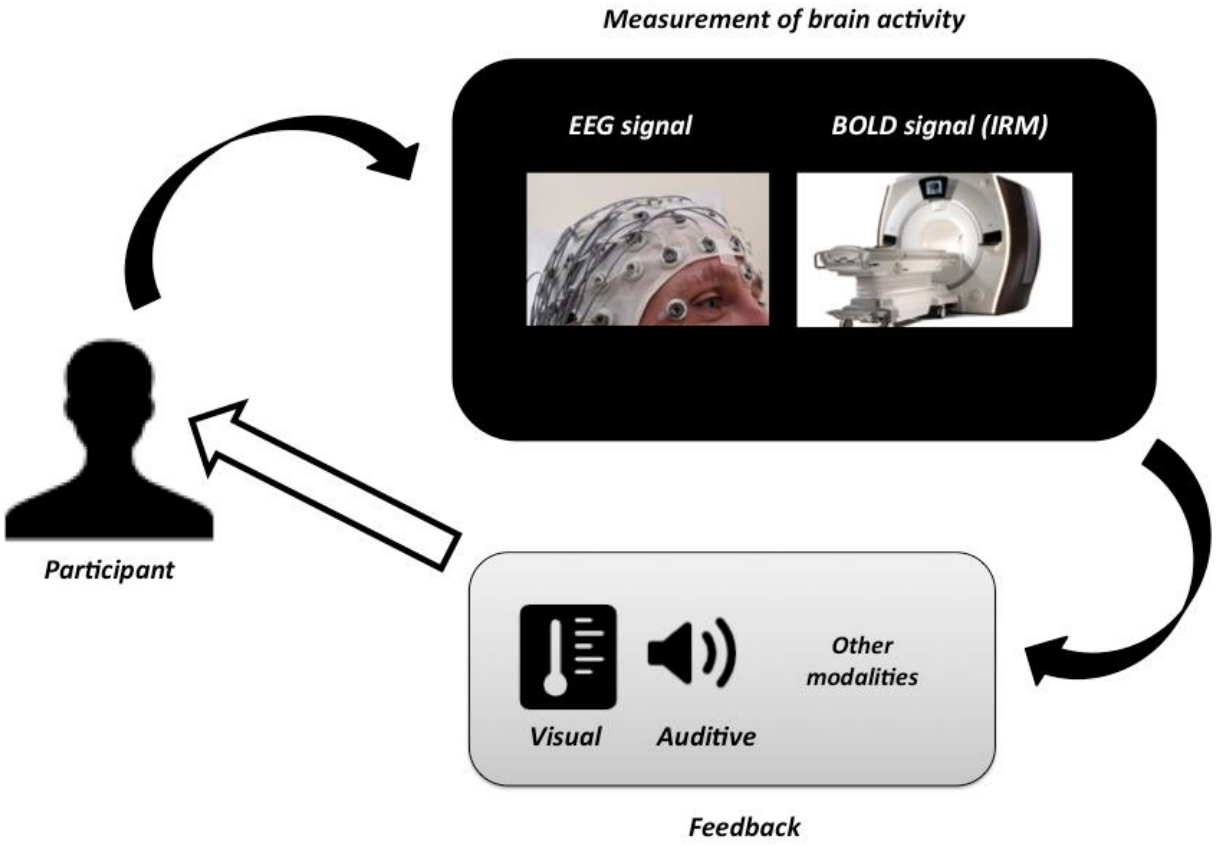


Figure 1. Principle of neurofeedback.

Electroencephalographic neurofeedback (EEG-NF)

Level of Evidence

Most trials on the efficacy of EEG neurofeedback in psychiatric disorders have significant methodological weaknesses (in particular: size of the population studied, none randomized or none blinded protocol, inadequate control group, low quality of the EEG neurofeedback session) [13]. This point could explain the skepticism of many researchers and clinicians concerning the effectiveness of EEG neurofeedback to treat psychiatric disorders [12]. However, a number of studies have presented good methodological criteria (studies designed with controlled, randomized, and open or blind protocols, a primary endpoint related to the treated disorder and assessed using standardized measurement tools, and an identifiable EEG neurophysiological target) particularly in the field of attentional-deficit/hyperactivity disorder (ADHD) [9, 12, 14].

Attentional-Deficit/Hyperactivity Disorder, the emblematic disorder

Four meta-analyses discussed the therapeutic interest of EEG neurofeedback in ADHD [15-18]. Computed effect size (ES) in the meta-analyses can be considered as small between 0.2 and 0.5, medium between 0.5 and 0.8 and large above 0.8. The first meta-analysis conducted by Arns et al. (2009) found an effect size (ES) that was more larger for the domain of inattention (ES=0.81, 95% CI=0.39-1.23) than for the domain of hyperactivity (ES=0.39, 95% CI=0.05-0.75) in ADHD [16]. The second meta-analysis of Sonuga-Barke et al. (2013) found a significant ES using parent ratings in randomized controlled trials (RCTs) (ES=0.59, 95% CI=0.31-0.87), but this result was no longer significant (ES=0.29, 95% CI=-0.02-0.61, though trend, p=0.07) when looking at “probably blinded” teacher ratings [17]. The third meta-analysis of Micoulaud-Franchi et al. (2014) found an ES that was significantly higher than in the control group on “probably blinded” teacher ratings for the inattention dimension of ADHD in RCTs (ES=0.30, 95% CI=0.03-0.58) [18]. The fourth meta-analysis of Cortese et al (2016) is the updated Sonuga-Barke et al. meta-analysis and reported similar results (ADHD total symptoms, ES=0.35, 95% CI=0.11-0.59; inattention, ES= 0.36, 95% CI= 0.09-0.63; hyperactivity/impulsivity, ES=0.26, 95% CI=0.08-0.43 for parent ratings, but non significant ES for “probably blinded” teacher ratings) [15]. However, a sub-analysis in this meta-analysis focused on standard neurofeedback protocols (based on the Arns et al. criteria [12]), and for this sub-analysis a significant ES for probably blinded ratings was found (ADHD total symptoms ES=0.35, 95% CI=0.04-0.69) [12]. RCTs that have compared EEG neurofeedback with medication found that methylphenidate was not superior to EEG neurofeedback training [19, 20]. In the study of Meisel et al. (2013), significant pre-post academic performance improvements were obtained only in the neurofeedback group [19]. However, studies that added EEG neurofeedback to methylphenidate treatment did not report ‘add-on’ improvements on clinical symptoms [21, 22] or cognitive function [23].

Other psychiatric disorders

There has been too limited research (i.e. lack of RCTs and independent replications) on the following indications to warrant its use in clinical practice: Depression [24], Addictions [25, 26], Anxiety disorders [27, 28].

Advantages and pitfalls of EEG neurofeedback

Despite the meta-analyses presented before, the effectiveness of EEG neurofeedback in treating ADHD remains debated because of the studies that were included [12, 29-34]. These choices warrant some explanations. For example, in the meta-analysis of Micoulaud-Franchi et al. (2014), the well-controlled, randomized and blinded study conducted by Arnold et al. (2013) [35] was not included because the EEG neurofeedback protocol was not based on the basic learning theory used in standard EEG neurofeedback protocols (particularly because of the type of reinforcement chosen) [1]. Moreover, the EEG recording was carried out using an unconventional setup, with electrodes placed on the forehead, a region known to be problematic for recording because of muscular artefacts. The study by Arnold et al. thus highlights the need to avoid some pitfalls regarding technical issues of electrophysiology [36] and technical issues of learning [1, 37] when a study on neurofeedback is conducted. In further support of this notion is the above reported result from the Cortese et al. (2016) meta-analysis, who reported that when focusing on 'standard neurofeedback protocols' significant effects are found for both parent as well as teacher rated symptoms. Further emphasizing the need to evaluate neurofeedback not as a singular phenomenon (neurofeedback as an umbrella term i.e. medication) but evaluate it based on the specific protocol used (specific protocol i.e. antidepressant, psychostimulant) [15]. These aspects are too rarely discussed in the debate of EEG neurofeedback efficacy. Considering the absence of a current consensus [12, 38-40], these points will be crucial in the next years to gradually improve the practice of EEG neurofeedback in psychiatry [41].

Two groups of technical issues can be identified in EEG neurofeedback protocols: i) electrophysiology because the practice of EEG neurofeedback requires high quality recordings of EEG signal [9, 36]; ii) learning because the practice of EEG neurofeedback requires attention to some important technical aspects as described below and in **Table 1**.

The number of sessions is the first technical aspect, which is usually between 20 and 30, one to three times per week, but the ideal number and the optimum inter-session duration have not been defined yet [42]. It should be noted that efficacy with regards to the inattention dimension in ADHD is proportional to the number of neurofeedback sessions [16] and seemed to be maintained over time [43].

Second is the choice of the threshold of reward, which is essential. Adjusting a threshold (and a given occupation time) determines the number of positive reinforcements required to strengthen the subject in a type of neurocognitive strategy. The threshold may be set automatically or manually. When the threshold is determined automatically there is a continuous updating of a threshold in order to give positive reinforcement to the subject for a given percentage of occupation time below or above the threshold. The threshold is continuously calculated according to signal just before. When the threshold is determined manually, the professional determines the threshold based on a baseline recorded before the neurofeedback session. If the number of positive reinforcement is too high or too low during the session, the professional can adjust the threshold. The manual threshold seems to lead to better learning [1, 42]. Indeed, if the subject is being asked to increase the amplitude of a given brain activity and the threshold is calculated automatically, he will always be getting a percentage of feedback even if the amplitudes are decreased across time. However, the manual threshold requires performing a

baseline measurement before each session and the adjustment during the session by the professional complicates the standardization of neurofeedback protocol.

Third is the type of positive reinforcement. This can be visual or auditory, proportioned (graduated) or binary (present or absent), immediate or delayed, simple or complex, and frequent or rare. Visual feedback, which is proportionate, immediate and simple, seems to allow for better learning [42]. The number of reinforcements must be sufficient to maintain the motivation of the subject. However, if the number of reinforcement is too high the learning process can be altered [39, 42]. Note that positive reinforcement incorporated in an entertaining interface (such as video games) may increase the motivation of the subject but could impair learning according to some authors [1, 14].

Fourth is the evaluation of the training parameter during one session (evolution of the performance), and the evaluation of the learning curve across the sessions (evolution of the training parameter) that should be determined to ensure that a learning process occurs during neurofeedback treatment. Lastly, the “transfer sessions” allow for the generalization of skills learned in daily life [12, 14, 40].

Aim of the learning during neurofeedback	
Learnability	The parameters of interest can be regulated by the learner
Perceptibility	The parameter of interest can be perceived by the learner without exceeding his/her perception capabilities
Mastery	The learner gains progressively control over the sessions
Motivation	The learner should be preserved from boredom and not experience disengagement from the task
Autonomy	The learner achieves progressive independence from the feedback and can self-regulate the brain signal of interest without feedback
Technical aspects related to the learning	
Quality of signal recording	Quality of the signal-to-noise ratio / Method to avoid artefact
Signal processing	Signal processing method to compute the parameter of interest
Occupation time	Time above or below a threshold until a reward is given
Threshold	Automatically adapted or manually
Number of positive reinforcements	Number of positive reinforcements above or below a certain number until the threshold is modified
Perceptual modality of feedback	Type of cue used to provide feedback (e.g. visual, auditory or tactile)
Mode of feedback presentation	Continuous or intermittent
Complexity of the feedback	e.g., for visual feedback, a thermometer display or more complex scenes based on virtual reality
Number of sessions	Number of session to obtain a learning
Duration of a session	Duration of a session and number of block per session
Inter session duration	Duration between two sessions
Training curve	Evaluation of the training parameter during the session
Learning curve	Evolution of the training across the sessions
Role of the professional	Task instructions and motivation given to the subject before, during and after the session
Transfer sessions	Generalization of learned skills to activities of daily living i.e. in an ecologically relevant setting

Table 1. Principles and technical aspects of learning during neurofeedback.

EEG neurofeedback and the vigilance system

Neurophysiological targets for EEG neurofeedback in ADHD are underpinned by pathophysiological relevance related to the vigilance system. EEG neurofeedback traditionally records a limited amount of information provided by a single electrode placed on the scalp. This information concerns the EEG power in certain spectral bands: the beta band (12-21 Hz) and the theta band (4-8 Hz) [44, 45]. In a simple manner, an increase in the central frontal beta band can be related to an increase in vigilance [46], and an increase in central frontal theta band is related to a decrease in vigilance with subjective diurnal sleepiness and possibly entering the first stage of sleep [45, 47]. Interestingly, an increase in theta power and a decrease in beta power were observed in a subgroup of ADHD patients (greater theta/beta (TBR) ratio) [48]. These EEG patterns suggest a link between the vigilance system, sleep problems and ADHD (particularly in the subgroup with the greater TBR ratio) [49]. As a result, decreasing TBR can be a potentially interesting target for EEG neurofeedback [50-52]. Indeed, it was shown that TBR neurofeedback is more effective in the subgroup of patients with the greater TBR ratio [53].

Several studies have also demonstrated that sensori-motor rhythm neurofeedback (SMR), a frequency that overlaps with to the TBR protocol, results in increased sleep spindle density during sleep [54, 55], decreased sleep latency [54] and increased total sleep time [54, 56]. More specifically, it was recently demonstrated that SMR neurofeedback in ADHD resulted in reduced inattention, hyperactivity and impulsivity, and these effects were mediated by reduced sleep onset latency [50], further demonstrating a causal link between delayed sleep onset latency and ADHD symptoms, specifically inattention. The TBR neurofeedback overlaps with the SMR protocol, with clinical effects on ADHD indistinguishable from SMR neurofeedback. However, the effect of TBR neurofeedback was not be mediated via sleep onset latency normalization [50]. The effect of TBR neurofeedback could be mediated via a reduction in diurnal sleepiness [49], but further research is needed to investigate the exact working mechanism of TBR neurofeedback in ADHD [14].

EEG neurofeedback and new target methods

The major limitation of “traditional” neurofeedback resides in the limited information provided by a single electrode placed on the scalp, which is a differentially measured potential with respect to a reference electrode. It is known that the EEG signal reflects mainly the superposition of the electric potential created by ionic charge oscillation (due to postsynaptic potentials) around the pyramidal cells found in the neocortex [57]. The potential generated from a large population of neurons beneath the electrode are superimposed to create the measurable EEG. Put differently, the response of the electrode is highly spatially unspecific. It has been suggested that this lack of spatial specificity may impede the ability of subjects to acquire control over the region of interest (ROI), i.e., the brain structures to be trained [58]. Another limitation of traditional neurofeedback is the filtering resulting from the choice of the reference electrode placement; depending on the position of the active and reference electrode on the scalp, the measurement is sensitive to current flowing in the ROI along one direction only. Therefore, a considerable improvement in the neurofeedback technique can be obtained considering spatial-specific brain activity, solving implicitly the issue of the chosen reference. Two possible improvements in this sense have been proposed, namely, basing the neurofeedback not on the signal captured by the two

scalp electrodes but on EEG inverse solutions or on EEG blind source separation. Both methods require the use of multiple electrodes (a minimum of eight); it is indeed the spatial information contained in such a multivariate EEG recording that allow for better estimates of the ROI's current.

EEG neurofeedback based on inverse solutions

An EEG inverse solution is a mathematical method used to estimate the intracranial current generated in the observed scalp potential. Once the current is estimated in the ROI, its density (energy) provides an appropriate feedback signal. By acquiring data from 19 electrodes, Congedo, Lubar and Joffe (2004) demonstrated learned control of the cognitive division of the anterior cingulate cortex using the inverse solution known as *low resolution electromagnetic tomography* (LORETA) [59, 60]. Subsequent studies confirmed the viability and further explored the correlates of LORETA-neurofeedback of the anterior cingulate cortex [61, 62]. This preliminary work was replicated and reiterated later by several other research groups using other inverse solutions in proof-of-concept studies [63, 64].

EEG neurofeedback based on BSS/ICA

Over the past 20 years, research on blind source separation (BSS) has developed into a burgeoning signal processing method with applications across a wide variety of fields. It has since been proven valuable in identifying cortical sources of brain activity associated with cognitive task performance [65]. Such a spatial filtering technique may provide an ideal way to train specific brain regions or networks in a neurofeedback setting. In fact, a blind source separation filter can estimate both the location and the direction of current, thus yielding a sharper filter compared to an inverse solution [66]. Further advantages of such spatial filters are that they are computationally inexpensive (important for 'real-time' feedback) and potentially more robust in the presence of artefacts. The viability of BSS neurofeedback has been explored in two studies; the first aimed to suppress excessive theta in deep frontal medial regions for the treatment of obsessive-compulsive disorder [67], the second aimed to enhance theta activity on a source localized into deep medial-temporal regions associated with spatial-navigation abilities [68].

EEG neurofeedback based on stereotactic EEG

As early as the 1960s, the important work by Fetz (1969) on primates showed the operant conditioning of single cell spike trains in the motor cortex [69]*. The motor cortex is probably the most obvious place to search for cortical signals directly associated with volitional movement* [70]*. This may be one of the reasons why a substantial part of invasive neurofeedback research has been conducted on paralyzed or lock-in patients, recognizing the need of people with disabilities and aiming to restore their communicative or motor functions. In this context, brain-computer interfaces (BCIs) were tested in amyotrophic lateral sclerosis, brain stem stroke and spinal cord lesions using cortical neuronal activity recorded by implanted electrodes* [71]*. Nevertheless,* conscious control has also been shown to be possible at the cellular level *in human temporal lobe structures* [72]. *The successful cases in these applications encouraged the usage of invasive neurofeedback for other neurological and neuropsychiatric conditions. Such a technique has been called BrainTV [73]. The technique enables to combine the spatial resolution of fMRI neurofeedback and the temporal resolution of scalp-level EEG*

neurofeedback [74]. Thus, despite the invasive nature of BrainTV, these protocols could be a response to some limitations of neurofeedback protocols in the future.

In this context, *neurofeedback can indeed be performed* in patients with drug resistant epilepsy undergoing long-term monitoring, where depth electrodes are implanted for clinical diagnostics. The effects of self-induced intracortical oscillatory activity (4-8 Hz) were studied in several neurosurgical patients. It was found that subjects learned to robustly and specifically induce oscillations in the target frequency, confirmed by increased oscillatory event density [75]. As controls improved during learning, induced oscillatory activity at the target electrode became functionally decoupled from distant sites, which predicted the individual session-to-session performance variability. Furthermore, in another study [75], patients were trained to up-regulate the relative proportion of the gamma rhythm at different fronto-temporal cortical locations. In line with previous findings, on monkeys using direct cortical recordings [76], it was found that most subjects learned to specifically increase local cortical gamma power. These findings suggest that the effects of voluntary control of intracortical oscillations can be exploited to specifically target plasticity processes to reconfigure network activity, with a particular relevance for memory function or skill acquisition [77]. In particular, abnormalities in gamma oscillations exist in a number of neurologic and psychiatric diseases [78]. Thus, the specific rectification of gamma oscillations could ameliorate some of the deficits caused by these pathological conditions [77].

Functional magnetic resonance imagery and neurofeedback

Real-time functional magnetic resonance imaging neurofeedback (fMRI neurofeedback) is a rather recent development for providing neurofeedback training based on blood oxygenation contrasts (blood-oxygen level dependent, BOLD) [79]. fMRI neurofeedback training can overcome some limitations of more traditional forms of neurofeedback, such as EEG-neurofeedback, because of its better spatial resolution and whole brain coverage. In particular, the whole brain coverage makes fMRI neurofeedback a promising technique for non-invasive psychiatric rehabilitation because it allows for training patients in self-regulating subcortical brain areas [80]. Depending on the disease model of interest, patients can be either trained to increase or decrease the activity of relevant brain areas [10].

Level of Evidence

Due to the novelty of the technique, the studies that have so far provided evidence for the clinical use of fMRI neurofeedback are limited. This section will focus on recent developments in the field and on clinical and translational applications. A more comprehensive review on relevant designs and training paradigms can be found elsewhere [10].

<u>Major Depressive Disorder, the emblematic disorder</u>

The psychiatric disorder most studied in the context of fMRI neurofeedback is major depressive disorder. The use of fMRI neurofeedback in treating depression is based on the pathophysiological model of emotional dysfunction during a depressive episode [81, 82]. Therefore, published studies have so far mainly focused on the up-regulation of brain areas or even on specific structures that are involved in emotions, including parts of the limbic system (e.g., the amygdala) and the ventral prefrontal cortex [83]. To date, no randomized control trials (RCTs) have been published, and the current literature consists

exclusively of open label and pilot studies [84-86]. These studies have demonstrated the feasibility of the technique and suggested that patients are able to self-regulate their brain activity in target areas. Further, improvements in mood were only found in the group that received fMRI neurofeedback training but not in a control group, suggesting a link between neurofeedback success, positive emotions (as accessed by self-reports in autobiographic memory recall and happiness ratings) and clinical improvement (e.g., HDRS-17). To rule out the unspecific effects (e.g., regression to the mean) of these pilot findings, RCTs are needed that are based on larger samples and appropriate clinical control conditions, including randomization and blinded assessments. Two ongoing (Young, clinicaltrials.gov: NCT02709161; Moll et al., NCT01920490), one completed single blind (Linden et al., NCT01544205), and one completed double blind (Young et al., NCT02079610) RCTs are currently listed.

Other psychiatric disorders

For other psychiatric conditions, such as schizophrenia, addiction, obsessive compulsive disorder and eating disorder, the feasibility of fMRI neurofeedback training has been investigated in pilot studies with small sample sizes (for review [10]). These studies used different target areas such as the insula in schizophrenia (based on a facial emotion recognition paradigm) and in psychopathic personality disorder (regulation of fear circuitry) and the anterior cingulate cortex in controlling cravings in nicotine addiction. The Collaborative Research Project BRAINTRAIN is a European consortium that focuses on the improvement and translation of real-time fMRI neurofeedback protocols for clinical applications (braintrainproject.eu). Current registered RCTs investigate therapeutic effects of fMRI neurofeedback in alcohol addiction (Linden et al., NCT02486900), Anxiety in adolescents (Cohen-Kadosh et al., NCT02440451) and autism spectrum disorder (Castelo-Branco et al., NCT02440451). Finally, an independent RCT is focusing on training the functional connectivity between reward- and impulse-related brain areas in eating disorders (Hallschmid et al., NCT02148770).

Advantages and pitfalls of fMRI neurofeedback

The gold standard for evaluating a therapeutic technique requires assessing its efficacy in a double-blind randomized and placebo-controlled trial. However, some of these requirements can pose a challenge for the evaluation of fMRI neurofeedback training. First, implementing a double-blind design can be limited because most current training protocols require (at least in the early learning phase) that patients engage in specific conscious processes in the form of explicit mental strategies.

Second, designing an appropriate placebo-controlled condition for neurofeedback protocols requires careful consideration depending on the study type. Three main types of controls have been proposed and tested so far:

- Transfer runs, during which patients are instructed to engage in the same cognitive strategies in or outside the scanner but without being provided with neurofeedback.
- “Sham” neurofeedback, which entails either random or yoked feedback based on some other patient’s brain activity. However, sham feedback bears the risk that patients notice the non-contingency of the feedback [10].

- An active control group that receives veridical feedback from target areas of another functional system that is neither involved in the pathophysiology of the respective condition nor in the task (i.e., cognitive strategy) of interest. However, a recent study has demonstrated that neurofeedback training itself involves various brain regions besides the individual target areas, including structures of reward circuitry (basal ganglia, striatum) and parts of the prefrontal cortex [87]. Further, such a control group cannot control for potential unspecific effects due to the high-tech laboratory setting. Including a third treatment as a usual control group that receives standard therapy could address this problem at the expense of increased trial costs.

Third, it remains to be tested how to optimize neurofeedback protocols for psychiatric conditions. This includes:

- Defining effective target areas or the networks for a particular psychiatric condition based on a pathophysiological model. Target areas can either be chosen a priori based on anatomical landmarks, or they can be functionally defined using a so-called “localizer” task (e.g., presenting emotionally valenced visual stimuli in a neurofeedback protocol for depression [86]). Similarly, target areas for functional connectivity-based neurofeedback are determined by the correlation of activity among brain areas that belong to a network of interest.
- Determining efficient study designs with regard to the duration and number of sessions to exploit regarding the learning capacities of patients who have cognitive impairments (e.g., attention and memory deficits).
- The nature of task instructions for patients, either given explicit strategies at hand (e.g., imaging positive autobiographical memories) or task instructions that rather focus on the goal to achieve a certain target level in the feedback while patients learn implicitly the effect of various strategies [88].
- The design of the interface, such as the modality of feedback (e.g., visual, auditory or tactile), the mode of feedback presentation (e.g., continuous or intermittent) and the complexity of the presented feedback (e.g., for visual feedback, a thermometer display or more complex scenes based on virtual reality)

fMRI neurofeedback and new target method

As previously described, different strategies exist to optimally define the brain target, or the region(s) of interest (ROI), in fMRI neurofeedback protocols [10]. This ROI can be localized using structural information but can also be functionally defined. In the latter, the patient is asked to perform a specific task in the scanner, and the highlighted areas can be used as the ROI for the fMRI-neurofeedback in a second step (e.g., in [86]).

For fMRI neurofeedback with a therapeutic purpose, both of these methods rely on our a priori knowledge of the underlying neural mechanisms of the disorder/symptom we want to relieve. Such strategies appear very relevant for disorders with persistent (or tonic) symptoms, i.e., symptoms that do not change much over time (e.g., depressive mood) but pose special challenges for more acute symptoms, characterized by intrusiveness and phasic activity (e.g., hallucinations in schizophrenia or

obsessions in obsessive compulsive disorder). For the latter symptoms, which are associated with transitory brain-states, strategies using pre-defined anatomical targets appear poorly appropriate. On the contrary, training patients to self-regulate the activity of brain regions that re-activate during the occurrence of subjective symptoms could be an interesting alternative.

To address this issue, a first method could be to induce symptoms while scanning to localize functional activations associated with the targeted subjective experience that can then be used as the ROI for fMRI neurofeedback. However, in some cases (such as hallucinations [89]), symptom provocation may not be possible, and another method to detect the onset of symptoms together with the associated brain activation patterns is needed.

Machine-learning, and particularly the recent development for fMRI analysis of linear support vector machines (lSVMs), offers several advantages in this context. Such techniques classify functional or anatomical patterns using a multivariate strategy and thus allow for decoding and capturing the fine-grained spatial pattern of BOLD activity to predict future mental states, such as perception or free choices [90]. In the same way, it is now possible to develop classifiers able to quickly detect the emergence of subjective symptoms by detecting specific patterns of brain activity identified during symptomatic periods [91, 92]. Such fine-grained activity patterns can be used as the signal that is fed back to the patient during neurofeedback protocols. However, to be eligible for this strategy, the patient's symptoms must exhibit some specific features, such as frequent occurrence (i.e., the symptom must occur several times during the fMRI session) [93].

Combining lSVM (or other advanced machine learning classifiers) and fMRI neurofeedback could constitute a promising way to develop fMRI neurofeedback for the treatment of phasic psychiatric symptoms. However, considering the potential cost necessary to implement fMRI neurofeedback, proof-of-concept studies are urgently required.

Human learning and neurofeedback

The learning process is crucial in neurofeedback and requires models to understand the mechanism of feedback learning [94]. A good practice guide is also of critical importance for the evaluation of these interventions and to reach higher standards in clinical practice [9]. Learning during neurofeedback can be either explicit or implicit [94]. In the explicit learning process, the user observes a feedback signal, which is a direct correlate of the neurosignal to be regulated. In the implicit learning process, the signal is not explicitly presented to the subject but instead changes some detail(s) of the experimental conditions. For example, a person using a videogame whose content (e.g., changing levels of difficulty or access to bonus items) evolves depending upon his frontal alpha rhythm is receiving implicit feedback; he/she does not know directly that his brainwaves have changed, but he/she experiences indirect effects of this physiological change.

Theory of human learning

From the perspective of the experimenter, operant conditioning has historically been the dominant interpretation of neurofeedback mechanisms; in this case the feedback is modeled as an implicit infra-cognitive reinforcement learning (RL) signal [1]. Such an approach is indeed supported by animal studies: for example, prefrontal cortical neurons can be controlled by rhesus monkeys through an operant conditioning paradigm [95]. The problem lies with the definition of the reward: the interpretation of the biosignal depends upon the motivational state of the subject. Furthermore, RL has two possible mechanisms [96]:

- either the subject is in a goal-directed setup and supports his learning from an internal model, in which case learning is termed as model-based RL;
- or the subject has no model of the outside events and learning arises from simple associations, termed as model-free RL.

The two issues associated with operant conditioning are therefore to determine the reward mechanisms and the type of RL.

From the perspective of the subject, neurofeedback relies on two specific biofeedback skills [97]:

- discrimination, which is the aptitude to achieve an inner perception of the biological variable,
- and self-maintenance, which is the ability to affect the biological variable and to effectively change it in the intended direction.

The acquisition of these skills could be either explicit or implicit, depending on the type of neurofeedback.

During an implicit neurofeedback procedure, learning is more likely to follow a model-free RL mechanism. The subject scans the different percepts available to him/her at a given time. Several levels of salience filters attribute weights to both external and internal percepts based on their physical, temporal, motivational, and emotional properties [98]. The resulting neural representations then go through a competitive selection process to determine which information enters working memory (WM). This filtering layer is referred to as bottom-up attention and will, for example, allow a loud, unexpected sound to enter almost anyone's WM (in addition to triggering subcortical responses).

During an explicit neurofeedback procedure, a model-based RL is triggered: the subject seeks to reach a goal (regulating the feedback signal). Top-down signals may therefore alter the bottom-up selection process by modifying the behavior of salience filters (e.g., emotional regulation) or by enhancing or inhibiting a neural representation that has already entered WM and has gained or lost salience through high-level processing (voluntary attention and percept inhibition, respectively). The subjects will then manipulate their different neural representations to determine if a correlation between the feedback and the neural representation can be established with the feedback, which is a typical set-shifting task. Set-shifting indeed refers to the ability to switch between different high-level neural representations of a percept on the basis of a feedback [99]. Sustained attention is another top-down component of attention and refers to the ability to maintain neural representations in WM over time [100], which is necessary for long-lasting neurofeedback sessions.

The interaction between these top-down and bottom-up processes lead to the dual-process theory for neurofeedback mechanisms [101] (**Figure 2**), a theory that categorizes the cognitive functions supporting neurofeedback into two main types of processing:

- more automatic and capacity-free processes
- vs. more controlled and capacity-limited processes.

These two processes lead to opposing perspectives on proper feedback designs:

- one based on bottom-up operant conditioning strategies [102];
- and another based on a top-down cognitive paradigm where higher cognitive functions percolate down from large-scale oscillations to small-scale and single-neuronal activities [77].

Recent models of explicit neurofeedback learning are based on a top-down skill learning paradigm [42]. Skill learning is a paradigm that describes the mechanisms involved in the acquisition of complex perceptual, cognitive, or motor skills. One can identify two significant properties of a motor action [103]:

- its performance, i.e., the quality of the subject's own movement (how to do the action);
- and its result, i.e., the success or failure of the action (what shall be performed).

The subject can learn about these two properties either by himself or with external help. When the subject has direct access to these two observables, it is termed "intrinsic feedback." When the information comes from an external source (for example, a sports coach or a device), it is termed "external feedback." Extrinsic feedback helps to accelerate and facilitate the learning process [104], especially when it is not redundant with internal feedback. It has informational functions and motivational properties with important influences on learning [105]. Successful feedback learning is an adaptation of internal feedback in a way that incorporates the external feedback [106]. Neurofeedback provides scaffolding for the subject, helping him/her to acquire or improve task-related discrimination and self-maintenance skills.

A possible resolution of the apparent contradiction between top-down and bottom-up models would be to postulate the existence of interactions between these two types of processing. Model-free RL and model-based RL form two cooperative systems with model-free RL driving online behavior and model-based RL working offline in the background to continuously adjust model-free RL. Once the subject becomes proficient with the task, model-free RL progressively dominates with time. As a consequence, early explicit neurofeedback learning can become implicit with time, and there is a continuum between the two learning mechanisms [1].

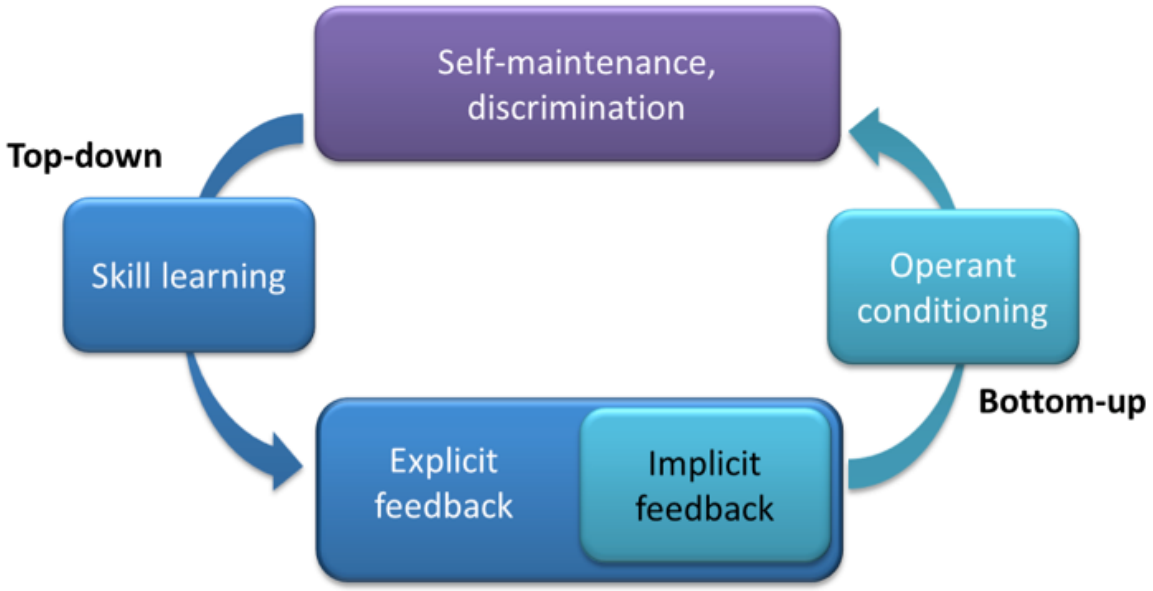


Figure 2. Dual process theory of neurofeedback. Bottom-up operant conditioning and top-down skill learning processes improve self-maintenance and discrimination skills. Implicit feedbacks interact mostly with the bottom-up system, whereas explicit feedbacks first interact with the top-down system, before becoming progressively integrated as the subject becomes independent from the feedback, which becomes then mostly a bottom-up reinforcement signal, migrating towards the operant conditioning mechanism.

Human learning and Brain Computer Interface

A brain-computer interface (BCI) can be defined as a system that translates the brain activity patterns of a user into messages or commands for an interactive application, this activity being measured and processed by the system [107]. With a BCI, the user's brain activity is usually measured via EEG and processed by the system. For instance, a BCI can enable a user to move a cursor to the left or to the right of a computer screen by imagining left or right hand movements, respectively. Because they make computer control possible without any physical activity, EEG-based BCIs have revolutionized many applications areas, notably enabling severely motor-impaired users to control assistive technologies, e.g., to control text input systems or wheelchairs, as a rehabilitation device for stroke patients, or as new gaming input device, for example [108-110].

Such BCI-based systems are used for communication and control applications in which the user voluntarily sends mental commands to the application. These types of BCIs are known either as active BCI (or explicit), when the user performs mental tasks (e.g., imagining movement), or as reactive BCI, when the users have to attend to stimuli (e.g., flickering visual images) [111, 112]. There is yet another category of BCI: passive BCI (or implicit), for which the mental state of the user is passively estimated, without any voluntary mental command from the user, to adapt the application in real-time to this mental state [111, 112].

BCIs, similarly to neurofeedback, thus rely on a closed loop that exploits brain activity in real time, specifically by acquiring EEG signals, preprocessing them (filtering), extracting relevant features describing the user's state or intent and translating them into feedback to close the loop. Although both BCIs and neurofeedback share similar technological tools, their original purposes were very different: BCIs enable users to control an external object, such as a computer or an orthosis, whereas neurofeedback enables their users to acquire control of themselves. Although some BCIs, e.g., BCIs based on mental imagery tasks, involve a learning process, and thus require the user to perform self-regulation, self-regulation is not the final objective [113]. As such, it can be said that neurofeedback is used to train users to learn how to control a BCI.

It should be noted though that the boundaries between BCI and neurofeedback remain blurry and are a subject of debate (see [114] for more detailed discussions). For instance, recently, active BCI systems that can detect imagined movements of the hands have been used to perform stroke rehabilitation by guiding users to self-regulate their brain activity in motor brain areas damaged by stroke [115], similar to neurofeedback. Passive BCIs can also be used to give feedback to a user regarding his own high-level mental states, such as mental stress or attention, to implicitly help him/her to self-regulate those states [115], again, similar to neurofeedback.

In these examples above, there are nonetheless differences between BCIs and neurofeedback. Indeed, contrary to classical neurofeedback approaches, BCIs usually heavily rely on machine learning tools to estimate some specific mental states [116]. BCIs typically use a set of example of EEG data that are recorded while the target user is in the mental state to be detected. Such data are used to calibrate a classifier to recognize this mental state using machine learning. Most neurofeedback approaches do not use a data-driven approach or machine learning to provide feedback to the user. Nevertheless, there is no fundamental constraint preventing neurofeedback from using machine learning as BCIs do, and future neurofeedback approaches could benefit from machine learning algorithms initially developed for BCI to provide more specific and robust feedback.

Overall, BCIs (both active/explicit and passive/implicit) and neurofeedback are clearly related approaches and technologies. Although they are primarily studied separately, they could both benefit from one another, notably in terms of EEG signal processing, feedback design and user training. In the future, it is not unlikely that BCI and neurofeedback share similar research paths.

Conclusion

This review highlights the growing body of evidence for use of neurofeedback in the field of psychiatry. Neurofeedback remains a very promising technique thanks to the progress of i) the techniques used (such as multivariate EEG recording for a better ROI localization, or coupled EEG-fMRI neurofeedback protocols), ii) signal processing (such as EEG-low resolution electromagnetic tomography or linear support vector machines in fMRI for phasic psychiatric disorders), and iii) understanding of the learning skills (both model-free and model-based reinforcement learning).

Thus, neurofeedback is a today's technique that is largely inspired by the original works of Durup and Fessard. However, it remains to be clarified whether the therapeutic effect of neurofeedback is clinically meaningful and how to optimally perform neurofeedback in a clinical setting. The respective place of neurofeedback techniques in the clinical armamentarium has to be defined. The field of neurofeedback involves psychiatrists, neurophysiologists and researchers in the field of brain-computer-interfaces. Future studies should determine the criteria for optimizing neurofeedback sessions. A better understanding of the learning processes underpinning neurofeedback could be a key element to develop the use of this technique in clinical practice.

REFERENCES

[1] Sherlin LH, Arns M, Lubar J, et al. Neurofeedback and basic lerning therory: implications for research and practice. Journal of Neurotherapy 2011;15:292-304.

[2] Micoulaud-Franchi JA, Quiles C, Fond G, et al. The covariation of independent and dependant variables in neurofeedback: a proposal framework to identify cognitive processes and brain activity variables. Conscious Cogn 2014;26:162-8.

[3] Durup G, Fessard A. L'électroencéphalogramme de l'homme. Observations psycho-physiologiques relatives à l'action des stimuli visuels et auditifs. L'année psychologique 1935;36:1-32.

[4] Jasper H, Shagass C. Conditionning of the occipital alpha rythm in man. Journal of Experimental Psychology 1941;28:373-88.

[5] Milstein V. Contingent Alpha Blocking: Conditioning or Sensitization? Electroencephalogr Clin Neurophysiol 1965;18:272-7.

[6] Thibault RT, Lifshitz M, Raz A. The self-regulating brain and neurofeedback: Experimental science and clinical promise. Cortex 2016;74:247-61.

[7] Micoulaud-Franchi JA, Fovet T. Neurofeedback: time needed for a promising non-pharmacological therapeutic method. Lancet Psychiatry 2016;3:e16.

[8] deCharms RC. Reading and controlling human brain activation using real-time functional magnetic resonance imaging. Trends Cogn Sci 2007;11:473-81.

[9] Micoulaud-Franchi JA, McGonigal A, Lopez R, et al. Electroencephalographic neurofeedback: Level of evidence in mental and brain disorders and suggestions for good clinical practice. Neurophysiol Clin 2015;45:423-33.

[10] Fovet T, Jardri R, Linden D. Current Issues in the Use of fMRI-Based Neurofeedback to Relieve Psychiatric Symptoms. Curr Pharm Des 2015;21:3384-94.

[11] Thibault RT, Raz A. When can neurofeedback join the clinical armamentarium? Lancet Psychiatry 2016;3:497-8.

[12] Arns M, Heinrich H, Strehl U. Evaluation of neurofeedback in ADHD: The long and winding road. Biol Psychol 2014;95:108-15.

[13] Schoenberg PL, David AS. Biofeedback for psychiatric disorders: a systematic review. Appl Psychophysiol Biofeedback 2014;39:109-35.

[14] Gevensleben H, Rothenberger A, Moll GH, et al. Neurofeedback in children with ADHD: validation and challenges. Expert Rev Neurother 2012;12:447-60.

[15] Cortese S, Ferrin M, Brandeis D, et al. Neurofeedback for Attention-Deficit/Hyperactivity Disorder: Meta-Analysis of Clinical and Neuropsychological Outcomes From Randomized Controlled Trials. J Am Acad Child Adolesc Psychiatry 2016;55:444-55.

[16] Arns M, de Ridder S, Strehl U, et al. Efficacy of neurofeedback treatment in ADHD: the effects on inattention, impulsivity and hyperactivity: a meta-analysis. Clin EEG Neurosci 2009;40:180-9.

[17] Sonuga-Barke EJ, Brandeis D, Cortese S, et al. Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. Am J Psychiatry 2013;170:275-89.

[18] Micoulaud Franchi J, Geoffroy P, Fond G, et al. EEG Neurofeedback treatments in children with ADHD: An updated meta-analysis of Randomized Controlled Trials. Front Hum Neurosc 2014;doi: 10.3389/fnhum.2014.00906.

[19] Meisel V, Servera M, Garcia-Banda G, et al. Reprint of "Neurofeedback and standard pharmacological intervention in ADHD: a randomized controlled trial with six-month follow-up". Biol Psychol 2014;95:116-25.

[20] Duric NS, Assmus J, Gundersen D, et al. Neurofeedback for the treatment of children and adolescents with ADHD: a randomized and controlled clinical trial using parental reports. BMC Psychiatry 2012;12:107.

[21] Bink M, van Nieuwenhuizen C, Popma A, et al. Behavioral effects of neurofeedback in adolescents with ADHD: a randomized controlled trial. Eur Child Adolesc Psychiatry 2015;24:1035-48.

[22] Ogrim G, Hestad KA. Effects of neurofeedback versus stimulant medication in attention-deficit/hyperactivity disorder: a randomized pilot study. J Child Adolesc Psychopharmacol 2013;23:448-57.

[23] Bink M, van Nieuwenhuizen C, Popma A, et al. Neurocognitive effects of neurofeedback in adolescents with ADHD: a randomized controlled trial. J Clin Psychiatry 2014;75:535-42.

[24] Choi SW, Chi SE, Chung SY, et al. Is alpha wave neurofeedback effective with randomized clinical trials in depression? A pilot study. Neuropsychobiology 2011;63:43-51.

[25] Scott WC, Kaiser D, Othmer S, et al. Effects of an EEG biofeedback protocol on a mixed substance abusing population. Am J Drug Alcohol Abuse 2005;31:455-69.

[26] Graap K, Freides D. Regarding the database for the Peniston Alpha-Theta EEG Biofeedback protocol. Applied Psychophysiology and Biofeedback 1998;23:265-72.

[27] Rice KM, Blanchard EB, Purcell M. Biofeedback treatments of generalized anxiety disorder: preliminary results. Biofeedback Self Regul 1993;18:93-105.

[28] Agnihotri H, Paul M, Singh Sandhu J. Biofeedback approach in the treatment of Generalized Anxiety Disorder. Iran Journal of Psychiatry 2007;2:90-5.

[29] Arns M, Strehl U. Evidence for efficacy of neurofeedback in ADHD? Am J Psychiatry 2013;170:799-800.

[30] Sonuga-Barke E, Brandeis D, Cortese S, et al. Response to Chronis-Tuscano et al. and Arns and Strehl. Am J Psychiatry 2013;170:800-2.

[31] van Dongen-Boomsma M. Dr van Dongen-Boomsma replies. J Clin Psychiatry 2014;75:779.

[32] van Dongen-Boomsma M, Vollebregt MA, Slaats-Willemse D, et al. Dr van Dongen-Boomsma and colleagues reply. J Clin Psychiatry 2014;75:290.

[33] Cannon RL, Pigott HE, Surmeli T, et al. The problem of patient heterogeneity and lack of proper training in a study of EEG neurofeedback in children. J Clin Psychiatry 2014;75:289-90.

[34] Dagenais E, Leroux-Boudreault A, El-Baalbaki G, et al. Doubting the efficacy/effectiveness of electroencephalographic neurofeedback in treating children with attention-deficit/hyperactivity disorder is as yet unjustified. J Clin Psychiatry 2014;75:778-9.

[35] Arnold LE, Lofthouse N, Hersch S, et al. EEG neurofeedback for ADHD: double-blind sham-controlled randomized pilot feasibility trial. J Atten Disord 2013;17:410-9.

[36] Deuschl G, Eisen A. Recommendations for the Practice of Clinical Neurophysiology: Guidelines of the International Federation of Clinical Physiology. Amsterdam: Elsevier; 1999.

[37] Vollebregt MA, van Dongen-Boomsma M, Slaats-Willemse D, et al. What future research should bring to help resolving the debate about the efficacy of EEG-neurofeedback in children with ADHD. Front Hum Neurosci 2014;8:321.

[38] Zuberer A, Drandeis D, Drechsler R. Are treatment effects of neurofeedback training in children with ADHD related to the successful regulation of brain activity? A review on the learning of regulation of brain activity and a contribution to the discussion on specificity. Front Hum Neurosc 2015;doi: 10.3389/fnhum.2015.00135.

[39] Gruzelier JH. EEG-neurofeedback for optimising performance. III: A review of methodological and theoretical considerations. Neurosci Biobehav Rev 2013.

[40] Mayer K, Wyckoff SN, Strehl U. One size fits all? Slow cortical potentials neurofeedback: a review. J Atten Disord 2012;17:393-409.

[41] Hammond D, Bodenhamer-Davis G, CGluck G, et al. Standards of Practice for Neurofeedback and Neurotherapy: A Position Paper of the International Society for Neurofeedback & Research. Journal of Neurotherapy 2011;15:54-64.

[42] Strehl U. What learning theories can teach us in designing neurofeedback treatments. Front Hum Neurosci 2014;8:894.

[43] Leins U, Goth G, Hinterberger T, et al. Neurofeedback for children with ADHD: a comparison of SCP and Theta/Beta protocols. Appl Psychophysiol Biofeedback 2007;32:73-88.

[44] Oken BS, Salinsky MC, Elsas SM. Vigilance, alertness, or sustained attention: physiological basis and measurement. Clin Neurophysiol 2006;117:1885-901.

[45] Hegerl U, Hensch T. The vigilance regulation model of affective disorders and ADHD. Neurosci Biobehav Rev 2012;44:45-57.

[46] Haenschel C, Baldeweg T, Croft RJ, et al. Gamma and beta frequency oscillations in response to novel auditory stimuli: A comparison of human electroencephalogram (EEG) data with in vitro models. Proc Natl Acad Sci U S A 2000;97:7645-50.

[47] Strijkstra AM, Beersma DG, Drayer B, et al. Subjective sleepiness correlates negatively with global alpha (8-12 Hz) and positively with central frontal theta (4-8 Hz) frequencies in the human resting awake electroencephalogram. Neurosci Lett 2003;340:17-20.

[48] Arns M, Conners CK, Kraemer HC. A decade of EEG Theta/Beta Ratio Research in ADHD: a meta-analysis. J Atten Disord 2013;17:374-83.

[49] Bioulac S, Micoulaud-Franchi JA, Philip P. Excessive daytime sleepiness in patients with ADHD--diagnostic and management strategies. Curr Psychiatry Rep 2015;17:608.

[50] Arns M, Feddema I, Kenemans JL. Differential effects of theta/beta and SMR neurofeedback in ADHD on sleep onset latency. Front Hum Neurosci 2014;8:1019.

[51] Arns M, Kenemans JL. Neurofeedback in ADHD and insomnia: Vigilance stabilization through sleep spindles and circadian networks. Neurosci Biobehav Rev 2013.

[52] Monastra VJ, Monastra DM, George S. The effects of stimulant therapy, EEG biofeedback, and parenting style on the primary symptoms of attention-deficit/hyperactivity disorder. Appl Psychophysiol Biofeedback 2002;27:231-49.

[53] Arns M, Drinkenburg W, Leon Kenemans J. The effects of QEEG-informed neurofeedback in ADHD: an open-label pilot study. Appl Psychophysiol Biofeedback 2012;37:171-80.

[54] Hoedlmoser K, Pecherstorfer T, Gruber G, et al. Instrumental conditioning of human sensorimotor rhythm (12-15 Hz) and its impact on sleep as well as declarative learning. Sleep 2008;31:1401-8.

[55] Sterman MB, Howe RC, Macdonald LR. Facilitation of spindle-burst sleep by conditioning of electroencephalographic activity while awake. Science 1970;167:1146-8.

[56] Cortoos A, De Valck E, Arns M, et al. An exploratory study on the effects of tele-neurofeedback and tele-biofeedback on objective and subjective sleep in patients with primary insomnia. Appl Psychophysiol Biofeedback 2010;35:125-34.

[57] Congedo M, Gouy-Pailler C, Jutten C. On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. Clin Neurophysiol 2008;119:2677-86.

[58] Philippens IH, Vanwersch RA. Neurofeedback training on sensorimotor rhythm in marmoset monkeys. Neuroreport 2010;21:328-32.

[59] Pascual-Marqui RD, Michel CM, Lehmann D. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int J Psychophysiol 1994;18:49-65.

[60] Congedo M, Lubar JF, Joffe D. Low-resolution electromagnetic tomography neurofeedback. IEEE Trans Neural Syst Rehabil Eng 2004;12:387-97.

[61] Cannon R, Congedo M, Lubar J, et al. Differentiating a network of executive attention: LORETA neurofeedback in anterior cingulate and dorsolateral prefrontal cortices. Int J Neurosci 2009;119:404-41.

[62] Cannon R, Lubar J, Congedo M, et al. The effects of neurofeedback training in the cognitive division of the anterior cingulate gyrus. Int J Neurosci 2007;117:337-57.

[63] Liechti MD, Maurizio S, Heinrich H, et al. First clinical trial of tomographic neurofeedback in attention-deficit/hyperactivity disorder: evaluation of voluntary cortical control. Clin Neurophysiol 2012;123:1989-2005.

[64] Bauer H, Pllana A. EEG-based local brain activity feedback training-tomographic neurofeedback. Front Hum Neurosci 2014;8:1005.

[65] Onton J, Delorme A, Makeig S. Frontal midline EEG dynamics during working memory. Neuroimage 2005;27:341-56.

[66] Congedo M, Sherlin L. EEG Source Analysis: Methods and Clinical Implications. In: Press A, editor. Neurofeedback and Neuromodulation Techniques and Applications. New York: Coben, R.

Evans, J.R.; 2010.

[67] Koprivova J, Congedo M, Raszka M, et al. Prediction of treatment response and the effect of independent component neurofeedback in obsessive-compulsive disorder: a randomized, sham-controlled, double-blind study. Neuropsychobiology 2013;67:210-23.

[68] White DJ, Congedo M, Ciorciari J. Source-based neurofeedback methods using EEG recordings: training altered brain activity in a functional brain source derived from blind source separation. Front Behav Neurosci 2014;8:373.

[69] Fetz EE. Operant conditioning of cortical unit activity. Science 1969;163:955-8.

[70] Fetz EE. Volitional control of neural activity: implications for brain-computer interfaces. J Physiol 2007;579:571-9.

[71] Hochberg LR, Serruya MD, Friehs GM, et al. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature 2006;442:164-71.

[72] Cerf M, Thiruvengadam N, Mormann F, et al. On-line, voluntary control of human temporal lobe neurons. Nature 2010;467:1104-8.

[73] Lachaux JP, Jerbi K, Bertrand O, et al. BrainTV: a novel approach for online mapping of human brain functions. Biol Res 2007;40:401-13.

[74] Petitmengin C, Lachaux J. Microcognitive science: bridging experiential and neuronal microdynamics. Front Hum Neurosci 2013;27:617.

[75] Bagdasaryan J, Valderrama M, Navarrete M, et al. Reconfiguration of network activity through voluntary control of intracranial oscillations. Scientific Reports 2016;in revision.

[76] Engelhard B, Ozeri N, Israel Z, et al. Inducing gamma oscillations and precise spike synchrony by operant conditioning via brain-machine interface. Neuron 2013;77:361-75.

[77] Bagdasaryan J, Le Van Quyen M. Experiencing your brain: neurofeedback as a new bridge between neuroscience and phenomenology. Front Hum Neurosci 2013;7:680.

[78] Uhlhaas PJ, Singer W. Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. Neuron 2006;52:155-68.

[79] Weiskopf N. Real-time fMRI and its application to neurofeedback. Neuroimage 2012;62:682-92.

[80] Weiskopf N. Real-time fMRI and its application to neurofeedback. Neuroimage 2011.

[81] Giacobbe P, Mayberg HS, Lozano AM. Treatment resistant depression as a failure of brain homeostatic mechanisms: implications for deep brain stimulation. Exp Neurol 2009;219:44-52.

[82] Phillips ML, Ladouceur CD, Drevets WC. A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. Mol Psychiatry 2008;13:829, 33-57.

[83] Micoulaud-Franchi JA, Fakra E, Cermolacce M, et al. [Towards a new approach of neurophysiology in clinical psychiatry: Functional magnetic resonance imaging neurofeedback applied to emotional dysfunctions]. Neurophysiol Clin 2012;42:79-94.

[84] Hamilton JP, Glover GH, Bagarinao E, et al. Effects of salience-network-node neurofeedback training on affective biases in major depressive disorder. Psychiatry Res 2016;249:91-6.

[85] Young KD, Zotev V, Phillips R, et al. Real-time FMRI neurofeedback training of amygdala activity in patients with major depressive disorder. PLoS One 2014;9:e88785.

[86] Linden DE, Habes I, Johnston SJ, et al. Real-time self-regulation of emotion networks in patients with depression. PLoS One 2012;7:e38115.

[87] Emmert K, Kopel R, Sulzer J, et al. Meta-analysis of real-time fMRI neurofeedback studies using individual participant data: How is brain regulation mediated? Neuroimage 2016;124:806-12.

[88] Marxen M, Jacob MJ, Muller DK, et al. Amygdala Regulation Following fMRI-Neurofeedback without Instructed Strategies. Front Hum Neurosci 2016;10:183.

[89] Fovet T, Orlov N, Dyck M, et al. Translating neurocognitive models of auditory-verbal hallucinations into therapy : using real-time fMRI neurofeedback to treat voices. Front Psychiatry 2016;7.

[90] Kriegeskorte N, Kreiman G. Visual Population Codes: Toward a Common Multivariate Framework for Cell Recording and Functional Imaging: MIT Press; 2012.

[91] Sitaram R, Lee S, Ruiz S, et al. Real-time support vector classification and feedback of multiple emotional brain states. Neuroimage 2010;56:753-65.

[92] LaConte SM, Peltier SJ, Hu XP. Real-time fMRI using brain-state classification. Hum Brain Mapp 2007;28:1033-44.

[93] Jardri R, Thomas P, Delmaire C, et al. The neurodynamic organization of modality-dependent hallucinations. Cereb Cortex 2013;23:1108-17.

[94] Gaume A, Vialatte A, Mora-Sánchez A, et al. A psychoengineering paradigm for the neurocognitive mechanisms of biofeedback and neurofeedback. Neurosci Biobehav Rev 2016;doi: 10.1016/j.neubiorev.2016.06.012.

[95] Schafer RJ, Moore T. Selective attention from voluntary control of neurons in prefrontal cortex. Science 2011;332:1568-71.

[96] Dayan P, Berridge KC. Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. Cogn Affect Behav Neurosci 2014;14:473-92.

[97] Epstein LH, Blanchard EB. Biofeedback, self-control, and self-management. Biofeedback Self Regul 1977;2:201-11.

[98] Menon V, Uddin LQ. Saliency, switching, attention and control: a network model of insula function. Brain Struct Funct 2010;214:655-67.

[99] Kehagia AA, Murray GK, Robbins TW. Learning and cognitive flexibility: frontostriatal function and monoaminergic modulation. Curr Opin Neurobiol 2010;20:199-204.

[100] Gazzaley A, Nobre AC. Top-down modulation: bridging selective attention and working memory. Trends Cogn Sci 2012;16:129-35.

[101] Wood G, Kober SE, Witte M, et al. On the need to better specify the concept of "control" in brain-computer-interfaces/neurofeedback research. Front Syst Neurosci 2014;8:171.

[102] Sterman MB, Egner T. Foundation and Practice of Neurofeedback for the Treatment of Epilepsy. Applied Psychophysiology and Biofeedback 2006;31:21-5.

[103] Salmoni AW, Schmidt RA, Walter CB. Knowledge of results and motor learning: a review and critical reappraisal. Psychol Bull 1984;95:355-86.

[104] Poole JL. Application of motor learning principles in occupational therapy. Am J Occup Ther 1991;45:531-7.

[105] Wulf G, Shea C, Lewthwaite R. Motor skill learning and performance: a review of influential factors. Med Educ 2010;44:75-84.

[106] Synofzik M, Thier P, Lindner A. Internalizing agency of self-action: perception of one's own hand movements depends on an adaptable prediction about the sensory action outcome. J Neurophysiol 2006;96:1592-601.

[107] Lotte F, Bougrain L, Clerc M. Electroencephalography (EEG)-based Brain-Computer Interfaces. Wiley Encyclopedia on Electrical and Electronices Engineering 2015.

[108] Birbaumer N, Cohen LG. Brain-computer interfaces: communication and restoration of movement in paralysis. J Physiol 2007;579:621-36.

[109] Daly JJ, Wolpaw JR. Brain-computer interfaces in neurological rehabilitation. Lancet Neurol 2008;7:1032-43.

[110] Wolpaw JR, Birbaumer N, McFarland DJ, et al. Brain-computer interfaces for communication and control. Clin Neurophysiol 2002;113:767-91.

[111] George L, Lécuyer A. An overview of research on "passive" brain-computer interfaces for implicit human-computer interaction. International Conference on Applied Bionics and Biomechanics - Workshop W1 "Brain-Computer Interfacing and Virtual Reality": ICABB 2010 2010.

[112] Zander TO, Kothe C. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. J Neural Eng 2011;8:025005.

[113] Lotte F, Jeunet C. Towards Improved BCI based on Human Learning Principles. 3rd International Brain-Computer Interfaces Winter Conference 2015.

[114] Perronnet L, Lécuyer A, Lotte F, et al. Brain Training with Neurofeedback. In: Clerc M, Bougrain L, Lotte F, editors. Brain-Computer Interfaces2016.

[115] Ramos-Murguialday A, Broetz D, Rea M, et al. Brain-machine interface in chronic stroke rehabilitation: a controlled study. Ann Neurol 2013;74:100-8.

[116] Lotte F. A Tutorial on EEG Signal-processing Techniques for Mental-state Recognition in Brain--Computer Interfaces. Guide to Brain-Computer Music Interfacing 2014:133-61

ARTICLE 7

Neurofeedback research: a fertile ground for psychiatry?

Auteurs par ordre alphabétique : Jean-Marie BATAIL[1], Stéphanie BIOULAC[2], Francois CABESTAING[3], Christophe DAUDET[4], Dominique DRAPIER[1], Mélanie FOUILLIEN[5], Thomas FOVET[6], Aurore HAKOUN[7], Renaud JARDRI[6], Camille JEUNET[8,9], Fabien LOTTE[10], Emmanuel MABY[5], Jérémie MATTOUT[5], Takfarinas MEDANI[7], Jean-Arthur MICOULAUD FRANCHI[2], Jelena MLADENOVIC[10], Lorraine PERRONER[9], Lea PILETTE[10], Thomas ROS[11], François VIALATTE[7], The NExT group[12]

1. Academic Psychiatry Department, Centre Hospitalier Guillaume Régnier, Rennes, France; EA 4712 Behavior and Basal Ganglia, CHU Rennes, Rennes 1 University, France.
2. Univ. Bordeaux, SANPSY, USR 3413, F-33000 Bordeaux France, CNRS, SANPSY, USR 3413, F-Bordeaux, France.
3. LAGIS Laboratory, Lille1 University, Villeneuve d'Ascq, France
4. Univ. Bordeaux, F-33000 Bordeaux France
5. Brain Dynamics and Cognition Team, Lyon Neuroscience Research Center, INSERM U1028-CNRS UMR5292-University Lyon 1, Lyon F-69000, France.
6. Univ. Lille, CNRS UMR 9193, Laboratoire de Sciences Cognitives et Sciences Affectives (SCALab-PsyCHIC), F-59000 Lille, France & CHU Lille, Pôle de Psychiatrie, Unité CURE, F-59000 Lille, France
7. Laboratoire Plasticité du Cerveau, UMR 8249, ESPCI Paris Tech, PSL Research University, 10 rue Vauquelin 75005 Paris, France.
8. Ecole Polytechnique Fédérale de Lausanne, CNBI – Campus BioTech, Geneva, Switzerland
9. Inria Rennes Bretagne-Atlantique – Campus de Beaulieu, 35000 Rennes, France
10. Inria Bordeaux Sud-Ouest, Project-Team Potioc / LaBRI / CNRS / IPB / Universite de Bordeaux – 200 avenue de la vieille tour, 33405 – Talence Cedex, France.
11. Laboratory of Behavioral Neurology and Imaging of Cognition, Department of Neuroscience, University Medical Center and Campus Biotech, Geneva, Switzerland.
12. The NExT (Neurofeedback Evaluation & Training) group of the AFPBN, http://www.afpbn.org/section/next

L'Encéphale. Soumis.

Abstract

The clinical efficacy of neurofeedback is still a matter of debate. This paper analyzes the factors that should be taken into account in a transdisciplinary approach to evaluate the use of NFB as a therapeutic tool in psychiatry. Neurofeedback is a neurocognitive therapy based on human-computer interaction that enables subjects to train voluntarily and modify functional biomarkers that are related to a defined mental disorder. We investigate three kinds of factors related to this definition of neurofeedback. The first part of the paper investigates neurophysiological factors underlying the brain mechanisms driving NFB training and learning to modify a functional biomarker voluntarily. Two kinds of neuroplasticity involved in neurofeedback are analyzed: Hebbian neuroplasticity, i.e. long-term modification of neural membrane excitability and/or synaptic potentiation, and homeostatic neuroplasticity, i.e. homeostasis attempts to stabilize network activity. The second part investigates psychophysiological factors related to the targeted biomarker. It is demonstrated that neurofeedback involves clearly defining which kind of relationship between EEG biomarkers and clinical dimensions (symptoms or cognitive processes) is to be targeted. A nomenclature of accurate EEG biomarkers is proposed in the form of a short EEG encyclopedia (EEGcopia). The third part investigates human-computer interaction factors for optimizing NFB training and learning during the closed loop interaction. A model is proposed to summarize the different features that should be controlled to optimize learning. The need for accurate and reliable metrics of training and learning in line with human-computer interaction is also emphasized, including targeted biomarkers and neuroplasticity. All these factors related to neurofeedback show that it can be considered as a fertile ground for innovative research in psychiatry.

Keywords

Neurofeedback; EEG; Neurophysiology; Psychophysiology; Brain Computer Interface; Training; Learning

Introduction

Neurofeedback (NFB) is a neurocognitive therapy based on human-computer interaction. The objective of NFB is to enable subjects to voluntarily train and modify functional biomarkers that are specific to mental disorders, in order to improve symptoms or cognitive processes. In psychiatry, a biomarker is usually a psychophysiological variable that is objectively measured and evaluated as an indicator of pathogenic processes or therapeutic responses [1]. However, most of the current NFB protocols are not based on the modulation of disorder-specific biomarkers but on the modulation of a few spontaneous *brain rhythms*, mainly defined by the frequency of their oscillation [2-4]. This strategy is prevalent since spontaneous brain rhythms demonstrate a high signal-to-noise ratio in electroencephalogram (EEG) recordings, and because they can be disrupted in some mental disorders, *e.g.* increased theta and reduced beta power in patients with Attentional Deficit and Hyperactivity Disorder (ADHD) when compared to healthy controls [5]. However, the clinical efficacy of this approach remains a controversial and delicate issue even for well-investigated applications, such as the therapeutic use of NFB in ADHD [6, 7]. Indeed, the effectiveness of neurofeedback is largely debated [8-11]. In this paper, we propose that several factors related to the concept of biomarker may be responsible for the conflicting results in the NFB literature:

(i) Limited understanding of the brain mechanisms driving NFB learning to modify a functional biomarker voluntarily, *i.e. neurophysiological factors* [11],

(ii) The inconsistent relationship between EEG biomarkers and clinical dimensions (symptoms or cognitive processes), potentially due to the symptom-based classification of psychiatric disorders and the heterogeneity of diagnostic categories, *i.e. psychophysiological factors* [12]

(iii) Superficial knowledge of how best to measure and optimize NFB learning during the closed loop interaction, *i.e. human-computer interaction factors* [13].

This paper investigates these factors (*neurophysiological*, *psychophysiological* and *human-computer interaction*) in a critical review of the existing literature. The objective is to integrate these interdependent issues into a general NFB framework in order to demonstrate that NFB can be considered as fertile scientific ground for psychiatry and to provide a roadmap for future research in this field.

Neurofeedback and its neurophysiological foundations

From electroencephalographic oscillations to neurofeedback

The EEG may be recorded via non-invasive electrodes placed on the scalp as a result of intracranial fluctuations of electromagnetic field potentials, which are generated by ionic exchanges at cell membranes and synapses during neuronal activity. When neuronal activities occur in a circumscribed region and become temporally synchronized, their local field potentials (LFPs) are then spatially summated, giving rise to large fluctuations of the EEG signal [14]. Hence, changes in EEG oscillation amplitude essentially reflect the degree of synchronization of intracortical neuronal populations. Synchronization is influenced by both the intrinsic excitability of the neuronal population and the synaptic input it receives from other regions. Hence, intra- and inter- electrode EEG measures of amplitude and coherence indicate neuronal excitability within and functional segregation/integration between cortical

regions, respectively [15]. Moreover, this dynamic activity can occur simultaneously on different timescales (i.e. frequencies): infraslow (<1 Hz), delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), sigma (12–15 Hz), beta (15–30 Hz), and gamma (>30 Hz). Studies involving patients with mental disorders have reported significant deviations in a host of task-related and resting-state EEG parameters (e.g. amplitude, coherence) compared to healthy controls [16].

Thus, NFB has been developed in these patients mostly to correct notable deviations of cortical oscillations by training subjects to modify their EEG activities. In this perspective, the impact of NFB is thought to be based on the training and subsequent normalization of specific "targeted" neurophysiological signatures to reduce the clinical symptoms related to a given disorder. It has been also postulated that, to achieve therapeutic efficacy with NFB, it is important to demonstrate significant *online* self-regulation of the trained parameter(s) (i.e. during NFB). After which, long-term *offline* changes might be induced through mechanisms of neuroplasticity (i.e. of functionally persistent brain reorganization after termination of NFB training) [17]. Thus, in the simplest scenario, the incremental process of NFB "learning" can be seen as the direct sum of two principal factors: 1) the online component, i.e. the within-session change of the trained signal relative to its resting-state baseline, also called "performance" in the field of Brain Computer Interface - BCI, and 2) the offline component, i.e. the absolute change of the between-session resting-state baseline, which may be related to "skills acquisition". Surprisingly however, there is a scarcity of BCI/NFB studies that examine these online and offline criteria in combination. Moreover, a better definition of online/offline metrics would enable a more rigorous assessment of NFB protocols and BCI training [18] together with their impact on brain plasticity [19] (see last section on *human-computer interaction factors*). This first section focuses on the basis of neuroplasticity during NFB.

From electroencephalographic oscillations to neuroplasticity

The dynamic modulation of EEG oscillations using NFB may induce different types of neuroplasticity [19]. Neuroplasticity in general may be defined as a durable (i.e. long-term) change in neural function outlasting the training period itself, underpinned by long-term modification of neural membrane excitability and/or synaptic potentiation. In practice, one may expect long-term plasticity to manifest itself during resting-state EEG recording(s) *outside of* training sessions (i.e. offline), and/or as progressive changes *during* repeated training sessions (i.e. online). Based on the neuroscience literature, there are two main forms of neuroplasticity: the Hebbian type and the homeostatic type.

The underlying mechanism of Hebbian plasticity is *correlation-based*. Hence, NFB-induced Hebbian plasticity may be expected to produce functional changes that occur *in the same direction* as that dictated by the NFB protocol (e.g. long-term alpha increase following alpha-upregulation NFB) [20]. On the other hand, since homeostasis attempts to stabilize network activity within a bounded range, homeostatic plasticity is not correlation-based and may be expected to produce changes in the opposite direction of NFB training (e.g. long-term alpha increase following alpha-downregulation NFB) [21]. Generally, synaptic potentiation brain oscillations are closely linked, given that changes in neuronal coupling directly affect levels of neuronal synchronization, and vice-versa.

Hebbian plasticity and neurofeedback

Historically speaking, pioneering experiments in the 1960s that demonstrated self-regulation of the EEG [22] were followed by reports that NFB training of spindle oscillations during wakefulness may result in their stronger expression during sleep [23]. Recent studies provide convincing data that NFB can be used to induce plastic *increases* of *theta*, *alpha*, *beta*, and *gamma* rhythms, as well as their corresponding *decreases* [17]. However, the exact neurophysiological mechanism(s) behind the long-term conditioning of brain rhythms remain unclear.

Given common observations that plasticity manifests in the same direction/frequency targeted by the NFB protocol, Ros and colleagues proposed a mechanism based on associative (i.e. Hebbian) plasticity and encapsulated by the phrase [19]: “synapses that fire together wire together, and synapses that fire apart wire apart”. This type of correlation-based plasticity occurs when connectivity is reinforced by temporally-coincident neuronal activation. As explained in the section above, EEG oscillatory amplitude positively covaries with the degree of synchronized neurons/synapses. **Figure 1.** Hence, during amplified oscillations, synchronized neural populations involved in generating this oscillatory pattern would, after some time, strengthen the connections between themselves, and further facilitate the oscillation to emerge in the future. Conversely, maintaining a cortical region in a low-amplitude (“desynchronized”) state would reduce synaptic correlations and weaken the connections that give rise to synchronization. Encouragingly, recent experimental work provides support for this mechanism outside of NFB, reporting up-regulation and down-regulation of cortical oscillations using synchronizing and desynchronizing patterns of stimulation, respectively [17, 19].

Homeostatic plasticity and neurofeedback

Animal research has consistently revealed the presence of an additional form of plasticity referred to as ‘homeostatic’ plasticity, which actively counteracts the Hebbian type so as to prevent its unlimited expression [19]. Otherwise, unchecked Hebbian plasticity would inevitably lead to pathologically high or low neural connectivity, firing or synchronization. Hence, from the point of view of NFB, one would anticipate homeostatic forms of plasticity to produce changes opposite to the direction of training. Early observations within this context were made by Kluetsch and colleagues [21], who reported that following down-training of alpha rhythm, patients with Post Traumatic Stress Disorder (PTSD) displayed a paradoxical increase in alpha rhythm above and beyond its resting-state value. Since PTSD patients are found to exhibit significantly low alpha amplitude at baseline relative to healthy subjects, a recent framework proposed that this might reflect homeostatic regulation of the excitation/ inhibition balance [19, 24].

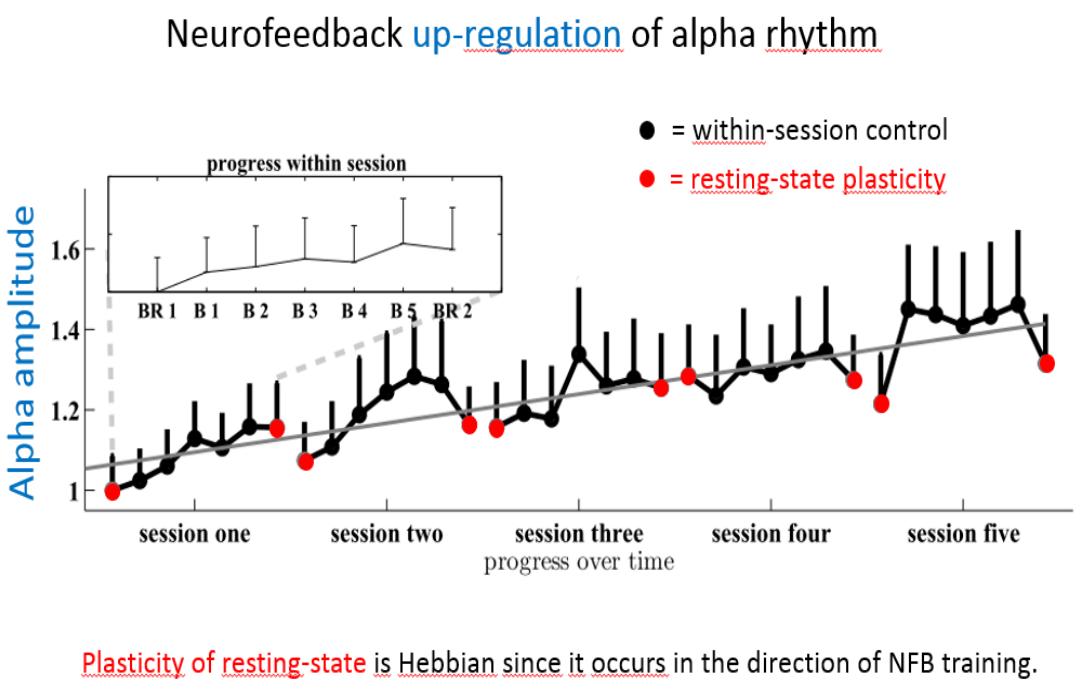


Figure 1. An example of Hebbian-type neuroplasticity mechanism subsuming neurofeedback training with experimental data on alpha rhythm up-regulation (adapted to experimental data from [20]).

Towards new neurophysiological measures of neuroplastic effects of neurofeedback

In addition to EEG-based measures, the neuroplastic effects of NFB have started to be explored using several other techniques, including transcranial magnetic stimulation (TMS), functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI). For example, a single 30-minute session of NFB alpha downregulation has been found to enhance cortical excitability, as measured by a plastic (>20 minute) increase in TMS-induced motor evoked potentials after training [25]. Of note is also the observation of reduced intracortical inhibition, in view of its established association as a cortical state that facilitates plasticity and learning [26]. Elsewhere, fMRI has shown that NFB may induce plastic changes in cortical hubs responsible for cognitive control such as the dorsal anterior cingulate [27], which was associated with improvements in symptoms of ADHD [28] or on-task mind wandering [29]. Lastly, data from a DTI study make an encouraging case for NFB affecting white matter and grey matter [30].

In closing, this first section has focused on the neurophysiological foundations of NFB, which enable it to be used as a unique therapeutic tool for targeting specific neural activities and inducing neuroplasticity. However, beyond basic up- or down- regulation of brain rhythms, the central challenge of NFB is to target clinically relevant biomarkers that are consistent with the psychophysiological foundations of mental disorders. The following section focuses on this challenge.

Neurofeedback and its psychophysiological foundations

Dimensional approach for neurofeedback in psychiatry

Because the psychiatric nosology has weak biological grounds, on the one hand, and because the link between biomarkers (electrophysiologic biomarkers in particular with EEG or metabolic biomarkers with functional neuroimagery) and cognitive processes remain mostly unraveled, on the other hand, it is impossible to confirm the functional specificity of current NFB biomarkers. **Figure 2.** In fact, contemporary psychiatry is undergoing a taxonomic crisis that is characterized by the poor diagnostic power of current nosology [31]. Interestingly, in 2010, the National Institute of Mental Health (NIMH) proposed a dimensional approach to circumvent this issue. For Insel et al., the current symptom-based classification probably does not reflect the pathophysiological mechanisms that underlie mental disorders [32]. The aim of the Research Domain Criteria (RDoC) project is to conceptualize mental illnesses as brain disorders with pathophysiological features represented by a reliable and validated continuum from the clinical to the genetic, all defined by tools from neuroscience [32].[1] Such an approach could be very useful in the field of NFB research applied to mental disorders. By targeting specific biomarkers related to well identified symptoms or cognitive processes, the psychophysiological rationale underlying NFB therapy should be stronger and its efficacy probably greater. Importantly, although the quality of EEG recordings and the design parameters of NFB protocols (e.g. the number of sessions per week) are essential variables to be optimized to foster training, their optimization will never overcome the putative deleterious effects of our current lack of precise knowledge about the underlying brain/mental processes. We advocate here that acknowledging this fundamental limitation is a useful starting point to guide the research and development of future NFB therapies. Furthermore, this limitation holds whatever the functional modality used to record brain activity (electrophysiology, fMRI, fNIRS, etc.).

As the first step to overcome this limitation, we consider it essential to inventory and refine the existing list of EEG biomarkers and associated cognitive functions. In the following section, we propose an “EEGcopia” to illustrate the need to rely on EEG biomarkers that are strongly linked to symptoms or cognitive processes. We discuss this concept of EEGcopia below and provide a preliminary list that highlights the need to link psychiatric nosology and putative biomarkers with clinical dimensions such as executive function, emotion regulation and reward processing (see **Supplementary material**). The opportunity to construct new therapeutic hypotheses based on other EEG and putatively more specific biomarkers than those used so far in NFB is illustrated in two concrete and very topical fields of NFB/psychiatric research: depression and ADHD.

[1] https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2011/treatment-development-the-past-50-years.shtml

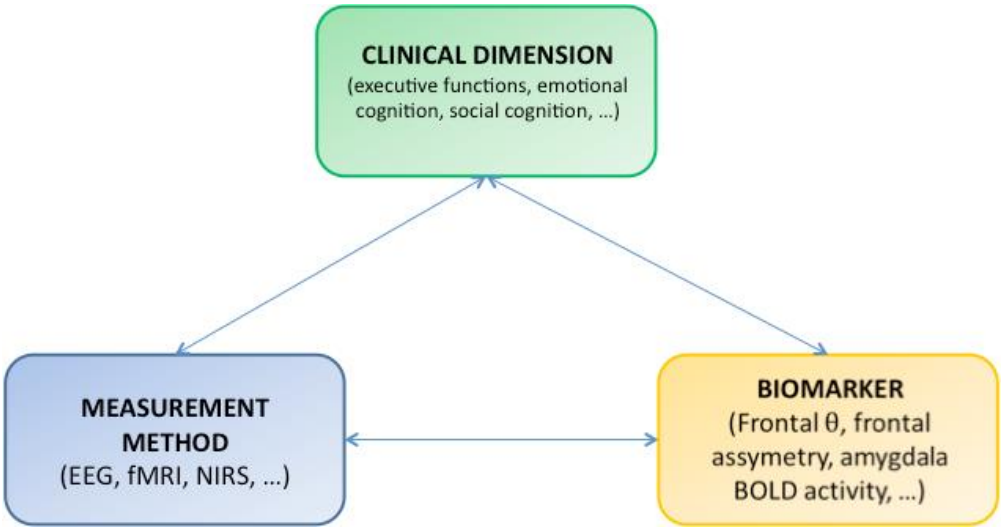


Figure 2. Psychophysiological issues related to neurofeedback in psychiatry

A proposed EEGcopia for neurofeedback in psychiatry

Most NFB investigations to date have focused on a limited set of EEG frequency ranges (the two most famous being the $\frac{\theta}{/\beta}$ ratio and the Sensory Motor Rhythm - SMR). However, there are several other known correlates of cognitive functions that could be used as potential target biomarkers. Indeed, one can extract numerous biomarkers from EEG signals such as discrete EEG events like event-related potentials (ERP), measures of complexity, or local and long distance neural synchrony, which could have potential NFB applications. The use of these EEG biomarkers for NFB has received little attention until now. We introduce here a brief nomenclature of cognitive functions (see **Supplementary material**), together with their known EEG biomarkers. Dimensional EEG biomarkers of cognitive functions with known neural correlates of sensory processing, executive functions, emotional cognition, memory, embodied cognition and social cognition are presented. This short EEG encyclopedia (EEGcopia) reflects the main theories linking EEG and cognitive dimensions in neurophysiology. A more complete and exhaustive EEGcopia would be of great help to the NFB community.

Among the different biomarkers listed, Event Related Potentials (ERP) for NFB open up new avenues for application. The numerous publications on BCI based on the real-time detection of P300 demonstrate the feasibility of this approach [33]. Recent studies have generalized these results to other ERP components, such as error negativity (ERN) [34] and auditory mismatch negativity (MMN) [35]. However, each ERP has its specific properties, such as differences in refractoriness [36], which may limit their detection rate for real-time applications and make them usable only for discrete delayed feedback. Another promising candidate is the use of classification algorithms targeting specific dimensions. For instance, arousal detection using the VIGALL algorithm [37] was recently used to investigate brain mechanisms, and it can also be used to design efficient NFB strategies [38].

Linking brain / mental processes and psychiatric disorders

The emblematic research field of depression

Which innovative biomarker could be relevant to treat depression? EEG biomarkers of cognitive functions could be affected during a mood depressive episode. In line with this approach, a cognitive description of depression as an applied reflection on the choice of the most relevant target for NFB could be proposed.

Recently, Rayner et al. published a comprehensive review of cognition-brain related networks of depression. Two main networks are involved: autobiographical memory (AMN) and cognitive control networks (CCN) [39]. The former is involved in self-referential cognitive processing and the latter in the ability to perform goal-directed tasks. The authors postulated that AMN is hyperactivated (self-referential cogitation and congruent emotional processing) over the CCN, which is deactivated during a mood depressive episode. This model highlights the central role of cognition (and its neural substrates) in depression [39]. This cognitive dimension could be a promising therapeutic target for NFB instead of more conventional therapeutics. However, the best psychophysiological signal related to this cognitive dimension remains to be determined.

Most of the literature on EEG-NFB has focused on alpha asymmetry but with controversial results concerning its efficacy. In fact, EEG-NFB protocols on depression enhance cognitive functioning [40] but have failed to have any effect on emotional and mood features (for review; see Arns et al., 2017 [3]). Based on the cognitive dimension of depression [39], it can be hypothesized that the ultimate NFB should disengage the emotional cognitive processes of AMN, strengthen cognitive processes oriented to external stimuli (CCN), and strengthen working memory. Therefore, NFB targeting both AMN and CCN should fit this issue well. Some recent work on NFB has proposed to combine EEG and fMRI in order to provide a more specific self-regulation of these targets [41, 42]. These studies highlight the fact that bimodal/simultaneous NFB could be more specific and more engaging than EEG-NFB alone. Zotev et al. have demonstrated its feasibility and potential in depression [43, 44]. This perspective seems to be of great interest for targeting complex psychophysiological processes involved in mental disorders such as depression.

The emblematic research field of ADHD and P300-based training

Which innovative biomarker could be relevant to treat ADHD? The P300 is a large positive complex that reaches its peak at approximately 300 milliseconds after stimulus onset and is composed of two subcomponents, a frontal P3a reflecting attentional capture by some external stimulation, followed by a parietal P3b elicited by the voluntary orientation of attention [45]. The amplitude of the P300 grows with the amount of attentional resources engaged in processing the external event [46]. Although this biomarker has never been used for NFB, it is very much used online for controlling BCI applications such as the P300 speller [47]. With this interface, items are selected on screen based on the orientation of spatial attention. Interestingly, the same principle can be used in engaging EEG-controlled video games [48]. Such games offer a motivating training environment, may include strategic components (e.g. “Connect Four”) and rely on clear instructions about the requested mental effort to be produced in

order to control the game and possibly win (e.g. focus spatial attention and avoid being distracted). Interestingly, the P300 is known to be altered in children with ADHD [49]. It is also a marker of treatment efficacy as P300 amplitude has been shown to return to normal levels in patients who respond positively to methylphenidate [50]. This has led to an ongoing clinical trial to evaluate the usefulness of P300-based training in children with ADHD [51]. If successful, this trial will support the extension of this kind of training to other pathological states associated with impairment in selective attention.

This second section has focused on the psychophysiological foundations of NFB applied to mental disorder and has demonstrated how it should be related to a better definition of biomarker in order to target neural activities specific to symptoms or cognitive processes. However, even if the chosen biomarker is strictly related to symptoms or cognitive processes, it should also be verified that it is effectively modified during the NFB sessions. Surprisingly, this domain on which the following section focuses remains a major challenge for NFB, and the field of BCI is of great interest to enhance knowledge on optimized training and learning for NFB in psychiatry [11].

Neurofeedback and its human-computer interaction foundations

A human computer interaction model for neurofeedback

To globally improve NFB efficacy in patients, it is necessary to understand and then reduce its variability. To this end, Sitaram et al. (2016) and Gaume et al. (2016) have reviewed the neurophysiological [17] and neuropsychological [12] mechanisms underlying NFB training procedures. In addition, Enriquez-Geppert et al. (2017) have proposed a tutorial explaining how to design rigorous NFB training protocols [52]. While Sitaram et al. (2016) and Gaume et al. (2016) adopted a standpoint purely centered on "human learning" (*i.e.* centered on the psychological and neurophysiological mechanisms that enable patients to learn how to self-regulate specific neural substrates), Enriquez-Geppert et al. (2017) focused on "machine learning" (*i.e.* centered on the technological factors, especially signal processing and machine learning, potentially impacting performance). These papers offer insightful elements to understand and reduce the variability of clinical NFB efficacy. Nonetheless, as illustrated in **Figure 3**, these uni-centered approaches are not sufficient to reach a deep understanding of the NFB training process. "A human-computer interaction/human-factor standpoint", like the one proposed by Alkoby et al. (2017) [53] and Jeunet et al. (2017) [54], is also needed to understand how, depending on their profile (i.e., psychological, cognitive and neurophysiological states and traits), patients interact with the training protocol and what the consequences of this interaction on learning and on clinical efficacy are. In fact, we have proposed a model combining factors that influence learning in Brain Computer Interface (BCI) and NFB (NF) [54]. The model is based substantially on the BCI literature and more specifically on Mental-Imagery-based BCIs (MI-BCIs) [13, 55]. MI-BCIs are neurotechnologies that enable a user to control an application through the completion of mental-imagery tasks such as imagining movements, *i.e*., motor-imagery, that are associated with a specific modulation of the user's brain activity. Therefore, as is the case in NFB applications, MI-BCI users have to learn to modulate a target neurophysiological substrate. Consequently, the literature on BCI is of interest to better understand the factors influencing learning in NFB.

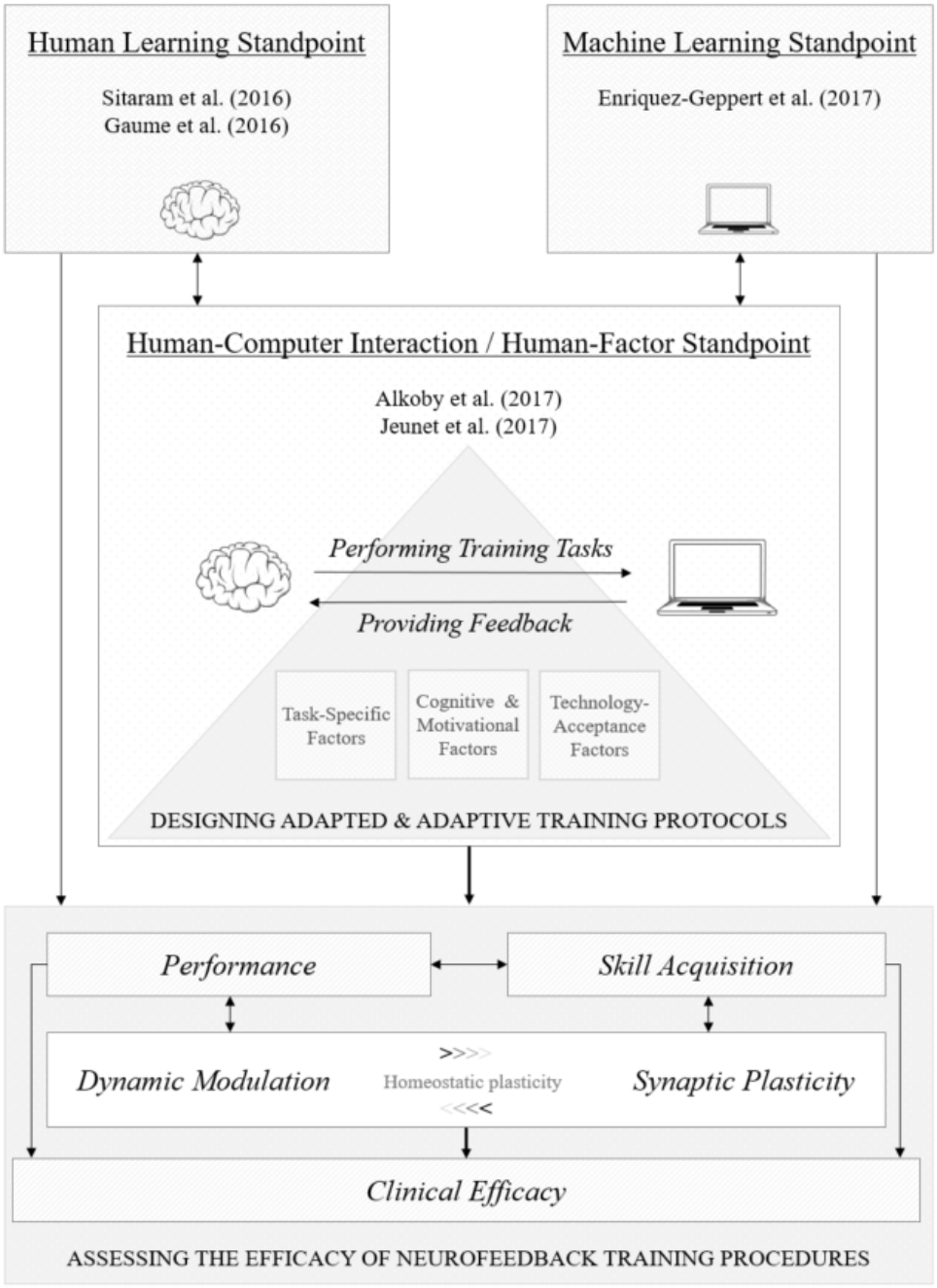


Figure 3. Schematic representation of proposed approach. While some studies contribute to improving the efficacy of neurofeedback procedures by adopting either purely "human-learning" or "machine learning" standpoints, we posit that a "human-computer interaction / human-factor" approach would enable deeper understanding of the processes subsuming neurofeedback-related performance and skill acquisition, and thus improve its clinical efficacy. This would provide insights into how users' traits and states impact the efficacy of neurofeedback, notably through three types of factors, and allow training tasks and feedback to be adapted in order to better grasp the interaction and improve the efficacy of neurofeedback. For an extensive description of the factors involved in the model, see [54, 55]. Moreover, we believe that neuroplasticity indicators are important intermediate variables to be considered between NFB training/learning and clinical efficacy. We distinguish two kinds of neuroplasticity indicators: dynamic modulation indicators and synaptic plasticity (also called Hebbian plasticity) indicators. For an extensive discussion on neuroplasticity and neurofeedback, see [17, 19].

The model in **Figure 3** includes three categories of factors: task-specific, cognitive/motivational and technology-acceptance related factors. As this model focuses on MI-BCIs, the task-specific factors refer to spatial abilities, *i.e*., the ability to produce, transform and manipulate mental images. It is likely that in other kinds of BCI or NFB paradigms, different task-specific factors related to the targeted neurophysiological will have to be identified. The other two families of factors are more generic and do not depend upon the BCI/NFB paradigm used. They include, on the one hand, factors related to cognitive and motivational traits and states, and on the other hand, factors related to patients' acceptance of the technology, *i.e*., the way they perceive the technology and consequently the way they will interact with it, *e.g*., to what extent they feel in control as well as their anxiety or confidence. The model suggests that the learning process during BCI or NFB training procedures is influenced by patients' traits and states, which in turn are modulated by the perception of the technology. By considering these factors, one could design training protocols and feedback adapted to the profile of each patient and adaptive to the evolution of their states and skills as they evolve during the course of BCI or NF. Both the training tasks and the feedback can be adapted (*i.e*., specific to the patient's profile - traits and states - estimated at the beginning of training) and adaptive (*i.e*., modified dynamically during training to fit the evolving state of the patient) in order to optimize the learning process. The first subsection is dedicated to a review of the literature on how to design efficient adapted and adaptive training tasks and feedback. Then, to evaluate the efficacy of NFB training procedures, relevant metrics of performance, skills acquisition and clinical efficacy are needed. However, to date such relevant metrics have received little attention. Thus, the second subsection describes some metrics dedicated to assessing users' performance and skills and then discusses the relationship between these metrics and the clinical efficacy of NFB procedures.

BCI principles to adapt training tasks and feedback in neurofeedback

Based on an analysis of the literature, the following paragraphs present insights on how a training protocol may be adapted. The protocol comprises two main parts: training tasks and feedback. Indeed, during BCI/NFB training, the patient performs different training tasks according to the instructions provided by the system or experimenter, so as to self-regulate their EEG. They are then provided with feedback from the machine to inform them about the quality of their EEG self-regulation (see **Figure 3**). Thus, training tasks are neurocognitive exercises that the patient will perform, such as trying different mental strategies or trying to self-regulate the targeted EEG feature with various levels of difficulty, *e.g.*, thresholds to reach. The feedback is the information provided by the machine to represent real-time variations in the EEG feature and/or to guide the patient in the training task, *e.g*., towards a modification of their strategy. For instance, feedback can be a visual gauge or an audio sound of which the size or amplitude varies according to the EEG feature value. The following sections first present various training tasks that have been explored for BCI training, and then present different types of feedback that have been used for the same purpose. They also describe which of these tasks and feedback types are adapted and adaptive according to the users' traits and states, or how they could be made so.

Towards adapted and adaptive BCI/Neurofeedback training tasks

This subsection analyzes a training task that can be adapted and adaptive in order to optimize the learning process. The type of the task and its difficulty can be adapted [56]. The type of the task comprises the psychophysiological parameter that the user is asked to modulate. This modulation can be used to control various applications. For instance, with motor imagery, the different exercise types would be the possible mental commands; *e.g*., motor imagery of hands, feet or tongue. The instructions serve to guide the user in knowing which exact mental command he is supposed to perform in real time (trial-by-trial). The type of the task can be adapted or adaptive. So far in the literature, adapted types do not seem to have been explored. However, adaptive BCI/NFB task types have been explored. For instance, the machine could automatically identify which psychophysiological parameter works best for the users to assist them to more easily manipulate the system. For instance, machine learning (Bandit algorithm) has been used to select the MI task type within runs (among hands, feet and tongue) in order to identify as quickly as possible for which one the user has the best performance [57]. The same could apply for NFB tasks, where the user is asked to regulate different EEG patterns from the initial ones if he is unable to regulate or produce them.

The difficulty of the task may be defined by the amount of mental resources that the patient needs to engage in it in order to complete it successfully. This is related to the skills of the user at EEG self-regulation. Ideally, to ensure efficient learning, the task difficulty should match the user's skills in order to be neither too easy - which would be boring - nor too difficult - which would be frustrating. The difficulty of the task can be adapted or adaptive, *i.e*., increased or decreased according to the user's profile and the speed at which he acquires skills. Traditionally, adapted and adaptive task difficulty has been set by using a threshold initially adapted to the user's physiology and regularly updated between sessions. It has not yet been adapted to the user's cognitive profile, which thus remains to be explored. Additionally, recent research is now exploring other ways to dynamically adapt the difficulty instead of changing the threshold between sessions. For instance, in McFarland et al. (2010) motor-imagery task difficulty was increased from 1D, then to 2D, and finally to 3D cursor control within sessions [58]. Another way to increase user performance and motivation is to adapt the perceived task difficulty by providing a feedback which does not comply with the real performance of the user but is positively biased or is adaptively biased [59]. Finally, the difficulty in an experimental context can differ from an ecological one, so virtual reality coupled with NF/BCI could be useful to train the subject in a more realistic environment [60]. Indeed, in these types of protocol, the level of the environmental distractors and therefore difficulty can be controlled, *e.g*., by increasing the speed of instructions or adding distracting, real-life, environmental noise.

Adaptive difficulty can be further explored by educational theories. Indeed, instructional design theories and flow theory show that to promote progress and intrinsic motivation, a task should be engaging, often ludic and adapted to the user's skills [59, 61]. This suggests that NFB training tasks could also follow educational theories to foster learning and intrinsic motivation. Moreover, the cognitive strategy of the user, which refers to the way the user tries to modulate the psychophysiological parameter used in the exercise, could be influenced by the instructions as well as by various feedback.

Towards adapted and adaptive feedback for BCI/Neurofeedback

This subsection analyzed the feedback that can be adapted and adaptive in order to optimize the learning process. Feedback is an indication provided to users that allows them to learn to modulate their brain activity. However, providing feedback that is appropriate and informative is a great challenge [61]. A substantial number of studies on BCI have focused on feedback modality, content and social features. Concerning the feedback modality, the effects of adapted and adaptive classic visual feedback, auditory feedback, tactile feedback or even multiple sensory modalities feedback have been studied. Such feedback can improve control display mapping to further enhance the sense of agency which influences the technology acceptance factor presented in **Figure 3**. Adapting the modality of the feedback also makes it possible to take general cognitive principles into account, e.g. the presentation of information on different modalities enables a faster response, related to the “redundant signal effect”, but it also makes it possible to adapt to the sensorial impairments of patients [62]. Moreover, virtual reality can be used to improve training by providing motivating and immersive feedback [60].

Concerning the content of feedback, some task-specific elements have been studied. For example, a key element for controlling BCI is for users to understand how their brain activity is modified when performing a task. Such representation of their brain activity can be provided by new visualization tools, e.g., TEEGI [63]. These can show users an engaging visualization of their own brain activity in real time to help them to understand which EEG patterns should be produced.

Lastly, concerning social features, some original studies have provided adapted and adaptive emotional support as well as a social presence to compensate for the lack of interaction during BCI/NFB sessions by using a learning companion, see **Figure 4** [64]. Each of the companion interventions was composed of an animation of its face and a spoken sentence. The feedback provided took the performances and progression of the user into account. It focused on the subject’s effort and strategy and on reinforcing good performances and progress. Results showed a beneficial impact on the user’s experience and might also indicate a differential effect on users that is yet to be verified. These results are encouraging and require further investigation.

A key objective for future research should be to focus on making feedback more informative by better understanding learning processes and improving measures of performances of BCI. Moreover, a challenge arises from enriching the feedback without overloading users with more information than they can process given their capacities. Assessing cognitive abilities such as attention and providing related adaptive feedback would provide interesting insights into this issue. Overall, BCI/NFB would benefit from studies combining several of these factors and assessing the interactions between them. The goal is to provide feedback that is both adapted and adaptive to training tasks, users’ profiles, and their social and physical environment, a criterion often forgotten but which should be given more consideration by doing more ecological experiments, *e.g*. by using virtual reality.

Redefining the assessment of BCI/Neurofeedback training efficacy

The assessment of NFB training efficacy is essential to better understand the clinical efficacy of such therapeutics. Indeed, most studies that investigated the clinical efficacy of NFB did not evaluate or even report the efficacy of training [65]. Thus, it cannot be concluded whether patients gained control over their brain activity during the NFB training procedure or not. However, as learning is the most immediate result of NFB training according to the principle of NFB, it seems essential to measure the learning that takes place across sessions. As Rémond & Rémond stressed: "*Doubting the effectiveness of a biofeedback treatment on a physiological variable when this treatment is carried out without previously testing the modification of this variable, is the equivalent of doubting the effectiveness of a drug to cure a disease when the drug has not been absorbed by the patient*" [66]. Thus, the following subsections first present how to assess NFB and BCI user learning by distinguishing: (i) how well users can self-regulate their EEG activity at a given time, which represents their current "performance", and (ii) how well they acquire new skills across sessions to improve this EEG self-regulation, which represents their EEG self-regulation "skill". The following subsection describes the issues involved in redefining such metrics in order to both (i) improve the design of adapted and adaptive training tasks and feedback in NF, and (ii) better link such metrics to neurophysiological and neuroplasticity indicators.

Figure 4. Brain Computer Interface training during which PEANUT (on the left) provides user with social presence and emotional support adapted to his performance and progression [64].

Towards new performance and skill metrics in BCI/Neurofeedback

Performance is typically assessed by using success rates as metrics, *i.e*., how often a) users' NFB features successfully crossed the threshold, or b) users' mental tasks are successfully recognized by the BCI. In both cases, a threshold is used: generally, a univariate one for classical NFB analysis in mental disorders (*i.e*., a single value to be crossed by the unidimensional feature value) [3], usually defined manually, or a multivariate one for BCI, the EEG classifier typically used being a multidimensional threshold on all the features used by the BCI to recognize each mental task. While success rates are typically used in NFB/BCI, it can be argued that they are a poor performance metric of user learning. Indeed, success rates are discrete and depend on the data used to determine the threshold/classifier, whereas users' skills at EEG self-regulation are continuous and threshold/classifier-independent. This means that an improvement in EEG self-regulation might not translate into an improvement in success rates, e.g. if the threshold is too high. This also means that if the threshold or classifier is calibrated on data of poor quality, this will result in poor feedback and in a poor measure of performance based on them. To date, only a few studies have evaluated the relevance of performance metrics in BCI/NFB during a session. Recently, new metrics were proposed to study BCI user training that provide a continuous and threshold-free measure of how stable and distinct EEG patterns for each mental task are [18]. Comparisons showed that such metrics could reveal fast learning of EEG self-regulation in several BCI subjects whereas success rates sometimes did not. NFB success rates very likely have the same limitation and should thus be reconsidered when assessing NFB interventions. In any case, research into more specific and learning-related metrics of performance is needed.

Skill metrics are computed to quantify learning across sessions. They are typically based on relevant performance metrics estimated on each session/run. They estimate whether these performance metrics increase over time and sessions, which would indicate learning. An example of such a metric could be the difference between performances obtained during the previous sessions and those obtained during the first ones, or the slope of the regression line passing by the performances across sessions (the steeper the regression line, assuming increasing performances, the faster the learning). Nonetheless, so far there is no gold standard in skill metrics and the ones currently used suffer from several limitations. For instance, the metrics mentioned above are very sensitive to outliers, and a single failed session (*e.g.*, due to a failing sensor or a tired patient) or an overly good one (due, *e.g.*, to chance) may lead to an inadequate corresponding skill metric. Skill metrics also depend typically on the threshold used in the performance metrics. If the threshold changes across sessions, which is typically the case in NF as in BCI if the classifiers are adaptive or recalibrated regularly, then performances are not comparable between sessions and the resulting skill metric may be meaningless. Finally, performance metrics also depend on rest/baseline EEG, such baseline values typically changing at each session. As such, the performance metrics used to compute skill metrics may not be comparable with each other. Overall, there is thus a need for new relevant skill metrics that are stable, meaningful and robust to outliers, as well as for investigation into their impact on clinical efficacy.

Towards optimizing clinical efficacy based on new metrics and neuroplastic approaches

We need to improve our knowledge about the relevant performance and skill metrics in order to optimize the clinical efficacy of NFB. Indeed, such metrics are essential for designing adapted and adaptive training tasks and feedback in NFB. At present, the task and the feedback are adapted by NFB practitioners before and during the training procedure. An important step for NFB practitioners is determining a threshold and the kind of feedback [67-69]. Adjusting a threshold and a given occupation time determines the number of positive reinforcements. Traditionally, the threshold may be set automatically or manually. When the threshold is determined automatically, it is continuously updated in order to provide patients with a positive reinforcement for a given percentage of occupation time below or above the threshold. The threshold is continuously estimated according to the signal recorded just before. However, the limitation is that the patient is rewarded only for changing his/her brain signal based on the previous averaged time period and not from the starting point, which drastically reduces the chance of learning across NFB sessions [67]. When the threshold is set manually by the professional, it is based on a baseline recorded before the NFB session. If the number of positive reinforcements is too high or too low during the session, the threshold can be adjusted [67]. However, there is a risk of inconsistency between different NFB practitioners, as each one will adapt the task according to their own clinical experience. Moreover, different practitioners will typically take the profile of each patient into account (*i.e.*, psychological, cognitive and neurophysiological states and traits) subjectively according to their global feeling and not according to evidence and objective features. Moreover, the clinician may not be able to evaluate a state or a trait evolution that would be crucial to adapt the training task. Strehl (2014) stressed that "*the therapist will need to know the laws of learning as well as how to apply NFB training in order to be a competent partner*". However, the limitation of this standpoint is that these skills currently rely on clinical experience [70] rather than on scientific knowledge related to NFB learning processes [65, 67, 69]. Thus, the remaining challenge for assessing the efficacy of NFB therapies is to develop rigorous standards that ensure the consistency (*a.k.a*., fidelity - Gevensleben et al., 2012) of NFB training protocols in order to optimize the potential positive effects of NFB on learning. However, no "optimal" NFB training procedure has yet been defined, and one research challenge is to design and evaluate optimal NFB training based on relevant performance and skill metrics.

The second challenge is to improve understanding about how these metrics and neuroplasticity indicators are linked in order to grasp the underlying neurophysiological mechanisms that explain EEG self-regulation and skills acquisition. If this relationship could be established, it would go a long way to validating such metrics. Indeed, as shown in **Figure 3**, performance and skills metrics should be understood not only in terms of the training BCI/NFB task but also with regard to indicators of neuroplasticity specific to the trained neural substrate [17]. Furthermore, this relationship could be considered as an important intermediate variable between NFB training/learning and clinical efficacy. As described in the first section of this paper, there are two kinds of indicators: dynamic modulation indicators based on EEG oscillation and Hebbian-type neuroplasticity indicators [19]. Thus, as EEG-based BCI/NFB tasks generally tend to modify EEG oscillations, performance metrics need to be related to dynamic modulation indicators. Maintaining the brain in a persistent oscillatory pattern improves the

brain circuit so that it can produce the same pattern with a higher probability in the future [19]. Thus, as BCI/NFB trains the brain to maintain certain oscillatory patterns, skills metrics need to be related to Hebbian neuroplasticity. See **Figure 3**. Very few studies dedicated to the clinical efficacy of NFB have investigated such neurophysiological indicators. Thus, in NFB, the neurophysiological relationship between dynamic modulation and deserves further attention [20, 71].

In conclusion, the human-computer interaction foundations of NFB demonstrates that training and learning are central to designing rigorous NFB protocols. Such protocols should be designed so that the induction of neuroplasticity is optimized *i.e*. it produces a lasting change after the training session. The relationship between NFB training performance, skills metrics and neuroplasticity induction is very exciting new ground that must now be explored in order to find new means of optimizing the clinical effect of NFB in the long term.

Conclusion

This paper investigated the neurophysiological, psychophysiological and human computer interaction foundations of neurofeedback. A transdisciplinary approach is now needed to evaluate rigorously the use of NFB as a therapeutic tool in psychiatry (**Figure 5)**. Notwithstanding the debate on the efficacy of NFB for treating mental disorders, this field of research remains fertile ground for innovative research in psychiatry. Neurophysiology, psychophysiology and human-computer interaction approaches of NFB pave the way for innovative research on two levels: for fundamental research attempting to define the mechanisms subsuming NFB training; and for clinical research aiming to establish better designed NFB protocols, control/active groups and clinical criteria that define efficacy in terms of targeted biomarkers.

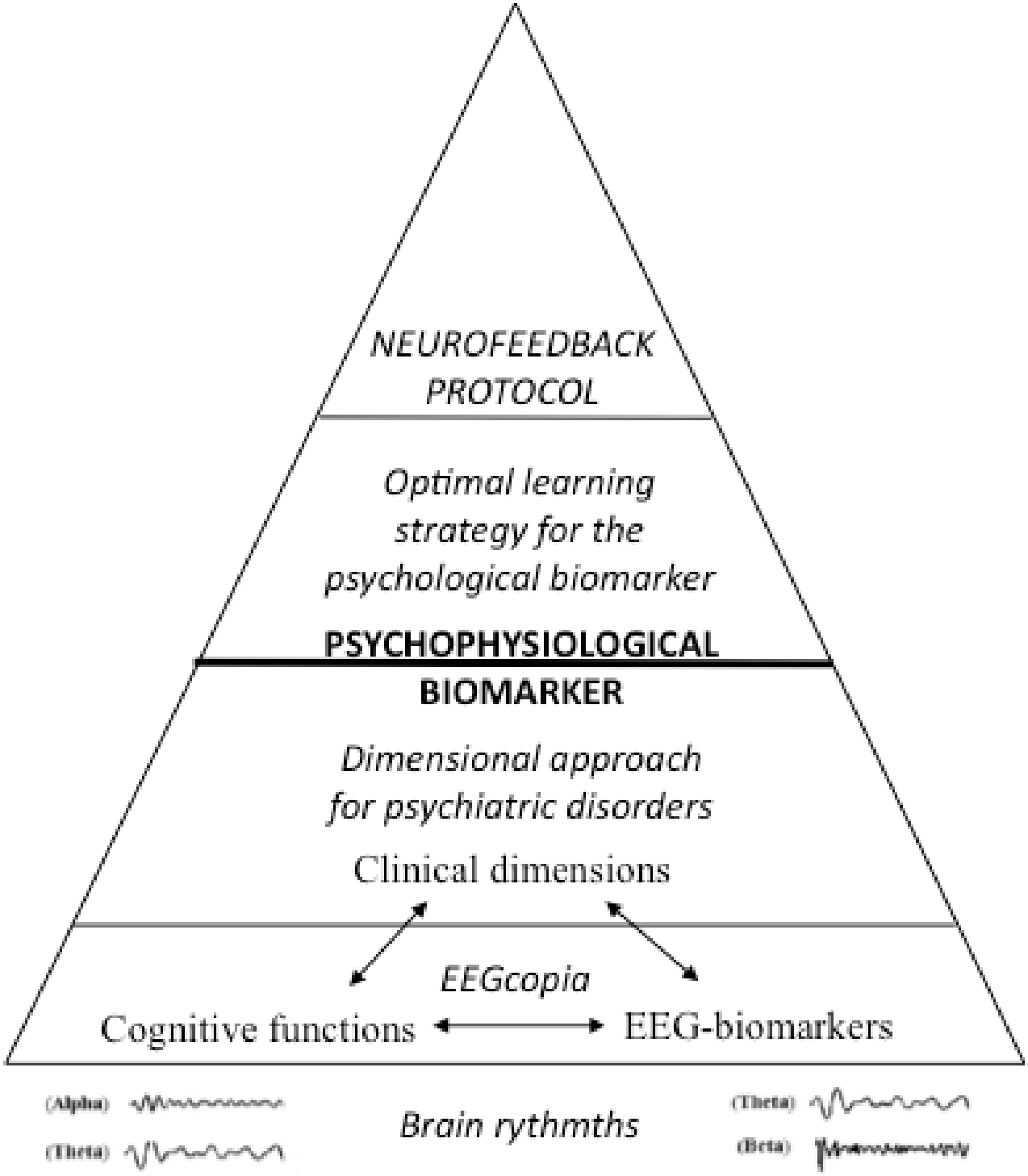


Figure 5. The quest to optimize neurofeedback protocol according to a transdisciplinary approach taking into account the neurophysiological, psychophysiological and human computer interaction bases of neurofeedback.

Acknowledgments

We thank Anatole Lecuyer for his participation to the second French congress on neurofeedback organized by NExT.

This work was supported by the French National Research Agency within the REBEL project (grant ANR-15-CE23-0013-01), the European Research Council with the BrainConquest project (grant ERC-2016-STG-714567), the Inria Project-Lab BCI-Lift and the EPFL/Inria International Lab.

Conflict of interest

None to declare concerning this paper.

References

[1] Savitz JB, Rauch SL, Drevets WC. Clinical application of brain imaging for the diagnosis of mood disorders: the current state of play. Mol Psychiatry 2013;18:528-39.

[2] Micoulaud-Franchi JA, McGonigal A, Lopez R, et al. Electroencephalographic neurofeedback: Level of evidence in mental and brain disorders and suggestions for good clinical practice. Neurophysiol Clin 2015;45:423-33.

[3] Arns M, Batail JM, Bioulac S, et al. Neurofeedback: One of today's techniques in psychiatry? Encephale 2017;43:135-45.

[4] Micoulaud Franchi JA, Pallanca O. Neurofeedback. In: Vion Dury J, Balzani C, Micoulaud Franchi JA, editors. Neurophysiologie clinique en psychiatrie. Paris: Elsevier Masson; 2015. p. 185-212.

[5] Arns M, Conners CK, Kraemer HC. A decade of EEG Theta/Beta Ratio Research in ADHD: a meta-analysis. J Atten Disord 2013;17:374-83.

[6] Micoulaud Franchi J, Geoffroy P, Fond G, et al. EEG Neurofeedback treatments in children with ADHD: An updated meta-analysis of Randomized Controlled Trials. Front Hum Neurosc 2014;doi: 10.3389/fnhum.2014.00906.

[7] Cortese S, Ferrin M, Brandeis D, et al. Neurofeedback for Attention-Deficit/Hyperactivity Disorder: Meta-Analysis of Clinical and Neuropsychological Outcomes From Randomized Controlled Trials. J Am Acad Child Adolesc Psychiatry 2016;55:444-55.

[8] Thibault RT, Raz A. When can neurofeedback join the clinical armamentarium? Lancet Psychiatry 2016;3:497-8.

[9] Thibault RT, Raz A. Neurofeedback: the power of psychosocial therapeutics. Lancet Psychiatry 2016;3:e18.

[10] Micoulaud-Franchi JA, Fovet T. Neurofeedback: time needed for a promising non-pharmacological therapeutic method. Lancet Psychiatry 2016;3:e16.

[11] Fovet T, Micoulaud-Franchi JA, Vialatte FB, et al. On assessing neurofeedback effects: should double-blind replace neurophysiological mechanisms? Brain 2017;140:e63.

[12] Gaume A, Vialatte A, Mora-Sánchez A, et al. A psychoengineering paradigm for the neurocognitive mechanisms of biofeedback and neurofeedback. Neurosci Biobehav Rev 2016;doi: 10.1016/j.neubiorev.2016.06.012.

[13] Jeunet C, N'Kaoua B, Lotte F. Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates. Prog Brain Res 2016;228:3-35.

[14] Vion-Dury J, Balzani C, Cermolaccce M. Modalités d'acquisition et d'analyse du signal EEG. In: Vion Dury J, Balzani C, Micoulaud Franchi JA, editors. Neurophysiologie clinique en psychiatrie. paris: Elsevier Masson; 2015.

[15] Engel AK, Gerloff C, Hilgetag CC, et al. Intrinsic coupling modes: multiscale interactions in ongoing brain activity. Neuron 2016;80:867-86.

[16] Coburn KL, Lauterbach EC, Boutros NN, et al. The value of quantitative electroencephalography in clinical psychiatry: a report by the Committee on Research of the American Neuropsychiatric Association. J Neuropsychiatry Clin Neurosci 2006;18:460-500.

[17] Sitaram R, Ros T, Stoeckel L, et al. Closed-loop brain training: the science of neurofeedback. Nat Rev Neurosci 2016;18:86-100.

[18] Lotte F, Jeunet C. Online classification accuracy is a poor metric of mental imagery-based BCI (MI-BCI) performances: an experimental demonstration and new metrics. Submitted.

[19] Ros T, B JB, Lanius RA, et al. Tuning pathological brain oscillations with neurofeedback: a systems neuroscience framework. Front Hum Neurosci 2014;8:1008.

[20] Zoefel B, Huster RJ, Herrmann CS. Neurofeedback training of the upper alpha frequency band in EEG improves cognitive performance. Neuroimage 2011;54:1427-31.

[21] Kluetsch RC, Ros T, Theberge J, et al. Plastic modulation of PTSD resting-state networks and subjective wellbeing by EEG neurofeedback. Acta Psychiatr Scand 2014;130:123-36.

[22] Kamiya J. Biofeedback training in voluntary control of EEG alpha rhythms. Calif Med 1971;115:44.

[23] Sterman MB, Howe RC, Macdonald LR. Facilitation of spindle-burst sleep by conditioning of electroencephalographic activity while awake. Science 1970;167:1146-8.

[24] Ros T, B JB, Lanius RA, et al. Tuning pathological brain oscillations with neurofeedback: a systems neuroscience framework. Front Hum Neurosci 2016;8:1008.

[25] Ros T, Munneke MA, Ruge D, et al. Endogenous control of waking brain rhythms induces neuroplasticity in humans. Eur J Neurosci 2010;31:770-8.

[26] Butefisch CM, Davis BC, Wise SP, et al. Mechanisms of use-dependent plasticity in the human motor cortex. Proc Natl Acad Sci U S A 2000;97:3661-5.

[27] Heilbronner SR, Hayden BY. Dorsal Anterior Cingulate Cortex: A Bottom-Up View. Annu Rev Neurosci 2016;39:149-70.

[28] Levesque J, Beauregard M, Mensour B. Effect of neurofeedback training on the neural substrates of selective attention in children with attention-deficit/hyperactivity disorder: a functional magnetic resonance imaging study. Neurosci Lett 2006;394:216-21.

[29] Ros T, Theberge J, Frewen PA, et al. Mind over chatter: plastic up-regulation of the fMRI salience network directly after EEG neurofeedback. Neuroimage 2013;65:324-35.

[30] Ghaziri J, Tucholka A, Larue V, et al. Neurofeedback training induces changes in white and gray matter. Clin EEG Neurosci 2013;44:265-72.

[31] Cuthbert BN. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry 2014;13:28-35.

[32] Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry 2010;167:748-51.

[33] Wolpaw JR, Birbaumer N, McFarland DJ, et al. Brain-computer interfaces for communication and control. Clin Neurophysiol 2002;113:767-91.

[34] Buttfield A, Ferrez P, Del R Millan J. owards a Robust BCI: Error Potentials and Online Learning. EEE Trans Neural Syst Rehabil Eng 2006;14:164–68.

[35] Brandmeyer A, Sadakata M, Spyrou L, et al. Decoding of single-trial auditory mismatch responses for online perceptual monitoring and neurofeedback. Front Neurosci 2013;7:265.

[36] Woodman GF. A brief introduction to the use of event-related potentials in studies of perception and attention. Atten Percept Psychophys 2010;72:2031-46.

[37] Olbrich S, Mulert C, Karch S, et al. EEG-vigilance and BOLD effect during simultaneous EEG/fMRI measurement. Neuroimage 2009;45:319-32.

[38] Hegerl U, Hensch T. The vigilance regulation model of affective disorders and ADHD. Neurosci Biobehav Rev 2012;44:45-57.

[39] Rayner G, Jackson G, Wilson S. Cognition-related brain networks underpin the symptoms of unipolar depression: Evidence from a systematic review. Neurosci Biobehav Rev 2016;61:53-65.

[40] Escolano C, Navarro-Gil M, Garcia-Campayo J, et al. A controlled study on the cognitive effect of alpha neurofeedback training in patients with major depressive disorder. Front Behav Neurosci 2014;8:296.

[41] Mano M, Lecuyer A, Bannier E, et al. How to Build a Hybrid Neurofeedback Platform Combining EEG and fMRI. Front Neurosci 2017;11:140.

[42] Perronnet L, Lecuyer A, Mano M, et al. Unimodal Versus Bimodal EEG-fMRI Neurofeedback of a Motor Imagery Task. Front Hum Neurosci 2017;11:193.

[43] Zotev V, Phillips R, Yuan H, et al. Self-regulation of human brain activity using simultaneous real-time fMRI and EEG neurofeedback. Neuroimage 2014;85 Pt 3:985-95.

[44] Zotev V, Yuan H, Misaki M, et al. Correlation between amygdala BOLD activity and frontal EEG asymmetry during real-time fMRI neurofeedback training in patients with depression. Neuroimage Clin 2016;11:224-38.

[45] Polich J. Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol 2007;118:2128-48.

[46] Johnson R, Jr. Scalp-recorded P300 activity in patients following unilateral temporal lobectomy. Brain 1988;111 (Pt 6):1517-29.

[47] Mattout J, Perrin M, Bertrand O, et al. Improving BCI performance through co-adaptation: applications to the P300-speller. Ann Phys Rehabil Med 2015;58:23-8.

[48] Maby E, Perrin M, Bertrand O, et al. BCI Could Make Old Two-Player Games Even More Fun: A Proof of Concept with “Connect Four”. . Adv Hum-Comput Interact 2012.

[49] Johnstone SJ, Barry RJ, Clarke AR. Ten years on: a follow-up review of ERP research in attention-deficit/hyperactivity disorder. Clin Neurophysiol 2013;124:644-57.

[50] Seifert J, Scheuerpflug P, Zillessen KE, et al. Electrophysiological investigation of the effectiveness of methylphenidate in children with and without ADHD. J Neural Transm (Vienna) 2003;110:821-9.

[51] Fouillen M, Maby E, Le Carrer L, et al. ERP-based BCI training for children with ADHD: motivations and trial design. Proceeding 7th Int Brain-Comput Interface Conf Graz 2017.

[52] Enriquez-Geppert S, Huster RJ, Herrmann CS. EEG-Neurofeedback as a Tool to Modulate Cognition and Behavior: A Review Tutorial. Front Hum Neurosci 2017;11:51.

[53] Alkoby O, Abu-Rmileh A, Shriki O, et al. Can we predict who will respond to neurofeedback? A review of the inefficacy problem and existing predictors for successful EEG neurofeedback learning. Neuroscience 2017.

[54] Jeunet C, Lotte F, Batail J-M, et al. How to improve clinical neurofeedback using a human-factor centered standpoint? A short review of the insights provided by the literature on BCI. Neuroscience In revision.

[55] Jeunet C, Jahanpour E, Lotte F. Why Standard Brain-Computer Interface (BCI) Training Protocols Should be Changed: An Experimental Study. Journal of Neural Engineering 2016;13.

[56] Mladenovic J, Mattout J, Lotte F. A Generic Framework for EEG based BCI training and operation. BCI Handbook: Technological and Theoretical Advances 2917.

[57] Fruitet J, Carpentier A, Munos R, et al. Automatic motor task selection via a bandit algorithm for a brain-controlled button. J Neural Eng 2013;10:016012.

[58] McFarland DJ, Sarnacki WA, Wolpaw JR. Electroencephalographic (EEG) control of three-dimensional movement. J Neural Eng 2010;7:036007.

[59] Mladenovic J, Frey J, Bonnet-Save M, et al. The Impact of Flow in an EEG-based Brain Computer Interface. 7th International Graz BCI Conference 2017.

[60] Lotte F, Faller J, Guger C, et al. Combining BCI with virtual reality: towards new applications and improved BCI. Towards Practical Brain-Computer Interfaces 2012:197-220.

[61] Lotte F, Larrue F, Muhl C. Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design. Front Hum Neurosci 2013;7:568.

[62] Kubler A, Furdea A, Halder S, et al. A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. Ann N Y Acad Sci 2009;1157:90-100.

[63] Frey J, Gervais R, Fleck S, et al. Teegi: tangible EEG interface. Proceedings of the 27th annual ACM symposium on User interface software and technology 2014:301-8.

[64] Pillette L, Jeunet C, Mansencal B, et al. PEANUT: Personalised Emotional Agent for Neurotechnology User-Training. 7th International BCI Conference 2017.

[65] Zuberer A, Drandeis D, Drechsler R. Are treatment effects of neurofeedback training in children with ADHD related to the successful regulation of brain activity? A review on the learning of regulation of brain activity and a contribution to the discussion on specificity. Front Hum Neurosc 2015;doi: 10.3389/fnhum.2015.00135.

[66] Rémond A, Rémond A. Biofeedback : principes et applications. Paris: Masson; 1997.

[67] Sherlin LH, Arns M, Lubar J, et al. Neurofeedback and basic lerning therory: implications for research and practice. Journal of Neurotherapy 2011;15:292-304.

[68] Strehl U. What learning theories can teach us in designing neurofeedback treatments. Front Hum Neurosci 2014;8:894.

[69] Vollebregt MA, van Dongen-Boomsma M, Slaats-Willemse D, et al. What future research should bring to help resolving the debate about the efficacy of EEG-neurofeedback in children with ADHD. Front Hum Neurosci 2014;8:321.

[70] Gevensleben H, Rothenberger A, Moll GH, et al. Neurofeedback in children with ADHD: validation and challenges. Expert Rev Neurother 2012;12:447-60.

[71] Engelbregt HJ, Keeser D, van Eijk L, et al. Short and long-term effects of sham-controlled prefrontal EEG-neurofeedback training in healthy subjects. Clin Neurophysiol 2016;127:1931-7.

3.1.3 Polémiques actuelles sur l'efficacité du neurofeedback

L'efficacité des techniques de neurofeedback reste à ce jour très discutée, certains auteurs soulignant l'absence d'essais contrôlés randomisés rigoureusement menés dans le domaine et l'importance potentielle de l'effet placebo avec ce type de techniques, ce qui a pu donner lieu récemment à des débats dans des revues internationales comme le *Lancet Psychiatry* (223,225–227).

Même les utilisations pour lesquelles de nombreuses études d'évaluation sont disponibles comme l'utilisation du neurofeedback guidé par EEG dans la prise en charge du trouble déficit de l'attention / hyperactivité, restent sujettes à polémiques (228–230). Notamment, la dernière méta-analyse sur le sujet montre une absence d'efficacité (228). Toutefois, certains choix méthodologiques y apparaissent discutables puisque la précédente méta-analyse avait mis en évidence une efficacité du neurofeedback dans cette indication et qu'aucune étude supplémentaire n'a été publiée entre la réalisation des deux méta-analyses (228,231,232).

Des études dans d'autres champs comme la prise en charge de l'insomnie montrent à quel point l'évaluation des protocoles de neurofeedback est complexe (233). C'est cette complexité qui a fait l'objet de l'**Article 8** publié dans *Brain* en réponse aux résultats négatifs d'un essai contrôlé randomisé en double-aveugle évaluant l'intérêt du neurofeedback guidé par EEG dans la prise en charge de l'insomnie (233). Cet article était accompagné d'un commentaire généralisant les résultats de cette étude à l'ensemble du champs du neurofeedback (234).

ARTICLE 8

On assessing neurofeedback effects:

Should double-blind replace neurophysiological mechanisms?

Thomas FOVET[1*], Jean-Arthur MICOULAUD-FRANCHI[2], François-Benoît VIALATTE[3,4], Fabien LOTTE[5], Christophe DAUDET[6], Jean-Marie BATAIL[7], Jérémie MATTOUT[8], Guilherme WOOD[9], Renaud JARDRI[1], Stefanie ENRIQUEZ-GEPPERT[10], Tomas ROS[11]

1. Univ. Lille, CNRS UMR 9193, Laboratoire de Sciences Cognitives et Sciences Affectives (SCALab-PsyCHIC), F-59000 Lille, France & CHU Lille, Pôle de Psychiatrie, Unité CURE, F-59000 Lille, France

2. Services d'explorations fonctionnelles du système nerveux, Clinique du sommeil, CHU de Bordeaux, Place Amélie Raba-Leon, 33076 Bordeaux, France & USR CNRS 3413 SANPSY, CHU Pellegrin, Université de Bordeaux, France.

3. CNRS UMR 8249, Brain Plasticity Unit, Brain-Computer Interface team, 10 rue Vauquelin, 75231 Paris Cedex 05, France.

4. ESPCI Paris, PSL Research University, 10 rue Vauquelin, 75231 Paris Cedex 05, France.

5. Inria Bordeaux Sud-Ouest / LaBRI / Univ. Bordeaux / CNRS / IPB – 200 avenue de la vieille tour, 33405, Talence Cedex, France

6. Université de Bordeaux, 33000 Bordeaux, France.

7. Academic Psychiatry Department, Centre Hospitalier Guillaume Régnier, Rennes, France ; Inserm, Visages U746, IRISA, Campus de Beaulieu, F-35042 Rennes, France ; "Behavior and Basal Ganglia" research unit (EA 4712), University of Rennes 1, Rennes, France

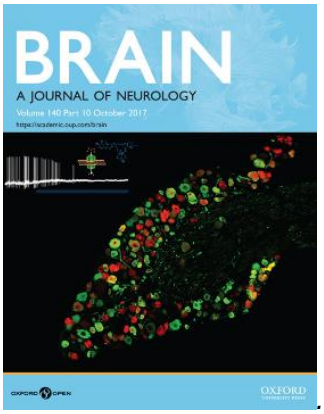
8. Brain Dynamics and Cognition Team, Lyon Neuroscience Research Center, INSERM U1028-CNRS UMR5292-University Lyon 1, Lyon F-69000, France

9. Department of Psychology, Karl-Franzens-University Graz Graz, Austria ; BioTechMed Graz, Austria.

10. Department of Clinical and Developmental Neuropsychology, University of Groningen, Groningen, the Netherlands

11. Laboratory Behavioral Neurology and Imaging of Cognition, Department of Neuroscience, University Medical Center and Campus Biotech, Geneva, Switzerland.

Brain. 2017 Oct 1;140(10):e63



ABBREVIATED SUMMARY

In their article published in *Brain*, Schabus et al. (2017) conclude that neurofeedback does not have any specific efficacy in primary insomnia. Fovet et al. report that inconsistencies exist between earlier and current results published by the same team, emphasizing the need for a deeper exploration of the neural mechanisms underlying neurofeedback.

Sir,

We read with great interest the recent article of Schabus et al. entitled "Better than sham? A double-blind placebo-controlled neurofeedback study in primary insomnia" published in *Brain* (Schabus *et al.*, 2017) and its commentary “Neurofeedback or neuroplacebo?” (Thibault *et al.*, 2017). In recent years, EEG-neurofeedback (NFB) has benefited from a revival of interest, although its clinical efficacy remains a controversial and delicate issue (Micoulaud-Franchi and Fovet, 2016; Thibault and Raz, 2016; Sitaram *et al.*, 2017). In the context of a general reproducibility crisis in science (Baker, 2016), we can only be delighted that negative results are being recognised in leading neurology journals such as *Brain*. The findings of Schabus et al. are of great scientific interest, contributing to a stimulating debate in the domain of EEG-NFB. However, we believe that caution is needed before generalizing these results to SMR frequency training as well as the entire field. This is because irrespective of whether results from a single study are positive or negative, “one swallow does not a summer make”. Moreover, despite the study’s double-blind design, inconsistencies exist between earlier and current results published by the same research group.

It may at first appear surprising that Schabus et al were not able to replicate some key neurophysiological relationships that have consistently emerged in their previous work. In 2008, with a sample of 27 healthy subjects, the same team demonstrated that 10 NFB sessions of sensorimotor rhythm (SMR, 12-15 Hz) up-regulation was successful in: (1) conditioning an increase in relative SMR amplitude; (2) eliciting positive changes in sleep parameters (sleep spindle number and sleep onset latency); (3) eliciting changes in declarative memory performance (enhancement in retrieval score computed at immediate cued report) (Hoedlmoser *et al.*, 2008). In the same vein, in 2014, Schabus et al. reported in 24 patients with insomnia disorder (4) a positive correlation between SMR-NFB training enhancement, overnight memory consolidation and sleep spindle changes; (5) a significant effect of SMR-NFB on objective sleep quality (a decrease in the number of awakenings, a trend towards decreased sleep onset latency and an increase in slow wave sleep) (Schabus *et al.*, 2014).

Paradoxically, in their double-blind study involving 16 patients with insomnia disorder and 9 patients with misperception insomnia, they did not reproduce a significant effect of SMR-NFB on objective measures of sleep quality neither in the active nor in the placebo group (Schabus *et al.*, 2017). However, in a well-controlled double-blind study unrelated to insomnia, Kober et al. replicated finding (1) and extended (3) from immediate recall to 24-hours delayed response (Kober *et al.*, 2015a). Therefore, the use of double-blind control alone cannot explain all the null-results of Schabus et al. (2017). Definite conclusions can only be drawn on the basis of more solid data.

In particular, we think any discussion of NFB is incomplete without considering its basis from the point of view of *neurophysiological mechanisms*. Indeed, the previously established correlations between post-NFB sleep spindle generation and within-session NFB control highlight key neurophysiological mechanisms that are difficult to be reduced to simple placebo processes and/or a single-blind design (Hoedlmoser *et al.*, 2008; Schabus *et al.*, 2014). From a historical perspective, seminal experiments by Sterman and colleagues (Sterman *et al.*, 1970) were the first to show that waking SMR activity may be operantly-conditioned to be more strongly expressed during subsequent sleep. Since long-term effects manifest in the same direction dictated by the training, a candidate mechanism may be Hebbian plasticity (see (Ros *et al.* 2014) for a review). In fact, a host of other NFB studies indicate a similar Hebbian relationship between within-session and post-session EEG changes (*e.g.* Cho *et al.*, 2008; Zoefel *et al.*, 2011; Engelbregt *et al.*, 2016). Importantly, online control of spectral power is necessary but insufficient as a demonstration of brain plasticity induction (*i.e.* a lasting change outside of the training session). Hence, a crucial question of mechanistic importance is why there was no association between the NFB training in the current study with any change in offline spindle activity and offline SMR activity; as observed twice in the team's previous work (Hoedlmoser *et al.*, 2008; Schabus *et al.*, 2014)?

In that respect, we noticed a potentially important methodological change between their 2008, 2014 and 2017 studies. Each time the authors used a different rule for setting and adapting the reward threshold. They decreased the threshold based on no particular rule in 2008, following "< 5 success in a 3 min block" in 2014, and "< 13 success in a 5 min block" in 2017. We wonder why such changes have been made, since the last two rules are not proportionally related, while in contrast the proportionality between the rules governing the threshold increase were preserved. Since this modification effectively makes the NFB task less challenging, it may have led to a decrease in the SMR activity during the 3-second baseline periods preceding each trial and used as references for each trial when compared with the spontaneous *resting-state* SMR activity of each subject prior to NFB. Hence, given that *Figure 2* only displays percent evoked power *from* this respective 3-second baseline, it is hard to verify whether absolute SMR power during NFB actually exceeded the values of the resting-state activity recorded before the start of NFB training. From a mechanistic standpoint, altering EEG activity significantly from its spontaneous value may be a critical determinant of plasticity induction (Ros *et al.*, 2014). We therefore invite Schabus et al to clarify this issue by submitting supplementary data providing absolute SMR changes during NFB relative to the initial 2-minute resting-state, unreferenced to the (potentially variable) 3-second baseline(s) before each trial. Such differences (Hoedlmoser *et al.*, 2008; Schabus *et al.*, 2014, 2017) may arise by simply using different learning indices, and/or the influence of the experimenter's knowledge/un-blindness; however, they might also reflect differences in NFB learning and ultimately contribute to a reduced impact on brain plasticity.

Finally, and regardless of whether brain plasticity was actually induced by Schabus et al's (2017) paradigm, logical reasoning should have restricted Thibault et al to a simpler claim: that this specific NFB training, for this specific application, was not better than placebo. Instead, they appear to overgeneralize the null findings in the treatment of insomnia to the greater field of EEG-NFB. Such a position appears to us to be more ideological than scientific and contradicts the overall spirit necessary

for advancing medical research. By jumping to conclusions, Thibault et al. forget that caution needs to be applied to both positive and negative findings. NFB should be considered as a unique tool for targeting specific neural activities rather than as a panacea for all brain disorders. Successful deployment of NFB critically depends on our knowledge of the brain's inner-workings, which still remains incomplete. Hence, to properly exploit this approach, there is an urgent need for more research in order to both optimize NFB learning (e.g.number of trials in a NFB block, threshold rules, number of training sessions) and (Micoulaud-Franchi *et al.*, 2016; Enriquez-Geppert *et al.*, 2017) to select the most appropriate training protocol for each disorder (Kober *et al.*, 2015b).

In conclusion, the negative findings published by Schabus et al. excitingly generate more questions than answers for the field. Moreover, the last fifty years since NFB's discovery cannot be considered a homogenous record, given several decades of relative dormancy in terms of research output, before a resurgence in the early 2000s. Hence, these results call for more research rather than less, including a deeper exploration of the neural mechanisms and methodological nuances emerging from this embryonic field - preferably before premature launches of double-blind clinical studies. It is more conceivable that the story of NFB is a simple reflection of the general scientific process, which, through its twists and turns, remains in the safer judgement of posterity.

REFERENCES

Baker M. Reproducibility: Seek out stronger science. Nature 2016; 537: 703–704.

Cho MK, Jang HS, Jeong S-H, Jang I-S, Choi B-J, Lee M-GT. Alpha neurofeedback improves the maintaining ability of alpha activity. Neuroreport 2008; 19: 315–317.

Engelbregt HJ, Keeser D, van Eijk L, Suiker EM, Eichhorn D, Karch S, et al. Short and long-term effects of sham-controlled prefrontal EEG-neurofeedback training in healthy subjects. Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. 2016; 127: 1931–1937.

Enriquez-Geppert S, Huster RJ, Herrmann CS. EEG-Neurofeedback as a Tool to Modulate Cognition and Behavior: A Review Tutorial. Front. Hum. Neurosci. 2017; 11.

Hoedlmoser K, Pecherstorfer T, Gruber G, Anderer P, Doppelmayr M, Klimesch W, et al. Instrumental conditioning of human sensorimotor rhythm (12-15 Hz) and its impact on sleep as well as declarative learning. Sleep 2008; 31: 1401–1408.

Kober SE, Witte M, Stangl M, Väljamäe A, Neuper C, Wood G. Shutting down sensorimotor interference unblocks the networks for stimulus processing: An SMR neurofeedback training study. Clin. Neurophysiol. 2015a; 126: 82–95.

Kober SE, Schweiger D, Witte M, Reichert JL, Grieshofer P, Neuper C, et al. Specific effects of EEG based neurofeedback training on memory functions in post-stroke victims. J. Neuroengineering Rehabil. 2015b; 12: 107.

Micoulaud-Franchi J-A, Fovet T. Neurofeedback: time needed for a promising non-pharmacological therapeutic method. Lancet Psychiatry 2016; 3: e16.

Micoulaud-Franchi J-A, Salvo F, Bioulac S, Fovet T. Neurofeedback in Attention-Deficit/Hyperactivity Disorder: Efficacy. J. Am. Acad. Child Adolesc. Psychiatry 2016; 55: 1091–1092.

Ros T, J Baars B, Lanius RA, Vuilleumier P. Tuning pathological brain oscillations with neurofeedback: a systems neuroscience framework. Front. Hum. Neurosci. 2014; 8: 1008.

Schabus M, Griessenberger H, Gnjezda M-T, Heib DPJ, Wislowska M, Hoedlmoser K. Better than sham? A double-blind placebo-controlled neurofeedback study in primary insomnia. Brain J. Neurol. 2017; 140: 1041–1052.

Schabus M, Heib DPJ, Lechinger J, Griessenberger H, Klimesch W, Pawlizki A, et al. Enhancing sleep quality and memory in insomnia using instrumental sensorimotor rhythm conditioning. Biol. Psychol. 2014; 95: 126–134.

Sitaram R, Ros T, Stoeckel L, Haller S, Scharnowski F, Lewis-Peacock J, et al. Closed-loop brain training: the science of neurofeedback. Nat. Rev. Neurosci. 2017; 18: 86–100.

Sterman MB, Howe RC, Macdonald LR. Facilitation of spindle-burst sleep by conditioning of electroencephalographic activity while awake. Science 1970; 167: 1146–1148.

Thibault RT, Lifshitz M, Raz A. Neurofeedback or neuroplacebo? Brain 2017; 140: 862–864.

Thibault RT, Raz A. When can neurofeedback join the clinical armamentarium? Lancet Psychiatry 2016; 3: 497–498.

Zoefel B, Huster RJ, Herrmann CS. Neurofeedback training of the upper alpha frequency band in EEG improves cognitive performance. NeuroImage 2011; 54: 1427–1431.

3.2 Le neurofeedback guidé par IRMf : utilisation en psychiatrie

Après une description générale du champ du neurofeedback, nous proposons de nous focaliser ici sur le neurofeedback guidé par IRM fonctionnelle et à ses applications possibles en psychiatrie.

Ce domaine a fait l'objet d'une revue systématique de la littérature présentée en **Article 9**.

ARTICLE 9

Current issues in the use of fMRI-based neurofeedback to relieve psychiatric symptoms

Thomas FOVET[1,2*], Renaud JARDRI[1,2], David LINDEN[3]

1. Pôle de psychiatrie, Hôpital Fontan, CS-70001, CHRU de Lille, F-59000 Lille, France

2. SCA-Lab, CNRS UMR 9193, Université de Lille, F-59000 Lille, France

3. School of Psychology and MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, UK

Current Pharmaceutical Design. 2015; 21(23):3384-94.

ABSTRACT:

fMRI-based neurofeedback (fMRI-NF) is a non-invasive technique that allows participants to achieve control of their own brain activity using the real-time feedback of the activity (measured indirectly based on the BOLD signal) of a particular brain region or network. The feasibility of fMRI-NF in healthy subjects has been documented for a variety of brain areas and neural systems, and this technique has also been proposed for the treatment of psychiatric disorders in recent years. Through a systematic review of the scientific literature this paper probes the rationale and expected applications of fMRI-NF in psychiatry, discusses issues that must be addressed in the use of this technique to treat mental disorders. Six relevant references and five ongoing studies were identified according to our inclusion criteria. These studies show that in most psychiatric disorders (*major depressive disorder*, *schizophrenia*, *personality disorders*, *addiction*), patients are able to learn voluntary control of the neuronal activity of the targeted brain region(s). Interestingly, in some cases, this learning is associated with clinical improvement, showing that fMRI-NF can potentially be developed into a therapeutic tool. However, only low-level evidence is available to support the use of this relatively new technique in clinical practice. Notably, no randomized, controlled trial is currently available in this field of research. Finally, methodological issues and clinical perspectives (especially the potential use of pattern recognition in fMRI-NF protocols) are discussed.

KEY WORDS: neurofeedback, real-time fMRI, psychiatric disorder, pattern recognition, machine learning, self-efficacy.

Introduction

Mental disorders pose a major worldwide public health concern. Because of their high prevalence and poor long-course evolution (1), these conditions are among the leading causes of disability and health costs in developed countries (2). Despite the increasing sophistication of biological and psychosocial research methods, our understanding of these disorders remains fragmented and recent progress in genetics and neuroimaging has not yet been converted into new therapeutic applications (3). The evolution of psychopharmacological research is a good example, because this is an area where no significant therapeutic breakthrough has occurred in the last three decades. Furthermore, recent evidence suggests that this trend is set to continue as many pharmaceutical industries are drastically cutting down their investments in research on mental disorders (4).

Non-drug treatments have therefore been proposed as an option for addressing the unmet medical needs described above. Previous studies have reported some promising results, for example, for deep-brain stimulation (DBS) (5), repetitive transcranial magnetic stimulation (rTMS) (6), vagus nerve stimulation (VNS) (7) or transcranial direct current stimulation (8). Crucially, the principles of these techniques emerged from the recent progress in deciphering the neural underpinnings of psychiatric symptoms and disorders, allowing the definition of specific targets. However, all of these therapeutic strategies require physical interventions, which can be invasive for some of these approaches (especially VNS or DBS), thus limiting a wider use in clinical practice.

In this context, fMRI-neurofeedback appears to provide a both innovative and non-invasive tool for use in psychiatry (9–11). Indeed, several promising results were recently obtained using neurofeedback techniques to relieve subjective symptoms such as pain (12) or tinnitus (13), and it appears logical to consider the translation of this tool to psychiatry (i.e., to alleviate subjective psychiatric symptoms).

What is fMRI-based neurofeedback?

The principles of neurofeedback

Neurofeedback is a non-invasive technique that applies the principles of biofeedback to brain activity (7). The principle of biofeedback is simple: (i) a physiological variable (e.g., heart rate or electrodermal skin conductance) is measured, and information on this variable is provided to the subject from time to time; (ii) the subject is then asked to modify this variable based on the real-time feedback. In the case of neurofeedback strategies, the physiological variable is neural activity (see **figure 1**) (8). This methodology provides a closed loop, allowing the participant to learn to control the neuronal activity of a particular brain region (i.e., the target).

Neurofeedback using electroencephalography (EEG) signals emerged in the 1960s and 1970s (14,15), but its development was slowed by technical limitations. Despite this limitation, these studies provided promising findings. For example, a recent meta-analysis of randomized controlled trials showed that EEG-based neurofeedback strategies result in an improvement of the inattention dimension of *attention deficit hyperactivity disorder* (ADHD) symptoms in children (16).

The rapid development of functional magnetic resonance imaging (fMRI) techniques and especially the emergence of real-time functional MRI (rtf-MRI) in the early 2000s (17) spawned a revolution in the development of neurofeedback, notably including the emergence of new neurofeedback brain-computer interfaces (see **Box 1**).

fMRI-based neurofeedback (fMRI-NF)

fMRI is a non-invasive high-spatial-resolution technique that allows recording of the hemodynamic response in the whole brain, referred to as the *blood oxygenation level-dependent* (BOLD) signal, which is considered an indirect measure of neuronal activity (18). Considering that the pathophysiology of psychiatric disorders mainly involves large cortical-subcortical networks that are notably implicated in emotional regulation processes, the high spatial resolution and the whole-brain information obtained with fMRI are valuable in designing new neurofeedback protocols (19).

It is now possible to obtain immediate access to experimental results by analyzing data online, which is referred to as rtf-MRI (9). However, fMRI has a much lower temporal resolution than electrophysiological techniques, and the hemodynamic delay of about five seconds between neural activity and the vascular response must be considered in fMRI-NF protocols (i.e., participants should be aware of this delay (20)). Despite this limitation, the emergence of rtf-MRI allows the development of fMRI-NF protocols (see **figure 1**).

Consequently, the last ten years have seen a true “rebirth” of this forgotten neurofeedback technique in the field of neuroscience as well as in the field of psychiatry (21), which is paralleled by a resurgent interest in methodological and clinical developments of EEG-based neurofeedback.

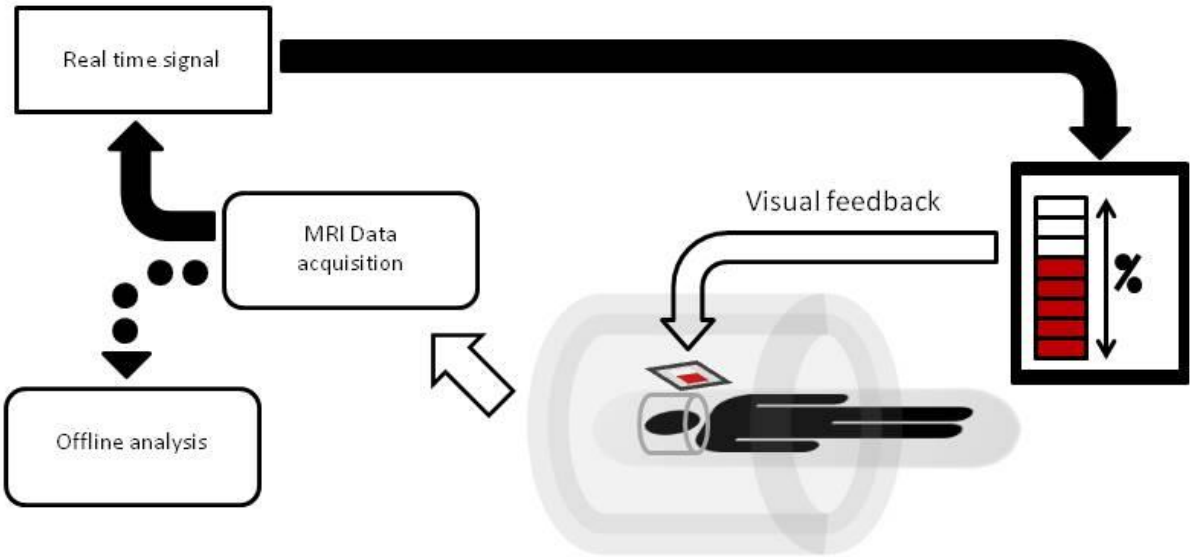


Figure 1: diagram of an fMRI-based neurofeedback system. The ongoing BOLD signals (i.e., *Blood Oxygen Level-Dependent* signals) from the participant are acquired and processed in two ways: through online analysis, providing a real-time signal that is presented to the participant, and offline analysis, providing information to explore changes in the whole brain. The participant, lying in the MR scanner, is provided with visual feedback (other modalities are also possible) of his or her own brain activity (different strategies can be used to define the target from which the BOLD signal will be processed online: see figure 3).

BOX 1: What is a Brain-Computer Interface?

A *brain-computer interface* (BCI) is a system that measures the activity of the central nervous system and converts it into an artificial output that replaces, restores, enhances, supplements, or improves its natural outputs (22). In other words, changes in brain activity are recorded and used to control a computer. In the last several years, BCI systems have been implemented in rehabilitation strategies for amyotrophic lateral sclerosis, Parkinson's disease, spinal cord injury, stroke, and disorders of consciousness with promising results (23). However, changes in brain activity can also be transformed by the computer system into sensory inputs (visual, auditory or other types) and fed-back to the subject. This last application corresponds to the definition of neurofeedback, which can be considered a subtype of BCI.

The use of fMRI-NF to treat mental disorders: state-of-the-art

The first studies that attempted to apply neurofeedback recruited healthy subjects. Ruiz et al. recently published an exhaustive review of these experiments (10), emphasizing the feasibility of fMRI-NF and showing that healthy individuals were able to learn volitional control of the activity of specific brain regions (24–29). Though their results may not yet be fully understood, a sufficient number of fMRI-NF proof-of-concept experiments have been published (30). Moreover, some studies revealed that the learning sessions were associated with behavioral changes (e.g., the response to aversive stimuli in (31)).

For many psychiatric symptoms, functional abnormalities have been identified in neuroimaging studies (e.g., for auditory hallucinations (32)). Normalizing these aberrant activations should be a key issue in new treatments. It is for this reason that neurofeedback could constitute an interesting and innovative therapeutic tool, offering the possibility for patients to normalize the activity level or connectivity strength of specific targets to reduce symptom severity (33). The general design of a neurofeedback protocol is described in **figure 2**.

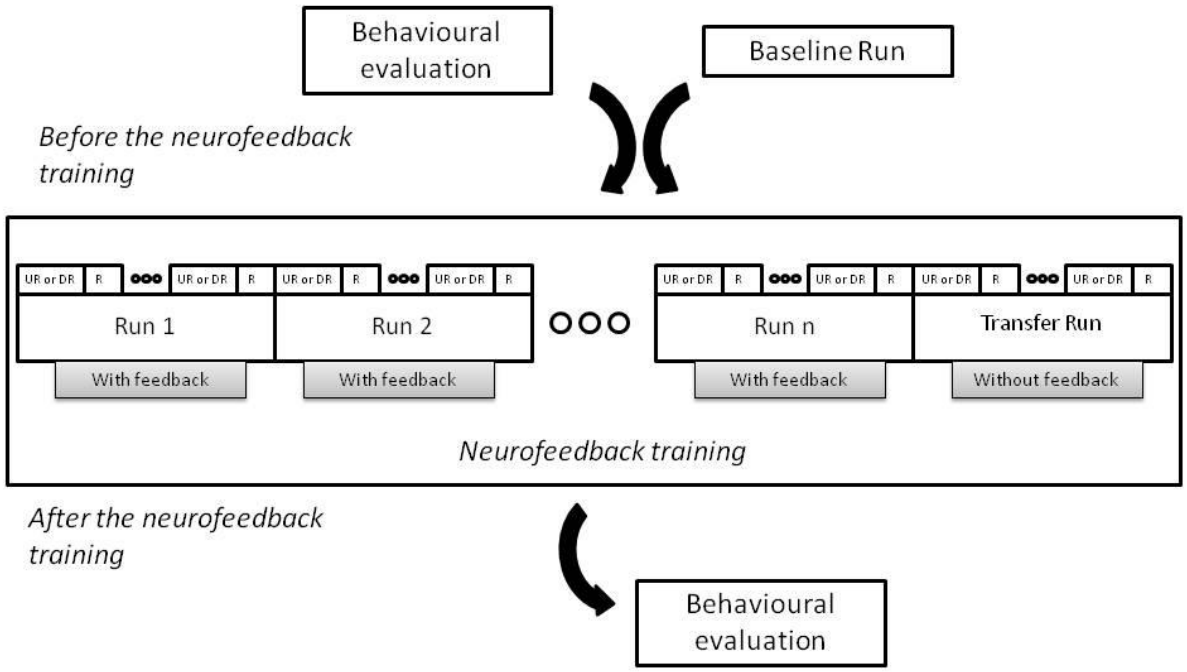


Figure 2. Example of an experimental design for the evaluation of an fMRI-based neurofeedback protocol to treat a psychiatric symptom or disorder. During each run, the participant alternates between periods of active regulation (up- or down-regulation) using the neurofeedback signal and rest periods (with the possibility of an intercurrent task, such as counting). The final run, during which the participant is asked to realize the same task without any feedback, is referred to as the "transfer run" and is important for testing the generalizability of the learning. Behavioral evaluations (clinical measures with validated scales, depending on the symptom studied) are conducted before and after the protocol. A baseline run is conducted to evaluate the activity of the target region before the neurofeedback training.

UR: upregulation; DR: downregulation; R: Rest

Here, we review the studies investigating the use of fMRI-NF in the treatment of psychiatric disorders. To identify eligible studies, a search was conducted in the PubMed electronic database to identify articles published until January 2015. We used the following MeSH headings: "neurofeedback" and "fMRI". The article titles and abstracts of the identified studies were then screened. The exclusion criteria were as follows: (i) papers written in a language other than English; (ii) opinion or hypothesis papers; (iii) populations suffering from a non-psychiatric disorder; or (iv) neurofeedback not based on an fMRI signal. Moreover, a search was conducted in the Clinical Trials electronic database using the terms “neurofeedback” and “fMRI” to identify current experiments that have not yet been published. The same exclusion criteria were applied.

The literature search identified 51 articles, from which 5 were eligible. One article was identified from an additional source (Google Scholar). The results are presented in **table 1**. The search conducted in the Clinical Trials electronic database identified 16 ongoing studies, from which 5 were eligible. The results are presented in **table 2**.

Major depressive disorder

Major depressive disorder is a highly prevalent psychiatric disorder, for which current treatments may remain insufficient (with frequent residual symptoms: more than 90% of remitters had at least one residual depressive symptom in a recent study (34)). This situation could explain why several studies have focused on fMRI-NF to treat this disorder: 2 studies have been published (see **table 1**), and 3 trials are currently ongoing (see **table 2**).

In a pilot study, Linden et al. functionally localized brain areas that were responsive to positively valenced visual stimuli (adapted from the International Affective Pictures System - IAPS) to train a group of patients to upregulate the activity in these target regions (mostly areas in the ventral prefrontal cortex and limbic system) over 4 sessions of fMRI-NF (35). The same team, which includes one of the authors of this review, had previously shown that using this procedure, healthy participants were able to learn to control the levels of activation in the areas involved in emotional processing (36,37). The depressed patients enrolled in the neurofeedback group (n=8) successfully learned to upregulate the BOLD signal of the target area and showed significant clinical improvement on the HDRS-17 (approximately four points on the 17-item HDRS (30% in their symptom score over the 1-month trial); effect size: *Cohen’s d* = 1.5), compared with a control group (in which 8 patients were instructed to engage in positive imagery strategies outside the scanner). Interestingly, no specific strategy was suggested to the participants, and most of them reported at debriefing that the best strategy was to evoke positive images related to themselves (for example, memories of happy events) and not remembering the pictures. The same team has just finished a randomized controlled trial (*clinicaltrials.gov*: NCT01544205) to replicate these results as well as to compare the effect of modulating different targets (i.e., upregulation of emotion networks vs. upregulation of a higher visual area) and, finally, to test the sustained benefits of such protocols (follow-up assessment one month after the intervention).

Based on a previous study (38), Young et al. proposed an fMRI-NF design to address depression that used a specific anatomical target, the amygdala. These authors taught 14 patients suffering from *major depressive disorder* to enhance the left amygdala (localized with structural MRI) response to positive autobiographical memory recall. Thus, through an fMRI-NF procedure targeting the left amygdala, depressed subjects were able to self-regulate their amygdala response. This functional effect persisted in the transfer run. Moreover, significant pre-post scan decreases in anxiety ratings and increases in happiness ratings were observed in the experimental group vs the control group (7 patients receiving sham feedback in the scanner) (39). An ongoing randomized double-blind, controlled trial of fMRI-NF using the same strategy focusing on the amygdala is currently being conducted by Young et al. (*clinicaltrials.gov*: NCT02079610) to replicate these findings in a larger sample (n=60).

In a third ongoing randomized double-blind, controlled trial, Moll et al. are testing the efficiency of fMRI-NF in reducing the risk of recurrent episodes in people with an antecedent of a depressive episode who are currently exhibiting remission (*clinicaltrials.gov*: NCT01920490). The investigators have previously shown that decoupling of brain networks (notably the anterior temporal and subgenual cortices) during experiences of guilty feelings may be a functional brain imaging biomarker of an increased risk of major depression. This biomarker remains detectable during remission periods (40). Moll et al. chose to use this biomarker as a potential target for fMRI-NF strategies. In this pilot study, the investigators seek to treat self-blame-selective neural decoupling by increasing the correlation between the anterior temporal and subgenual frontal fMRI signal for guilt relative to indignation.

Brain imaging research, and particularly fMRI research, has yielded several potential disease-relevant targets for major depressive disorder. However, there is still a long road to be travelled before an ideal fMRI-NF paradigm can be defined for depressive patients. Notably, it is still unclear whether self-regulation training might address primary abnormal processes (such as hyper- or hypoactivation of specific brain areas or networks involved in depressive symptomatology), or act in a different manner, by activating or suppressing circuits that are not primarily abnormal, but whose modulation may nevertheless produce clinical benefits (21).

Schizophrenia

In a proof-of-concept study published in 2013, Ruiz et al. demonstrated the feasibility of using fMRI-NF for patients with *schizophrenia*. In this work, 9 patients suffering from *schizophrenia* learned to regulate the hemodynamic response of the bilateral anterior insula (41). Interestingly, several behavioral effects were found to be associated with this training (patients recognized disgusted faces more accurately and happy faces less accurately). Even if no clinical applications follow directly, this study showed for the first time that patients with *schizophrenia* can learn volitional brain regulation through rtfMRI feedback training, paving the way for the development of such strategies to treat this highly prevalent and impairing disorder.

Personality disorders

Personality disorders are highly prevalent (10 percent of the general population suffer from a diagnosable *personality disorder* (42)), particularly in community mental health care (43). Given the high degree of comorbidities of *personality disorders*, ranging from depressive disorder to suicide (44), many brain imaging (45) or genetics (46) studies have attempted to better understand their neural underpinnings. Notably, the fact that personality disorders are no longer considered as fixed and stable over time (47,48) allows the development of innovative therapeutic strategies, including neurofeedback.

In a recent publication, Sitaram and colleagues proposed the use of fMRI-NF to treat patients with *psychopathic personality disorder* (49). In individuals with *psychopathy*, the absence of conditioned fear is reflected in a virtually complete lack of activation of the fear circuitry in the brain (50), i.e., hypoactivation of the insular cortex during classical fear conditioning compared with healthy controls. The same team previously showed that healthy subjects were able to learn to self-regulate the activity of the left anterior insula and that this regulation could modulate the emotional response specific to aversive pictures (31). Sitaram et al. explored whether patients suffering from *psychopathic personality disorder* could learn to regulate the activity of anterior insula. Only one of the four participants learned to up-regulate the BOLD signal of the insular cortex across training runs, but he was also the only participant to have completed all 12 training sessions (4 sessions for the 3 other participants). Interestingly, subjects with higher *Psychopathy Checklist Revised* (PCL: SV) scores were less able to increase the signal. In the participant who learned self-regulation, the volitional up-regulation of the insula was associated with changes in the subjective ratings of valence and arousal related to aversive stimuli.

This pilot study, despite several limitations (particularly the number of participants), indicates that some individuals with *psychopathic personality disorder* could learn to volitionally control brain activity pertaining to emotion, with behavioral consequences.

Addiction

Two studies exploring the potential use of fMRI-NF in the field of *addiction* medicine have been published in the last years. Both were conducted by the same team and focused on craving in nicotine-dependent cigarette smokers (51,52). The authors compared 2 strategies: “to reduce craving” by decreasing anterior cingular cortex (ACC) activity, which has been shown to be involved in craving (53), vs. “to increase resistance” by increasing middle prefrontal cortex (mPFC) activity. Both studies used cue-induced craving to preliminary define the regions of interest. For the “reduce craving” region of interest, a no-feedback session, during which the participants were instructed to “allow yourself to crave when you see the smoking-related pictures”, was used. For the “increase resistance” strategy, the same task was used, but the participants were instructed to “resist the urge to smoke when you see the smoking pictures by any means you find helpful”. In Hanlon’s work, the participants received simultaneous visual feedback (via two thermometers) (52), whereas in Li’s study, two successive paradigms were used (51).

The results showed that patients may be more able to learn to exert voluntary control over ACC activation than deliberate control of the mPFC to reduce craving. Indeed, Li et al. found that there was a significant correlation between decreased ACC activation and reduced craving ratings (subjectively reduced cue-induced craving ratings during rtfMRI feedback were significantly correlated with a decreased BOLD signal in the ACC) during the “reduce craving” session, whereas there was no modulation of the BOLD signal in the mPFC during the “increase resistance” session (51).

Obsessive compulsive disorder (OCD)

An ongoing controlled randomized trial to test the potential use of fMRI-NF in *OCD* is being conducted by Michelle Hampson’s team at Yale University (*clinicaltrials.gov*: NCT02206945). In a previous study (54), the same team demonstrated that healthy individuals with significant, but subclinical contamination anxiety were able to reduce the activity of a sub-region of the orbitofrontal cortex using fMRI-NF (the methodology is described in (55)). They also showed that after training, the participants exhibited reduced resting-state connectivity in the limbic circuitry and increased connectivity in the dorsolateral prefrontal cortex. These changes persisted several days after the neurofeedback training. Moreover, these changes in connectivity were correlated with the subjects’ control over contamination anxiety (evaluated in assessment sessions before and after the neurofeedback protocol, during which subjects viewed 25 contamination-related images and were instructed to attempt to minimize their anxiety in response to each image and to indicate the anxiety they experienced on a scale of 1–5). None of these improvements was observed in the control group, which received sham feedbacks. Contamination anxiety is a common symptom in OCD, and there is strong evidence to support a role of the orbitofrontal cortex in the pathophysiology of OCD (56), supporting the notion of testing fMRI-NF in OCD, targeting the orbitofrontal cortex.

Eating disorders

Although there is interest in using neurofeedback strategies to treat eating disorders (57), no published study is currently available. We can mention one randomized trial that is currently in the recruiting phase (*clinicaltrials.gov*: NCT02148770), in which the investigators seek to examine the effect of fMRI-NF on the up-regulation of functional connectivity between reward- and impulse-related brain areas in obese individuals. Indeed, altered activation patterns in reward-processing networks have been observed in populations suffering from obesity (58). This trial is based on a preliminary study, showing that obese people possess an improved capacity to self-regulate the anterior insula, a brain system tightly related to bodily awareness and gustatory functions (59).

Study	Disorder	N	Method	Target	N of runs	Control group	Method to define region(s) of interest	Main finding
Linden et al. (2012) (35)	Major depressive disorder	8 vs. 8	Upregulation	Brain areas involved in the generation of positive emotions	4	Cognitive strategies without neurofeedback outside the scanner (n=8)	Functional localizer (fMRI block-based paradigm: positive vs. neutral images)	Successful regulation Significant clinical improvement (*17 item-Hamilton Rating Scale for Depression* (HDRS))
Hanlon et al. (2013) (52)	Cue induced craving in nicotine users	9	Upregulation of "resist" regions Downregulation of "crave" regions (Simultaneous feedback)	"resist" regions (mPFC) and "crave" regions (ACC)	3	No	Functional localizer: (fMRI block-based paradigm with cue-induced craving. 2 conditions: "allow yourself to crave" and "resist the urge to smoke")	Successful regulation of activity in the "crave" ROIs but not the "resist" ROIs
Li et al. (2013) (51)	Nicotine-dependence (≥10 cigarettes/day)	10	Upregulation of "resist" regions Downregulation of "crave" regions (Two successive feedbacks)	"resist" regions (mPFC) and "crave" regions (ACC)	4	No	Functional localizer: (fMRI block-based paradigm with cue-induced craving. 2 conditions: "allow yourself to crave" and "resist the urge to smoke")	Successful regulation of activity in the "crave" ROIs with instructions to reduce their subjective craving to smoke but not the "resist" ROIs
Ruiz et al. (2013) (41)	Schizophrenia	9	Upregulation	Left and right anterior insula	12	No	Functional localizer (fMRI block-based paradigm: imagery to recall emotionally relevant experiences vs. rest)	Successful regulation Enhanced accuracy of the recognition of disgust faces and reduced accuracy for recognizing happy faces.
Young et al. (2014) (39)	Major depressive disorder	21 (14 vs. 7)	Upregulation during recall of positive autobiographical memories	Amygdale	3	Neurofeedback of non interest region (Intraparietal sulcus)(n=7)	MRI : defined in the stereotaxic array of Talairach and Tournoux	Successful regulation Significant pre-post scan decreases in anxiety ratings and increases in happiness ratings in the experimental vs control group
Sitaram et al. (2014) (49)	Psychopathic personality disorder	4	Upregulation	Left anterior insula	12	No	Functional localizer: mental imagery to recall emotionally relevant personal experiences vs. rest	Successful regulation in only one of the four participants (but 3 participants only completed 4 runs/12)

Table 1: studies investigating the use of fMRI-based neurofeedback in the treatment of psychiatric disorders

Investigation center	ClinicalTrials.gov Identifier	Disorder	Start date	Control Group	ROI	Masking	Expected final sample size
Yale School of Medicine	NCT02206945	Obsessive-compulsive disorder	November 2014	Sham-neurofeedback (non-interest region)	Orbitofrontal cortex	Double-blind	54
D'Or Institute for Research and Education	NCT01920490	Recurrent episodes of major depressive disorder	May 2013	Stabilize vs. increase	Anterior temporal lobe (ATL)-septal/subgenual cingulate (SCSR) coupling	Double-blind	30
Cardiff University	NCT01544205	Major depressive disorder	January 2012	Place processing network up regulation	Emotion network	Single-blind	30
University Hospital Tuebingen	NCT02148770	Eating disorders	November 2014	Sham neurofeedback	Upregulation of the functional connectivity between the ventral prefrontal medial cortex and the dorsal prefrontal cortex	Single-blind	120
Laureate Institute for Brain Research	NCT02079610	Major depressive disorder	April 2014	Sham neurofeedback (left horizontal segment of the intraparietal sulcus)	Upregulation of amygdala	Double-blind	60

Table 2: Ongoing trials investigating the use of fMRI-based neurofeedback in the treatment of psychiatric disorders, taken from Clinicaltrials.gov (January 27, 2015)

Methodological issues

Several methodological issues have emerged regarding the optimization of fMRI-NF but also concerning the best way to evaluate such protocols (see **table 3**).

Evaluation of fMRI-NF

The standard option for evaluating the potential therapeutic effect of a new treatment is a double-blind, randomized controlled trial. Double-blind protocols are relatively easy to design and implement for pharmaceutical trials but is more difficult when non-drug interventions are tested. An additional difficulty arises in the testing of psychotherapeutic tools such as neurofeedback, where active collaboration of the patient is needed. In this case, complete "blindness" cannot be achieved in the patients. This partially explains why 4 of the 6 studies identified in the present review did not use a comparative group (41,49,51,52). In one study, individuals from the "placebo group" underwent a training procedure using cognitive strategies (without neurofeedback) outside the scanner (35). Some investigators recommend using a placebo neurofeedback intervention as comparator, which is based on random or "yoked" feedback (from another participant's brain activation). However, this poses problems of frustration and patient retention. Patients receiving fake feedback may discover the non-contingency of the feedback signal, which would effectively unblind them and result in reduced engagement with the intervention. Moreover, patients in the active group may falsely assume that they are in the placebo group and give up too early. Neurofeedback from a non-interest region should thus be the "least bad" solution for a control condition. However, this option presents other challenges, such as the selection of the non-interest region. Mental disorders generally involve very complex patterns of activation, and the chosen non-interest region must therefore not be involved in these wide networks. Furthermore, the experience of gaining control over the brain (increased "self-efficacy") may be a nonspecific component of neurofeedback that contributes to improvement across disorders (21). Thus, if individuals in the control group fail to gain voluntary control of the activity of the non-interest region, their motivation will likely decline rapidly, introducing additional bias for comparison.

Testing whether brain self-regulation persists after the neurofeedback protocol also appears to be a crucial issue for future work. The "transfer session" included in many studies may provide information about the capacity of participants to self-regulate the target region(s) without feedback. However, it will be important to determine how long this ability persists after the training and if specificity exists depending on the psychiatric disorder considered. Hence, long-term evaluation of clinical modifications will have to be conducted in controlled randomized clinical trials with well-defined clinical groups.

Neurofeedback protocol

Even if numerous studies in healthy individuals lead to better comprehension of the learning processes involved in fMRI-NF protocols, optimization of the parameters of neurofeedback protocols will be a critical challenge for future studies.

In February 2012, the first international conference on rtf-MRI neurofeedback allowed the synthesis of open questions, perspectives on study design and the future outlook of fMRI-based-neurofeedback (11).

Task design

Considering task design, most of the relevant studies have used a block design, i.e., alternating periods of regulation with rest periods during neurofeedback sessions. However, the optimal duration of regulation blocks, the best number of blocks per session, and even the ideal number of sessions to maximize learning remain unknown. Indeed, the attention span and the tiredness of patients appear to be important limiting factors. It is still considerably unclear whether these parameters may undergo special arrangements in individuals suffering from psychiatric disorders. For example, subjects with schizophrenia exhibit cognitive impairments that may disturb learning processes related to the effectiveness of fMRI-NF, for which specific adaptations are required. It is furthermore still unclear whether learning to control local cerebrovascular regulation plays a role in the training of the BOLD signal which is, after all, primarily a vascular rather than a neural signal.

Instructions

The instructions given to participants are crucial in fMRI-NF protocols.

First, the subject must be informed of the hemodynamic delay in BOLD response, which may appear surprising during the first session. Indeed, due to this delay, the participant must wait five seconds to have an updated feedback signal about the potential efficacy of the chosen mental strategy. Moreover, instructions must be given to minimize behaviors causing artifacts (e.g., head motion, irregular breathing). A training session may be necessary to ensure that the subject fully understands these instructions.

The instructions given to participants related to mental imagery need further consideration: should we prefer explicit instructions (the participant is asked to use specific strategies for self-regulation) or implicit instructions (the participant is only asked to up- or down-regulate with the feedback provided)? Implicit strategies have numerous advantages: for particular regions or when a functional connectivity signal is used, the identification of specific strategies may be complex. Moreover, explicit strategies may be different for each individual and, thus, difficult to generalize and to summarize in simple instructions. However, participants can achieve voluntary control more quickly with explicit strategies in some cases. Indeed, providing a specific explicit strategy could enhance learning because participants do not spend time finding a strategy but focus on training and improving the suggested strategy. For patients suffering from psychiatric disorders, it is also a way to ensure that patients do not use pathological strategies during the run (such as increasing tension and anxiety to increase amygdala activity).

Interface

Visual feedback is the most common type of feedback, but other modalities are possible as well (e.g., auditory, tactile). Combinations of different types of feedback could also be used. Visual feedback can reflect the time-series of BOLD curves (intermittent or continuous) after various types of formatting: from a thermometer display to more complex interfaces, for example the level of brain activation can be converted into the size of a motivational cue (60). The feedback can even be implemented through virtual reality or serious games (61). The use of gaming interfaces has been proposed to improve the motivation of participants in neurofeedback protocols and, thus, increase their attention span and reduce tiredness (62). However, the effect of more complex interfaces on the learning processes involved in fMRI-NF has not yet been tested.

Definition of the fMRI-NF target

In current fMRI-NF protocols, the target brain area is localized using several methods (see **figure 3, A-B-C**).

First, the neurofeedback target can be localized using structural MRI information (e.g., (39)), based on brain atlases (e.g., using the Talairach and Tournoux coordinates) or according to macroscopic anatomical landmarks. This method is the easiest to implement but supposes a good understanding of the underlying neural mechanisms (11).

Second, the neurofeedback target can be functionally defined. In this case, the patient is asked to conduct a task within the scanner. Highlighted areas are then used as the *region(s) of interest* (ROI) for the neurofeedback protocol (e.g., (35)). Identical to what was described for the “anatomical” method, the choice in the “functional localizer” task is based on an *a priori* hypothesis that must be well validated in previous studies. A combination of anatomical and functional data can also be used (e.g., (51)).

In these two methods, the signal fed-back to the patient is generally the average BOLD response in the target area but may also be more complex (e.g., differential activity in two regions of interest (20,63), a method assumed to correct the physiological noise because common artifacts are identical in the two regions).

Finally, functional connectivity has been recently proposed as a third target definition method (64). The feedback signal is then based on the correlation of activation between the brain regions belonging to the network of interest. However, this method still needs to be tested in clinical populations.

Evaluation
What intervention should be used for the control group?
How long does the brain self-regulation persist? How long do behavioral changes persist?
Protocol
What is the ideal duration for a training session? How many sessions are needed? How long must the inter-session interval be? Are adaptations necessary depending on the psychiatric disorder targeted?
Should we prefer explicit or implicit instructions to the subject? Are adaptations necessary depending on the psychiatric disorder targeted?
What is the best interface for the feedback display? Are adaptations necessary depending on the psychiatric disorder targeted?

Table 3: Methodological issues for future research regarding the use of fMRI-based neurofeedback to treat psychiatric disorders

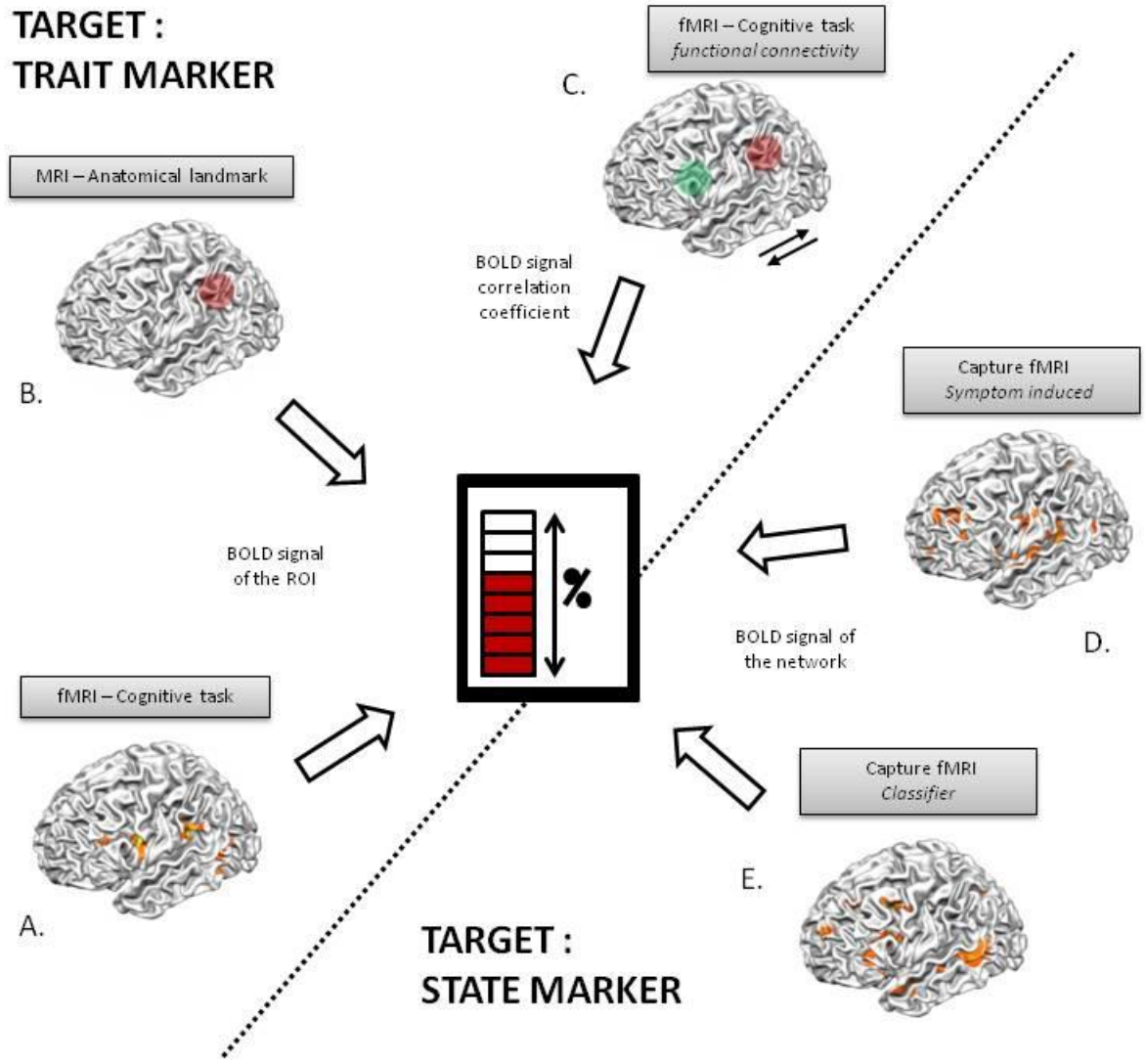


Figure 3: Different methods for localizing the fMRI-based neurofeedback target region(s)

ROI: region of interest; BOLD: blood oxygenation level-dependent

Perspectives

Clinical perspectives for future research regarding the use of fMRI-NF to treat psychiatric disorders are presented in **table 4**.

Trait-target, state-target or more?

One approach in the quest for strong *a priori* strategies for target localization relies on imaging-based "trait markers" (i.e., persistent traits or vulnerability markers that can also be detected in the pre-symptomatic and remitted phases of mental disorders). The patient will then learn to control this area's aberrant activity. Thus, this method is highly relevant for disorders with persistent (or tonic) symptoms, i.e., symptoms that do not change greatly over time. The occurrence of acute symptoms in the scanner during the neurofeedback session is not necessary. *Major depressive disorder* falls under this category. Finding stable functional imaging traits of *major depressive disorder* would greatly enhance the scope for the development of neurofeedback protocols for this disorder.

However, these strategies appear less relevant when considering acute symptoms, characterized by intrusiveness and phasic activity (such as obsessions or hallucinations). In these cases, strategies focused on "state markers" (correlates of symptomatic states) appear more appropriate to train the patient to self-regulate the activity of brain areas that re activated during symptomatic states (see **figure 3, D-E**).

Finally, neurofeedback protocols can also be developed without strong evidence for aberrant (tonic or phasic) activation in specific brain areas or circuits as long as one has a well-founded model of compensatory activation or resilience processes that need to be strengthened.

Pattern recognition and fMRI-NF

A first method for addressing state targets for fMRI-NF consists of inducing symptoms while the patient lies in the scanner (i.e., (51)). Such symptom provocation was shown to be possible for symptoms such as craving, panic attack or exposition in phobic disorders. It is necessary to assess the validity of the experimental symptom provocation using psychological and/or physiological methods (65).

Machine-learning and pattern recognition could also offer innovative strategies for state-marker identification (see **box 2**). Using fMRI classifiers, it is now possible to detect the onset of subjective symptoms together with the associated brain activation patterns (66,67). Such fine-grained activity patterns can then be used as the signal fed-back to the patient. However, to be eligible for this strategy, the patient's symptoms must exhibit some specific features. The most important criterion is frequent occurrence. Indeed, the symptom must occur several times during the fMRI session. Moreover, data analysis and patient interviews must allow the identification of "symptomatic" and "asymptomatic" periods to build an efficient classifier.

BOX 2: What is a classifier? How can pattern recognition be used for psychiatric symptoms?

The recent development of machine learning strategies such as the "*linear Support-Vector-Machine*" (lSVM) strategy could allow the extrapolation of results from neuroimaging studies exploring psychiatric disorders to everyday clinical practice (from the group level to the individual level). Indeed, this technique classifies functional or anatomical patterns using a multivariate strategy. A training session allows the optimal classifier to be built on the basis of a training dataset, for which the periods of interest (e.g., symptomatic vs. asymptomatic) have been identified and provided (68). A validation session is then needed to test the performance and possible generalization of this classifier to new data based on an independent sample. Several interesting results for diagnosis or therapeutic response prediction purposes have been published, notably in bipolar disorder (69) or schizophrenia (70). However, this is not the only way to use such tools in psychiatry. Classifiers can quickly detect the emergence of subjective symptoms by detecting specific patterns of brain activity identified during symptomatic periods.

Toward brain imaging-guided psychotherapy?

Although the clinical applications of neurofeedback may be in their infancy, many arguments encourage us to develop such non-invasive strategies (65). One of the most important features of this type of treatment is its “active” nature, which confers on the patients a central role in their care. Indeed, unlike physical treatments developed recently in which the patient remains in a passive condition (such as rTMS or TDCS), neurofeedback protocols require the active participation of the patients: neurofeedback enables the patients themselves to control their brain activity. In this way, neurofeedback could contribute to reinforcement of the experience of self-efficacy, which may be an important therapeutic factor (71). However, the “active” nature of fMRI-NF also constitutes a source of important limitations and questions. It remains unclear how motivation (and its evolution over time within the scanner during the session) may impact the efficacy of neurofeedback. In particular, there are still many questions about what arrangements must be made in patients suffering from disorders in which motivation is altered.

It appears essential to take into account the patients’ experiences during neurofeedback protocols. Moreover, developing specific scales to be applied post-session will be a major challenge in future years, which could lead to the identification of specific cognitive coping strategies in various mental disorders. Such individualized strategies could then be applied in psychotherapy, leading to the development of neuroimaging-guided programs. Even if these strategies are identified at an individual level, a rigorous evaluation could allow common strategies found in patients treated with fMRI-NF to be highlighted. If effective, these strategies could then more easily be implemented in general psychotherapy programs. This maintains the two-way relationship between conventional psychotherapy and fMRI-NF because a better understanding of the neuronal correlates of successful psychotherapy (72) could lead to the definition of new neurofeedback targets.

The major barrier in the translation of fMRI-NF protocols to clinical practice remains the accessibility and cost of the equipment. It appears essential to develop less complex devices in parallel that could be employed in second-line treatment (i.e., after a limited number of fMRI-NF sessions). Neurofeedback using signals other than BOLD signals offers one interesting possibility. EEG-based neurofeedback presents many advantages (e.g., wide availability, ambulatory use), and advances in simultaneous EEG–fMRI have made it possible to combine the two approaches (73). However, this method measures electric activity, but not a hemodynamic signal, as obtained in fMRI. *Near-infrared spectroscopy* (NIRS) measures a hemodynamic signal that could be used in neurofeedback protocols more easily after fMRI sessions (74,75).

Who might benefit from neurofeedback strategies?

An important question is related to the specific population who may benefit from fMRI-NF. The identification of predictive response markers represents a key challenge because the cognitive and motivational factors underlying the response to fMRI-NF treatment remain unknown.

Recently, Scheinost et al. showed that resting state functional connectivity predicted the fMRI-NF response in a study focusing on contamination anxiety (76). However, apart from this work, little evidence has emerged from published studies in the field. In patients suffering from *schizophrenia*, Ruiz and colleagues showed that negative symptoms were negatively correlated with the capacity of patients to learn volitional control of the anterior insula BOLD signal (41). Zotev and colleagues highlighted a relationship between the capability for self-regulation using fMRI-NF and psychological trait measures (26). In a study conducted by Sitaram, subjects with higher *Psychopathic Checklist-Revised* (PCL:SV) scores were less able to increase the BOLD signal in the anterior insula compared with their lower PCL:SV counterparts (49). Although these studies may not allow the establishment of rigorous criteria to distinguish between potential “responders” and “non-responders” to fMRI-NF, they indicate that some individuals are more likely to be able to undergo and benefit from this treatment. Clarifying this issue may lead to the definition of the optimal place for fMRI-NF within the medical toolbox: as a first intention treatment, a treatment dedicated to resistant disorders, a specific treatment for residual symptoms, or a treatment to prevent relapse.

The question of combining fMRI-NF with other treatments also appears to be a key issue, given the potential synergic effects with drugs or non-drug treatments (such as rTMS (77)). For example, in the case of nicotine craving, previous work showed that ACC is a very interesting target because of the link between the ACC activity level and craving intensity (52). A significant correlation of decreased ACC activation with fMRI-NF and reduced craving ratings was observed in Li’s work. (51) Interestingly, cigarette cue-induced ACC activation and self-reported craving are attenuated by the smoking cessation medication bupropion (78), thus leading to a potential interest in developing combined strategies.

Finally, many neuroimaging studies have shown that the genetic susceptibility to mental illness could be significantly mediated by intermediate connectivity phenotypes (79). Neuroplasticity might therefore be greatly modulated by genetics. Thus, identifying genetic susceptibility regarding the response to neurofeedback treatment could allow this treatment to be proposed to patients who are the most likely to respond. Although no study on this topic is currently available, combining genetic and imaging data to predict the response to fMRI-NF appears to be a promising strategy in this field.

Development of fMRI-NF protocols using pattern recognition to focus on the neural correlates of symptoms
Development of specific post-session scales to focus on patients' experiences within the scanner.
Development of less complex devices that could be used after a limited number of fMRI-NF sessions (e.g., EEG-based neurofeedback, NIRS-based neurofeedback)
Identification of predictive markers of the response to fMRI-NF (clinical, neuroimaging, genetics)

Table 4: Clinical perspectives for future research regarding the use of fMRI-based neurofeedback (fMRI-NF) to treat psychiatric disorders

NIRS: Near-Infrared Spectroscopy

Conclusion

fMRI-NF is a new and ambitious technique for the self-regulation of brain activity, which opens up attractive potential applications in psychiatry. Promising but preliminary results have emerged in the treatment of major depressive disorder and addiction, and several trials in depression, anxiety, addiction and eating disorders are currently under way. These trials will provide further evidence about the efficacy of fMRI-NF in psychiatry and elucidate the most promising candidates in terms of mental disorders and tailored neurofeedback protocols. Other projects would probably emerge soon (i.e. potential use of fMRI-NF in autism (80)). The next step would be the investigation of the benefit/cost balance of this methodology in pragmatic trials, which will be particularly important because of the considerable costs of the machinery involved. However, all the targeted disorders entail huge socioeconomic costs at the individual and societal level and thus further investment in research in this field seems justified.

References

1. Kessler RC, Aguilar-Gaxiola S, Alonso J, Chatterji S, Lee S, Ormel J, et al. The global burden of mental disorders: An update from the WHO World Mental Health (WMH) Surveys. Epidemiol Psichiatr Soc. 2009;18(1):23-33.
2. Hyman SE. A glimmer of light for neuropsychiatric disorders. Nature. 16 oct 2008;455(7215):890-3.
3. Linden D. Biological psychiatry: time for new paradigms. Br J Psychiatry. mars 2013;202(3):166-7.
4. Hyman SE. Revitalizing psychiatric therapeutics. Neuropsychopharmacology. janv 2014;39(1):220-9.
5. Williams NR, Okun MS. Deep brain stimulation (DBS) at the interface of neurology and psychiatry. J Clin Invest. 1 nov 2013;123(11):4546-56.
6. Demeulemeester M, Amad A, Bubrovszky M, Pins D, Thomas P, Jardri R. What is the real effect of 1-Hz repetitive transcranial magnetic stimulation on hallucinations? Controlling for publication bias in neuromodulation trials. Biol Psychiatry. 15 mars 2012;71(6):e15-6.
7. Bajbouj M, Merkl A, Schlaepfer TE, Frick C, Zobel A, Maier W, et al. Two-year outcome of vagus nerve stimulation in treatment-resistant depression. J Clin Psychopharmacol. juin 2010;30(3):273-81.
8. Agarwal SM, Shivakumar V, Bose A, Subramaniam A, Nawani H, Chhabra H, et al. Transcranial direct current stimulation in schizophrenia. Clin Psychopharmacol Neurosci. déc 2013;11(3):118-25.
9. deCharms RC. Applications of real-time fMRI. Nat Rev Neurosci. sept 2008;9(9):720-9.
10. Ruiz S, Buyukturkoglu K, Rana M, Birbaumer N, Sitaram R. Real-time fMRI brain computer interfaces: self-regulation of single brain regions to networks. Biol Psychol. janv 2014;95:4-20.
11. Sulzer J, Haller S, Scharnowski F, Weiskopf N, Birbaumer N, Blefari ML, et al. Real-time fMRI neurofeedback: progress and challenges. Neuroimage. 1 août 2013;76:386-99.
12. deCharms RC, Maeda F, Glover GH, Ludlow D, Pauly JM, Soneji D, et al. Control over brain activation and pain learned by using real-time functional MRI. Proc Natl Acad Sci U S A. 20 déc 2005;102(51):18626-31.
13. Haller S, Birbaumer N, Veit R. Real-time fMRI feedback training may improve chronic tinnitus. Eur Radiol. mars 2010;20(3):696-703.
14. Miller NE. Editorial: Biofeedback: evaluation of a new technic. N Engl J Med. 21 mars 1974;290(12):684-5.
15. Linden D. Brain Control: Developments in Therapy and Implications for Society. Palgrave Macmillan; 2014. 205 p.
16. Micoulaud-Franchi J-A, Geoffroy PA, Fond G, Lopez R, Bioulac S, Philip P. EEG neurofeedback treatments in children with ADHD: an updated meta-analysis of randomized controlled trials. Front Hum Neurosci. 2014;8:906.
17. Cohen MS. Real-time functional magnetic resonance imaging. Methods. oct 2001;25(2):201-20.
18. Boynton GM, Engel SA, Glover GH, Heeger DJ. Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci. 1 juill 1996;16(13):4207-21.
19. Micoulaud-Franchi J-A, Fakra E, Cermolacce M, Vion-Dury J. [Towards a new approach of neurophysiology in clinical psychiatry: functional magnetic resonance imaging neurofeedback applied to emotional dysfunctions]. Neurophysiol Clin. avr 2012;42(3):79-94.
20. Weiskopf N, Mathiak K, Bock SW, Scharnowski F, Veit R, Grodd W, et al. Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI). IEEE Trans Biomed Eng. juin 2004;51(6):966-70.
21. Linden DEJ. Neurofeedback and networks of depression. Dialogues Clin Neurosci. mars 2014;16(1):103-12.
22. Wolpaw J, Wolpaw EW, éditeurs. Brain-Computer Interfaces: Principles and Practice. OUP USA; 2012. 424 p.
23. Burns A, Adeli H, Buford JA. Brain-Computer Interface after Nervous System Injury. Neuroscientist. déc 2014;20(6):639-51.
24. Posse S, Fitzgerald D, Gao K, Habel U, Rosenberg D, Moore GJ, et al. Real-time fMRI of temporolimbic regions detects amygdala activation during single-trial self-induced sadness. Neuroimage. mars 2003;18(3):760-8.
25. Caria A, Veit R, Sitaram R, Lotze M, Weiskopf N, Grodd W, et al. Regulation of anterior insular cortex activity using real-time fMRI. Neuroimage. 15 avr 2007;35(3):1238-46.
26. Zotev V, Krueger F, Phillips R, Alvarez RP, Simmons WK, Bellgowan P, et al. Self-regulation of amygdala activation using real-time FMRI neurofeedback. PLoS ONE. 2011;6(9):e24522.
27. Lee J-H, Ryu J, Jolesz FA, Cho Z-H, Yoo S-S. Brain-machine interface via real-time fMRI: preliminary study on thought-controlled robotic arm. Neurosci Lett. 23 janv 2009;450(1):1-6.
28. Yoo S-S, Lee J-H, O'Leary H, Panych LP, Jolesz FA. Neurofeedback fMRI-mediated learning and consolidation of regional brain activation during motor imagery. Int J Imaging Syst Technol. 13 juin 2008;18(1):69-78.
29. Rota G, Sitaram R, Veit R, Erb M, Weiskopf N, Dogil G, et al. Self-regulation of regional cortical activity using real-time fMRI: the right inferior frontal gyrus and linguistic processing. Hum Brain Mapp. mai 2009;30(5):1605-14.
30. Birbaumer N, Ruiz S, Sitaram R. Learned regulation of brain metabolism. Trends Cogn Sci (Regul Ed). juin 2013;17(6):295-302.
31. Caria A, Sitaram R, Veit R, Begliomini C, Birbaumer N. Volitional control of anterior insula activity modulates the response to aversive stimuli. A real-time functional magnetic resonance imaging study. Biol Psychiatry. 1 sept 2010;68(5):425-32.
32. Jardri R, Thomas P, Delmaire C, Delion P, Pins D. The neurodynamic organization of modality-dependent hallucinations. Cereb Cortex. mai 2013;23(5):1108-17.
33. Kim S, Birbaumer N. Real-time functional MRI neurofeedback: a tool for psychiatry. Curr Opin Psychiatry. sept 2014;27(5):332-6.
34. Nierenberg AA, Husain MM, Trivedi MH, Fava M, Warden D, Wisniewski SR, et al. Residual symptoms after remission of major depressive disorder with citalopram and risk of relapse: a STAR*D report. Psychol Med. janv 2010;40(1):41-50.
35. Linden DEJ, Habes I, Johnston SJ, Linden S, Tatineni R, Subramanian L, et al. Real-time self-regulation of emotion networks in patients with depression. PLoS ONE. 2012;7(6):e38115.

36. Johnston S, Linden DEJ, Healy D, Goebel R, Habes I, Boehm SG. Upregulation of emotion areas through neurofeedback with a focus on positive mood. Cogn Affect Behav Neurosci. mars 2011;11(1):44-51.
37. Johnston SJ, Boehm SG, Healy D, Goebel R, Linden DEJ. Neurofeedback: A promising tool for the self-regulation of emotion networks. Neuroimage. 1 janv 2010;49(1):1066-72.
38. Johnston S, Linden DEJ, Healy D, Goebel R, Habes I, Boehm SG. Upregulation of emotion areas through neurofeedback with a focus on positive mood. Cogn Affect Behav Neurosci. mars 2011;11(1):44-51.
39. Young KD, Zotev V, Phillips R, Misaki M, Yuan H, Drevets WC, et al. Real-time FMRI neurofeedback training of amygdala activity in patients with major depressive disorder. PLoS ONE. 2014;9(2):e88785.
40. Green S, Lambon Ralph MA, Moll J, Deakin JFW, Zahn R. Guilt-selective functional disconnection of anterior temporal and subgenual cortices in major depressive disorder. Arch Gen Psychiatry. oct 2012;69(10):1014-21.
41. Ruiz S, Lee S, Soekadar SR, Caria A, Veit R, Kircher T, et al. Acquired self-control of insula cortex modulates emotion recognition and brain network connectivity in schizophrenia. Hum Brain Mapp. janv 2013;34(1):200-12.
42. Lenzenweger MF. Current status of the scientific study of the personality disorders: an overview of epidemiological, longitudinal, experimental psychopathology, and neurobehavioral perspectives. J Am Psychoanal Assoc. août 2010;58(4):741-78.
43. Beckwith H, Moran PF, Reilly J. Personality disorder prevalence in psychiatric outpatients: a systematic literature review. Personal Ment Health. mai 2014;8(2):91-101.
44. Skodol AE, Gunderson JG, Shea MT, McGlashan TH, Morey LC, Sanislow CA, et al. The Collaborative Longitudinal Personality Disorders Study (CLPS): overview and implications. J Pers Disord. oct 2005;19(5):487-504.
45. Jiang W, Liu H, Liao J, Ma X, Rong P, Tang Y, et al. A functional MRI study of deception among offenders with antisocial personality disorders. Neuroscience. 6 août 2013;244:90-8.
46. Amad A, Ramoz N, Thomas P, Jardri R, Gorwood P. Genetics of borderline personality disorder: systematic review and proposal of an integrative model. Neurosci Biobehav Rev. mars 2014;40:6-19.
47. Morey LC, Hopwood CJ. Stability and change in personality disorders. Annu Rev Clin Psychol. 2013;9:499-528.
48. Roberts BW, Walton KE, Viechtbauer W. Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. Psychol Bull. janv 2006;132(1):1-25.
49. Sitaram R, Caria A, Veit R, Gaber T, Ruiz S, Birbaumer N. Volitional control of the anterior insula in criminal psychopaths using real-time fMRI neurofeedback: a pilot study. Front Behav Neurosci. 2014;8:344.
50. Birbaumer N, Veit R, Lotze M, Erb M, Hermann C, Grodd W, et al. Deficient fear conditioning in psychopathy: a functional magnetic resonance imaging study. Arch Gen Psychiatry. juill 2005;62(7):799-805.
51. Li X, Hartwell KJ, Borckardt J, Prisciandaro JJ, Saladin ME, Morgan PS, et al. Volitional reduction of anterior cingulate cortex activity produces decreased cue craving in smoking cessation: a preliminary real-time fMRI study. Addict Biol. juill 2013;18(4):739-48.
52. Hanlon CA, Hartwell KJ, Canterberry M, Li X, Owens M, Lematty T, et al. Reduction of cue-induced craving through realtime neurofeedback in nicotine users: the role of region of interest selection and multiple visits. Psychiatry Res. 30 juill 2013;213(1):79-81.
53. Hartwell KJ, Johnson KA, Li X, Myrick H, LeMatty T, George MS, et al. Neural correlates of craving and resisting craving for tobacco in nicotine dependent smokers. Addict Biol. oct 2011;16(4):654-66.
54. Scheinost D, Stoica T, Saksa J, Papademetris X, Constable RT, Pittenger C, et al. Orbitofrontal cortex neurofeedback produces lasting changes in contamination anxiety and resting-state connectivity. Transl Psychiatry. 2013;3:e250.
55. Hampson M, Stoica T, Saksa J, Scheinost D, Qiu M, Bhawnani J, et al. Real-time fMRI biofeedback targeting the orbitofrontal cortex for contamination anxiety. J Vis Exp. 2012;(59).
56. Mataix-Cols D, Wooderson S, Lawrence N, Brammer MJ, Speckens A, Phillips ML. Distinct neural correlates of washing, checking, and hoarding symptom dimensions in obsessive-compulsive disorder. Arch Gen Psychiatry. juin 2004;61(6):564-76.
57. Bartholdy S, Musiat P, Campbell IC, Schmidt U. The potential of neurofeedback in the treatment of eating disorders: a review of the literature. Eur Eat Disord Rev. nov 2013;21(6):456-63.
58. Kaye WH, Fudge JL, Paulus M. New insights into symptoms and neurocircuit function of anorexia nervosa. Nat Rev Neurosci. août 2009;10(8):573-84.
59. Frank S, Lee S, Preissl H, Schultes B, Birbaumer N, Veit R. The Obese Brain Athlete: Self-Regulation of the Anterior Insula in Adiposity. PLoS ONE. 8 août 2012;7(8):e42570.
60. Sokunbi MO, Linden DEJ, Habes I, Johnston S, Ihssen N. Real-time fMRI brain-computer interface: development of a « motivational feedback » subsystem for the regulation of visual cue reactivity. Front Behav Neurosci. 2014;8:392.
61. Goebel R, Sorger B, Kaiser J, Birbaumer N, Weiskopf N. BOLD brain pong: self-regulation of local brain activity during synchronously scanned, interacting subjects. 34th Annual Meeting of the Society for Neuroscience. 2004.
62. Friedrich EVC, Suttie N, Sivanathan A, Lim T, Louchart S, Pineda JA. Brain-computer interface game applications for combined neurofeedback and biofeedback treatment for children on the autism spectrum. Front Neuroeng. 2014;7:21.
63. Chiew M, LaConte SM, Graham SJ. Investigation of fMRI neurofeedback of differential primary motor cortex activity using kinesthetic motor imagery. Neuroimage. 15 mai 2012;61(1):21-31.
64. Ruiz S, Rana M, Sass K, Kircher T, Birbaumer N, Sitaram R. Brain network connectivity and behaviour enhancement: a fMRI-BCI study. 17th Annual Meeting of the Organization for Human Brain Mapping [Internet]. 2011 [cité 9 janv 2015]. Disponible sur: http://www.bme.ufl.edu/labs/wp-content/blogs.dir/12/files/2013/10/3.Brain-Network-Connectivity-and-Behaviour-Enhancement.pdf
65. Linden DEJ. Brain imaging and psychotherapy: methodological considerations and practical implications. Eur Arch Psychiatry Clin Neurosci. nov 2008;258 Suppl 5:71-5.
66. LaConte SM. Decoding fMRI brain states in real-time. Neuroimage. 15 mai 2011;56(2):440-54.
67. Sitaram R, Lee S, Ruiz S, Rana M, Veit R, Birbaumer N. Real-time support vector classification and feedback of multiple emotional brain states. Neuroimage. 15 mai 2011;56(2):753-65.

68. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev. avr 2012;36(4):1140-52.
69. Ellison-Wright I, Bullmore E. Anatomy of bipolar disorder and schizophrenia: a meta-analysis. Schizophr Res. mars 2010;117(1):1-12.
70. Smieskova R, Fusar-Poli P, Allen P, Bendfeldt K, Stieglitz RD, Drewe J, et al. Neuroimaging predictors of transition to psychosis--a systematic review and meta-analysis. Neurosci Biobehav Rev. juill 2010;34(8):1207-22.
71. Bandura A. Self-Efficacy: The Exercise of Control. Worth Publishers; 1997. 628 p.
72. Linden DEJ. How psychotherapy changes the brain--the contribution of functional neuroimaging. Mol Psychiatry. juin 2006;11(6):528-38.
73. Zotev V, Phillips R, Yuan H, Misaki M, Bodurka J. Self-regulation of human brain activity using simultaneous real-time fMRI and EEG neurofeedback. Neuroimage. 15 janv 2014;85 Pt 3:985-95.
74. Kober SE, Wood G, Kurzmann J, Friedrich EVC, Stangl M, Wippel T, et al. Near-infrared spectroscopy based neurofeedback training increases specific motor imagery related cortical activation compared to sham feedback. Biol Psychol. janv 2014;95:21-30.
75. Mihara M, Miyai I, Hattori N, Hatakenaka M, Yagura H, Kawano T, et al. Neurofeedback using real-time near-infrared spectroscopy enhances motor imagery related cortical activation. PLoS ONE. 2012;7(3):e32234.
76. Scheinost D, Stoica T, Wasylink S, Gruner P, Saksa J, Pittenger C, et al. Resting state functional connectivity predicts neurofeedback response. Front Behav Neurosci. 2014;8:338.
77. Micoulaud Franchi J-A, Fond G, Dumas G. Cyborg psychiatry to ensure agency and autonomy in mental disorders. A proposal for neuromodulation therapeutics. Front Hum Neurosci. 2013;7:463.
78. Brody AL, Mandelkern MA, Lee G, Smith E, Sadeghi M, Saxena S, et al. Attenuation of cue-induced cigarette craving and anterior cingulate cortex activation in bupropion-treated smokers: a preliminary study. Psychiatry Res. 30 avr 2004;130(3):269-81.
79. Tost H, Bilek E, Meyer-Lindenberg A. Brain connectivity in psychiatric imaging genetics. Neuroimage. 1 oct 2012;62(4):2250-60.
80. Caria A, de Falco S. Anterior insular cortex regulation in autism spectrum disorders. Front Behav Neurosci. 2015;9:38.

3.3 Le neurofeedback dans la prise en charge des hallucinations auditives

Comme nous l'avons vu en introduction, les corrélats neuronaux des hallucinations auditives chez les sujets souffrant de schizophrénie sont aujourd'hui beaucoup mieux appréhendés. Cependant, ces avancées n'ont à ce jour pas été accompagnées d'une « révolution » en terme de moyens thérapeutiques. Dans ces conditions, le neurofeedback offre une perspective extrêmement intéressante qui pourrait permettre une translation entre recherche fondamentale et application clinique (235).

3.3.1 Neurofeedback guidé par EEG

De nombreux travaux ont pu établir que les patients souffrant de schizophrénie sont capables, comme les sujets sans pathologie, d'apprendre à réguler certains signaux d'activité cérébrale qui leur sont renvoyés par neurofeedback que ce soit en EEG (236–240) ou en IRMf (241–243). Ceci a permis le développement d'un certain nombre de protocoles à visée thérapeutique dans le champ de la schizophrénie. Citons par exemple les résultats prometteurs d'une étude récemment publiée montrant une amélioration des performances cognitives (vitesse de traitement de l'information, attention, mémoire de travail, apprentissage visuel et verbal) après entrainement par neurofeedback (10 sessions) chez les patients souffrant de schizophrénie dans un essai contrôlé (8 patients et 12 sujets contrôles) (244). D'autres essais sont actuellement en cours, par exemple celui mené par Fiza Singh à San Diego visant à tester l'effet d'un entrainement de la bande EEG gamma sur la mémoire de travail (enregistrement NCT03260257 sur la base *Clinical Trials*).

Toutefois, très peu d'études ont rigoureusement testé des protocoles de neurofeedback guidé par EEG pour la prise en charge spécifique des hallucinations. Plusieurs pistes ont été évoquées (235), par exemple, l'entraînement de la bande gamma (pour lequel un effet sur les fonctions mnésiques, notamment mémoire de source impliquée dans la physiopathologie des hallucinations auditives (245), a été mis en évidence chez les sujets sans trouble psychiatrique (246)). Certains auteurs ont également pu évoquer le rôle de la composante N100 en potentiels évoqués auditifs qui est réduite chez les patients souffrant de schizophrénie notamment au cours des hallucinations auditives, pour proposer des protocoles de neurofeedback basés sur un entrainement visant à augmenter cette composante (247). En effet, cette réduction pourrait résulter d'une activation dysfonctionnelle des aires auditives primaires par le discours interne aboutissant à une diminution de la réactivité aux stimuli auditifs externes. L'entrainement vise donc à augmenter la réactivité du cortex auditif primaire à ces stimuli auditifs externes. Cependant, cette hypothèse n'a, à l'heure actuelle pas été validée chez des patients, seule

une étude de faisabilité chez les sujets sans trouble psychiatrique a été réalisée (247). Enfin, un entrainement par neurofeedback visant à augmenter la présence des micro-états de type D en EEG a également été proposé (248). En effet, le modèle des micro-états décrit les signaux EEG par des suites de topographies associées à des états cérébraux demeurant stables durant quelques dizaines de millisecondes (249). Des anomalies au niveau des micro-états de type D apparaissent corrélées aux symptômes positifs dans la schizophrénie (250). Cependant, comme pour le protocole précédent, seule une étude de faisabilité a été, à ce jour, réalisée (248).

En conclusion, bien que des pistes soient à l'étude, il n'existe actuellement aucun protocole de neurofeedback guidé par EEG validé pour la prise en charge des hallucinations auditives dans la schizophrénie.

3.3.2 Neurofeedback guidé par IRMf

Les nombreux résultats prometteurs pour la prise en charge des symptômes psychiatriques (voir **Article 9**) encouragent le développement de protocoles de neurofeedback guidés par IRMf pour le traitement des hallucinations auditives. Pour cela, plusieurs stratégies sont possibles.

L'**Article 10** réalisé en collaboration avec plusieurs équipes travaillant sur ce sujet, propose une synthèse des différentes stratégies possibles pour le développement de protocoles de neurofeedback guidé par IRMf dans la prise en charge des hallucinations auditives. Il constitue également le rationnel sur lequel s'appuie le choix de notre stratégie : une stratégie basée sur la capture en temps réel des patterns d'activation associés à l'hallucination auditive (voir 2. Détection automatisée des hallucinations auditives en IRMf). La stratégie n°3 exposée dans cet article, s'appuyant notamment sur les méthodes de détection automatique des hallucinations (**Article 2**) et de la périodes pré-hallucinatoires (**Article 3**), a récemment été soumise à l'appel à projet de l'Agence Nationale de la Recherche. Ce projet a été positivement évalué et financé (projet INTRUDE, ANR-16-CE37-0015, porté par le Professeur Renaud Jardri).

ARTICLE 10

Translating Neurocognitive Models of Auditory-Verbal Hallucinations into Therapy:

Using real-time fMRI-neurofeedback to treat voices

Thomas FOVET[1], Natasza ORLOV[2], Miriam DYCK[3], Paul ALLEN[2,4], Klaus MATHIAK[3],

Renaud JARDRI[1].

1. Univ Lille, CNRS, UMR-9193, SCA-Lab, psyCHIC team & CHU Lille, Psychiatry Dpt (CURE), Fontan hospital, F-59000 Lille, France

2. Department of Psychosis Studies, Institute of Psychiatry Psychology & Neuroscience, King's College London, United Kingdom

3. Department of Psychiatry, Psychotherapy and Psychosomatics, JARA-Brain, RWTH Aachen University, Aachen, Germany

4. Department of Psychology, University of Roehampton, London, United Kingdom

Frontiers in Psychiatry. 2016; 7: 103.

ABSTRACT

Auditory-verbal hallucinations (AVHs) are frequent and disabling symptoms, which can be refractory to conventional psychopharmacological treatment in more than 25% of the cases. Recent advances in brain imaging allow for a better understanding of the neural underpinnings of AVHs. These findings strengthened transdiagnostic neurocognitive models that characterize these frequent and disabling experiences. At the same time, technical improvements in real-time functional magnetic resonance imaging (fMRI) enabled the development of innovative and non-invasive methods with the potential to relieve psychiatric symptoms, such as fMRI-based neurofeedback (fMRI-NF). During fMRI-NF, brain activity is measured and fed-back in real time to the participant in order to help subjects to progressively achieve voluntary control over their own neural activity. Precisely defining the target brain area/network(s) appears critical in fMRI-NF protocols. After reviewing the available neurocognitive models for AVHs, we elaborate on how recent findings in the field may help to develop strong *a priori* strategies for fMRI-NF target localization. The first approach relies on imaging-based "trait markers" (i.e., persistent traits or vulnerability markers that can also be detected in the pre-symptomatic and remitted phases of AVHs). The goal of such strategies is to target areas that show aberrant activations during AVHs or are known to be involved in compensatory activation (or resilience processes). Brain regions, from which the NF signal is derived, can be based on structural MRI and neurocognitive knowledge, or functional MRI information collected during specific cognitive tasks. Because hallucinations are acute and intrusive symptoms, a second strategy focuses more on "state markers". In this case, the signal of interest relies on fMRI capture of the neural networks exhibiting increased activity during AVHs occurrences, by means of multivariate pattern recognition methods. The fine-grained activity patterns concomitant to hallucinations can then be fed-back to the patients for therapeutic purpose. Considering the potential cost necessary to implement fMRI-NF, proof-of-concept studies are urgently required to define the optimal strategy for application in patients with AVHs. This technique has the potential to establish a new brain imaging-guided psychotherapy for patients that do not respond to conventional treatments, and take functional neuroimaging to therapeutic applications.

INTRODUCTION

Auditory-verbal hallucinations (AVHs), i.e., hearing voices in the absence of appropriate external stimuli, are frequent experiences in schizophrenia, with a lifetime prevalence of 60 to 80% (1,2). AVHs are often strongly disabling symptoms, which can be refractory to conventional psychopharmacological treatment in more than 25% of the cases (3). A recent meta-analysis supports the effectiveness of cognitive-behavioural therapy (CBT) in the treatment of AVHs (4). However, in the specific case of treatment-refractory symptoms, CBT seems to have modest and only short term benefits (5,6).

In recent years, the number of brain imaging studies in the field of AVHs has grown substantially, leading to a better understanding of this subjective phenomenon (7,8). Recent progress in deciphering the neural underpinnings of AVHs has strengthened transdiagnostic neurocognitive models that characterize AVHs, but more specifically these findings built the bases for new therapeutic strategies. Indeed, brain imaging now allows for the identification of therapeutic targets by determining the brain regions involved in the occurrence of AVHs. For example, based on findings implicating the left temporoparietal cortex in AVHs, *repetitive Transcranial Magnetic Stimulation* (rTMS), a non-invasive brain stimulation method, has been used to target this region and shown to have a significant, although moderate, effect in alleviating drug-resistant AVHs (9).

Recently, technical improvements in real-time fMRI has enabled the development of fMRI-based neurofeedback (fMRI-NF) (10). During fMRI-NF, brain activity is measured in real-time and fed-back to the participant, usually using visual or auditory information, in order to facilitate voluntary control over his own neural activity. Considering the advances in the identification of anatomical and functional changes linked with AVHs, fMRI-NF strategies constitute a promising tool, giving the possibility for patients to normalize their brain activity level or connectivity strength in the AVHs-specific brain regions, and thus reduce symptom severity. Precisely defining the target brain area/network(s) appears crucial for future fMRI-NF protocols designed to treat AVHs.

After briefly reviewing current literature about the neural basis of AVHs (mainly neurocognitive models and brain imaging findings) and providing an overview of how fMRI-NF can be used in psychiatry, the review will then elaborate on how recent advances in the field may help to develop strong *a priori* strategies for fMRI-NF target localization. Three different fMRI-NF strategies dedicated to AVHs' treatment will be proposed. Current limits, potential difficulties for patients with schizophrenia to benefit from fMRI-NF, as well as future directions will be critically discussed.

WHAT IS fMRI-NEUROFEEDBACK?

fMRI-neurofeedback: principles

Neurofeedback is a non-invasive technique enabling participants to achieve voluntary control over the neuronal activity of one or more brain regions (for a recent review on the technique, see (11)). In the case of fMRI-NF, this is accomplished by deriving and presenting Blood Oxygen Level Dependent (BOLD) signal derived from the target brain area(s) to the subject in real time (12). Visual feedback is primarily used, but neurofeedback derived from other or combination of different modalities is also

possible. Visual feedback can be presented in various formats: from a thermometer display to more complex interfaces (e.g. social feedback (13)). The participants use this feedback to self-regulate their neuronal response or adjust their cognitive strategy, during the experimental task in real time (see **Figure 1a**). They must be informed in detail(s) of the hemodynamic delay of four or five seconds (due to the BOLD response) to update the neurofeedback signal. The general experimental design of an fMRI-NF protocol is described in **Figure 1b**. This technique is currently being used in cognitive modification (14) and clinical trials (15).

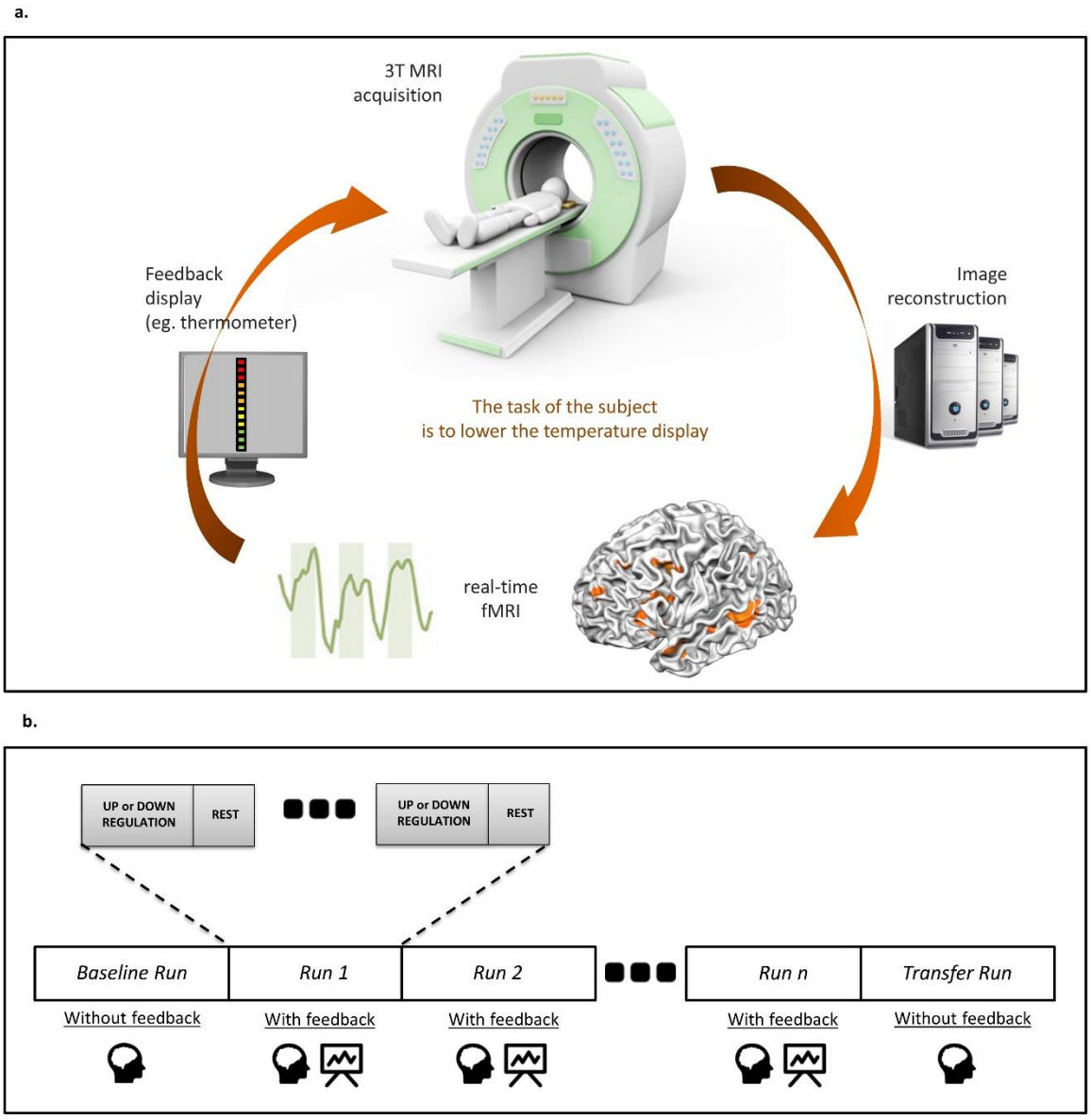


Figure 1: the principles of fMRI-neurofeedback.

Figure 1a: diagram of an fMRI-based neurofeedback system.

Figure 1b: the neurofeedback training in fMRI-neurofeedback protocols.

fMRI-neurofeedback in psychiatry

fMRI-NF could be a useful tool in psychiatry. Numerous studies have shown the benefits of fMRI-NF to relieve non-psychiatric clinical symptoms. Haller et al. demonstrated therapeutic effects of fMRI-NF (focusing on down–regulation of auditory cortex) in the treatment of chronic tinnitus (16). DeCharms et al. also published promising results for the management of chronic pain (17), although these results failed to be replicated (12).

Furthermore, recent progress in the field of brain imaging has allowed the identification of functional changes associated with a range of psychiatric symptoms (18). Because fMRI-NF can potentially be used to normalize the activity level of specific brain regions (which should be a key issue in new treatments), fMRI-NF could offer a new interesting way to treat mental health symptoms (19). To date, promising positive results have been already demonstrated in major depressive disorder (15,20) and addiction (21–23).

Why using fMRI-Neurofeedback for auditory-verbal hallucinations in schizophrenia?

In a paper published in 2012, Simon McCarthy-Jones stressed the potential interests of developing neurofeedback as a new treatment of AVHs (24). In the past decade, significant progress in identifying the neural underpinning of AVHs has been made. This knowledge can inform on a target region for fMRI-NF.

To date, only a few studies reporting the use of neurofeedback in patients with schizophrenia have been published. Most of these studies used electroencephalogram (EEG) based-neurofeedback (the principle is the same as fMRI-NF but the brain activity is measured with EEG) (e.g. (25–27) with one current study running a trial for the treatment of AVHs (28)). Two studies have used fMRI-NF in patients with schizophrenia. Ruiz et al. demonstrated that patients with schizophrenia (n=9), were able to achieve voluntary control of bilateral anterior insula cortex using an fMRI-NF protocol (29). Participants completed 4 training sessions spread over 2 weeks. Each training session comprised 3 runs of self-regulation training. Each run consisted of 6 upregulation and 7 baseline blocks (30 seconds blocks). Patients were instructed that the recall of emotionally relevant past experiences combined with the feedback could enable them to control the thermometer bars. No specific emotional cues or recall strategies were given. The gain in the voluntary control was associated with behavioural changes assessed on a facial emotion recognition task (i.e. patients recognized disgust faces more accurately and happy faces less accurately after the fMRI-NF training). Furthermore, the training was associated with an increase in the number of the incoming and outgoing effective connections in the anterior insula. This proof-of-concept study demonstrates that patients with schizophrenia can benefit from fMRI-NF and learn volitional brain regulation but also that such learning is accompanied with behavioural changes and neurophysiological changes in the underlying brain network (29). More recently, Cordes et al. showed that patients with schizophrenia (n=11) were also able to learn to control the activity of their anterior cingulate cortex (ACC) (30). Here, 3 fMRI-NF training sessions were completed in one week. Each session included 3 runs consisting of 8 regulation and 9 baseline blocks lasting 30 seconds each. During the fMRI-NF session, the participants were asked to upregulate the signal using individual mental

strategies. However, some template strategies from different cognitive domains were given: positive autobiographic memories, picturing oneself doing sports or playing an instrument, and concentration on given perceptions like feeling the temperature of one's own left foot. The results demonstrated that both patients with schizophrenia and healthy controls were able to develop control abilities. However, they used different neural strategies: patients activated more the dorsal and healthy control the rostral subdivision of ACC. They also used different mental strategies: patients mainly imagined of music whereas healthy controls used more imagined sports.

In summary, evidence suggests that patients with schizophrenia are able to learn voluntary control over their brain in spite of their pathology. All of this makes the fMRI-NF a promising tool to tackle frequent and disabling symptoms in this population, such as AVHs.

WHAT DO WE KNOW ABOUT THE NEURAL BASIS OF AUDITORY-VERBAL HALLUCINATIONS?

Neurocognitive models

Phonologically, AVHs are heterogeneous in form and content (31). They vary from acousmas (primitive sounds such as blowing, shooting), utterances or simple words, to full conversations, with defined characteristics such a pitch, volume and accent. They might consist of a single voice or a collection of voices that speak the individual's thought aloud, issuing commands and instructions, or provide a running commentary on the person's behaviour. The voices might be familiar or unknown (32). They often carry power, authority (33) and a negative quality (e.g. (2,34)) and persons experiencing them often feel that they have no or little control over their AVHs (2).

From a neurocognitive perspective, hallucinations are erroneous perceptions or sensory deceptions without the presence of external stimuli, and have been attributed to erroneous integration of sensory and cognitive processes (35) that may influence conscious perception (36). Brain regions that have been implicated in the experience of AVHs include the auditory cortex and the ventral attentional system that spontaneously orientates attention toward an incoming stimulus (37,38).

A number of neurocognitive models have been proposed to account for heterogenic phenomenology of AVHs (39,40). The current models are based on research findings that illustrate the following contributing factors to the experience of AVHs. These are: AVHs have clear perceptual qualities, AVHs are internally generated but are not attributed to an internal source, those experiencing AVHs have a reduced sense of control over the onset, content and frequency of AVHs, and AVHs carry often an emotional component.

Externalisation, or lack of agency was explained by a model proposed by Frith (41) which postulated the breakdown in a physiological process known as self-monitoring. This model is based on the assumption that in patients with schizophrenia, inner-speech and/thoughts fail to be recognised as self-generated due to a self-monitoring deficit; reflecting a dysfunction of the efference copy or corollary discharge mechanism that accompanies a motor action such as speech or movement (42,43).

In those experiencing hallucinations the efference copy of inner-speech does not produce a corollary discharge of the expected experience. Consequently, this failure in the corollary discharge

mechanism can produce confusion regarding the agency between one's own thoughts, and externally generated voice, potentially resulting in an external attribution of the experience and the experience of AVHs. At a neuronal level, this may result in greater activity in the auditory cortex when self-generated speech or inner speech is produced (43,44).

At a behavioural level, it has been shown that patients with schizophrenia and AVHs exhibit difficulty in identifying self-generated information (45–47). However, models based on the misattribution of inner-speech do easily account for observed phenomenology of AVH (48,49) and there is no evidence that the cancellation or suppression of reafference indicates the source of a sensory event: zero signal is not the same as self-generation (50).

Another early model postulates a deficit in source-monitoring or reality testing (51). Source-monitoring is a meta-cognitive (thinking about thinking) process that enables us to make attributions as to origins of beliefs and thoughts in order to form a cohesive representation of an experience (51). Bentall et al. suggested that patients with schizophrenia have deficits in discriminating between external (real) and internal (imagined) events, accompanied with as specific externalisation bias. For example is it has been demonstrated that patients with schizophrenia and AVHs were more prone to misattribute self-generated items to other sources (52). Further, the experience of AVHs has also been related to deficits in reality testing. Based on signal detection theory (SDT), it was suggested that patients with schizophrenia and AVHs show a shift in the decision criterion (the point at which a person decides they perceive a stimulus) (53). SDT proposes that detection of a stimulus is based on two premises: perceptual sensitivity - the general efficiency of the perceptual system, and response bias – the subjective decision criteria to deciding that a perceived event is a stimulus. For example, patients with schizophrenia and AVHs demonstrate higher perceptual sensitivity to detecting words or sounds embedded in white noise, as compared to non-hallucinating patients, but lower sensitivity compared to healthy controls (54). Further, patients with current AVHs also demonstrate a response bias, i.e. indicated that they were certain that a stimulus was presented, even when it was absent, suggesting that perception/signal detection is unimpaired in patients with AVHs, but there is uncertainty in the signal recognition. This uncertainty, accompanied by a misattribution bias and source/reality monitoring deficits, perpetuates the attribution of thoughts to an external source. This may result in perceptual hypervigilance (55,56) in responding to biases and lead to (strong) consolidation of such responses with time (57).

Substantial evidence supports the link between AVH and self-, source and reality monitoring and has been provided over the last two decades (57,58). Nonetheless, these early models alone cannot account for the presence of AVHs, as they fail to account for certain aspects of their phenomenology. AVHs are often experienced in the second and third person, they may consist of multiple voices which are not the voice of the experiencer, and the experiencers often converse with the AVHs (46,50).

More recently, a number of models have been developed in order to incorporate the complex phenomenology of AVHs, by integrating the available neurophysiological data and adapting the predictive processing framework (PPF). For example Allen et al. (35) proposed a neuroanatomical model founded upon a network of brain areas involved in both cognitive and perceptual processing; suggesting that hyperactivation of perceptual regions including the primary and secondary auditory

cortices evident during AVHs (38,50,59,60), and in related speech and language areas (44,59,61). Wilkinson et al. on the other hand (62), adopted the PPF (e.g. (63)) to account for the phenomenology of AVH. In the framework of PPF, neuronal systems have evolved to predict statistical regularities in the environment based on prior experiences (64). Through successfully encoding predictions in an accurate manner, they minimize prediction errors or deviations from these predictions, and these are seen as the neural systems demonstrating an attenuated response to these predictable events; permitting the serial updating of prediction to create a picture of the external world. This creates a dynamic internal model that can impact on neuronal activity in sensory systems, increasing activity to unpredicted events through a failure of this predictive mechanism, with consequent alterations in subjective perception and elaboration into delusional belief formation (65).

Finally, these recent models suggest a number of cortical and sub-cortical brain networks involved in the experience of AVH and that verbal hallucinations involve hyper activity in secondary and primary auditory cortex, accompanied by disrupted coupling with the cognitive processes associated with monitoring/reality testing.

Neuroimaging studies

Structural Brain Imaging

Structural imaging studies (i.e. studies investigating the brain morphology) have identified subtle but robust reductions in the grey matter volume (GMV) in patients with AVHs, particularly in areas involved in speech and language. Altered GMV in the superior temporal gyrus (STG) has been highlighted by both priori-defined region-of-interest (ROI) analyses (66) and voxel-based morphometry studies (67). Modinos et al. demonstrated that AVHs severity was significantly associated with GMV reduction in the left STG, including the Heschl's gyrus. Structural changes have also been identified in Broca's area and its homotopic contralateral area (68) and the primary auditory cortex (Heschl's gyrus) (69). In addition to reductions in GMV in language regions, numerous studies have reported modifications in other brain areas such as temporal and frontal regions (70), insular cortex (71,72), thalamus (73) and cerebellum (74).

In addition to these quantitative analyses, structural imaging also provides complementary qualitative measures of the cortical morphology, such as the shape of sulci and gyri (75). Indeed, gyrification is considered an indirect marker of brain development since cortical folding (i.e., gyrification and sulcation) begins in the tenth week of gestation and stabilizes by the end of the third trimester of pregnancy. The resulting complex sulcal/gyral patterns are then stable over life (76). Studying changes in cortical morphology associated with AVHs provides an novel way to assess the impact of developmental factors on this symptom (77). Significant reductions in the gyrification of language-related areas (e.g., the superior temporal ridges, the left middle frontal sulcus, Broca's area) have been identified in chronic schizophrenia patients with AVHs when compared with healthy controls (78). The phenomenology of AVHs has also been associated with morphological changes within the

language network. Indeed, the spatial location of AVHs (as internal or external percepts) have been associated with specific sucal deviations in the right temporoparietal junction (79).

Functional Brain Imaging

Functional brain imaging studies in patients with AVHs have provided information about the neural bases of the susceptibility to hallucinate (trait studies), and neural activation that is seen during AVHs (state studies).

Trait studies

Trait studies measure brain activity during specific tasks in patients who hallucinate and those who don't. Inquiring afterward for the absence of AVHs while scanning is necessary to avoid any "state" factor to interfere with this type of paradigm.

Trait studies have revealed altered functional activity in the temporal lobes of patients with AVHs (8,80). Altered activation is thought to emerge from a competition between AVHs and normal external speech for processing sites within the temporal cortex (81). Similarly designed studies have identified a decrease in the functional activity of the rostral dorsal anterior cingulate cortex, a structure known to be involved in the allocation of an internal or external origin for a given stimulus (82,83). These results are compatible with the misattribution models of AVHs (see **Neurocognitive models**) but also with recent structural data (84).

State studies

Functional brain imaging suggests that a distributed network of brain regions underlies the experience of AVHs (85). Speech production and comprehension areas have been shown to be involved, but in addition to this network, brain areas involved in contextual memory seem to play a role in AVHs. This was notably revealed by a coordinate-based meta-analysis of AVHs capture studies, which demonstrated increased activity in Broca's and Wernicke's areas, and also in the hippocampal complex (85), suggesting that hallucinations could result from the aberrant activation of memory traces within associative cortices (86,87). Although it is still a subject of debate, the activation of the primary auditory cortex does not appear to be necessary for the occurrence of AVHs. Nonetheless, its activation could be related to specific phenomenological aspects of the hallucinatory experience, such as the feeling of reality (88).

Connectivity studies

Brain connectivity can be studied using three different approaches: functional connectivity, effective connectivity and structural connectivity (89). Functional connectivity relies on correlation measures between spatially distant brain areas without information on the directionality or causality of the interaction. In contrast, effective connectivity explores the direct influence of one brain region on another, and thus provides information regarding the causal relationship between brain areas in a given network. Finally, structural connectivity is the measure of white matter tracts connecting different brain

regions, based on diffusion MRI and tractography algorithms. Many connectivity studies have confirmed the dysconnectivity hypothesis in schizophrenia patients and in particular those who report hallucinations. Indeed, abnormal connectivity between brain regions has been shown at rest (for review see (90,91)) and during verbal tasks by functional and effective connectivity studies (92,93). This dysconnectivity appears to play a major role in the emergence of hallucinations, but was also found to change according to the sensory-modality involved (94,95). Diffusion MRI studies comparing patients with schizophrenia who experience hallucinations, non-hallucinating patients with schizophrenia and healthy controls have found differences in the coherence of the white-matter bundles connecting language areas (69,96,97). This finding was particularly noteworthy in the arcuate fasciculus (98).

WHAT STRATEGY TO RELIEVE AUDITORY-VERBAL HALLUCINATIONS WITH fMRI-NEUROFEEDBACK?

In this section, we propose three different fMRI-NF strategies dedicated to AVHs' treatment on which our teams are currently working on (see **Figure 2)**. We focus on the localization of the target and the type of feedback used for each strategy.

Strategy 1: A priori target localized using structural MRI

Method used to localize the fMRI-neurofeedback target

During fMRI-NF protocols, the brain region(s) from which the NF signal is derived can be informed anatomically using structural MRI data and brain atlases (e.g., Talairach and Tournoux coordinates), or according to macroscopic anatomical landmarks. This method is the easiest to implement methodologically but assumes a good understanding of the underlying neural mechanisms and their anatomical location. The goal is to regulate neural activity in areas that show aberrant activations during AVHs (e.g. Broca's and Wernicke's areas) or to regulate activity in regions thought to be involved in compensatory or resilience processes (e.g. ACC). Below, we present an fMRI-NF protocol targeting the ACC (see **Figure 2 strategy 1**).

Why choose ACC as a target for fMRI-neurofeedback to relieve AVHs?

Disrupted connectivity between the temporal and cingulate cortices has been demonstrated in schizophrenia, with AVHs severity correlating with the connectivity strength between the ACC and the STG (99,100).

The ACC has a key role in regulating emotions, goal-directed behaviors, attentional processes, response selection, online source monitoring and cognitive control (101,102). Moreover, the ACC is involved in differentiating between self- and non-self related stimuli (83,103). Furthermore, a meta-analysis of trait studies conducted in patients with AVHs and healthy controls revealed decreased ACC activity in hallucinators (104). This finding is in line with cognitive models of AVHs (30,88).

A number of studies have demonstrated that the ACC can be reliably regulated using fMRI-based NF (13,105–109). Moreover, the successful up-regulation of the rostral ACC was associated with an increase in positive affect (106,109) and improved emotional perception of voices in healthy subjects (106).

Even though the theoretical accounts differ in the different studies (see **Neurocognitive models**), they all assume a failure of typical ACC functions. The monitoring of inner speech processes, the monitoring of retrieval processes and error detection, as well as the suppression of task-irrelevant stimuli are all classical ACC functions that should be fostered by an up-regulation of the ACC (110,111). Two previous studies however report increased ACC activation during hallucinations (59,112). It is possible that increased ACC activation may be related to default-mode fluctuations, considering simultaneous deactivations of auditory cortex and Wernicke's area in the former study and resting state activations without baseline subtraction in the latter study.

The fMRI-neurofeedback protocol

First, an anatomically predefined ACC mask is applied. From this ROI, the average signal is fed-back on a thermometer-like display after filtering and artefact reduction. A custom anatomical template mask of the ACC defines the ROI (details in (113)). This ACC mask is taken as a part of the cingulate cortex excluding parts inferior or posterior to the anterior fissure. The feedback signal is the average BOLD signal across this ACC mask for each volume with 1% representing the full scale. A custom toolbox conducts online processing comprising motion correction and co-registration to a template (114). Kalman filter reduces singular values and high-frequency components. An exponential moving average algorithm removes temporal drifts.

The patient performs three fMRI-NF training runs, each consisting of eight regulation blocks and nine baseline blocks (30 seconds each; see exemplary run in **Figure 2**). Increase of ACC signal makes a green bar moving up- and decrease downwards (see (30)). A fixed red bar in the regulation condition serves as a regulation target. It indicates the upper limit of ACC up-regulation. The baseline condition is indicated by a blue line display. Mental strategies should be tried to move the green line upwards to the red line. During the baseline blocks, the patient counts backwards from 100. Every repetition time (TR; 1 s), the display is updated. The NF procedure is explained to subjects, including the delay of the NF signal for 3-5 s due to the hemodynamic response and data processing (< 1 s).

Preliminary data

Patients with schizophrenia can learn to regulate the ACC to a comparable level than healthy controls, albeit involving different networks and cognitive strategies (30). Moreover, a recent article involving three schizophrenia patients suggests that even with ongoing AVHs, patients are able to learn ACC regulation (113). In this work, patients seemed to be very interested in the methodology and were eager to learn. Since the target groups were patients with longstanding symptoms, a core preposition was a good patient-therapist relationship and only limited impairments in cognitive functions. RWTH

Aachen University is just performing a clinical trial study investigating the effect of fMRI-NF training in schizophrenia patients with ongoing AVHs.

Previous fMRI-NF studies have demonstrated that up-regulation of a single area can elicit alterations of functional and effective connectivity (115,116). Further studies may elucidate whether ACC up-regulation also induces changes of the network dynamics. In the long-term, it may be even more effective if fMRI-NF could target several regions aiming to regulate the functional connectivity between these regions. This would allow fMRI-NF to address the neural dysconnectivity that is proposed to underpin AVHs (117). This approach would also enable regulation of connectivity and activity with the salience network, also proposed to be dysfunctional in people with AVHs (118–120).

Strategy 2: Region of interest defined using a functional localizer

Method used to localize the fMRI-neurofeedback target

The target chosen for fMRI-NF can also be functionally defined. In this case, the patient is asked to undertake a functional task within the scanner and activated areas are then used as the ROI(s) for fMRI-NF. The choice in the "functional localizer" task should be based on an *a priori* hypothesis that is well validated in previous studies.

Why use a functional localizer for fMRI-neurofeedback to relieve AVHs?

As already mentioned, in schizophrenia, both state and trait brain imaging studies have revealed aberrant neural activation in patients with AVHs. Resting-state or 'non-task' studies suggest that several speech-related areas are linked with such experiences, as well as the anterior cingulate cortex (ACC) and the hippocampal complex (see **The neural basis of AVHs**). Similarly, task-related paradigms have identified frontotemporal dysconnectivity in patients with AVHs, specifically between the left superior temporal gyrus and the dorsal ACC (100) and the medial prefrontal cortex (121), regions thought to be involved in self-other source monitoring. Disruption of these mechanisms is consistent with cognitive models which postulate aberrant bottom-up and top-down processes in AVHs.

Any of these regions could potentially be defined as a target to create a ROI mask for fMRI-NF. However, rather than using a structural or anatomically defined mask, a functional localiser task can be used to define the ROI (122). The choice of an appropriate task for the functional localizer should be informed by previous imaging studies, i.e. studies consistently discriminating the target ROI from other brain activity.

For example, two meta-analyses of AVHs in schizophrenia demonstrated that the human voice sensitive region of the left and right superior temporal gyrus (STG) is associated with the experience of AVHs (67,104). Therefore, this region could serve as a potential ROI mask (see **Figure 2 strategy 2**). The functional localiser task would need to be designed to specifically identify the human voice responsive auditory cortex (i.e. the task reported in (123)). This could be obtained by running of blocks of words (activation) and non-word speech analogues (baseline).

The fMRI-neurofeedback protocol

After completing data acquisition, the effective signal change measured within the functional localiser tasks is analysed with univariate fMRI methods, such as the General Linear Model. The difference between the average BOLD signal of the activation block and the baseline block should be used to create the ROI mask. Several programs offer tools for online analysis, e.g. the AFNI software (http://afni.nimh.nih.gov/afni/). Here, the mask is created by eyeballing the resulting 3D cluster and manually specifying the statistical thresholds until a cluster of the required size/shape is present in the target ROIs. Ideally, the cluster size choice should be informed by previous meta-analytic studies. The mask should also include a control region to serve the averaging out of non-specific brain activation. A randomised controlled trial should also include a control group utilising a control ROI mask, and each participant should complete both the target and control ROI localiser tasks, in spite of group assignment.

A new mask ROI can be created during each scan, or a retrospective method can be used, whereby the mask obtained during the first visit is used during subsequent neurofeedback trainings. The retrospective method requires the alignment of the different time-series data obtained from different scans. Some MRI scanners allow the re-alignment of previously obtained data with the current images. However, if this option is not available, most online analysis software have inbuilt algorithms that allow to realigning images obtained during different scanning sessions. The advantage of the retrospective method is the reduction of scanning time and therefore participant discomfort as well as global costs. In addition, the ROI mask does not change shape or size.

In terms of the neurofeedback training, this procedure remains the same as during anatomically masked ROI real time-fMRI, i.e. feedback is provided during the entire training run but remains static during rest (no-regulation blocks). Similarly, participants need to be informed about the inherent delay in feedback due to the hemodynamic response and adhere to standardized instructions. To enhance motivation and the likelihood of successful signal down-regulations, participants are instructed to devise their own strategy to down-regulate their signal (29,124).

Strategy 3: pattern recognition using a multivariate classifier

Method used to localize the fMRI-NF target

The two previous strategies rely on imaging-based “trait markers” (i.e., persistent traits or vulnerability markers that can also be detected in the pre-symptomatic and remitted phases of mental disorders). The patient is trained to gain control of areas known to be involved in the AVHs’ pathophysiology. When using such a methodology, the occurrence of hallucinations in the scanner during neurofeedback sessions is not necessary.

However, because hallucinations are acute symptoms, notably characterized by intrusiveness and phasic activity, they can also be targeted with a different type of strategy based on “state markers” (i.e. which correlate with symptomatic states). Here, the objective is to train the subject to self-regulate the activity of brain areas that re-activate during symptomatic states. Machine-learning, and particularly

the recent development of "linear Support-Vector-Machine" (lSVM), offers several advantages in this context. Indeed, this technique classifies functional or anatomical patterns using a multivariate strategy. A training session allows the optimal classifier to be built on the basis of a training dataset, for which the periods of interest (e.g., symptomatic vs. asymptomatic) have been identified and provided (125). A validation session is then needed to test the performance and possible generalization of this classifier to new data based on an independent sample. Several interesting results for diagnosis or therapeutic response prediction purposes have been published, notably in bipolar disorder (126) or schizophrenia (127). However, this is not the only way to use such tools in psychiatry. Classifiers can quickly detect the emergence of subjective symptoms by detecting specific patterns of brain activity identified during symptomatic periods (see **Figure 2 strategy 3**).

Why use classifiers for fMRI-NF to relieve AVHs?

Using fMRI classifiers, it is now possible to detect the onset of subjective symptoms together with the associated brain activation patterns (128,129). For example, our group developed such a classifier to detect AVHs occurrence while scanning a patient with a 71% accuracy (130). This algorithm is currently under optimization and already reaches 80% accuracy. Even if no data is currently available on the use of this kind of classifier in fMRI-NF protocols, the fine-grained activity patterns obtained could theoretically be used as the signal fed-back to the patient. Future studies should allow to specify the minimal necessary accuracy.

However, to be eligible for this strategy, the patient's hallucinations must exhibit some specific features. The most important criterion is frequent occurrence. Indeed, the symptom must occur several times during the fMRI session. Moreover, data analysis and patient interviews must allow the identification of "symptomatic" and "asymptomatic" periods to build an efficient classifier. In our case, we chose to build a subject-independent classifier based on the AVHs presence or absence, determined with the methodology described in (88). This strategy presents substantial benefits compared with a subject dependent pattern classification of fMRI signals, notably a considerable time saving (131).

The fMRI-neurofeedback protocol

Unlike the two methods described above, this strategy does not imply a block paradigm. Indeed, the visual feedback provide an information in real time about the current state of the participant (hallucinating or not) all along the session. The visual feedback may be a thermometer whose signal intensity is based on the level of activation in the regions of interest (given by the discriminative maps of the classifier). But other possibilities emerged from recent work on AVHs. Our team recently proposed a method to distinguish between the different periods in the occurrence of AVHs (120). Even if we are at a very preliminary stage, this could theoretically allow for the implementation of a multi-classifier strategy with the possibility to discriminate multiple "brain-states" as, in our case, (i) "No hallucination" ("Off" period on **Figure 2**), (ii) "Transition" ("Trans" period on **Figure 2**; i.e., period immediately preceding the AVHs occurrence) (iii) "Hallucination" ("On" period on **Figure 2**) (iv) "End" ("End" period on **Figure 2**; i.e., period immediately following the AVHs occurrence). This technique can provide a feedback

indicating which “brain-state” is identified. For example, (as presented in **Figure 2**), a four-part diagram presenting the four brain-states can be used. If the “hallucination” period or “transition” period is identified, he must adapt his mental strategy to go back to “end” or “no hallucination” periods. This kind of feedback could also be combined with a thermometer display (to provide both a continuous and a discrete variable to the subject).

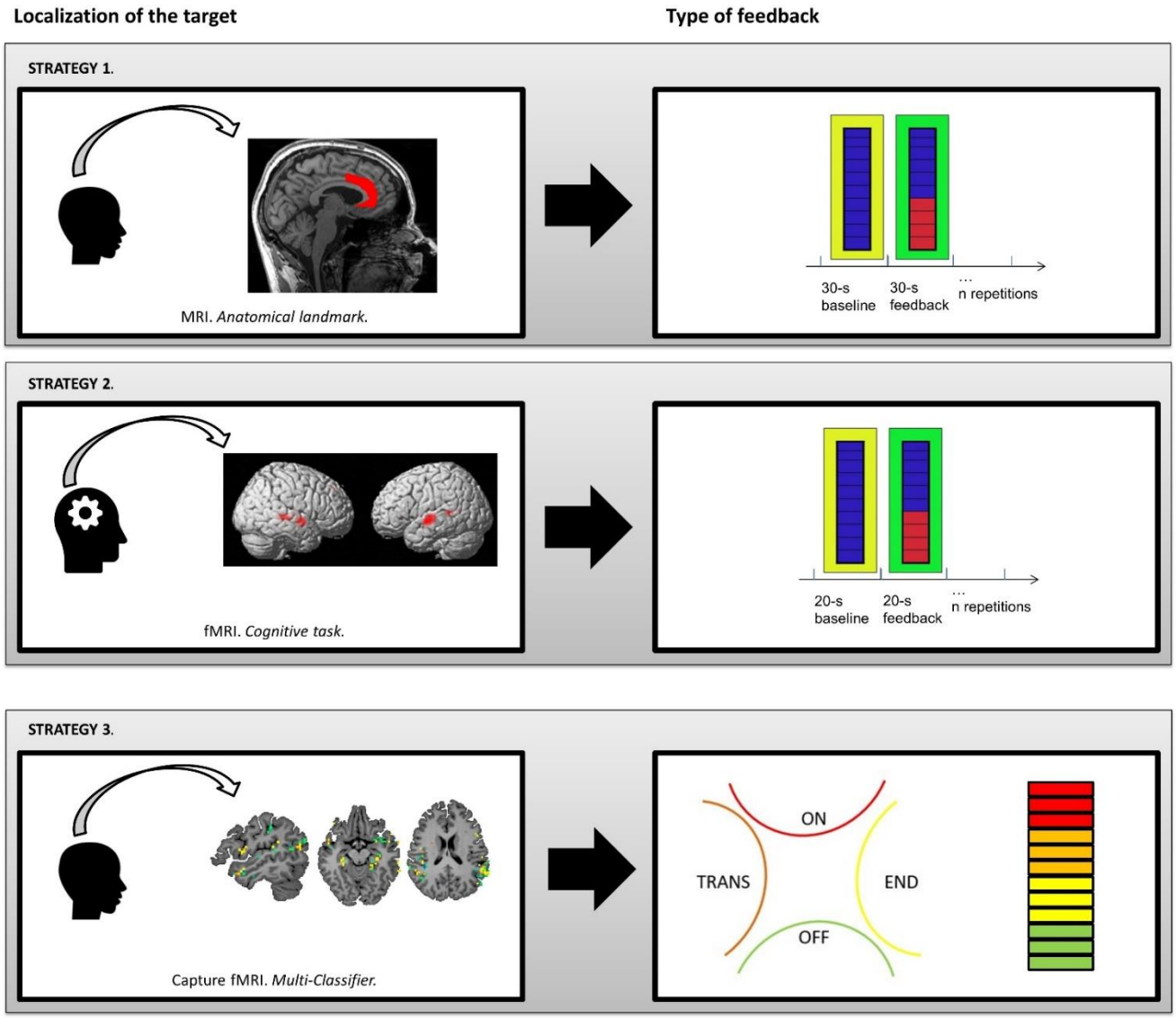


Figure 2: Presentation of 3 different strategies for fMRI-neurofeedback protocols to relieve AVHs.
Strategy 1: co-registration of anatomical template of anterior cingulate cortex.
Strategy 2: human voice responsive auditory cortex identified with a functional localizer.
Strategy 3: linear *Support Vector Machine discriminative maps of a classifier after recursive feature elimination steps, which is able to detect neural patterns associated with hallucinations in resting-state brain activity with a level of accuracy of 71%.*

LIMITS AND FUTURE DIRECTIONS

fMRI-neurofeedback experimental designs

The most obvious limitation of the available studies testing fMRI-NF protocols are their small sample sizes, making generalization difficult. For AVHs, no study assessing the efficacy of fMRI-NF is currently available. Nevertheless, the improved understanding of the neural underpinnings of AVHs seen in recent years, and the preliminary results presented here should inform future studies.

The gold-standard to assess new treatments is the double-blind, randomized controlled trial design. However, a major issue in fMRI-NF protocols is to achieve complete “blindness” in patients, because an active collaboration is needed during the sessions. This directly questions what could be an ideal control condition? Four kind of control conditions have been described in the literature (11): (i) mental task outside of the scanner, (ii) sham feedback using brain signal of interest from previous participant, (iii) sham feedback using inverse brain signal of interest, (iv) sham feedback using brain signal from an unrelated region. The first solution appears unsatisfactory because patients in the control group are not exposed to fMRI-NF. Using a brain signal of interest from previous participants may generate frustration and retention since participants may unravel the non-contingency of the feedback, which would unblind them and reduce their engagement with the intervention. Moreover, for patients with severe AVHs, this kind of feedback could increase anxiety, letting them think that they have no control on their neural activity. Using an inverse brain signal of interest is unethical in the specific case of AVHs treatment. Indeed, this kind of sham feedback aims to test if inverse brain modulation prompts opposite behavioral changes. As a consequence, the expected change would be a worsening of AVHs symptomatology. Neurofeedback from a non-interest region should be the “least bad” solution for a control condition in fMRI-NF protocols to treat AVHs. The selection of a non-interest region appears crucial here, and could be a difficult challenge given the complexity (and spread) of the brain networks involved in AVHs (unfortunately, no data are currently available on the potential non-interest ROI that could be used for protocol testing the efficiency of fMRI-NF in AVHs).

Another significant challenge to adequately assess neurofeedback effectiveness is to develop dedicated post-session scales, able to identify the specific cognitive coping strategies used by the patients during the session. Such individualized strategies could then be applied in psychotherapy, potentially leading to the development of neuroimaging-guided programs. A rigorous evaluation of the strategies used to cope with AVHs during the fMRI-NF sessions could then be helpful to optimize general hallucination-focused psychotherapy programs. We believe that this may constitute an interesting two-way relationship between conventional psychotherapy and fMRI-NF: fMRI-NF is a precious tool to optimize hallucination-focused psychotherapy programs while the identification of brain changes after psychotherapy allows for the identification of new neurofeedback-targets.

Finally, testing whether brain self-regulation persists after the fMRI-NF protocols is a crucial issue. The “transfer session” (see **Figure 1**) may provide information about the capacity of participants to self-regulate the target region(s) without feedback. Furthermore, it will be very important to determine how long this capacity persists after the fMRI-NF and how long the clinical improvement is maintained. To date, no formal follow-ups of symptoms were conducted with the patients. The question of the potential long-term effects of these treatments is clearly under-assessed in *fMRI-Brain Computer Interface* research in general (116) and no data are currently available for patients with schizophrenia.

fMRI-neurofeedback protocols

In addition to the non-invasive nature of fMRI-NF, one of its prominent features is to put the patient at the heart of the process. On the one hand, the active participation of the patient in fMRI-NF may contribute to the reinforcement of their feeling self-efficacy (which constitutes an important therapeutic factor (132)). On the other hand, this active nature may be source of limitations in schizophrenia patients with strong negative symptoms, who may lack motivation. Although some data seem to indicate that patients suffering from schizophrenia are able to achieve voluntary control of their own brain activity during fMRI-NF (29,30), these results need to be confirmed in studies with larger samples. Given the importance of motivation in neurofeedback protocols, it seems very relevant to consider factors interfering with reward processing such as negative symptoms and antipsychotic medication. The effort required by patients to undergo the fMRI-NF training should not be underestimated as well as the mixed motivation of the patients, since there can often exists some positive aspects to the hallucinatory experiences, which the patients may fear losing as a result of the training. Future research will have to determine what are the best experimental settings/ instructions for patients suffering from refractory AVHs together with severe negative symptoms.

Considering task design, the most of fMRI-NF studies use a block design (i.e., alternating periods of up or down regulation with rest periods during neurofeedback runs, see **Figure** 1). However, the optimal number of blocks per run, the ideal duration of regulation blocks, and the best number of sessions to obtain a maximal efficacy in the treatment of AVHs remain unknown. Future research should determine if patients suffering from schizophrenia (particularly those who exhibit severe cognitive impairment) may benefit from special arrangements in fMRI-NF protocols to minimize the attention span and the tiredness.

Informal reassessments during clinical visits that were conducted during the pilot study of the currently ongoing study with schizophrenia patients with AVHs (see above) revealed that none of the patients reported adverse events, and two of the patients claimed to have developed different strategies in dealing with their AVHs up to few weeks after the training (113). However, during contact and assessment occurring more than a month after the training, none of the patient had the impression that fMRI-NF training had had any influence on their symptoms. Individual variability and fluctuation in the disease course may override the – so far rather small – effects of the fMRI-NF training. This may change with better targeted fMRI-NF protocols. However, based on clinical impressions, we would suggest that at least monthly booster session would be advisable for clinical trials.

From a methodological point of view, uncertainly lies also about the instructions to be given before the session. It remains unknown if explicit (the participant is asked to use specific mental strategies for self-regulation) or implicit (the participant is only asked to up- or down-regulate with the feedback provided) instructions should be preferred. Implicit instructions are ideal in general population because they favor the development of individualized strategies to achieve voluntary control of the target region(s). However, the identification of an optimal strategy may be difficult for patients with severe AVHs which could lead to a rapid decline in motivation. That is why providing specific explicit strategies could be useful to enhance the efficacy of fMRI-NF to treat AVHs. Strategies inspired from CBT could allow the participant for achieving voluntary control more quickly.

Finally, considering the definition of the fMRI-NF target, many other neural networks may serve as target for the fMRI-NF training. Interestingly, one of the most robust effects of fMRI-NF training seems to be changes in connectivity (e.g. (115,133). Indeed, the first fMRI-NF studies attempt to train network connectivity directly (134,135). Considering the importance of network function on the AVHs phenotype, connectivity fMRI-NF will be one of the next targets for treatment approaches to AVHs.

CONCLUSION

Although a number of studies is currently investigating the efficacy of fMRI-NF for AVH, as for today, efficiency data from randomized controlled trials is lacking (11). In this paper, we focused on specific fMRI-NF strategies to treat AVHs and selected three of them that appear most feasible, emphasizing the need for preliminary studies. Indeed, considering the potential cost necessary to implement fMRI-NF, proof-of-concept studies are urgently required to define the optimal strategy for application in patients with AVHs. This technique has the potential to establish a new brain imaging-guided psychotherapy for patients that do not respond to conventional treatments, and take functional neuroimaging to therapeutic applications.

ACKNOWLEDGEMENTS

Thomas Fovet is partially supported by a doctoral fellowship from the *Pierre Deniker Foundation*.

REFERENCES

1. Andreasen NC, Flaum M. Schizophrenia: the characteristic symptoms. *Schizophr Bull* (1991) **17**:27–49.
2. Nayani TH, David AS. The auditory hallucination: a phenomenological survey. *Psychol Med* (1996) **26**:177–189.
3. Shergill SS, Murray RM, McGuire PK. Auditory hallucinations: a review of psychological treatments. *Schizophr Res* (1998) **32**:137–150.
4. van der Gaag M, Valmaggia LR, Smit F. The effects of individually tailored formulation-based cognitive behavioural therapy in auditory hallucinations and delusions: a meta-analysis. *Schizophr Res* (2014) **156**:30–37. doi:10.1016/j.schres.2014.03.016
5. Valmaggia LR, van der Gaag M, Tarrier N, Pijnenborg M, Slooff CJ. Cognitive-behavioural therapy for refractory psychotic symptoms of schizophrenia resistant to atypical antipsychotic medication. Randomised controlled trial. *Br J Psychiatry J Ment Sci* (2005) **186**:324–330. doi:10.1192/bjp.186.4.324
6. Durham RC, Guthrie M, Morton RV, Reid DA, Treliving LR, Fowler D, Macdonald RR. Tayside-Fife clinical trial of cognitive-behavioural therapy for medication-resistant psychotic symptoms. Results to 3-month follow-up. *Br J Psychiatry J Ment Sci* (2003) **182**:303–311.
7. Hugdahl K. Auditory hallucinations: A review of the ERC “VOICE” project. *World J Psychiatry* (2015) **5**:193–209. doi:10.5498/wjp.v5.i2.193
8. Allen P, Modinos G, Hubl D, Shields G, Cachia A, Jardri R, Thomas P, Woodward T, Shotbolt P, Plaze M, et al. Neuroimaging auditory hallucinations in schizophrenia: from neuroanatomy to neurochemistry and beyond. *Schizophr Bull* (2012) **38**:695–703. doi:10.1093/schbul/sbs066
9. Demeulemeester M, Amad A, Bubrovszky M, Pins D, Thomas P, Jardri R. What is the real effect of 1-Hz repetitive transcranial magnetic stimulation on hallucinations? Controlling for publication bias in neuromodulation trials. *Biol Psychiatry* (2012) **71**:e15-16. doi:10.1016/j.biopsych.2011.10.010
10. Fovet T, Jardri R, Linden D. Current Issues in the Use of fMRI-Based Neurofeedback to Relieve Psychiatric Symptoms. *Curr Pharm Des* (2015) **21**:3384–3394.
11. Thibault RT, Lifshitz M, Raz A. The self-regulating brain and neurofeedback: Experimental science and clinical promise. *Cortex J Devoted Study Nerv Syst Behav* (2015) **74**:247–261. doi:10.1016/j.cortex.2015.10.024
12. Sulzer J, Haller S, Scharnowski F, Weiskopf N, Birbaumer N, Blefari ML, Bruehl AB, Cohen LG, DeCharms RC, Gassert R, et al. Real-time fMRI neurofeedback: progress and challenges. *NeuroImage* (2013) **76**:386–399. doi:10.1016/j.neuroimage.2013.03.033
13. Mathiak KA, Koush Y, Dyck M, Gaber TJ, Alawi E, Zepf FD, Zvyagintsev M, Mathiak K. Social reinforcement can regulate localized brain activity. *Eur Arch Psychiatry Clin Neurosci* (2010) **260 Suppl**:S132-6. doi:10.1007/s00406-010-0135-9
14. Ruiz S, Buyukturkoglu K, Rana M, Birbaumer N, Sitaram R. Real-time fMRI brain computer interfaces: self-regulation of single brain regions to networks. *Biol Psychol* (2014) **95**:4–20. doi:10.1016/j.biopsycho.2013.04.010
15. Linden DEJ, Habes I, Johnston SJ, Linden S, Tatineni R, Subramanian L, Sorger B, Healy D, Goebel R. Real-time self-regulation of emotion networks in patients with depression. *PloS One* (2012) **7**:e38115. doi:10.1371/journal.pone.0038115
16. Haller S, Birbaumer N, Veit R. Real-time fMRI feedback training may improve chronic tinnitus. *Eur Radiol* (2010) **20**:696–703. doi:10.1007/s00330-009-1595-z
17. deCharms RC, Maeda F, Glover GH, Ludlow D, Pauly JM, Soneji D, Gabrieli JDE, Mackey SC. Control over brain activation and pain learned by using real-time functional MRI. *Proc Natl Acad Sci U S A* (2005) **102**:18626–18631. doi:10.1073/pnas.0505210102
18. Amad A, Fovet T, Geoffroy PA. Keep calm: Psychiatric disorders are organic! The power of words in medicine. *Aust N Z J Psychiatry* (2016) **50**:100–101. doi:10.1177/0004867415601730
19. Kim S, Birbaumer N. Real-time functional MRI neurofeedback: a tool for psychiatry. *Curr Opin Psychiatry* (2014) **27**:332–336. doi:10.1097/YCO.0000000000000087
20. Linden DEJ. Neurofeedback and networks of depression. *Dialogues Clin Neurosci* (2014) **16**:103–112.
21. Hanlon CA, Hartwell KJ, Canterberry M, Li X, Owens M, Lematty T, Prisciandaro JJ, Borckardt J, Brady KT, George MS. Reduction of cue-induced craving through realtime neurofeedback in nicotine users: the role of region of interest selection and multiple visits. *Psychiatry Res* (2013) **213**:79–81. doi:10.1016/j.pscychresns.2013.03.003
22. Karch S, Keeser D, Hümmer S, Paolini M, Kirsch V, Karali T, Kupka M, Rauchmann B-S, Chrobok A, Blautzik J, et al. Modulation of Craving Related Brain Responses Using Real-Time fMRI in Patients with Alcohol Use Disorder. *PloS One* (2015) **10**:e0133034. doi:10.1371/journal.pone.0133034
23. Kirsch M, Gruber I, Ruf M, Kiefer F, Kirsch P. Real-time functional magnetic resonance imaging neurofeedback can reduce striatal cue-reactivity to alcohol stimuli. *Addict Biol* (2015) doi:10.1111/adb.12278
24. McCarthy-Jones S. Taking back the brain: could neurofeedback training be effective for relieving distressing auditory verbal hallucinations in patients with schizophrenia? *Schizophr Bull* (2012) **38**:678–682. doi:10.1093/schbul/sbs006
25. Bolea AS. Neurofeedback Treatment of Chronic Inpatient Schizophrenia. *J Neurother* (2010) **14**:47–54. doi:10.1080/10874200903543971
26. Gruzelier J, Hardman E, Wild J, Zaman R. Learned control of slow potential interhemispheric asymmetry in schizophrenia. *Int J Psychophysiol Off J Int Organ Psychophysiol* (1999) **34**:341–348.
27. Surmeli T, Ertem A, Eralp E, Kos IH. Schizophrenia and the efficacy of qEEG-guided neurofeedback treatment: a clinical case series. *Clin EEG Neurosci* (2012) **43**:133–144. doi:10.1177/1550059411429531
28. ISRCTN - ISRCTN23930628: Neurofeedback treatment of auditory verbal hallucinations. Available at: http://www.isrctn.com/ISRCTN23930628?q=neurofeedback%20hallucinations&filters=&sort=&offset=1&totalResults=1&page=1&pageSize=10&searchType=basic-search [Accessed January 5, 2016]
29. Ruiz S, Lee S, Soekadar SR, Caria A, Veit R, Kircher T, Birbaumer N, Sitaram R. Acquired self-control of insula cortex modulates emotion recognition and brain network connectivity in schizophrenia. *Hum Brain Mapp* (2013) **34**:200–212. doi:10.1002/hbm.21427

30. Cordes JS, Mathiak KA, Dyck M, Alawi EM, Gaber TJ, Zepf FD, Klasen M, Zvyagintsev M, Gur RC, Mathiak K. Cognitive and neural strategies during control of the anterior cingulate cortex by fMRI neurofeedback in patients with schizophrenia. *Front Behav Neurosci* (2015) **9**:169. doi:10.3389/fnbeh.2015.00169
31. Aleman A, Larøi F. *Hallucinations: The science of idiosyncratic perception*. American Psychological Association (2008).
32. Larøi F, Sommer IE, Blom JD, Fernyhough C, Ffytche DH, Hugdahl K, Johns LC, McCarthy-Jones S, Preti A, Raballo A, et al. The characteristic features of auditory verbal hallucinations in clinical and nonclinical groups: state-of-the-art overview and future directions. *Schizophr Bull* (2012) **38**:724–733. doi:10.1093/schbul/sbs061
33. Birchwood M, Chadwick P. The omnipotence of voices: testing the validity of a cognitive model. *Psychol Med* (1997) **27**:1345–1353.
34. Johns LC, Hemsley D, Kuipers E. A comparison of auditory hallucinations in a psychiatric and non-psychiatric group. *Br J Clin Psychol Br Psychol Soc* (2002) **41**:81–86.
35. Allen P, Larøi F, McGuire PK, Aleman A. The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. *Neurosci Biobehav Rev* (2008) **32**:175–191.
36. Aleman A, Böcker KBE, Hijman R, de Haan EHF, Kahn RS. Cognitive basis of hallucinations in schizophrenia: role of top-down information processing. *Schizophr Res* (2003) **64**:175–185.
37. Belin P, Fecteau S, Bédard C. Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci* (2004) **8**:129–135. doi:10.1016/j.tics.2004.01.008
38. Hunter MD, Eickhoff SB, Miller TWR, Farrow TFD, Wilkinson ID, Woodruff PWR. Neural activity in speech-sensitive auditory cortex during silence. *Proc Natl Acad Sci U S A* (2006) **103**:189–194. doi:10.1073/pnas.0506268103
39. McCarthy-Jones S, Thomas N, Strauss C, Dodgson G, Jones N, Woods A, Brewin CR, Hayward M, Stephane M, Barton J, et al. Better than mermaids and stray dogs? Subtyping auditory verbal hallucinations and its implications for research and practice. *Schizophr Bull* (2014) **40 Suppl 4**:S275-284. doi:10.1093/schbul/sbu018
40. Belzeaux R, Cermolacce M, Jardri R. Hallucinations: toward a dialogue between phenomenology and brain imaging research. *J Conscious Stud* (In press)
41. Frith CD. *The Cognitive Neuropsychology of Schizophrenia*. Psychology Press (1992).
42. Feinberg I. Efference copy and corollary discharge: implications for thinking and its disorders. *Schizophr Bull* (1978) **4**:636–640.
43. Ford JM, Mathalon DH. Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *Int J Psychophysiol Off J Int Organ Psychophysiol* (2005) **58**:179–189. doi:10.1016/j.ijpsycho.2005.01.014
44. McGuire PK, Silbersweig DA, Wright I, Murray RM, Frackowiak RS, Frith CD. The neural correlates of inner speech and auditory verbal imagery in schizophrenia: relationship to auditory verbal hallucinations. *Br J Psychiatry J Ment Sci* (1996) **169**:148–159.
45. Johns LC, McGuire PK. Verbal self-monitoring and auditory hallucinations in schizophrenia. *Lancet Lond Engl* (1999) **353**:469–470. doi:10.1016/S0140-6736(98)05288-X
46. Allen PP, Johns LC, Fu CHY, Broome MR, Vythelingum GN, McGuire PK. Misattribution of external speech in patients with hallucinations and delusions. *Schizophr Res* (2004) **69**:277–287.
47. Johns LC, Gregg L, Allen P, McGuire PK. Impaired verbal self-monitoring in psychosis: effects of state, trait and diagnosis. *Psychol Med* (2006) **36**:465–474. doi:10.1017/S0033291705006628
48. McCarthy-Jones S, Fernyhough C. The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults. *Conscious Cogn* (2011) **20**:1586–1593. doi:10.1016/j.concog.2011.08.005
49. Langdon R, Jones SR, Connaughton E, Fernyhough C. The phenomenology of inner speech: comparison of schizophrenia patients with auditory verbal hallucinations and healthy controls. *Psychol Med* (2009) **39**:655–663. doi:10.1017/S0033291708003978
50. Cho R, Wu W. Mechanisms of auditory verbal hallucination in schizophrenia. *Front Psychiatry* (2013) **4**:155. doi:10.3389/fpsyt.2013.00155
51. Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. *Psychol Bull* (1993) **114**:3–28.
52. Brébion G, Amador X, David A, Malaspina D, Sharif Z, Gorman JM. Positive symptomatology and source-monitoring failure in schizophrenia--an analysis of symptom-specific effects. *Psychiatry Res* (2000) **95**:119–131.
53. Bentall RP, Slade PD. Reality testing and auditory hallucinations: a signal detection analysis. *Br J Clin Psychol Br Psychol Soc* (1985) **24 (Pt 3)**:159–169.
54. Vercammen A, de Haan EHF, Aleman A. Hearing a voice in the noise: auditory hallucinations and speech perception. *Psychol Med* (2008) **38**:1177–1184. doi:10.1017/S0033291707002437
55. Dodgson G, Gordon S. Avoiding false negatives: are some auditory hallucinations an evolved design flaw? *Behav Cogn Psychother* (2009) **37**:325–334. doi:10.1017/S1352465809005244
56. Dudley R, Dodgson G, Sarll G, Halhead R, Bolas H, McCarthy-Jones S. The effect of arousal on auditory threat detection and the relationship to auditory hallucinations. *J Behav Ther Exp Psychiatry* (2014) **45**:311–318. doi:10.1016/j.jbtep.2014.02.002
57. Waters F, Woodward T, Allen P, Aleman A, Sommer I. Self-recognition deficits in schizophrenia patients with auditory hallucinations: a meta-analysis of the literature. *Schizophr Bull* (2012) **38**:741–750. doi:10.1093/schbul/sbq144
58. Brookwell ML, Bentall RP, Varese F. Externalizing biases and hallucinations in source-monitoring, self-monitoring and signal detection studies: a meta-analytic review. *Psychol Med* (2013) **43**:2465–2475. doi:10.1017/S0033291712002760
59. Shergill SS, Brammer MJ, Williams SCR, Murray RM, McGuire PK. Mapping Auditory Hallucinations in Schizophrenia Using Functional Magnetic Resonance Imaging. *Arch Gen Psychiatry* (2000) **57**:1033. doi:10.1001/archpsyc.57.11.1033
60. Homan P, Kindler J, Hauf M, Walther S, Hubl D, Dierks T. Repeated measurements of cerebral blood flow in the left superior temporal gyrus reveal tonic hyperactivity in patients with auditory verbal hallucinations: a possible trait marker. *Front Hum Neurosci* (2013) **7**:304. doi:10.3389/fnhum.2013.00304

61. Sommer IEC, Diederen KMJ, Blom J-D, Willems A, Kushan L, Slotema K, Boks MPM, Daalman K, Hoek HW, Neggers SFW, et al. Auditory verbal hallucinations predominantly activate the right inferior frontal area. *Brain J Neurol* (2008) **131**:3169–3177. doi:10.1093/brain/awn251
62. Wilkinson S. Accounting for the phenomenology and varieties of auditory verbal hallucination within a predictive processing framework. *Conscious Cogn* (2014) **30**:142–155. doi:10.1016/j.concog.2014.09.002
63. Wolpert DM. Computational approaches to motor control. *Trends Cogn Sci* (1997) **1**:209–216. doi:10.1016/S1364-6613(97)01070-X
64. Friston K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* (2010) **11**:127–138. doi:10.1038/nrn2787
65. Horga G, Schatz KC, Abi-Dargham A, Peterson BS. Deficits in predictive coding underlie hallucinations in schizophrenia. *J Neurosci Off J Soc Neurosci* (2014) **34**:8072–8082. doi:10.1523/JNEUROSCI.0200-14.2014
66. Levitan C, Ward PB, Catts SV. Superior temporal gyral volumes and laterality correlates of auditory hallucinations in schizophrenia. *Biol Psychiatry* (1999) **46**:955–962.
67. Modinos G, Costafreda SG, van Tol M-J, McGuire PK, Aleman A, Allen P. Neuroanatomy of auditory verbal hallucinations in schizophrenia: a quantitative meta-analysis of voxel-based morphometry studies. *Cortex J Devoted Study Nerv Syst Behav* (2013) **49**:1046–1055. doi:10.1016/j.cortex.2012.01.009
68. García-Martí G, Aguilar EJ, Lull JJ, Martí-Bonmatí L, Escartí MJ, Manjón JV, Moratal D, Robles M, Sanjuán J. Schizophrenia with auditory hallucinations: a voxel-based morphometry study. *Prog Neuropsychopharmacol Biol Psychiatry* (2008) **32**:72–80. doi:10.1016/j.pnpbp.2007.07.014
69. Hubl D, Dougoud-Chauvin V, Zeller M, Federspiel A, Boesch C, Strik W, Dierks T, Koenig T. Structural analysis of Heschl's gyrus in schizophrenia patients with auditory hallucinations. *Neuropsychobiology* (2010) **61**:1–9. doi:10.1159/000258637
70. Shin S-E, Lee J-S, Kang M-H, Kim C-E, Bae J-N, Jung G. Segmented volumes of cerebrum and cerebellum in first episode schizophrenia with auditory hallucinations. *Psychiatry Res* (2005) **138**:33–42. doi:10.1016/j.pscychresns.2004.11.005
71. Crespo-Facorro B, Kim J, Andreasen NC, O'Leary DS, Bockholt HJ, Magnotta V. Insular cortex abnormalities in schizophrenia: a structural magnetic resonance imaging study of first-episode patients. *Schizophr Res* (2000) **46**:35–43.
72. Crespo-Facorro B, Roiz-Santiáñez R, Quintero C, Pérez-Iglesias R, Tordesillas-Gutiérrez D, Mata I, Rodríguez-Sánchez JM, Gutiérrez A, Vázquez-Barquero JL. Insular cortex morphometry in first-episode schizophrenia-spectrum patients: Diagnostic specificity and clinical correlations. *J Psychiatr Res* (2010) **44**:314–320. doi:10.1016/j.jpsychires.2009.08.014
73. Huang P, Xi Y, Lu Z-L, Chen Y, Li X, Li W, Zhu X, Cui L-B, Tan Q, Liu W, et al. Decreased bilateral thalamic gray matter volume in first-episode schizophrenia with prominent hallucinatory symptoms: A volumetric MRI study. *Sci Rep* (2015) **5**:14505. doi:10.1038/srep14505
74. Neckelmann G, Specht K, Lund A, Ersland L, Smievoll AI, Neckelmann D, Hugdahl K. Mr morphometry analysis of grey matter volume reduction in schizophrenia: association with hallucinations. *Int J Neurosci* (2006) **116**:9–23. doi:10.1080/00207450690962244
75. Cachia A, Amad A, Brunelin J, Krebs M-O, Plaze M, Thomas P, Jardri R. Deviations in cortex sulcation associated with visual hallucinations in schizophrenia. *Mol Psychiatry* (2014) doi:10.1038/mp.2014.140
76. Mangin J-F, Jouvent E, Cachia A. In-vivo measurement of cortical morphology: means and meanings. *Curr Opin Neurol* (2010) **23**:359–367. doi:10.1097/WCO.0b013e32833a0afc
77. Cachia A, Mangin JF, Rivière D, Kherif F, Boddaert N, Andrade A, Papadopoulos-Orfanos D, Poline JB, Bloch I, Zilbovicius M, et al. A primal sketch of the cortex mean curvature: a morphogenesis based approach to study the variability of the folding patterns. *IEEE Trans Med Imaging* (2003) **22**:754–765. doi:10.1109/TMI.2003.814781
78. Cachia A, Paillère-Martinot M-L, Galinowski A, Januel D, de Beaurepaire R, Bellivier F, Artiges E, Andoh J, Bartrés-Faz D, Duchesnay E, et al. Cortical folding abnormalities in schizophrenia patients with resistant auditory hallucinations. *NeuroImage* (2008) **39**:927–935. doi:10.1016/j.neuroimage.2007.08.049
79. Plaze M, Paillère-Martinot M-L, Penttilä J, Januel D, de Beaurepaire R, Bellivier F, Andoh J, Galinowski A, Gallarda T, Artiges E, et al. "Where do auditory hallucinations come from?"--a brain morphometry study of schizophrenia patients with inner or outer space hallucinations. *Schizophr Bull* (2011) **37**:212–221. doi:10.1093/schbul/sbp081
80. Kühn S, Gallinat J. Quantitative meta-analysis on state and trait aspects of auditory verbal hallucinations in schizophrenia. *Schizophr Bull* (2012) **38**:779–786. doi:10.1093/schbul/sbq152
81. Plaze M, Bartr s-Faz D, Martinot J-L, Januel D, Bellivier F, De Beaurepaire R, Chanraud S, Andoh J, Lefaucheur J-P, Artiges E. Left superior temporal gyrus activation during sentence perception negatively correlates with auditory hallucination severity in schizophrenia patients. *Schizophr Res* (2006) **87**:109–115.
82. Swick D, Turken AU. Dissociation between conflict detection and error monitoring in the human anterior cingulate cortex. *Proc Natl Acad Sci U S A* (2002) **99**:16354–16359. doi:10.1073/pnas.252521499
83. Northoff G, Bermpohl F. Cortical midline structures and the self. *Trends Cogn Sci* (2004) **8**:102–7. doi:10.1016/j.tics.2004.01.004
84. Garrison JR, Fernyhough C, McCarthy-Jones S, Haggard M, Australian Schizophrenia Research Bank, Simons JS. Paracingulate sulcus morphology is associated with hallucinations in the human brain. *Nat Commun* (2015) **6**:8956. doi:10.1038/ncomms9956
85. Jardri R, Pouchet A, Pins D, Thomas P. Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. *Am J Psychiatry* (2011) **168**:73–81. doi:10.1176/appi.ajp.2010.09101522
86. Rossell SL. "The role of memory retrieval and emotional salience in the emergence of auditory hallucinations.," in *The neuroscience of hallucinations.* (New York), 137–52.
87. Waters FAV, Badcock JC, Michie PT, Maybery MT. Auditory hallucinations in schizophrenia: intrusive thoughts and forgotten memories. *Cognit Neuropsychiatry* (2006) **11**:65–83. doi:10.1080/13546800444000191
88. Jardri R, Thomas P, Delmaire C, Delion P, Pins D. The Neurodynamic Organization of Modality-Dependent Hallucinations. *Cereb Cortex* (2013) **23**:1108–1117. doi:10.1093/cercor/bhs082
89. Benetti S, Pettersson-Yeo W, Mechelli A. "Connectivity issues of the hallucinating brain.," in *R Jardri, A Cachia, P Thomas & D Pins (Eds): The Neuroscience of Hallucinations.* (NY: Springer), 417–43.

90. Alderson-Day B, McCarthy-Jones S, Fernyhough C. Hearing voices in the resting brain: A review of intrinsic functional connectivity research on auditory verbal hallucinations. *Neurosci Biobehav Rev* (2015) **55**:78–87. doi:10.1016/j.neubiorev.2015.04.016
91. Pettersson-Yeo W, Allen P, Benetti S, McGuire P, Mechelli A. Dysconnectivity in schizophrenia: where are we now? *Neurosci Biobehav Rev* (2011) **35**:1110–1124. doi:10.1016/j.neubiorev.2010.11.004
92. Oertel-Knöchel V, Knöchel C, Matura S, Prvulovic D, Linden DEJ, van de Ven V. Reduced functional connectivity and asymmetry of the planum temporale in patients with schizophrenia and first-degree relatives. *Schizophr Res* (2013) **147**:331–338. doi:10.1016/j.schres.2013.04.024
93. Mou X, Bai F, Xie C, Shi J, Yao Z, Hao G, Chen N, Zhang Z. Voice recognition and altered connectivity in schizophrenic patients with auditory hallucinations. *Prog Neuropsychopharmacol Biol Psychiatry* (2013) **44**:265–270. doi:10.1016/j.pnpbp.2013.03.006
94. Amad A, Cachia A, Gorwood P, Pins D, Delmaire C, Rolland B, Mondino M, Thomas P, Jardri R. The multimodal connectivity of the hippocampal complex in auditory and visual hallucinations. *Mol Psychiatry* (2013) doi:10.1038/mp.2012.181
95. Rolland B, Amad A, Poulet E, Bordet R, Vignaud A, Bation R, Delmaire C, Thomas P, Cottencin O, Jardri R. Resting-State Functional Connectivity of the Nucleus Accumbens in Auditory and Visual Hallucinations in Schizophrenia. *Schizophr Bull* (2014) doi:10.1093/schbul/sbu097
96. Rotarska-Jagiela A, Oertel-Knoechel V, DeMartino F, van de Ven V, Formisano E, Roebroeck A, Rami A, Schoenmeyer R, Haenschel C, Hendler T, et al. Anatomical brain connectivity and positive symptoms of schizophrenia: a diffusion tensor imaging study. *Psychiatry Res* (2009) **174**:9–16. doi:10.1016/j.pscychresns.2009.03.002
97. Ćurčić-Blake B, Nanetti L, van der Meer L, Cerliani L, Renken R, Pijnenborg GHM, Aleman A. Not on speaking terms: hallucinations and structural network disconnectivity in schizophrenia. *Brain Struct Funct* (2015) **220**:407–418. doi:10.1007/s00429-013-0663-y
98. Geoffroy PA, Houenou J, Duhamel A, Amad A, De Weijer AD, Curčić-Blake B, Linden DEJ, Thomas P, Jardri R. The arcuate fasciculus in auditory-verbal hallucinations: A meta-analysis of diffusion-tensor-imaging studies. *Schizophr Res* (2014) doi:10.1016/j.schres.2014.07.014
99. Wolf ND, Sambataro F, Vasic N, Frasch K, Schmid M, Schönfeldt-Lecuona C, Thomann PA, Wolf RC. Dysconnectivity of multiple resting-state networks in patients with schizophrenia who have persistent auditory verbal hallucinations. *J Psychiatry Neurosci JPN* (2011) **36**:366–74. doi:10.1503/jpn.110008
100. Mechelli A, Allen P, Amaro E, Fu CHY, Williams SCR, Brammer MJ, Johns LC, McGuire PK. Misattribution of speech and impaired connectivity in patients with auditory verbal hallucinations. *Hum Brain Mapp* (2007) **28**:1213–1222. doi:10.1002/hbm.20341
101. Bush G, Luu P, Posner M. Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn Sci* (2000) **4**:215–222.
102. Carter CS, van Veen V. Anterior cingulate cortex and conflict detection: an update of theory and data. *Cogn Affect Behav Neurosci* (2007) **7**:367–79.
103. Allen P, Aleman A, McGuire PK. Inner speech models of auditory verbal hallucinations: evidence from behavioural and neuroimaging studies. *Int Rev Psychiatry Abingdon Engl* (2007) **19**:407–415. doi:10.1080/09540260701486498
104. Kühn S, Gallinat J. Quantitative meta-analysis on state and trait aspects of auditory verbal hallucinations in schizophrenia. *Schizophr Bull* (2012) **38**:779–86. doi:10.1093/schbul/sbq152
105. deCharms RC, Maeda F, Glover GH, Ludlow D, Pauly JM, Soneji D, Gabrieli JDE, Mackey SC. Control over brain activation and pain learned by using real-time functional MRI. *Proc Natl Acad Sci U S A* (2005) **102**:18626–31. doi:10.1073/pnas.0505210102
106. Gröne M, Dyck M, Koush Y, Bergert S, Mathiak KA, Alawi EM, Elliott M, Mathiak K. Upregulation of the rostral anterior cingulate cortex can alter the perception of emotions: fMRI-based neurofeedback at 3 and 7 T. *Brain Topogr* (2014)
107. Hamilton JP, Glover GH, Hsu J-J, Johnson RF, Gotlib IH. Modulation of subgenual anterior cingulate cortex activity with real-time neurofeedback. *Hum Brain Mapp* (2011) **32**:22–31. doi:10.1002/hbm.20997
108. Li X, Hartwell KJ, Borckardt J, Prisciandaro JJ, Saladin ME, Morgan PS, Johnson KA, Lematty T, Brady KT, George MS. Volitional reduction of anterior cingulate cortex activity produces decreased cue craving in smoking cessation: a preliminary real-time fMRI study. *Addict Biol* (2013) **18**:739–48. doi:10.1111/j.1369-1600.2012.00449.x
109. Weiskopf N, Veit R, Erb M, Mathiak K, Grodd W, Goebel R, Birbaumer N. Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fMRI): methodology and exemplary data. *NeuroImage* (2003) **19**:577–586. doi:10.1016/S1053-8119(03)00145-9
110. Jones SR, Fernyhough C. Neural correlates of inner speech and auditory verbal hallucinations: a critical review and theoretical integration. *Clin Psychol Rev* (2007) **27**:140–154. doi:10.1016/j.cpr.2006.10.001
111. Shergill SS, Bullmore ET, Brammer MJ, Williams SC, Murray RM, McGuire PK. A functional study of auditory verbal imagery. *Psychol Med* (2001) **31**:241–253.
112. Cleghorn JM, Franco S, Szechtman B, Kaplan RD, Szechtman H, Brown GM, Nahmias C, Garnett ES. Toward a brain map of auditory hallucinations. *Am J Psychiatry* (1992) **149**:1062–9.
113. Dyck MS, Mathiak KA, Bergert S, Sarkheil P, Koush Y, Alawi EM, Zvyagintsev M, Gaebler AJ, Shergill S, Mathiak K. Targeting treatment-resistant auditory verbal hallucinations in schizophrenia with fMRI-based neurofeedback – exploring different cases of schizophrenia. *Neuroimaging Stimul* (2016) **7**:37. doi:10.3389/fpsyt.2016.00037
114. Koush Y, Zvyagintsev M, Dyck M, Mathiak KA, Mathiak K. Signal quality and Bayesian signal processing in neurofeedback based on real-time fMRI. *NeuroImage* (2012) **59**:478–489. doi:10.1016/j.neuroimage.2011.07.076
115. Harmelech T, Preminger S, Wertman E, Malach R. The day-after effect: long term, Hebbian-like restructuring of resting-state fMRI patterns induced by a single epoch of cortical activation. *J Neurosci Off J Soc Neurosci* (2013) **33**:9488–97. doi:10.1523/JNEUROSCI.5911-12.2013
116. Ruiz S, Birbaumer N, Sitaram R. Abnormal Neural Connectivity in Schizophrenia and fMRI-Brain-Computer Interface as a Potential Therapeutic Approach. *Front Psychiatry* (2013) **4**:17. doi:10.3389/fpsyt.2013.00017

117. Tracy DK, Shergill SS. Mechanisms Underlying Auditory Hallucinations-Understanding Perception without Stimulus. *Brain Sci* (2013) **3**:642–69. doi:10.3390/brainsci3020642
118. Palaniyappan L, White TP, Liddle PF. The concept of salience network dysfunction in schizophrenia: from neuroimaging observations to therapeutic opportunities. *Curr Top Med Chem* (2012) **12**:2324–38.
119. Palaniyappan L, Simmonite M, White TP, Liddle EB, Liddle PF. Neural primacy of the salience processing system in schizophrenia. *Neuron* (2013) **79**:814–28. doi:10.1016/j.neuron.2013.06.027
120. Lefebvre S, Demeulemeester M, Leroy A, Delmaire C, Lopes R, Pins D, Thomas P, Jardri R. Network dynamics during the different stages of hallucinations in schizophrenia. (In revision)
121. Wang L, Metzak PD, Woodward TS. Aberrant connectivity during self-other source monitoring in schizophrenia. *Schizophr Res* (2011) **125**:136–142. doi:10.1016/j.schres.2010.11.012
122. Weiskopf N, Mathiak K, Bock SW, Scharnowski F, Veit R, Grodd W, Goebel R, Birbaumer N. Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI). *IEEE Trans Biomed Eng* (2004) **51**:966–970. doi:10.1109/TBME.2004.827063
123. Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. *Nature* (2000) **403**:309–312. doi:10.1038/35002078
124. Caria A, Veit R, Sitaram R, Lotze M, Weiskopf N, Grodd W, Birbaumer N. Regulation of anterior insular cortex activity using real-time fMRI. *NeuroImage* (2007) **35**:1238–1246. doi:10.1016/j.neuroimage.2007.01.018
125. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev* (2012) **36**:1140–1152. doi:10.1016/j.neubiorev.2012.01.004
126. Ellison-Wright I, Bullmore E. Anatomy of bipolar disorder and schizophrenia: a meta-analysis. *Schizophr Res* (2010) **117**:1–12. doi:10.1016/j.schres.2009.12.022
127. Smieskova R, Fusar-Poli P, Allen P, Bendfeldt K, Stieglitz RD, Drewe J, Radue EW, McGuire PK, Riecher-Rössler A, Borgwardt SJ. Neuroimaging predictors of transition to psychosis--a systematic review and meta-analysis. *Neurosci Biobehav Rev* (2010) **34**:1207–1222. doi:10.1016/j.neubiorev.2010.01.016
128. LaConte SM. Decoding fMRI brain states in real-time. *NeuroImage* (2011) **56**:440–454. doi:10.1016/j.neuroimage.2010.06.052
129. Sitaram R, Lee S, Ruiz S, Rana M, Veit R, Birbaumer N. Real-time support vector classification and feedback of multiple emotional brain states. *NeuroImage* (2011) **56**:753–765. doi:10.1016/j.neuroimage.2010.08.007
130. Jardri R, Pins D, Thomas P. Predictive functional brain imaging of hallucinations. *Biol Psychiatry* (2011)69(9) : s217.
131. Rana M, Gupta N, Dalboni Da Rocha JL, Lee S, Sitaram R. A toolbox for real-time subject-independent and subject-dependent classification of brain states from fMRI signals. *Front Neurosci* (2013) **7**: doi:10.3389/fnins.2013.00170
132. Bandura A. *Self-Efficacy: The Exercise of Control*. Worth Publishers (1997).
133. Scharnowski F, Rosa MJ, Golestani N, Hutton C, Josephs O, Weiskopf N, Rees G. Connectivity changes underlying neurofeedback training of visual cortex activity. *PloS One* (2014) **9**:e91090. doi:10.1371/journal.pone.0091090
134. Koush Y, Rosa MJ, Robineau F, Heinen K, W Rieger S, Weiskopf N, Vuilleumier P, Van De Ville D, Scharnowski F. Connectivity-based neurofeedback: dynamic causal modeling for real-time fMRI. *NeuroImage* (2013) **81**:422–430. doi:10.1016/j.neuroimage.2013.05.010
135. Kim D-Y, Yoo S-S, Tegethoff M, Meinlschmidt G, Lee J-H. The inclusion of functional connectivity information into fMRI-based neurofeedback improves its efficacy in the reduction of cigarette cravings. *J Cogn Neurosci* (2015) **27**:1552–1572. doi:10.1162/jocn_a_00802

4 . DISCUSSION

“Reality is that which, when you stop believing in it, doesn't go away.”

Philip K. Dick, I Hope I Shall Arrive Soon

Dans ce travail de thèse, je me suis attaché à montrer que la « capture » en IRMf des réseaux neuronaux impliqués dans la survenue des hallucinations auditives chez les patients souffrant de schizophrénie pouvait être automatisée grâce à des méthodes de *machine learning* (voir 2. Détection automatisée des hallucinations auditives en IRM fonctionnelle). L'utilisation en temps réel des classificateurs ainsi développés pourraient permettre une détection du phénomène hallucinatoire chez les patients souffrant de schizophrénie, au cours d'une session d'IRMf, ouvrant la voie à des applications thérapeutiques innovantes comme le neurofeedback (voir 3. Perspectives thérapeutiques dans la schizophrénie).

4.1 Quelle interface de neurofeedback ?

Dans l'**article 10**, nous avons pu présenter les différentes cibles thérapeutiques possibles dans la perspective d'une mise en route prochaine d'un protocole de neurofeedback à même de traiter les hallucinations auditives chez les patients souffrant de schizophrénie. Classiquement, le neurofeedback repose sur un entrainement visant à permettre au patient d'apprendre à réguler l'activité d'une (ou plusieurs) région(s) cérébrale(s). Dans la plupart des protocoles, la cible à neuromoduler est identifiée, soit par un repérage anatomique selon un référentiel (Talairach ou MNI), soit par un repérage fonctionnel (grâce à une tache réalisée en IRMf) (221). Toutefois, des techniques de neurofeedback basées sur un décodage du signal BOLD par analyse multivariée ont également pu être proposées (251,252). Ces dernières stratégies visent à *modifier des patterns* d'activité cérébrale et non plus à simplement augmenter ou diminuer l'amplitude du signal BOLD moyen dans une région d'intérêt choisie en fonction d'hypothèses *a priori* (253). Shibata et al. montrent par exemple dans un travail récent, qu'il est possible grâce à ce type de stratégie d'induire des *patterns* d'activation spécifiques dans le cortex cingulaire des participants, ceci étant associé à des modifications comportementales (dans une tache de préférence de visages avec émotion) (254). Pour décrire ce concept, certains auteurs ont proposé le terme de neurofeedback « décodé » (255). Cependant, dans les études utilisant ce type de neurofeedback « décodé », l'objectif est d'entrainer le sujet à induire des *patterns* d'activation déterminés (voir **Figure 12**). Dans le cadre de la prise en charge des hallucinations auditives, l'objectif est d'entrainer le sujet à développer des stratégies permettant de modifier leur activité cérébrale au moment où des *patterns* connus pour être associés aux hallucinations auditives sont détectés.

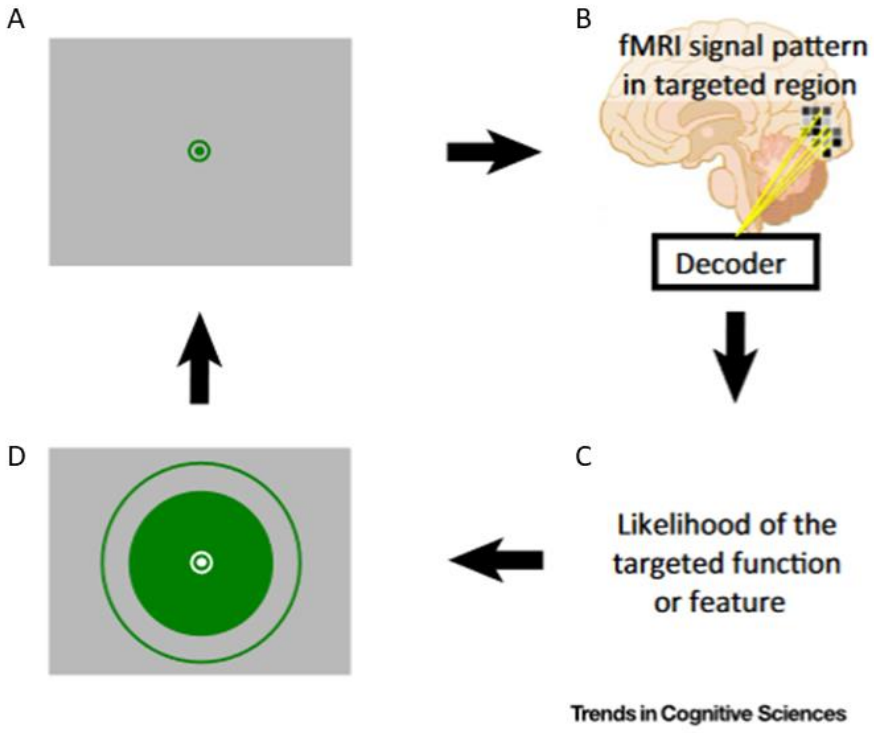


Figure 12. Principe du neurofeedback « décodé ».

Dans un premier temps, la fonction ciblée est prédéterminée et le pattern d'activité en IRMf qui lui est associé permet la construction d'un classificateur (ou décodeur). Les séances d'entrainement se déroulent ensuite en 4 étapes. A/ Il est demandé au participant, dans le scanner, de maintenir son attention sur le point central et de réguler son activité cérébrale. B/ Un pattern de voxels dans une région cible est mesuré. C/ En se basant sur le pattern mesuré, le classificateur fournit la probabilité que celui-ci corresponde au pattern cible. D/ La taille du disque présenté au participant est proportionnelle à cette probabilité et une récompense externe (argent) peut être donnée au participant selon la taille du disque. Adapté de (255).

Dans ce contexte, nous souhaitons à court-terme développer une interface de neurofeedback innovante basée sur une combinaison de plusieurs classificateurs. Comme nous l'avons vu dans l'**Article 2**, il est désormais possible de développer des classificateurs capables de détecter avec des performances satisfaisantes, la période hallucinatoire (dite « ON »), mais également la période précédant le symptôme hallucinatoire (dite « TRANS ») (voir **Article 3**). De la même façon, il serait tout à fait envisageable de développer un classificateur à même de détecter la période de « sortie » d'hallucination, c'est-à-dire, la période de quelques secondes suivant l'hallucination auditive (dite « END »).

Ainsi, la combinaison de 3 classificateurs (ON *Vs* Autres / TRANS *Vs* Autres / END *Vs* Autres) offrirait une interface multi-classificateurs (à choix forcé) qui permettrait de donner au sujet, une information en temps réel sur la phase (ON / TRANS / END / OFF) détectée. Cette information pourrait être associée à des instructions spécifiques proposant au sujet des stratégies de *coping* adaptées (exemple proposé en **Figure 13**). Toutefois, il est également envisageable d'utiliser uniquement l'information « brute » en demandant au sujet d'essayer d'obtenir un temps maximal en phase OFF au cours de la session d'IRMf sans lui indiquer à quoi correspondent les différents stades (qui pourraient être représentées au cours de la séance, uniquement par des couleurs). En effet, plusieurs études ont pu montrer qu'un apprentissage « implicite » (c'est-à-dire, sans que le sujet sache ce qu'il apprend à réguler) en neurofeedback, était possible (256,257).

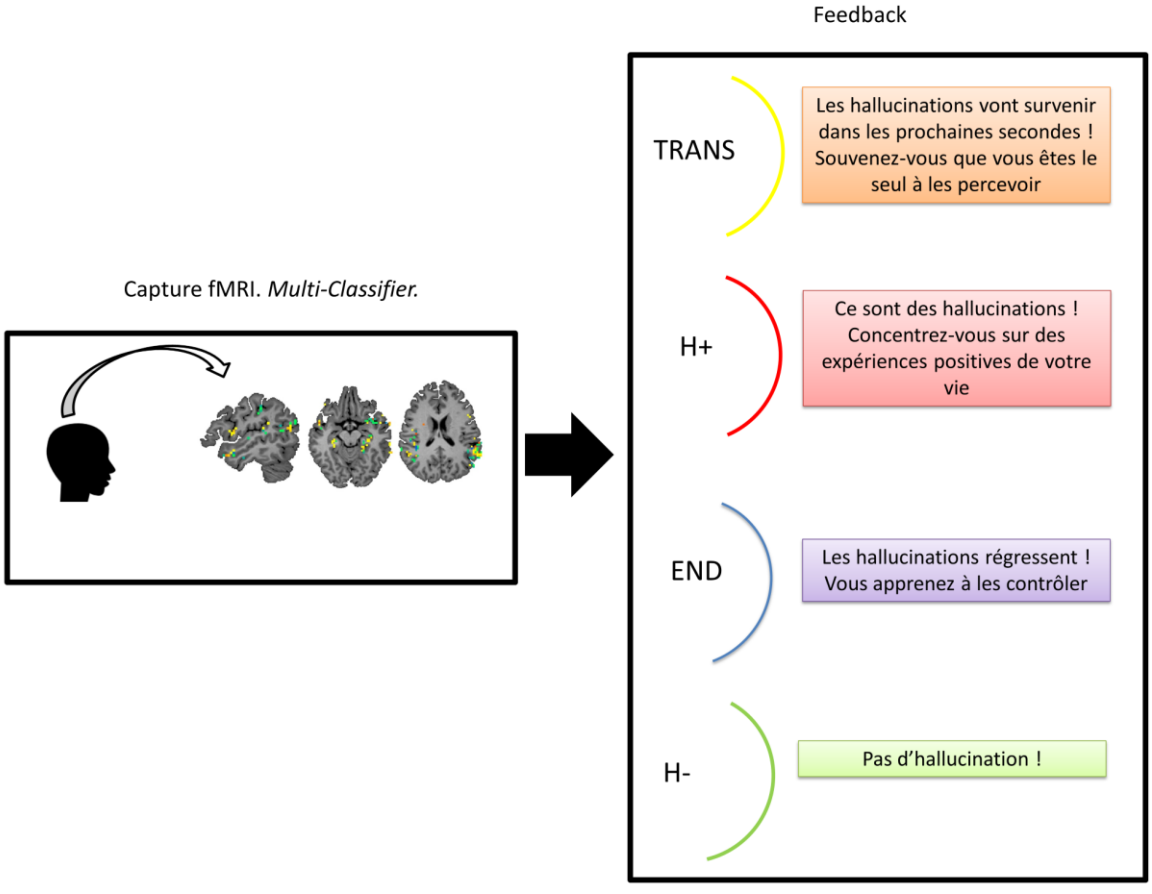


Figure 13. Proposition d'une interface multi-classificateurs pour la prise en charge des hallucinations auditives par neurofeedback guidé par IRMf.

4.2 Protocole d'entrainement et évaluation

Après la mise en place d'une étude de faisabilité sur 5 sujets, nous envisageons de tester l'efficacité de cette interface multi-classificateurs au cours d'un essai contrôlé randomisé dont le déroulement est résumé en **Figure 14**. La sévérité des hallucinations sera évaluée avant et après le protocole grâce à des échelles psychométriques validées dans l'évaluation des symptômes hallucinatoires : l'*Auditory Hallucinations Rating Scale* (AHRS) et la *Positive and Negative Syndrome Scale* (PANSS), ainsi qu'avec des échelles visuelles analogiques évaluant la fréquence et l'intensité des hallucinations. Notre étude de faisabilité permettra un calcul de puissance afin de déterminer le nombre de sujets à inclure.

Plusieurs objectifs sont ici visés :

- ***Objectif principal*** : montrer qu'un protocole de neurofeedback guidé par IRMf permet de diminuer la sévérité des hallucinations auditives chez les sujets souffrant de schizophrénie.
- ***Objectifs secondaires*** : (1) Déterminer les stratégies d'imagerie mentale utilisées par les patients au cours des séances au moyen d'un questionnaire post-IRM et d'entretiens filmés (il s'agira d'essayer de comprendre comment les patients parviennent à modifier le signal présenté) ; (2) Grâce à des analyses post-hoc, déterminer les régions dont l'activité fait l'objet d'une régulation au cours des séances et rechercher les modifications structurales, fonctionnelles et de connectivité associées à la réponse ou non-réponse au traitement (afin de dégager des pistes de marqueurs prédictifs de réponse au traitement).

Les critères d'inclusion seront les suivants :

- Age : 18 à 60 ans
- Droitier
- Diagnostic de schizophrénie selon les critères du DSM-5
- Hallucinations auditives fréquentes (caractérisées par SAPS (Item 1) ≥ 4)
- Traitement pharmacologique psychotrope non modifié dans les 30 jours précédents l'initiation du traitement par neurofeedback
- Pas de pathologie neurologique associée, dont l'épilepsie
- Consentement écrit libre après information complète sur l'étude, son principe, les bénéfices et les risques éventuels

Les critères de non-inclusion seront les suivants :

- Refus de participation après information claire et loyale sur l'étude
- Grossesse en cours (Date des dernières règles + possibilité de test urinaire si doute)
- Contre-indications à la passation d'une IRM : Pacemaker, clips vasculaires ferromagnétiques, corps étrangers ferromagnétiques
- Critères morphologiques : poids > 130 Kg, périmètre abdominal conditionné par l'ouverture de l'aimant, largeur des épaules
- Claustrophobie
- Refus d'être informé d'une particularité vue à l'IRM

Les patients inclus seront randomisés en 2 groupes :

- **un groupe "Neurofeedback actif" (NFAct),** qui bénéficiera d'un protocole de neurofeedback de 4 séances,
- **un groupe "Neurofeedback contrôle" (NFCont),** pour lequel le signal renvoyé au patient sera un signal aléatoire non corrélé à l'activité cérébrale du patient. Les séances seront réalisées à raison de 1 séance par jour sur 4 jours consécutifs.

Les évaluations proposées seront :

- *La sévérité des hallucinations auditives* : évaluée avant et après le protocole de neurofeedback (évaluation clinique initiale puis évaluation post-thérapeutique : J+1 après fin du traitement puis J+10 et J+30).
- *Une imagerie cérébrale anatomique et fonctionnelle de repos* sera également réalisée avant la mise en place du traitement et à J+1. Ainsi, il sera possible d'effectuer pour le groupe NFAct des comparaisons avant/après traitement afin de dégager des pistes à propos des mécanismes physiopathologiques sous-tendant la réponse thérapeutique à cette technique. Cela permettra également de s'intéresser plus particulièrement à l'imagerie cérébrale pré-thérapeutique des patients répondeurs, ceci afin d'identifier des marqueurs prédictifs de bonne réponse au traitement. Cette imagerie pré-thérapeutique permettra également de s'assurer de la survenue et la détection par le classificateur, d'un nombre suffisant d'épisodes hallucinatoires au cours de la séance de neurofeedback.
- Après chaque séance de neurofeedback : (1) *EVA fréquence/intensité des HA avant/après la séance* ; (2) *Entretien semi-structuré visant à identifier les stratégies d'imagerie mentale utilisées pendant la séance par les patients (voir Annexes 1, 2 et 3)* ; (3) *Effets indésirables potentiels* ; même si aucun effet indésirable n'est attendu, le patient sera interrogé après chaque séance sur d'éventuelles difficultés.

Le parcours du participant au cours de l'étude est résumé en **Figure 15**.

Pour d'évaluer l'efficacité de notre procédure, le critère de jugement principal sera une diminution significativement plus élevée ($p<0,05$) du score à l'AHRS après traitement dans le groupe NFAct par rapport au groupe NFCont. Les critères de jugement secondaires également utilisés seront les scores à la PANSS, la SAPS, la CGI, la GAF et de fréquence des (EVA) et intensité (EVA) des hallucinations. Toutes ces variables feront l'objet d'une comparaison entre groupes NFAct et NFCont. L'imagerie cérébrale fonctionnelle fera également l'objet d'une analyse off-line avant/après.

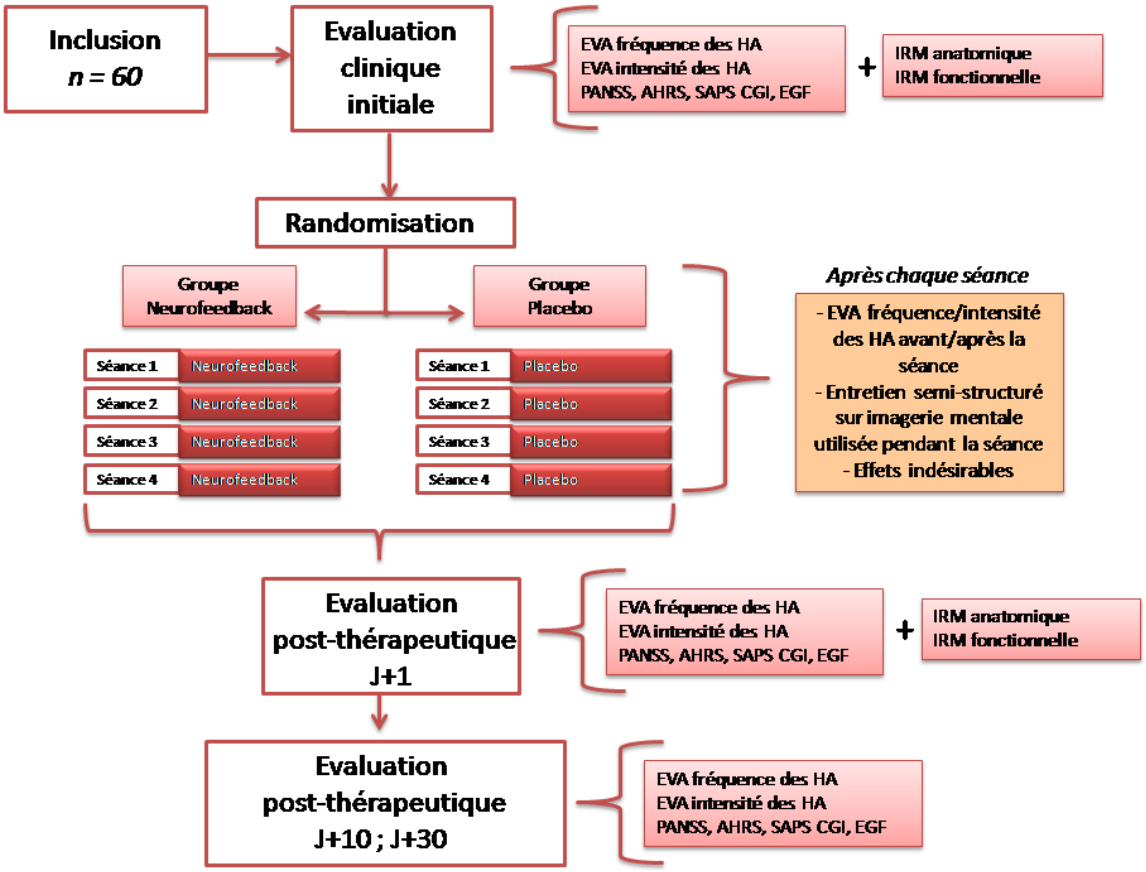


Figure 14. Présentation du *design* d'un essai contrôlé randomisé visant à évaluer l'efficacité d'un protocole de neurofeedback guidé par IRMf dans le traitement des hallucinations auditives chez les patients souffrant de schizophrénie.

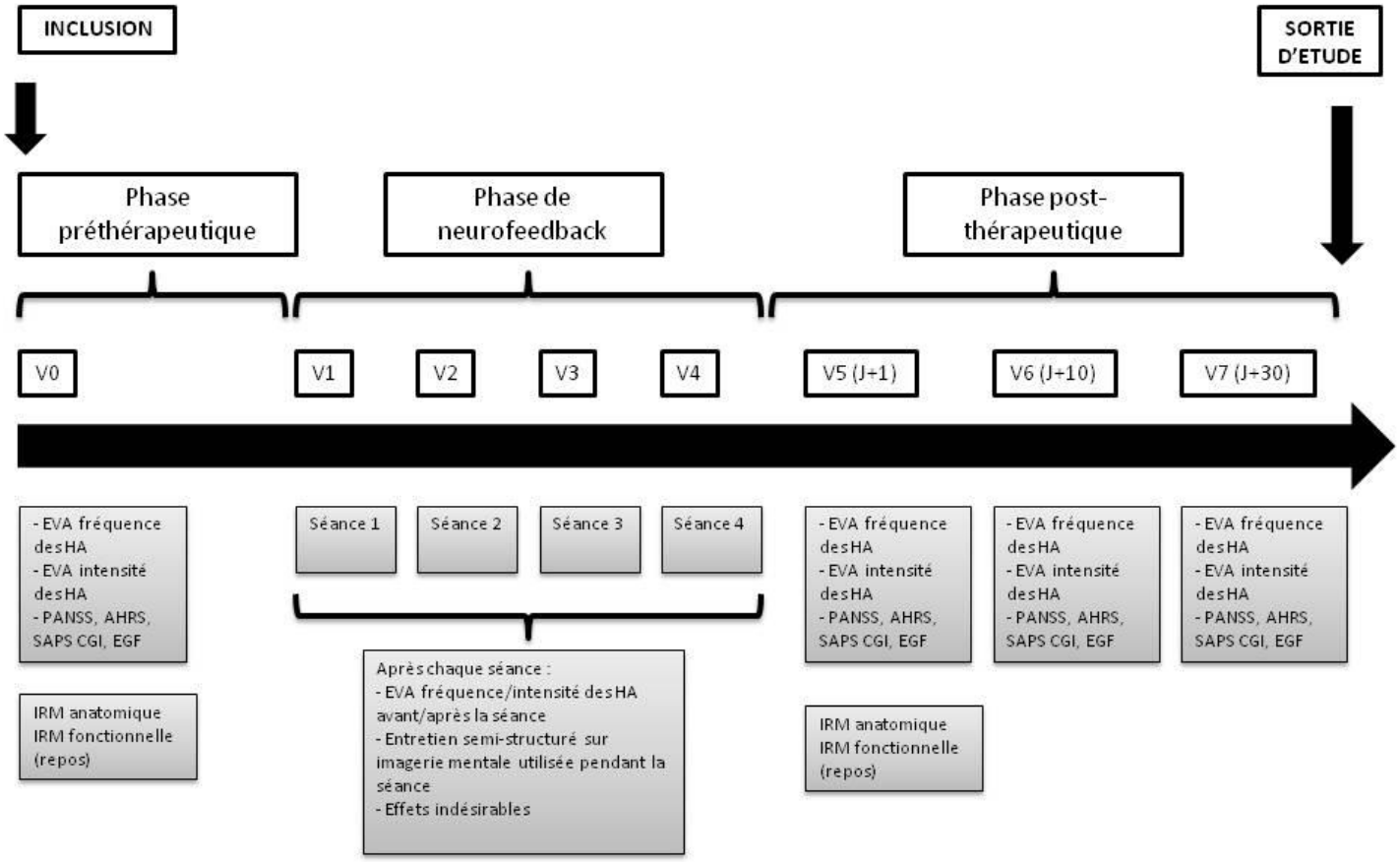


Figure 15. Résumé du parcours du participant à l'étude.

4.3 Une véritable psychothérapie guidée par l'imagerie ?

L'objectif de notre stratégie multi-classificateurs est de permettre aux participants de développer et de sélectionner des stratégies de *coping* personnalisées permettant de diminuer efficacement la fréquence des hallucinations auditives. L'entrainement par neurofeedback pourrait ainsi permettre au patient d'identifier des stratégies mentales alternatives lui permettant de sortir de l'état hallucinatoire (ON => END & OFF), voire de prévenir leur survenue lorsque la phase de transition survient (TRANS => END & OFF).

Par ailleurs, nous avons vu précédemment que les processus affectifs sont largement impliqués dans le phénomène hallucinatoire. Chez les patients souffrant de schizophrénie, les hallucinations auditives sont également associées à des niveaux élevés d'anxiété (258–260). En amont de l'expérience, les facteurs émotionnels peuvent être précurseurs, et en aval il peut s'agir de facteurs de maintien. Smith et collaborateurs proposent d'ailleurs l'hypothèse de l'instauration d'un cercle vicieux, les hallucinations auditives à valence émotionnelle négative ayant un impact négatif sur l'humeur et l'humeur dépressive pouvant être à l'origine d'une propension plus importante à halluciner (261). Dans cette perspective, la technique de neurofeedback proposée plus haut pourrait également permettre une augmentation des capacités de régulation émotionnelle des participants, notamment par détournement de l'attention allouée au phénomène hallucinatoire. En effet, en focalisant leur attention sur l'interface de neurofeedback plutôt que sur le symptôme, les participants auraient moins tendance à utiliser des stratégies de suppression expressive d'émotions à valence négative dont il a pu être montré qu'elles sont à la fois fréquentes chez les sujets souffrant de schizophrénie (262) mais également, paradoxalement, associées à un manque de contrôle émotionnel (263) et à une sévérité plus importante des hallucinations auditives chez les patients souffrant de schizophrénie (264). Ainsi, en diminuant l'allocation attentionnelle vers le symptôme et les affects négatifs qui y sont associés, le neurofeedback pourrait constituer un nouvel outil pour lutter contre ce facteur de maintien important des hallucinations auditives.

En ce qui concerne la dimension émotionnelle, d'autres perspectives intéressantes émergent de l'utilisation de la technique de neurofeedback, notamment de par les liens existants entre croyances métacognitives, hallucinations et stress généré. Par exemple, Koole met en avant l'importance de : (i) l'évaluation subjective de l'expérience hallucinatoire par le patient ; et (ii) ses croyances vis-à-vis de l'hallucination, sur le niveau de stress (265). Ainsi, il développe l'idée que les émotions générées par

les croyances sur les voix (*e.g.,* intention malveillante et omnipotence des voix notamment) et non les caractéristiques phénoménologiques du symptôme sont à l'origine du stress généré par l'hallucination auditive (40). C'est sur ce modèle théorique qu'un certain nombre de programmes de TCC s'appuient en proposant un schéma « ABC » (*activating event (hallucination) -> beliefs (croyances vis-à-vis des voix) -> conséquences émotionnelles et comportementales*) dans lequel la modification des croyances est centrale. En TCC, c'est un entrainement au raisonnement par hypothèse qui doit permettre cet assouplissement des croyances. Toutefois, en présentant une information sur l'activité cérébrale du patient, les techniques de neurofeedback pourraient également permettre de modifier la perception du symptôme dans ce sens. Notamment, en permettant au patient d'acquérir des stratégies de « sortie » de l'hallucination ou de repérage des périodes pré-hallucinatoires, les croyances vis-à-vis des hallucinations auditives pourraient être profondément modifiées par les séances de neurofeedback.

Un point clé devra cependant être précisé au cours de notre étude de faisabilité, celui de la consigne donnée au patient avant la séance. En effet, les mécanismes sous-tendant les processus d'apprentissage mis en jeu au cours des protocoles de neurofeedback restent mal connus (266,267). De grands principes comme le *conditionnement opérant*, la *neuroplasticité induite* par la répétition des séances et *l'implication consciente et volontaire* du sujet sont centraux mais, comme nous l'avons vu précédemment, un *apprentissage implicite* pourrait également être à l'œuvre et sous-tendre l'auto-régulation cérébrale acquise au cours du processus. Certains auteurs ont pu évoquer un double processus conscient et inconscient (268). Ces travaux pourraient avoir un impact direct sur la question des consignes à donner aux participants, à savoir : donner des stratégies mentales et des instructions explicites aux participants ou non pour optimiser l'apprentissage. De manière intéressante, il semblerait que le fait de ne pas donner d'informations sur des stratégies mentales spécifiques à utiliser aux participants, soit associée à de meilleurs résultats en terme d'apprentissage (269). Toutefois, très peu de données étant actuellement disponibles, notamment pour les sujets souffrant de schizophrénie, notre étude de faisabilité devra déterminer si les sujets parviennent à réaliser la tache de neurofeedback en l'absence d'instructions spécifiques. Il est possible que des adaptations soient nécessaires notamment car les régions impliquées dans le contrôle cognitif mis en œuvre lors de la tâche de régulation sont des régions dont on connait l'implication dans la schizophrénie. Ninaus et collaborateurs ont notamment pu montrer que, face à un signal de neurofeedback placebo (*i.e.* non corrélé à l'activité cérébrale du sujet), l'activité cérébrale des sujets à qui il est demandé de réguler volontairement le signal présenté est

caractérisée par une activation bilatérale du cortex insulaire, du cortex cingulaire antérieur, de l'aire motrice supplémentaire et du cortex pré-frontal dorso-médial et latéral (comparativement à une condition dans laquelle il est simplement demandé au sujet de regarder passivement le signal) (270) (voir 1. Introduction, pour présentation des travaux montrant les modifications d'activation au sein de ces régions chez les patients souffrant de schizophrénie avec hallucinations auditives).

Enfin, d'autres facteurs non spécifiques mais également susceptibles de modifier les performances des participants devront être pris en compte, notamment la motivation du sujet, son niveau de concentration, ses lieux de contrôle, etc. Ces facteurs pourront être explorés au moyen de questionnaires spécifiques pré-session et post-session, tels que ceux développés par Aurore Hakoun, Samy Chicki et François-Benoît Vialatte (voir **Annexe1, Annexe 2, Annexe 3**).

4.4 Perspectives

Je souhaiterais terminer cette discussion en évoquant un certain nombre de perspectives d'optimisation de l'interface multi-classificateurs évoquée plus-haut, notamment celles qui pourraient s'appuyer sur l'utilisation d'une *gamification* de cette interface. L'**Article 11** propose en effet un modèle basé sur la notion de feedback présentant comment le *serious game*, c'est-à-dire l'utilisation d'un jeu vidéo dans un objectif non-ludique (271), pourrait permettre le développement de programmes de psychothérapie innovants. Dans cette perspective, l'adaptation d'une interface de *gaming* pour le protocole proposé précédemment, pourrait présenter de nombreux intérêts, notamment pour renforcer la motivation des patients au cours de l'entrainement. Il pourrait s'agir d'un jeu très simple dans lequel l'objectif serait atteint lorsque la période OFF est atteinte. A titre purement illustratif, il est possible d'imaginer une interface basée sur le jeu de *bowling* : un ensemble de quilles apparaitraient lorsque la phase ON survient et pour faire tomber les quilles, le participant devrait passer en phase END ou en phase OFF. Toutefois, afin de garder un niveau d'attention suffisant et éviter un ennui des participants, un système de points ou de difficulté croissante devrait probablement être intégré pour éviter toute démotivation.

ARTICLE 11

Serious games: the future of psychotherapy?

Proposal of an integrative model

Thomas FOVET[1*], Jean-Arthur MICOULAUD-FRANCHI[2], Renaud JARDRI[1],

David E J LINDEN[3], Ali AMAD[1,4]

1. Univ Lille, CNRS, UMR-9193, SCA-Lab, équipe psyCHIC et CHU Lille, Pôle de psychiatrie (CURE), Hôpital Fontan, F-59000 Lille, France

2. Université de Bordeaux, CNRS, USR-3413 SANPSY et Service d'Explorations Fonctionnelles du Système Nerveux, Clinique du Sommeil, CHU de Bordeaux, 33076 Bordeaux, France

3. School of Psychology and MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, UK

4. King's College London, Institute of Psychiatry, Psychology and Neuroscience, London, UK

Psychotherapy and psychosomatics. 2017;86(3):187-188

A serious game (SG) is a digital application developed with a “serious” initial purpose that uses the game-playing aspect of video games (VGs). SGs constitute very promising tools in medicine and preliminary results have shown improvements in therapeutic education for chronic disorders, such as diabetes, or for rehabilitation programs [1].

Whereas the literature on mental health effects of VG initially focused on potential negative associations, for example aggressive thoughts and behaviours or depression [2], SGs have recently been proposed as innovative assessment instruments or non-pharmacological treatments for psychiatric disorders. Indeed, these disorders are related to alterations in the cognitive, affective, motivational, and social functions, which constitute relevant targets for SGs. For example, the potential benefit of SGs for adolescents with depression was shown by Merry et al. in a controlled randomized trial [3]. Despite these promising results, questions remain about the place that should be occupied by SGs in psychiatry.

Here, we propose a model in which SGs are integrated into the therapeutic toolbox for psychiatric disorders (see **Figure**). This model is based on the crucial concept of feedback. Three levels of feedback are identified: (i) game feedback, (ii), psychophysiological feedback and (iii) therapist feedback.

Game feedback

The most specific level of feedback regards the gaming interface. Feedback on a game itself is a core feature of SGs.

The choices and actions of the participant have direct consequences in this virtual environment, leading the patient to adapt his or her way of playing. This feedback allows for the reinforcement of select voluntary or involuntary behaviours through a reward/punishment system that is integrated in the game.

We would emphasize the importance of the ludic (i.e. characterized by playful outlook) features of SGs because these aspects can have a positive impact on motivation. Indeed, games are designed to be enjoyable, contrary to psychotherapeutic programs, which are designed to maximize efficiency. The playful aspects of a game may therefore help strengthen the patient's intrinsic motivation and maximize the therapeutic effects through better patient involvement. These “fun” features can be incorporated into software through scoring or quests challenging the player, which constitutes one of the main advantages of SGs over other media. However, the level of challenge and difficulty must be adapted to create and foster motivation (causing neither anxiety nor boredom). User needs and preferences must also be taken into account in user-centered and individualized game designs in order to maximize engagement [4]. Indeed, given the heterogeneity of psychiatric disorders and their pathophysiology, game designs have to be adapted [5]. For example, reward processes must be modified depending on the psychiatric disorder targeted. Different rewards should be implemented in the game (i.e. real-time scoring system, theme changes, prizes) and the therapist should be able to take into account the type and severity of the disorder, the patient preferences and any physiological parameters that may be available.

Psychophysiological feedback

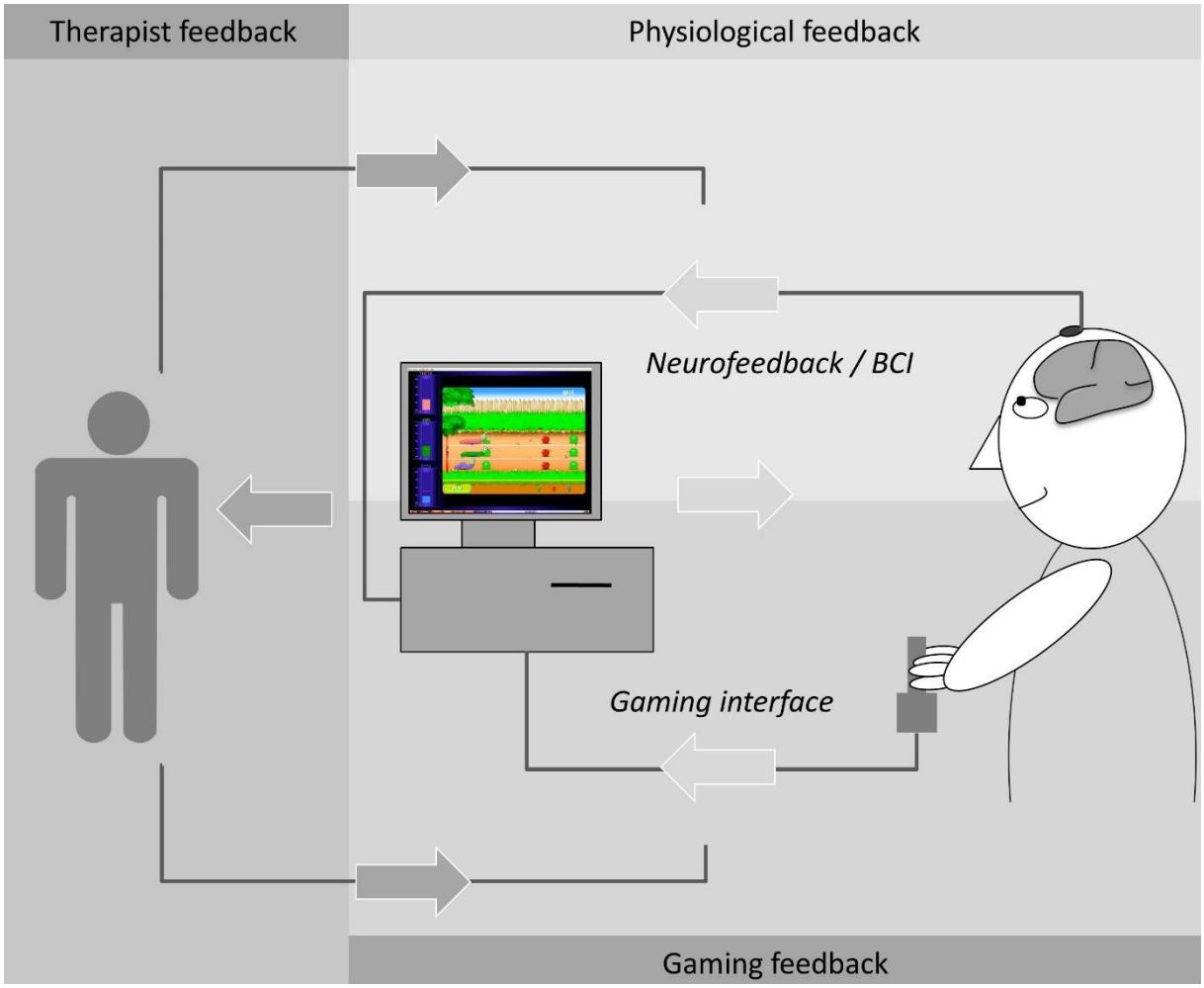
Feedback through psychophysiological information can be provided to the patient during his/her participation. This feedback refers to biofeedback, neurofeedback, or Brain Computer Interface (BCI) integration into SGs. The principle of such interfaces is simple: the subject receives information in real-time about a physiological variable (e.g. heart rate). In the case of neurofeedback or BCI, the physiological variable is neural activity. To date, two approaches have been described for this type of device: “active” and “passive”.

In the “active” scenario, the participant intentionally tries to control his/her cognitive activity to change his/her brain activity and control an external electronic device, in this case a SG. The goal is to enhance some voluntary or involuntary behaviour using a reward/punishment system. As such, a learning period is required. In the context of SGs, positive reinforcement can be achieved through scoring or unveiling clues in a quest. When a targeted neural activity is related to symptoms, these techniques may have therapeutic effects [6].

In contrast, for the “passive” approach, the real-time data streaming is used to optimize the user interface [7]. A passive BCI does not require a learning period, but it does improve the interaction between the subject and the game by adapting the content, structure, theme, and gameplay of a SG according to the variables measured. The ultimate goal is to increase the motivation of the participant and to improve the gaming experience. For example, the degree of difficulty of the game can be based on the electroencephalography (EEG) signal related to the level of attention of the subject [7].

Therapist feedback

The third type of feedback is provided by the therapist. Interestingly, recent meta-analyses identified a close relationship between alliance and the outcomes of individual psychotherapy [8] even for internet-based interventions [9]. SGs should then be considered as complementary tools to enrich the intersubjective relationship and not to substitute for the therapist. Indeed, the therapist could thus help the patient transfer skills and coping strategies acquired in the virtual environment to real life (*i.e.*, the “generalization” principle). The therapist also has a crucial role in reinforcing the motivation of the patient by adjusting the game settings (*e.g.*, levels of difficulty and reinforcing specific coping strategies) in a personalized way.



Serious game in psychotherapy: an integrative model

In conclusion, we propose an integrative framework for the use of SGs in psychotherapy, showing how patients could develop alternative coping strategies using this type of device. Promising initial results with SGs in psychiatry have already been obtained for depression, autism and attention deficit/hyperactivity disorder [10] but applications can probably extend to the entire field of mental and behavioural disorders. We recommend the use of the three different levels of feedback described above for the development of future SG software, which will also need validation through controlled randomized trials. A better understanding of how the challenging and ludic features of SGs can be optimized will be crucial in future studies. In complement, analysis of the three previously described levels of feedback could be considered an important research strategy for obtaining better understanding of the motivation and the coping strategies developed through therapy and, notably, the cognitive and neural underpinnings of such. In this regard, the integration of SGs into the psychiatric research framework offers new perspectives for innovative psychotherapy.

REFERENCES

1 Cheok G, Tan D, Low A, Hewitt J: Is Nintendo Wii an Effective Intervention for Individuals With Stroke? A Systematic Review and Meta-Analysis. J Am Med Dir Assoc 2015 Nov 1;16:923–932.

2 Tortolero SR, Peskin MF, Baumler ER, Cuccaro PM, Elliott MN, Davies SL, et al.: Daily violent video game playing and depression in preadolescent youth. Cyberpsychology Behav Soc Netw 2014 Sep;17:609–615.

3 Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG: The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012 Apr 19;344:e2598.

4 Fleming TM, Beurs D de, Khazaal Y, Gaggioli A, Riva G, Botella C, et al.: Maximizing the Impact of e-Therapy and Serious Gaming: Time for a Paradigm Shift. Front Psychiatry 2016;7. DOI: 10.3389/fpsyt.2016.00065

5 Suenderhauf C, Walter A, Lenz C, Lang UE, Borgwardt S: Counter striking psychosis: Commercial video games as potential treatment in schizophrenia? A systematic review of neuroimaging studies. Neurosci Biobehav Rev 2016 Sep;68:20–36.

6 Fovet T, Jardri R, Linden D: Current Issues in the Use of fMRI-Based Neurofeedback to Relieve Psychiatric Symptoms. Curr Pharm Des 2015;21:3384–3394.

7 Ewing KC, Fairclough SH, Gilleade K: Evaluation of an Adaptive Game that Uses EEG Measures Validated during the Design Process as Inputs to a Biocybernetic Loop. Front Hum Neurosci 2016;223.

8 Wampold BE: How important are the common factors in psychotherapy? An update. World Psychiatry 2015 Oct;14:270–277.

9 Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V: Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37:319–328.

10 Horne-Moyer HL, Moyer BH, Messer DC, Messer ES: The Use of Electronic Games in Therapy: a Review with Clinical Implications. Curr Psychiatry Rep 2014;16. DOI: 10.1007/s11920-014-0520-6

5 . CONCLUSION

“Any view of things that is not strange, is false.”

Neil Gaiman, Sandman

Ce travail de thèse constitue un pas, modeste mais significatif, vers le développement de stratégies thérapeutiques innovantes pour le traitement des hallucinations acoustico-verbales. Nos résultats sur la capture automatisée de l'hallucination auditive en IRMf permettent en effet de poser de solides fondations pour la mise en place de thérapies basées sur le décodage en temps-réel de l'activité cérébrale. En particulier, une interface de neurofeedback guidée par IRMf spécifique basée sur la dynamique temporelle du phénomène hallucinatoire est en cours de développement au sein de notre équipe et sera évaluée dans le cadre d'un essai contrôlé randomisé chez des patients souffrant de schizophrénie avec hallucinations acoustico-verbales pharmaco-résistantes (ANR-16-CE37-0015).

Si, comme nous en faisons l'hypothèse, ce protocole de neurofeedback permet l'identification de stratégies de coping spécifiques par les patients, cette nouvelle approche pourrait marquer l'avènement d'une véritable psychothérapie guidée par l'imagerie cérébrale pour les patients souffrant d'hallucinations auditives. Par ailleurs, l'absence d'intervention physique au cours des protocoles de neurofeedback (qui différencie cette technique des méthodes de neuromodulation comme la rTMS) pourrait permettre un entrainement au quotidien des patients, même après la fin du programme de neurofeedback.

Enfin, nous espérons que cette stratégie pourra avoir un impact majeur sur la qualité de vie des patients de par son impact sur le sentiment d'auto-efficacité. Il pourrait en effet s'avérer particulièrement intéressant d'évaluer l'impact du neurofeedback à travers le concept d'*empowerment*, une notion récente, issue de la psychologie de la santé et qui désigne l'augmentation de la capacité d'agir d'une personne souffrant de pathologie(s) par le biais du développement de son autonomie. Ainsi, le neurofeedback pourrait constituer une voie intéressante pour permettre aux patients de retrouver un sentiment de contrôle sur des symptômes extrêmement envahissants et stigmatisants.

6 . ANNEXES

Annexe 1 : questionnaire d'état pré-session

(© Aurore Hakoun, Samy Chikhi, François-Benoît Vialatte)

Pour chacune des propositions suivantes, indiquez la réponse qui correspond le plus à ce que vous éprouvez en ce moment entre les deux options proposées. Il n'y a pas de bonne ou de mauvaise réponse.

En ce moment :

Je me sens	Calme	□ □ □ □ □	Nerveux(se)
Je me sens	Endormi(e)	□ □ □ □ □	Réveillé(e)
Mon esprit a tendance spontanément à	S'évader	□ □ □ □ □	Rester dans le moment présent
A l'idée de faire cette tâche je me sens	Motivé	□ □ □ □ □	Ennuyé
Je me sens	Heureux(se)	□ □ □ □ □	Triste
Je me sens	Tendu(e)	□ □ □ □ □	Relaxé(e)
Lorsque mon esprit s'égare, j'arrive à me reconcentrer	Facilement	□ □ □ □ □	Difficilement
Je me sens	Satisfait(e)	□ □ □ □ □	Contrarié(e)

Annexe 2 : questionnaire d'état post-session

(© Aurore Hakoun, Samy Chikhi, François-Benoît Vialatte)

Pour chacune des propositions suivantes, indiquez la réponse qui correspond le plus à ce que vous éprouviez pendant cette session entre les deux options proposées. Il n'y a pas de bonne ou de mauvaise réponse.

Pendant cette session :

Le temps me semblait s'écouler	Rapidement	□ □ □ □ □	Lentement
Je me sentais	Calme	□ □ □ □ □	Nerveux(se)
Mon esprit avait tendance spontanément à	S'évader	□ □ □ □ □	Rester dans la tâche
Mon implication dans la tâche était	Légère	□ □ □ □ □	Résolue
Le signal de feedback me semblait	Déroutant	□ □ □ □ □	Prévisible
Mon niveau de confort pendant la tâche était	Elevé	□ □ □ □ □	Bas
Je me sentais	Endormi(e)	□ □ □ □ □	Réveillée
La tâche m'a paru	Ennuyeuse	□ □ □ □ □	Motivante
L'effort mental que j'ai fourni m'a paru	Élevé	□ □ □ □ □	Bas
De mon point de vue le feedback était un signal	Que j'observais	□ □ □ □ □	Qui venait de moi
Je me sentais	Heureux(se)	□ □ □ □ □	Triste
Lorsque mon esprit s'égarait, j'arrivais à me reconcentrer	Facilement	□ □ □ □ □	Difficilement
Je me sentais	Tendu(e)	□ □ □ □ □	Relaxé(e)
J'étais volontairement engagé dans la tâche	Assidument	□ □ □ □ □	Négligemment
Pendant la tâche, je me sentais	Accompagné	□ □ □ □ □	Seul
Le signal de feedback me semblait	Pertinent	□ □ □ □ □	Inadapté
Pendant la tâche les conditions (température, bruit, etc.) me semblaient	Défavorables	□ □ □ □ □	Favorables
Je me sentais	Satisfait(e)	□ □ □ □ □	Contrarié(e)
L'exigence mentale demandée par la tâche me semblait	Basse	□ □ □ □ □	Élevée
Le signal de feedback me semblait	Incontrôlable	□ □ □ □ □	Contrôlable
Je sentais que mes expérience et mes actions	Venaient de moi	□ □ □ □ □	Étaient contraintes

Annexe 3 : exploration métacognitions

(© Aurore Hakoun, Samy Chikhi, François-Benoît Vialatte)

<table>
<tr><td colspan="3">Avez-vous utilisé une (ou des) stratégie(s) pendant cette session ?</td><td>□ Oui
□ Non</td></tr>
<tr><td colspan="3">Avez-vous changé de stratégie en cours de session ?</td><td>□ Oui
□ Non</td></tr>
<tr><td colspan="2">Ce choix vous a-t-il paru efficace ?</td><td colspan="2">Inefficace □ □ □ □ □ Efficace</td></tr>
<tr><td>De quelle(s) stratégie(s) s’agissait-il ?</td><td colspan="3">Stratégie(s) :</td></tr>
<tr><td colspan="3">Souhaitez-vous utiliser une (des) stratégie(s) lors de la prochaine session ?</td><td>□ Oui
□ Non</td></tr>
<tr><td>Voici quelques pistes, laquelle (lesquelles) souhaiteriez-vous utiliser ?</td><td colspan="3">Stratégie(s) :</td></tr>
<tr><td>Cet entretien vous a-t-il paru utile ?</td><td>□ Oui
□ Non</td><td colspan="2">Pourquoi ? :</td></tr>
</table>

BIBLIOGRAPHIE

1. Ey H. Traité des Hallucinations. Masson; 1973.
2. Berrios GE, Dening TR. Pseudohallucinations: a conceptual history. Psychol Med. juill 1996;26(4):753-63.
3. Amad A, Bubrovszky M, Maitre E, Thomas P. La schizophrénie chez l'adulte. In: Pathologies schizophréniques. Médecine Sciences Publications; 2012. (Psychiatrie).
4. El Haj M, Roche J, Jardri R, Kapogiannis D, Gallouj K, Antoine P. Clinical and neurocognitive aspects of hallucinations in Alzheimer's disease. Neurosci Biobehav Rev. 21 févr 2017;
5. Fénelon G, Mahieux F, Huon R, Ziégler M. Hallucinations in Parkinson's disease: prevalence, phenomenology and risk factors. Brain J Neurol. avr 2000;123 (Pt 4):733-45.
6. Pezzoli S, Cagnin A, Bandmann O, Venneri A. Structural and Functional Neuroimaging of Visual Hallucinations in Lewy Body Disease: A Systematic Literature Review. Brain Sci. 15 juill 2017;7(7).
7. Baethge C, Baldessarini RJ, Freudenthal K, Streeruwitz A, Bauer M, Bschor T. Hallucinations in bipolar disorder: characteristics and comparison to unipolar depression and schizophrenia. Bipolar Disord. avr 2005;7(2):136-45.
8. Kingdon DG, Ashcroft K, Bhandari B, Gleeson S, Warikoo N, Symons M, et al. Schizophrenia and borderline personality disorder: similarities and differences in the experience of auditory hallucinations, paranoia, and childhood trauma. J Nerv Ment Dis. juin 2010;198(6):399-403.
9. Linscott RJ, van Os J. An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. Psychol Med. juin 2013;43(6):1133-49.
10. Johns LC. Hallucinations in the general population. Curr Psychiatry Rep. juin 2005;7(3):162-7.
11. Baumeister D, Sedgwick O, Howes O, Peters E. Auditory verbal hallucinations and continuum models of psychosis: A systematic review of the healthy voice-hearer literature. Clin Psychol Rev. févr 2017;51:125-41.
12. Larøi F, Sommer IE, Blom JD, Fernyhough C, Ffytche DH, Hugdahl K, et al. The characteristic features of auditory verbal hallucinations in clinical and nonclinical groups: state-of-the-art overview and future directions. Schizophr Bull. juin 2012;38(4):724-33.
13. Cutting J, Dunne F. Subjective experience of schizophrenia. Schizophr Bull. 1989;15(2):217-31.
14. Collectif. Référentiel de psychiatrie et addictologie : Psychiatrie de l'adulte, psychiatrie de l'enfant et de l'adolescent, addictologie. 2e édition. Tours: Presses universitaire François Rabelais; 2016. 583 p.
15. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry. juill 2010;167(7):748-51.
16. van Os J. « Salience syndrome » replaces « schizophrenia » in DSM-V and ICD-11: psychiatry's evidence-based entry into the 21st century? Acta Psychiatr Scand. nov 2009;120(5):363-72.
17. Ford JM, Morris SE, Hoffman RE, Sommer I, Waters F, McCarthy-Jones S, et al. Studying hallucinations within the NIMH RDoC framework. Schizophr Bull. juill 2014;40 Suppl 4:S295-304.
18. Lantéri-Laura G. Les hallucinations. Paris; Milan: Elsevier Masson; 1997. 178 p.
19. Peyroux E, Franck N. An Epistemological Approach: History of Concepts and Ideas About Hallucinations in Classical Psychiatry. In: The Neuroscience of Hallucinations. Springer, New York, NY; 2013. p. 3-20.
20. Parys Magalie. Introduction to mental illness in Ancient Mesopotamia. In: Rencontre Assyriologique Internationale 2014. Varsovie; 2014.
21. Amad A, Thomas P. Histoire de la maladie mentale dans le Moyen-Orient médiéval. Ann Méd-Psychol Rev Psychiatr. 1 juill 2011;169(6):373-6.
22. Foucault M. Histoire de la folie à l'âge classique. Paris: Gallimard; 1976. 688 p.
23. Deleuze G. Foucault. Paris: Editions de Minuit; 2004. 144 p.
24. Esquirol É. Des maladies mentales considérées sous les rapports médical, hygiénique et médico-légal. Lyon: Ch. Savy; 1838.
25. Gimenez G, Guimont M, Pedinielli J-L. Étude de l'évolution du concept d'hallucination dans la littérature psychiatrique classique. Evol Psychiatr (Paris). 68(2):289-98.
26. Lanteri-Laura G. Signification clinique et psychopathologique des hallucinations. LÉvolution Psychiatr. 1 avr 2000;65(2):245-54.
27. Asaad G, Shapiro B. Hallucinations: theoretical and clinical overview. Am J Psychiatry. sept 1986;143(9):1088-97.
28. Waters F, Collerton D, Ffytche DH, Jardri R, Pins D, Dudley R, et al. Visual hallucinations in the psychosis spectrum and comparative information from neurodegenerative disorders and eye disease. Schizophr Bull. juill 2014;40 Suppl 4:S233-245.
29. McCarthy-Jones S, Smailes D, Corvin A, Gill M, Morris DW, Dinan TG, et al. Occurrence and co-occurrence of hallucinations by modality in schizophrenia-spectrum disorders. Psychiatry Res. juin 2017;252:154-60.
30. Lim A, Hoek HW, Deen ML, Blom JD, GROUP Investigators. Prevalence and classification of hallucinations in multiple sensory modalities in schizophrenia spectrum disorders. Schizophr Res. oct 2016;176(2-3):493-9.
31. Shergill SS, Murray RM, McGuire PK. Auditory hallucinations: a review of psychological treatments. Schizophr Res. 17 août 1998;32(3):137-50.

32. Kane JM, Kishimoto T, Correll CU. Non-adherence to medication in patients with psychotic disorders: epidemiology, contributing factors and management strategies. World Psychiatry Off J World Psychiatr Assoc WPA. oct 2013;12(3):216-26.
33. Harkavy-Friedman JM, Kimhy D, Nelson EA, Venarde DF, Malaspina D, Mann JJ. Suicide attempts in schizophrenia: the role of command auditory hallucinations for suicide. J Clin Psychiatry. août 2003;64(8):871-4.
34. Upthegrove R, Birchwood M, Ross K, Brunett K, McCollum R, Jones L. The evolution of depression and suicidality in first episode psychosis. Acta Psychiatr Scand. sept 2010;122(3):211-8.
35. Birchwood M, Iqbal Z, Upthegrove R. Psychological pathways to depression in schizophrenia: studies in acute psychosis, post psychotic depression and auditory hallucinations. Eur Arch Psychiatry Clin Neurosci. juin 2005;255(3):202-12.
36. Rocca P, Castagna F, Mongini T, Montemagni C, Bogetto F. Relative contributions of psychotic symptoms and insight to quality of life in stable schizophrenia. Psychiatry Res. 15 mai 2010;177(1-2):71-6.
37. Aleman A, Larøi F. Hallucinations: The science of idiosyncratic perception. American Psychological Association; 2008.
38. Johns LC, Hemsley D, Kuipers E. A comparison of auditory hallucinations in a psychiatric and non-psychiatric group. Br J Clin Psychol Br Psychol Soc. mars 2002;41(Pt 1):81-6.
39. Nayani TH, David AS. The auditory hallucination: a phenomenological survey. Psychol Med. janv 1996;26(1):177-89.
40. Birchwood M, Chadwick P. The omnipotence of voices: testing the validity of a cognitive model. Psychol Med. nov 1997;27(6):1345-53.
41. Llorca PM, Pereira B, Jardri R, Chereau-Boudet I, Brousse G, Misdrahi D, et al. Hallucinations in schizophrenia and Parkinson's disease: an analysis of sensory modalities involved and the repercussion on patients. Sci Rep. 1 déc 2016;6:38152.
42. Upthegrove R, Broome MR, Caldwell K, Ives J, Oyebode F, Wood SJ. Understanding auditory verbal hallucinations: a systematic review of current evidence. Acta Psychiatr Scand. mai 2016;133(5):352-67.
43. Ratcliff K, Farhall J, Shawyer F. Auditory hallucinations: a review of assessment tools. Clin Psychol Psychother. déc 2011;18(6):524-34.
44. de Chazeron I, Pereira B, Chereau-Boudet I, Brousse G, Misdrahi D, Fénelon G, et al. Validation of a Psycho-Sensory hAllucinations Scale (PSAS) in schizophrenia and Parkinson's disease. Schizophr Res. févr 2015;161(2-3):269-76.
45. Larøi F, Marczewski P, Van der Linden M. Further evidence of the multi-dimensionality of hallucinatory predisposition: factor structure of a modified version of the Launay-Slade Hallucinations Scale in a normal sample. Eur Psychiatry J Assoc Eur Psychiatr. févr 2004;19(1):15-20.
46. Demeulemeester M, Kochman F, Fligans B, Tabet AJ, Thomas P, Jardri R. Assessing early-onset hallucinations in the touch-screen generation. Br J Psychiatry J Ment Sci. mars 2015;206(3):181-3.
47. Inzelberg R, Kipervasser S, Korczyn AD. Auditory hallucinations in Parkinson's disease. J Neurol Neurosurg Psychiatry. avr 1998;64(4):533-5.
48. Aleman A, Böcker KBE, Hijman R, de Haan EHF, Kahn RS. Cognitive basis of hallucinations in schizophrenia: role of top-down information processing. Schizophr Res. 15 nov 2003;64(2-3):175-85.
49. Allen P, Larøi F, McGuire PK, Aleman A. The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. Neurosci Biobehav Rev. 2008;32(1):175-91.
50. McCarthy-Jones S, Thomas N, Strauss C, Dodgson G, Jones N, Woods A, et al. Better than mermaids and stray dogs? Subtyping auditory verbal hallucinations and its implications for research and practice. Schizophr Bull. juill 2014;40 Suppl 4:S275-284.
51. Belzeaux R, Cermolacce M, Jardri R. Hallucinations: toward a dialogue between phenomenology and brain imaging research. J Consciousness Studies. In press;
52. Frith CD. The Cognitive Neuropsychology of Schizophrenia. Psychology Press; 1992. 192 p.
53. Feinberg I. Efference copy and corollary discharge: implications for thinking and its disorders. Schizophr Bull. 1978;4(4):636-40.
54. Ford JM, Mathalon DH. Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? Int J Psychophysiol Off J Int Organ Psychophysiol. déc 2005;58(2-3):179-89.
55. McGuire PK, Silbersweig DA, Wright I, Murray RM, Frackowiak RS, Frith CD. The neural correlates of inner speech and auditory verbal imagery in schizophrenia: relationship to auditory verbal hallucinations. Br J Psychiatry J Ment Sci. août 1996;169(2):148-59.
56. Johns LC, McGuire PK. Verbal self-monitoring and auditory hallucinations in schizophrenia. Lancet Lond Engl. 6 févr 1999;353(9151):469-70.
57. Allen PP, Johns LC, Fu CHY, Broome MR, Vythelingum GN, McGuire PK. Misattribution of external speech in patients with hallucinations and delusions. Schizophr Res. 1 août 2004;69(2-3):277-87.
58. Johns LC, Gregg L, Allen P, McGuire PK. Impaired verbal self-monitoring in psychosis: effects of state, trait and diagnosis. Psychol Med. avr 2006;36(4):465-74.
59. Cho R, Wu W. Mechanisms of auditory verbal hallucination in schizophrenia. Front Psychiatry. 2013;4:155.
60. Vercammen A, Knegtering H, Bruggeman R, Aleman A. Subjective loudness and reality of auditory verbal hallucinations and activation of the inner speech processing network. Schizophr Bull. sept 2011;37(5):1009-16.
61. Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. Psychol Bull. juill 1993;114(1):3-28.
62. Brébion G, Amador X, David A, Malaspina D, Sharif Z, Gorman JM. Positive symptomatology and source-monitoring failure in schizophrenia--an analysis of symptom-specific effects. Psychiatry Res. 21 août 2000;95(2):119-31.

63. Bentall RP, Slade PD. Reality testing and auditory hallucinations: a signal detection analysis. Br J Clin Psychol Br Psychol Soc. sept 1985;24 (Pt 3):159-69.
64. Vercammen A, de Haan EHF, Aleman A. Hearing a voice in the noise: auditory hallucinations and speech perception. Psychol Med. août 2008;38(8):1177-84.
65. Waters F, Woodward T, Allen P, Aleman A, Sommer I. Self-recognition deficits in schizophrenia patients with auditory hallucinations: a meta-analysis of the literature. Schizophr Bull. juin 2012;38(4):741-50.
66. Brookwell ML, Bentall RP, Varese F. Externalizing biases and hallucinations in source-monitoring, self-monitoring and signal detection studies: a meta-analytic review. Psychol Med. déc 2013;43(12):2465-75.
67. Allen P, Larøi F, McGuire PK, Aleman A. The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. Neurosci Biobehav Rev. 2008;32:175–191.
68. Wilkinson S. Accounting for the phenomenology and varieties of auditory verbal hallucination within a predictive processing framework. Conscious Cogn. nov 2014;30:142-55.
69. Friston K. The free-energy principle: a unified brain theory? Nat Rev Neurosci. févr 2010;11(2):127-38.
70. Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci. janv 2009;10(1):48-58.
71. Jardri R, Denève S. Circular inferences in schizophrenia. Brain J Neurol. nov 2013;136(Pt 11):3227-41.
72. Horga G, Schatz KC, Abi-Dargham A, Peterson BS. Deficits in predictive coding underlie hallucinations in schizophrenia. J Neurosci Off J Soc Neurosci. 11 juin 2014;34(24):8072-82.
73. García-Martí G, Aguilar EJ, Lull JJ, Martí-Bonmatí L, Escartí MJ, Manjón JV, et al. Schizophrenia with auditory hallucinations: a voxel-based morphometry study. Prog Neuropsychopharmacol Biol Psychiatry. 1 janv 2008;32(1):72-80.
74. Flaum M, O'Leary DS, Swayze VW 2nd, Miller DD, Arndt S, Andreasen NC. Symptom dimensions and brain morphology in schizophrenia and related psychotic disorders. J Psychiatr Res. août 1995;29(4):261-76.
75. Modinos G, Costafreda SG, van Tol M-J, McGuire PK, Aleman A, Allen P. Neuroanatomy of auditory verbal hallucinations in schizophrenia: a quantitative meta-analysis of voxel-based morphometry studies. Cortex J Devoted Study Nerv Syst Behav. avr 2013;49(4):1046-55.
76. Hubl D, Dougoud-Chauvin V, Zeller M, Federspiel A, Boesch C, Strik W, et al. Structural analysis of Heschl's gyrus in schizophrenia patients with auditory hallucinations. Neuropsychobiology. 2010;61(1):1-9.
77. Crespo-Facorro B, Kim J, Andreasen NC, O'Leary DS, Bockholt HJ, Magnotta V. Insular cortex abnormalities in schizophrenia: a structural magnetic resonance imaging study of first-episode patients. Schizophr Res. 30 nov 2000;46(1):35-43.
78. Crespo-Facorro B, Roiz-Santiáñez R, Quintero C, Pérez-Iglesias R, Tordesillas-Gutiérrez D, Mata I, et al. Insular cortex morphometry in first-episode schizophrenia-spectrum patients: Diagnostic specificity and clinical correlations. J Psychiatr Res. avr 2010;44(5):314-20.
79. Wylie KP, Tregellas JR. The role of the insula in schizophrenia. Schizophr Res. nov 2010;123(2-3):93-104.
80. Huang P, Xi Y, Lu Z-L, Chen Y, Li X, Li W, et al. Decreased bilateral thalamic gray matter volume in first-episode schizophrenia with prominent hallucinatory symptoms: A volumetric MRI study. Sci Rep. 2015;5:14505.
81. Neckelmann G, Specht K, Lund A, Ersland L, Smievoll AI, Neckelmann D, et al. Mr morphometry analysis of grey matter volume reduction in schizophrenia: association with hallucinations. Int J Neurosci. janv 2006;116(1):9-23.
82. Cachia A, Amad A, Brunelin J, Krebs M-O, Plaze M, Thomas P, et al. Deviations in cortex sulcation associated with visual hallucinations in schizophrenia. Mol Psychiatry. sept 2015;20(9):1101-7.
83. Mangin J-F, Jouvent E, Cachia A. In-vivo measurement of cortical morphology: means and meanings. Curr Opin Neurol. août 2010;23(4):359-67.
84. Cachia A, Mangin JF, Rivière D, Kherif F, Boddaert N, Andrade A, et al. A primal sketch of the cortex mean curvature: a morphogenesis based approach to study the variability of the folding patterns. IEEE Trans Med Imaging. juin 2003;22(6):754-65.
85. Cachia A, Paillère-Martinot M-L, Galinowski A, Januel D, de Beaurepaire R, Bellivier F, et al. Cortical folding abnormalities in schizophrenia patients with resistant auditory hallucinations. NeuroImage. 1 févr 2008;39(3):927-35.
86. Plaze M, Paillère-Martinot M-L, Penttilä J, Januel D, de Beaurepaire R, Bellivier F, et al. « Where do auditory hallucinations come from? »--a brain morphometry study of schizophrenia patients with inner or outer space hallucinations. Schizophr Bull. janv 2011;37(1):212-21.
87. Rapoport JL, Giedd JN, Gogtay N. Neurodevelopmental model of schizophrenia: update 2012. Mol Psychiatry. déc 2012;17(12):1228-38.
88. Gaser C, Nenadic I, Volz H-P, Büchel C, Sauer H. Neuroanatomy of « hearing voices »: a frontotemporal brain structural abnormality associated with auditory hallucinations in schizophrenia. Cereb Cortex N Y N 1991. janv 2004;14(1):91-6.
89. Jardri R, Sommer IE. Functional brain imaging of hallucinations: Symptom capture studies. In: R Jardri, A Cachia, P Thomas & D Pins (Eds): The Neuroscience of Hallucinations. NY: Springer; 2013. p. 375-91.
90. Ford JM, Palzes VA, Roach BJ, Potkin SG, van Erp TGM, Turner JA, et al. Visual hallucinations are associated with hyperconnectivity between the amygdala and visual cortex in people with a diagnosis of schizophrenia. Schizophr Bull. janv 2015;41(1):223-32.
91. Kühn S, Gallinat J. Quantitative meta-analysis on state and trait aspects of auditory verbal hallucinations in schizophrenia. Schizophr Bull. juin 2012;38(4):779-86.

92. Allen P, Modinos G, Hubl D, Shields G, Cachia A, Jardri R, et al. Neuroimaging auditory hallucinations in schizophrenia: from neuroanatomy to neurochemistry and beyond. Schizophr Bull. juin 2012;38(4):695-703.
93. Belin P, Fecteau S, Bédard C. Thinking the voice: neural correlates of voice perception. Trends Cogn Sci. mars 2004;8(3):129-35.
94. Yao B, Belin P, Scheepers C. Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. J Cogn Neurosci. oct 2011;23(10):3146-52.
95. Plaze M, Bartr s-Faz D, Martinot J-L, Januel D, Bellivier F, De Beaurepaire R, et al. Left superior temporal gyrus activation during sentence perception negatively correlates with auditory hallucination severity in schizophrenia patients. Schizophr Res. 2006;87(1-3):109–115.
96. Swick D, Turken AU. Dissociation between conflict detection and error monitoring in the human anterior cingulate cortex. Proc Natl Acad Sci U S A. 10 déc 2002;99(25):16354-9.
97. Northoff G, Bermpohl F. Cortical midline structures and the self. Trends Cogn Sci. mars 2004;8(3):102-7.
98. Garrison JR, Fernyhough C, McCarthy-Jones S, Haggard M, Australian Schizophrenia Research Bank, Simons JS. Paracingulate sulcus morphology is associated with hallucinations in the human brain. Nat Commun. 2015;6:8956.
99. McGuire PK, Silbersweig DA, Wright I, Murray RM, David AS, Frackowiak RS, et al. Abnormal monitoring of inner speech: a physiological basis for auditory hallucinations. Lancet. 2 sept 1995;346(8975):596-600.
100. Allen P, Amaro E, Fu CHY, Williams SCR, Brammer MJ, Johns LC, et al. Neural correlates of the misattribution of speech in schizophrenia. Br J Psychiatry J Ment Sci. févr 2007;190:162-9.
101. Jardri R, Pins D, Lafargue G, Very E, Ameller A, Delmaire C, et al. Increased overlap between the brain areas involved in self-other distinction in schizophrenia. PloS One. 2011;6(3):e17500.
102. Jardri R, Pouchet A, Pins D, Thomas P. Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. Am J Psychiatry. janv 2011;168(1):73-81.
103. Kompus K, Westerhausen R, Hugdahl K. The « paradoxical » engagement of the primary auditory cortex in patients with auditory verbal hallucinations: a meta-analysis of functional neuroimaging studies. Neuropsychologia. oct 2011;49(12):3361-9.
104. Zmigrod L, Garrison JR, Carr J, Simons JS. The neural mechanisms of hallucinations: A quantitative meta-analysis of neuroimaging studies. Neurosci Biobehav Rev. oct 2016;69:113-23.
105. Shergill SS, Brammer MJ, Williams SC, Murray RM, McGuire PK. Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. Arch Gen Psychiatry. nov 2000;57(11):1033-8.
106. Shergill SS, Brammer MJ, Williams SCR, Murray RM, McGuire PK. Mapping Auditory Hallucinations in Schizophrenia Using Functional Magnetic Resonance Imaging. Arch Gen Psychiatry. 1 nov 2000;57(11):1033.
107. Hunter MD, Eickhoff SB, Miller TWR, Farrow TFD, Wilkinson ID, Woodruff PWR. Neural activity in speech-sensitive auditory cortex during silence. Proc Natl Acad Sci U S A. 3 janv 2006;103(1):189-94.
108. Homan P, Kindler J, Hauf M, Walther S, Hubl D, Dierks T. Repeated measurements of cerebral blood flow in the left superior temporal gyrus reveal tonic hyperactivity in patients with auditory verbal hallucinations: a possible trait marker. Front Hum Neurosci. 2013;7:304.
109. Sommer IEC, Diederen KMJ, Blom J-D, Willems A, Kushan L, Slotema K, et al. Auditory verbal hallucinations predominantly activate the right inferior frontal area. Brain J Neurol. déc 2008;131(Pt 12):3169-77.
110. McGuire PK, Shah GM, Murray RM. Increased blood flow in Broca's area during auditory hallucinations in schizophrenia. Lancet Lond Engl. 18 sept 1993;342(8873):703-6.
111. Copolov DL, Seal ML, Maruff P, Ulusoy R, Wong MTH, Tochon-Danguy HJ, et al. Cortical activation associated with the experience of auditory hallucinations and perception of human speech in schizophrenia: a PET correlation study. Psychiatry Res. 1 avr 2003;122(3):139-52.
112. Rossell SL. The role of memory retrieval and emotional salience in the emergence of auditory hallucinations. In: The neuroscience of hallucinations. Springer. New York; 2012. p. 137-52.
113. Waters FAV, Badcock JC, Michie PT, Maybery MT. Auditory hallucinations in schizophrenia: intrusive thoughts and forgotten memories. Cognit Neuropsychiatry. janv 2006;11(1):65-83.
114. Dierks T, Linden DE, Jandl M, Formisano E, Goebel R, Lanfermann H, et al. Activation of Heschl's gyrus during auditory hallucinations. Neuron. mars 1999;22(3):615-21.
115. van de Ven VG, Formisano E, Röder CH, Prvulovic D, Bittner RA, Dietz MG, et al. The spatiotemporal pattern of auditory cortical responses during verbal hallucinations. NeuroImage. sept 2005;27(3):644-55.
116. Jardri R, Thomas P, Delmaire C, Delion P, Pins D. The Neurodynamic Organization of Modality-Dependent Hallucinations. Cereb Cortex. 5 janv 2013;23(5):1108-17.
117. Leroy A, Foucher JR, Pins D, Delmaire C, Thomas P, Roser MM, et al. fMRI capture of auditory hallucinations: Validation of the two-steps method. Hum Brain Mapp. 28 juin 2017;
118. van Lutterveld R, Diederen KMJ, Koops S, Begemann MJH, Sommer IEC. The influence of stimulus detection on activation patterns during auditory hallucinations. Schizophr Res. avr 2013;145(1-3):27-32.
119. Benetti S, Pettersson-Yeo W, Mechelli A. Connectivity issues of the hallucinating brain. In: R Jardri, A Cachia, P Thomas & D Pins (Eds): The Neuroscience of Hallucinations. NY: Springer; 2013. p. 417-43.
120. Sporns O. Brain connectivity. Scholarpedia. 28 oct 2007;2(10):4695.
121. Alderson-Day B, McCarthy-Jones S, Fernyhough C. Hearing voices in the resting brain: A review of intrinsic functional connectivity research on auditory verbal hallucinations. Neurosci Biobehav Rev. août 2015;55:78-87.

122. Pettersson-Yeo W, Allen P, Benetti S, McGuire P, Mechelli A. Dysconnectivity in schizophrenia: where are we now? Neurosci Biobehav Rev. avr 2011;35(5):1110-24.

123. Oertel-Knöchel V, Knöchel C, Matura S, Prvulovic D, Linden DEJ, van de Ven V. Reduced functional connectivity and asymmetry of the planum temporale in patients with schizophrenia and first-degree relatives. Schizophr Res. juill 2013;147(2-3):331-8.

124. Mou X, Bai F, Xie C, Shi J, Yao Z, Hao G, et al. Voice recognition and altered connectivity in schizophrenic patients with auditory hallucinations. Prog Neuropsychopharmacol Biol Psychiatry. juill 2013;44:265-70.

125. Amad A, Cachia A, Gorwood P, Pins D, Delmaire C, Rolland B, et al. The multimodal connectivity of the hippocampal complex in auditory and visual hallucinations. Mol Psychiatry. 15 janv 2013;

126. Rolland B, Amad A, Poulet E, Bordet R, Vignaud A, Bation R, et al. Resting-State Functional Connectivity of the Nucleus Accumbens in Auditory and Visual Hallucinations in Schizophrenia. Schizophr Bull. 21 juill 2014;

127. Curcic-Blake B, Liemburg E, Vercammen A, Swart M, Knegtering H, Bruggeman R, et al. When Broca goes uninformed: reduced information flow to Broca's area in schizophrenia patients with auditory hallucinations. Schizophr Bull. sept 2013;39(5):1087-95.

128. Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. Ann N Y Acad Sci. mars 2008;1124:1-38.

129. Broyd SJ, Demanuele C, Debener S, Helps SK, James CJ, Sonuga-Barke EJS. Default-mode brain dysfunction in mental disorders: a systematic review. Neurosci Biobehav Rev. mars 2009;33(3):279-96.

130. Menon V, Uddin LQ. Saliency, switching, attention and control: a network model of insula function. Brain Struct Funct. juin 2010;214(5-6):655-67.

131. Alderson-Day B, Diederen K, Fernyhough C, Ford JM, Horga G, Margulies DS, et al. Auditory Hallucinations and the Brain's Resting-State Networks: Findings and Methodological Observations. Schizophr Bull. sept 2016;42(5):1110-23.

132. Bentall RP, Haddock G, Slade PD. Cognitive behavior therapy for persistent auditory hallucinations: From theory to therapy. Behav Ther. 1 déc 1994;25(1):51-66.

133. Lefebvre S, Demeulemeester M, Leroy A, Delmaire C, Lopes R, Pins D, et al. Network dynamics during the different stages of hallucinations in schizophrenia. Hum Brain Mapp. juill 2016;37(7):2571-86.

134. Wheeler AL, Voineskos AN. A review of structural neuroimaging in schizophrenia: from connectivity to connectomics. Front Hum Neurosci. 2014;8:653.

135. Rotarska-Jagiela A, Oertel-Knoechel V, DeMartino F, van de Ven V, Formisano E, Roebroeck A, et al. Anatomical brain connectivity and positive symptoms of schizophrenia: a diffusion tensor imaging study. Psychiatry Res. 30 oct 2009;174(1):9-16.

136. Ćurčić-Blake B, Nanetti L, van der Meer L, Cerliani L, Renken R, Pijnenborg GHM, et al. Not on speaking terms: hallucinations and structural network disconnectivity in schizophrenia. Brain Struct Funct. janv 2015;220(1):407-18.

137. Geoffroy PA, Houenou J, Duhamel A, Amad A, De Weijer AD, Curčić-Blake B, et al. The arcuate fasciculus in auditory-verbal hallucinations: A meta-analysis of diffusion-tensor-imaging studies. Schizophr Res. 9 août 2014;

138. Seok J-H, Park H-J, Chun J-W, Lee S-K, Cho HS, Kwon JS, et al. White matter abnormalities associated with auditory hallucinations in schizophrenia: a combined study of voxel-based analyses of diffusion tensor imaging and structural magnetic resonance imaging. Psychiatry Res. 15 nov 2007;156(2):93-104.

139. de Weijer AD, Mandl RCW, Diederen KMJ, Neggers SFW, Kahn RS, Hulshoff Pol HE, et al. Microstructural alterations of the arcuate fasciculus in schizophrenia patients with frequent auditory verbal hallucinations. Schizophr Res. août 2011;130(1-3):68-77.

140. McCarthy-Jones S, Oestreich LKL, Australian Schizophrenia Research Bank, Whitford TJ. Reduced integrity of the left arcuate fasciculus is specifically associated with auditory verbal hallucinations in schizophrenia. Schizophr Res. mars 2015;162(1-3):1-6.

141. de Weijer AD, Neggers SFW, Diederen KMS, Mandl RCW, Kahn RS, Hulshoff Pol HE, et al. Aberrations in the arcuate fasciculus are associated with auditory verbal hallucinations in psychotic and in non-psychotic individuals. Hum Brain Mapp. mars 2013;34(3):626-34.

142. Shergill SS, Kanaan RA, Chitnis XA, O'Daly O, Jones DK, Frangou S, et al. A diffusion tensor imaging study of fasciculi in schizophrenia. Am J Psychiatry. mars 2007;164(3):467-73.

143. Catani M, Craig MC, Forkel SJ, Kanaan R, Picchioni M, Toulopoulou T, et al. Altered integrity of perisylvian language pathways in schizophrenia: relationship to auditory hallucinations. Biol Psychiatry. 15 déc 2011;70(12):1143-50.

144. Mulert C, Kirsch V, Whitford TJ, Alvarado J, Pelavin P, McCarley RW, et al. Hearing voices: a role of interhemispheric auditory connectivity? World J Biol Psychiatry Off J World Fed Soc Biol Psychiatry. févr 2012;13(2):153-8.

145. Milev P, Ho B-C, Arndt S, Nopoulos P, Andreasen NC. Initial magnetic resonance imaging volumetric brain measurements and outcome in schizophrenia: a prospective longitudinal study with 5-year follow-up. Biol Psychiatry. 15 sept 2003;54(6):608-15.

146. Kindler J, Homan P, Jann K, Federspiel A, Flury R, Hauf M, et al. Reduced neuronal activity in language-related regions after transcranial magnetic stimulation therapy for auditory verbal hallucinations. Biol Psychiatry. 15 mars 2013;73(6):518-24.

147. Homan P, Kindler J, Hauf M, Hubl D, Dierks T. Cerebral blood flow identifies responders to transcranial magnetic stimulation in auditory verbal hallucinations. Transl Psychiatry. 2012;2:e189.

148. Kahn RS, Fleischhacker WW, Boter H, Davidson M, Vergouwe Y, Keet IPM, et al. Effectiveness of antipsychotic drugs in first-episode schizophrenia and schizophreniform disorder: an open randomised clinical trial. Lancet Lond Engl. 29 mars 2008;371(9618):1085-97.

149. Sommer IEC, Slotema CW, Daskalakis ZJ, Derks EM, Blom JD, van der Gaag M. The treatment of hallucinations in schizophrenia spectrum disorders. Schizophr Bull. juin 2012;38(4):704-14.

150. Buchanan RW, Kreyenbuhl J, Kelly DL, Noel JM, Boggs DL, Fischer BA, et al. The 2009 schizophrenia PORT psychopharmacological treatment recommendations and summary statements. Schizophr Bull. janv 2010;36(1):71-93.

151. Tiihonen J, Haukka J, Taylor M, Haddad PM, Patel MX, Korhonen P. A nationwide cohort study of oral and depot antipsychotics after first hospitalization for schizophrenia. Am J Psychiatry. juin 2011;168(6):603-9.

152. Kane JM, Honigfeld G, Singer J, Meltzer H. Clozapine in treatment-resistant schizophrenics. Psychopharmacol Bull. 1988;24(1):62-7.

153. Wang C-Y, Xiang Y-T, Cai Z-J, Weng Y-Z, Bo Q-J, Zhao J-P, et al. Risperidone maintenance treatment in schizophrenia: a randomized, controlled trial. Am J Psychiatry. juin 2010;167(6):676-85.

154. Leucht C, Heres S, Kane JM, Kissling W, Davis JM, Leucht S. Oral versus depot antipsychotic drugs for schizophrenia--a critical systematic review and meta-analysis of randomised long-term trials. Schizophr Res. avr 2011;127(1-3):83-92.

155. Goozée R, Handley R, Kempton MJ, Dazzan P. A systematic review and meta-analysis of the effects of antipsychotic medications on regional cerebral blood flow (rCBF) in schizophrenia: association with response to treatment. Neurosci Biobehav Rev. juin 2014;43:118-36.

156. Vita A, De Peri L, Deste G, Barlati S, Sacchetti E. The Effect of Antipsychotic Treatment on Cortical Gray Matter Changes in Schizophrenia: Does the Class Matter? A Meta-analysis and Meta-regression of Longitudinal Magnetic Resonance Imaging Studies. Biol Psychiatry. 15 sept 2015;78(6):403-12.

157. Radua J, Borgwardt S, Crescini A, Mataix-Cols D, Meyer-Lindenberg A, McGuire PK, et al. Multimodal meta-analysis of structural and functional brain changes in first episode psychosis and the effects of antipsychotic medication. Neurosci Biobehav Rev. nov 2012;36(10):2325-33.

158. Mojtabai R, Nicholson RA, Carpenter BN. Role of psychosocial treatments in management of schizophrenia: a meta-analytic review of controlled outcome studies. Schizophr Bull. 1998;24(4):569-87.

159. Jardri R, Favrod J, Larøi F. Psychothérapies des hallucinations. Issy-les-Moulineaux: Elsevier Masson; 2016. 352 p.

160. Mazmanian S, Mirabel-Sarron C, Dardennes R. Les thérapies cognitivo-comportementales face aux hallucinations auditives dans la schizophrénie : une revue de la littérature. J Thérapie Comport Cogn. 1 nov 2015;25(4):177-88.

161. Wykes T, Steel C, Everitt B, Tarrier N. Cognitive behavior therapy for schizophrenia: effect sizes, clinical models, and methodological rigor. Schizophr Bull. mai 2008;34(3):523-37.

162. Trower P, Birchwood M, Meaden A, Byrne S, Nelson A, Ross K. Cognitive therapy for command hallucinations: randomised controlled trial. Br J Psychiatry J Ment Sci. avr 2004;184:312-20.

163. Jenner JA, Nienhuis FJ, van de Willige G, Wiersma D. « Hitting » voices of schizophrenia patients may lastingly reduce persistent auditory hallucinations and their burden: 18-month outcome of a randomized controlled trial. Can J Psychiatry Rev Can Psychiatr. mars 2006;51(3):169-77.

164. Thomas N, Rossell S, Farhall J, Shawyer F, Castle D. Cognitive behavioural therapy for auditory hallucinations: effectiveness and predictors of outcome in a specialist clinic. Behav Cogn Psychother. mars 2011;39(2):129-38.

165. Turner DT, van der Gaag M, Karyotaki E, Cuijpers P. Psychological interventions for psychosis: a meta-analysis of comparative outcome studies. Am J Psychiatry. mai 2014;171(5):523-38.

166. Wykes T, Parr AM, Landau S. Group treatment of auditory hallucinations. Exploratory study of effectiveness. Br J Psychiatry J Ment Sci. août 1999;175:180-5.

167. Penn DL, Meyer PS, Evans E, Wirth RJ, Cai K, Burchinal M. A randomized controlled trial of group cognitive-behavioral therapy vs. enhanced supportive therapy for auditory hallucinations. Schizophr Res. avr 2009;109(1-3):52-9.

168. Newman Taylor K, Harper S, Chadwick P. Impact of mindfulness on cognition and affect in voice hearing: evidence from two case studies. Behav Cogn Psychother. juill 2009;37(4):397-402.

169. Jenner JA, Nienhuis FJ, Wiersma D, van de Willige G. Hallucination focused integrative treatment: a randomized controlled trial. Schizophr Bull. 2004;30(1):133-45.

170. Wiersma D, Jenner JA, Nienhuis FJ, van de Willige G. Hallucination focused integrative treatment improves quality of life in schizophrenia patients. Acta Psychiatr Scand. mars 2004;109(3):194-201.

171. Antal A, Alekseichuk I, Bikson M, Brockmöller J, Brunoni AR, Chen R, et al. Low intensity transcranial electric stimulation: Safety, ethical, legal regulatory and application guidelines. Clin Neurophysiol Off J Int Fed Clin Neurophysiol. sept 2017;128(9):1774-809.

172. Lawrie SM, Buechel C, Whalley HC, Frith CD, Friston KJ, Johnstone EC. Reduced frontotemporal functional connectivity in schizophrenia associated with auditory hallucinations. Biol Psychiatry. 15 juin 2002;51(12):1008-11.

173. Jardri R, Pouchet A, Pins D, Thomas P. Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. Am J Psychiatry. janv 2011;168(1):73-81.

174. Sanfilipo M, Lafargue T, Rusinek H, Arena L, Loneragan C, Lautin A, et al. Volumetric measure of the frontal and temporal lobe regions in schizophrenia: relationship to negative symptoms. Arch Gen Psychiatry. mai 2000;57(5):471-80.

175. Brunelin J, Mondino M, Haesebaert F, Saoud M, Suaud-Chagny MF, Poulet E. Efficacy and safety of bifocal tDCS as an interventional treatment for refractory schizophrenia. Brain Stimulat. juill 2012;5(3):431-2.

176. Brunelin J, Mondino M, Gassab L, Haesebaert F, Gaha L, Suaud-Chagny M-F, et al. Examining transcranial direct-current stimulation (tDCS) as a treatment for hallucinations in schizophrenia. Am J Psychiatry. juill 2012;169(7):719-24.

177. Mondino M, Brunelin J, Palm U, Brunoni AR, Poulet E, Fecteau S. Transcranial Direct Current Stimulation for the Treatment of Refractory Symptoms of Schizophrenia. Current Evidence and Future Directions. Curr Pharm Des. 2015;21(23):3373-83.

178. Bose A, Shivakumar V, Narayanaswamy JC, Nawani H, Subramaniam A, Agarwal SM, et al. Insight facilitation with add-on tDCS in schizophrenia. Schizophr Res. juin 2014;156(1):63-5.

179. Mondino M, Haesebaert F, Poulet E, Suaud-Chagny M-F, Brunelin J. Fronto-temporal transcranial Direct Current Stimulation (tDCS) reduces source-monitoring deficits and auditory hallucinations in patients with schizophrenia. Schizophr Res. févr 2015;161(2-3):515-6.

180. Mondino M, Jardri R, Suaud-Chagny M-F, Saoud M, Poulet E, Brunelin J. Effects of Fronto-Temporal Transcranial Direct Current Stimulation on Auditory Verbal Hallucinations and Resting-State Functional Connectivity of the Left Temporo-Parietal Junction in Patients With Schizophrenia. Schizophr Bull. mars 2016;42(2):318-26.

181. Andrade C. Once- to twice-daily, 3-year domiciliary maintenance transcranial direct current stimulation for severe, disabling, clozapine-refractory continuous auditory hallucinations in schizophrenia. J ECT. sept 2013;29(3):239-42.

182. Pondé PH, de Sena EP, Camprodon JA, de Araújo AN, Neto MF, DiBiasi M, et al. Use of transcranial direct current stimulation for the treatment of auditory hallucinations of schizophrenia - a systematic review. Neuropsychiatr Dis Treat. 2017;13:347-55.

183. McClintock SM, Reti IM, Carpenter LL, McDonald WM, Dubin M, Taylor SF, et al. Consensus Recommendations for the Clinical Application of Repetitive Transcranial Magnetic Stimulation (rTMS) in the Treatment of Depression. J Clin Psychiatry. 23 mai 2017;

184. Slotema CW, Blom JD, Hoek HW, Sommer IEC. Should we expand the toolbox of psychiatric treatment methods to include Repetitive Transcranial Magnetic Stimulation (rTMS)? A meta-analysis of the efficacy of rTMS in psychiatric disorders. J Clin Psychiatry. juill 2010;71(7):873-84.

185. Kubera KM, Barth A, Hirjak D, Thomann PA, Wolf RC. Noninvasive brain stimulation for the treatment of auditory verbal hallucinations in schizophrenia: methods, effects and challenges. Front Syst Neurosci. 2015;9:131.

186. Lefaucheur J-P, André-Obadia N, Poulet E, Devanne H, Haffen E, Londero A, et al. [French guidelines on the use of repetitive transcranial magnetic stimulation (rTMS): safety and therapeutic indications]. Neurophysiol Clin Clin Neurophysiol. déc 2011;41(5-6):221-95.

187. Demeulemeester M, Amad A, Bubrovszky M, Pins D, Thomas P, Jardri R. What is the real effect of 1-Hz repetitive transcranial magnetic stimulation on hallucinations? Controlling for publication bias in neuromodulation trials. Biol Psychiatry. 15 mars 2012;71(6):e15-16.

188. Sparing R, Buelte D, Meister IG, Paus T, Fink GR. Transcranial magnetic stimulation and the challenge of coil placement: a comparison of conventional and stereotaxic neuronavigational strategies. Hum Brain Mapp. janv 2008;29(1):82-96.

189. Multimodal MRI-guided rTMS to Treat Refractory Hallucinations (The MULTIMODHAL project) [Internet]. Disponible sur: https://clinicaltrials.gov/ct2/show/NCT01373866)

190. Silbersweig DA, Stern E, Frith C, Cahill C, Holmes A, Grootoonk S, et al. A functional neuroanatomy of hallucinations in schizophrenia. Nature. 9 nov 1995;378(6553):176-9.

191. Lennox BR, Park SB, Medley I, Morris PG, Jones PB. The functional anatomy of auditory hallucinations in schizophrenia. Psychiatry Res. 20 nov 2000;100(1):13-20.

192. Sommer IEC, Diederen KMJ, Blom J-D, Willems A, Kushan L, Slotema K, et al. Auditory verbal hallucinations predominantly activate the right inferior frontal area. Brain J Neurol. déc 2008;131(Pt 12):3169-77.

193. Shergill SS, Cameron LA, Brammer MJ, Williams SC, Murray RM, McGuire PK. Modality specific neural correlates of auditory and somatic hallucinations. J Neurol Neurosurg Psychiatry. nov 2001;71(5):688-90.

194. Bazán PR, Biazoli CE, Sato JR, Amaro E. Motor Readiness Increases Brain Connectivity Between Default-Mode Network and Motor Cortex: Impact on Sampling Resting Periods from fMRI Event-Related Studies. Brain Connect. déc 2015;5(10):631-40.

195. Tan B-L. Profile of cognitive problems in schizophrenia and implications for vocational functioning. Aust Occup Ther J. août 2009;56(4):220-8.

196. Formisano E, Esposito F, Di Salle F, Goebel R. Cortex-based independent component analysis of fMRI time series. Magn Reson Imaging. déc 2004;22(10):1493-504.

197. Jardri R, Pins D, Delmaire C, Goeb J-L, Thomas P. Activation of bilateral auditory cortex during verbal hallucinations in a child with schizophrenia. Mol Psychiatry. avr 2007;12(4):319.

198. Jardri R, Pins D, Bubrovszky M, Lucas B, Lethuc V, Delmaire C, et al. Neural functional organization of hallucinations in schizophrenia: multisensory dissolution of pathological emergence in consciousness. Conscious Cogn. juin 2009;18(2):449-57.

199. Jardri R, Thomas P, Delmaire C, Delion P, Pins D. The neurodynamic organization of modality-dependent hallucinations. Cereb Cortex N Y N 1991. mai 2013;23(5):1108-17.

200. Vapnik VN. The Nature of Statistical Learning Theory [Internet]. New York, NY: Springer New York; 2000 [cité 18 févr 2016]. Disponible sur: http://link.springer.com/10.1007/978-1-4757-3264-1

201. Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. NeuroImage. mars 2009;45(1 Suppl):S199-209.

202. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev. avr 2012;36(4):1140-52.

203. Veronese E, Castellani U, Peruzzo D, Bellani M, Brambilla P. Machine learning approaches: from theory to application in schizophrenia. Comput Math Methods Med. 2013;2013:867924.

204. Mourao-Miranda J, Reinders A a. TS, Rocha-Rego V, Lappin J, Rondina J, Morgan C, et al. Individualized prediction of illness course at the first psychotic episode: a support vector machine MRI study. Psychol Med. mai 2012;42(5):1037-47.

205. Dazzan P, Arango C, Fleischacker W, Galderisi S, Glenthøj B, Leucht S, et al. Magnetic resonance imaging and the prediction of outcome in first-episode schizophrenia: a review of current evidence and directions for future research. Schizophr Bull. mai 2015;41(3):574-83.

206. Ellison-Wright I, Glahn DC, Laird AR, Thelen SM, Bullmore E. The anatomy of first-episode and chronic schizophrenia: an anatomical likelihood estimation meta-analysis. Am J Psychiatry. août 2008;165(8):1015-23.

207. Bloemen OJN, de Koning MB, Schmitz N, Nieman DH, Becker HE, de Haan L, et al. White-matter markers for psychosis in a prospective ultra-high-risk cohort. Psychol Med. août 2010;40(8):1297-304.

208. Liu L, Fu L, Zhang X, Zhang J, Zhang X, Xu B, et al. Combination of dynamic (11)C-PIB PET and structural MRI improves diagnosis of Alzheimer's disease. Psychiatry Res. 30 août 2015;233(2):131-40.

209. Diederen KMJ, Neggers SFW, Daalman K, Blom JD, Goekoop R, Kahn RS, et al. Deactivation of the parahippocampal gyrus preceding auditory hallucinations in schizophrenia. Am J Psychiatry. avr 2010;167(4):427-35.

210. Hoffman RE, Anderson AW, Varanko M, Gore JC, Hampson M. Time course of regional brain activation associated with onset of auditory/verbal hallucinations. Br J Psychiatry J Ment Sci. nov 2008;193(5):424-5.

211. Gonsalves BD, Kahn I, Curran T, Norman KA, Wagner AD. Memory strength and repetition suppression: multimodal imaging of medial temporal cortical contributions to recognition. Neuron. 1 sept 2005;47(5):751-61.

212. Lennox BR, Park SB, Jones PB, Morris PG, Park G. Spatial and temporal mapping of neural activity associated with auditory hallucinations. Lancet Lond Engl. 20 févr 1999;353(9153):644.

213. Shergill SS, Brammer MJ, Amaro E, Williams SCR, Murray RM, McGuire PK. Temporal course of auditory hallucinations. Br J Psychiatry J Ment Sci. déc 2004;185:516-7.

214. Lefebvre S, Demeulemeester M, Leroy A, Delmaire C, Lopes R, Pins D, et al. Network dynamics during the different stages of hallucinations in schizophrenia. Hum Brain Mapp. juill 2016;37(7):2571-86.

215. Duchesnay E, Bourgin J, Fouad HS, Frouin V. Méthodes de prédiction multivariées basées sur la neuroimagerie : application aux maladies psychiatriques. Ann Méd-Psychol Rev Psychiatr. 23 mars 2015;173.

216. Sterman MB, Friar L. Suppression of seizures in an epileptic following sensorimotor EEG feedback training. Electroencephalogr Clin Neurophysiol. juill 1972;33(1):89-95.

217. Lubar JF, Shouse MN. EEG and behavioral changes in a hyperkinetic child concurrent with training of the sensorimotor rhythm (SMR): a preliminary report. Biofeedback Self-Regul. sept 1976;1(3):293-306.

218. Elbert T, Rockstroh B, Lutzenberger W, Birbaumer N. Biofeedback of slow cortical potentials. I. Electroencephalogr Clin Neurophysiol. mars 1980;48(3):293-301.

219. Rockstroh B, Elbert T, Birbaumer N, Wolf P, Düchting-Röth A, Reker M, et al. Cortical self-regulation in patients with epilepsies. Epilepsy Res. janv 1993;14(1):63-72.

220. deCharms RC. Applications of real-time fMRI. Nat Rev Neurosci. sept 2008;9(9):720-9.

221. Fovet T, Jardri R, Linden D. Current Issues in the Use of fMRI-Based Neurofeedback to Relieve Psychiatric Symptoms. Curr Pharm Des. 2015;21(23):3384-94.

222. Clerc M, Bougrain L, Lotte F. Les interfaces cerveau-ordinateur 1. Sciences cognitives. London: ISTE éditions; 2016. 310 p.

223. Thibault RT, Raz A. When can neurofeedback join the clinical armamentarium? Lancet Psychiatry. 1 juin 2016;3(6):497-8.

224. Arns M. A NExT Step for neurofeedback in France. L'Encephale. avr 2017;43(2):97-8.

225. Thibault RT, Raz A. Neurofeedback: the power of psychosocial therapeutics. Lancet Psychiatry. 1 nov 2016;3(11):e18.

226. Micoulaud-Franchi J-A, Fovet T. Neurofeedback: time needed for a promising non-pharmacological therapeutic method. Lancet Psychiatry. sept 2016;3(9):e16.

227. Schönenberg M, Wiedemann E, Schneidt A, Scheeff J, Logemann A, Keune PM, et al. Neurofeedback, sham neurofeedback, and cognitive-behavioural group therapy in adults with attention-deficit hyperactivity disorder: a triple-blind, randomised, controlled trial. Lancet Psychiatry. sept 2017;4(9):673-84.

228. Cortese S, Ferrin M, Brandeis D, Holtmann M, Aggensteiner P, Daley D, et al. Neurofeedback for Attention-Deficit/Hyperactivity Disorder: Meta-Analysis of Clinical and Neuropsychological Outcomes From Randomized Controlled Trials. J Am Acad Child Adolesc Psychiatry. juin 2016;55(6):444-55.

229. Micoulaud-Franchi J-A, Geoffroy PA, Fond G, Lopez R, Bioulac S, Philip P. EEG neurofeedback treatments in children with ADHD: an updated meta-analysis of randomized controlled trials. Front Hum Neurosci. 2014;8:906.

230. Arns M, Strehl U. Evidence for efficacy of neurofeedback in ADHD? Am J Psychiatry. juill 2013;170(7):799-800.

231. Cortese S, Brandeis D, Holtmann M, Sonuga-Barke EJS, Asherson P, Banaschewki T, et al. The European ADHD Guidelines Group replies: J Am Acad Child Adolesc Psychiatry. 1 déc 2016;55(12):1092-3.

232. Micoulaud-Franchi J-A, Salvo F, Bioulac S, Fovet T. Neurofeedback in Attention-Deficit/Hyperactivity Disorder: Efficacy. J Am Acad Child Adolesc Psychiatry. déc 2016;55(12):1091-2.

233. Schabus M, Griessenberger H, Gnjezda M-T, Heib DPJ, Wislowska M, Hoedlmoser K. Better than sham? A double-blind placebo-controlled neurofeedback study in primary insomnia. Brain J Neurol. 1 avr 2017;140(4):1041-52.

234. Thibault RT, Lifshitz M, Raz A. Neurofeedback or neuroplacebo? Brain J Neurol. 1 avr 2017;140(4):862-4.

235. McCarthy-Jones S. Taking back the brain: could neurofeedback training be effective for relieving distressing auditory verbal hallucinations in patients with schizophrenia? Schizophr Bull. juin 2012;38(4):678-82.

236. Gruzelier J, Hardman E, Wild J, Zaman R. Learned control of slow potential interhemispheric asymmetry in schizophrenia. Int J Psychophysiol Off J Int Organ Psychophysiol. déc 1999;34(3):341-8.

237. PhD ASB. Neurofeedback Treatment of Chronic Inpatient Schizophrenia. J Neurother. 3 mars 2010;14(1):47-54.

238. Surmeli T, Ertem A, Eralp E, Kos IH. Schizophrenia and the efficacy of qEEG-guided neurofeedback treatment: a clinical case series. Clin EEG Neurosci. avr 2012;43(2):133-44.

239. Nan W, Wan F, Chang L, Pun SH, Vai MI, Rosa A. An Exploratory Study of Intensive Neurofeedback Training for Schizophrenia. Behav Neurol [Internet]. 2017 [cité 27 sept 2017];2017. Disponible sur: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5497641/

240. Schneider F, Rockstroh B, Heimann H, Lutzenberger W, Mattes R, Elbert T, et al. Self-regulation of slow cortical potentials in psychiatric patients: Schizophrenia. Biofeedback Self-Regul. 1 déc 1992;17(4):277-92.

241. Dyck MS, Mathiak KA, Bergert S, Sarkheil P, Koush Y, Alawi EM, et al. Targeting Treatment-Resistant Auditory Verbal Hallucinations in Schizophrenia with fMRI-Based Neurofeedback - Exploring Different Cases of Schizophrenia. Front Psychiatry. 2016;7:37.

242. Cordes JS, Mathiak KA, Dyck M, Alawi EM, Gaber TJ, Zepf FD, et al. Cognitive and neural strategies during control of the anterior cingulate cortex by fMRI neurofeedback in patients with schizophrenia. Front Behav Neurosci. 2015;9:169.

243. Ruiz S, Lee S, Soekadar SR, Caria A, Veit R, Kircher T, et al. Acquired self-control of insula cortex modulates emotion recognition and brain network connectivity in schizophrenia. Hum Brain Mapp. janv 2013;34(1):200-12.

244. Gomes JS, Ducos DV, Gadelha A, Ortiz BB, Deusen AMV, Akiba HT, et al. Hemoencephalography self-regulation training and its impact on cognition: A study with schizophrenia and healthy participants. Schizophr Res [Internet]. 4 sept 2017 [cité 12 sept 2017];0(0). Disponible sur: http://www.schres-journal.com/article/S0920-9964(17)30520-0/fulltext

245. Waters FAV, Badcock JC, Michie PT, Maybery MT. Auditory hallucinations in schizophrenia: intrusive thoughts and forgotten memories. Cognit Neuropsychiatry. janv 2006;11(1):65-83.

246. Keizer AW, Verment RS, Hommel B. Enhancing cognitive control through neurofeedback: a role of gamma-band activity in managing episodic retrieval. NeuroImage. 15 févr 2010;49(4):3404-13.

247. Rieger K, Rarra M-H, Moor N, Diaz Hernandez L, Baenninger A, Razavi N, et al. Neurofeedback-Based Enhancement of Single Trial Auditory Evoked Potentials: Feasibility in Healthy Subjects. Clin EEG Neurosci. 1 mai 2017;1550059417708935.

248. Diaz Hernandez L, Rieger K, Baenninger A, Brandeis D, Koenig T. Towards Using Microstate-Neurofeedback for the Treatment of Psychotic Symptoms in Schizophrenia. A Feasibility Study in Healthy Participants. Brain Topogr. mars 2016;29(2):308-21.

249. Isaac Y, Barthélemy Q, Gouy-Pailler C, Atif J, Sebag M. Généralisation des micro-états EEG par apprentissage régularisé temporellement de dictionnaires topographiques. In: Colloque Gretsi XXV [Internet]. Lyon, France; 2015 [cité 27 sept 2017]. Disponible sur: https://hal.archives-ouvertes.fr/hal-01200542

250. Rieger K, Diaz Hernandez L, Baenninger A, Koenig T. 15 Years of Microstate Research in Schizophrenia - Where Are We? A Meta-Analysis. Front Psychiatry. 2016;7:22.

251. Shibata K, Watanabe T, Sasaki Y, Kawato M. Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. Science. 9 déc 2011;334(6061):1413-5.

252. LaConte SM, Peltier SJ, Hu XP. Real-time fMRI using brain-state classification. Hum Brain Mapp. oct 2007;28(10):1033-44.

253. deBettencourt MT, Cohen JD, Lee RF, Norman KA, Turk-Browne NB. Closed-loop training of attention with real-time brain imaging. Nat Neurosci. mars 2015;18(3):470-5.

254. Shibata K, Watanabe T, Kawato M, Sasaki Y. Differential Activation Patterns in the Same Brain Region Led to Opposite Emotional States. PLoS Biol. sept 2016;14(9):e1002546.

255. Watanabe T, Sasaki Y, Shibata K, Kawato M. Advances in fMRI Real-Time Neurofeedback. Trends Cogn Sci. 11 oct 2017;

256. Cortese A, Amano K, Koizumi A, Kawato M, Lau H. Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. Nat Commun. 15 déc 2016;7:ncomms13669.

257. Amano K, Shibata K, Kawato M, Sasaki Y, Watanabe T. Learning to Associate Orientation with Color in Early Visual Areas by Associative Decoded fMRI Neurofeedback. Curr Biol CB. 25 juill 2016;26(14):1861-6.

258. P. Morrison A. The interpretation of intrusions in psychosis: An integrative cognitive approach to hallucinations and delusions. Behav Cogn Psychother. 1 juill 2001;29:257-76.

259. Langer ÁI, Stanghellini G, Cangas AJ, Lysaker PH, Nieto-Muñoz L, Moriana JA, et al. Interpretation, emotional reactions and daily life implications of hallucination-like experiences in clinical and nonclinical populations. Psicothema. 2015;27(1):19-25.

260. Freeman D, Garety PA. Connecting neurosis and psychosis: the direct influence of emotion on delusions and hallucinations. Behav Res Ther. août 2003;41(8):923-47.

261. Smith B, Fowler DG, Freeman D, Bebbington P, Bashforth H, Garety P, et al. Emotion and psychosis: links between depression, self-esteem, negative schematic beliefs and delusions and hallucinations. Schizophr Res. sept 2006;86(1-3):181-8.

262. Larøi F, Haan S de, Jones S, Raballo A. Auditory verbal hallucinations: Dialoguing between the cognitive sciences and phenomenology. Phenomenol Cogn Sci. 1 juin 2010;9(2):225-40.

263. Dodgson G, Gordon S. Avoiding false negatives: are some auditory hallucinations an evolved design flaw? Behav Cogn Psychother. mai 2009;37(3):325-34.

264. Badcock JC, Paulik G, Maybery MT. The role of emotion regulation in auditory hallucinations. Psychiatry Res. 28 févr 2011;185(3):303-8.

265. Koole SL. The psychology of emotion regulation: An integrative review. Cogn Emot. 1 janv 2009;23(1):4-41.

266. Birbaumer N, Ruiz S, Sitaram R. Learned regulation of brain metabolism. Trends Cogn Sci. juin 2013;17(6):295-302.

267. Gaume A, Vialatte A, Mora-Sánchez A, Ramdani C, Vialatte FB. A psychoengineering paradigm for the neurocognitive mechanisms of biofeedback and neurofeedback. Neurosci Biobehav Rev. sept 2016;68:891-910.

268. Strehl U, Birkle SM, Wörz S, Kotchoubey B. Sustained reduction of seizures in patients with intractable epilepsy after self-regulation training of slow cortical potentials - 10 years after. Front Hum Neurosci. 2014;8:604.

269. Kober SE, Witte M, Ninaus M, Neuper C, Wood G. Learning to modulate one's own brain activity: the effect of spontaneous mental strategies. Front Hum Neurosci. 2013;7:695.

270. Ninaus M, Kober SE, Witte M, Koschutnig K, Stangl M, Neuper C, et al. Neural substrates of cognitive control under the belief of getting neurofeedback training. Front Hum Neurosci. 2013;7:914.

271. Fovet T, Micoulaud-Franchi J-A, Vaiva G, Thomas P, Jardri R, Amad A. [Serious game as a therapeutic tool in psychiatry: A systematic review]. L'Encephale. oct 2016;42(5):463-9.

Auteur : Thomas Fovet
Date de Soutenance : 15 décembre 2017
Titre de la thèse : Détection automatisée des hallucinations auditives en IRM fonctionnelle et perspectives thérapeutiques dans la schizophrénie.
Thèse d'Université – Ecole Doctorale Biologie Santé de Lille
Mots–clés : Hallucination, IRM fonctionnelle, imagerie de capture, apprentissage machine, temps-réel, neurofeedback.

RESUME

L'hallucination est une expérience subjective vécue en pleine conscience consistant en une perception impossible à distinguer d'une perception réelle, mais survenant en l'absence de tout stimulus en provenance de l'environnement externe. Les symptômes hallucinatoires, qui peuvent concerner toutes les modalités sensorielles, sont retrouvés dans divers troubles neurologiques et psychiatriques mais également chez certains sujets indemnes de toute pathologie. Dans le champ de la psychiatrie, la pathologie la plus fréquemment associée aux hallucinations reste la schizophrénie et la modalité auditive est la plus représentée, puisque 60 à 80% des patients souffrant de ce trouble sont concernés. Le retentissement fonctionnel des hallucinations auditives peut être important, altérant significativement la qualité de vie des patients.

Dans ce contexte, la prise en charge de ce type de symptômes s'avère un enjeu considérable pour les personnes souffrant de schizophrénie. Pourtant, les moyens thérapeutiques actuellement disponibles (traitements médicamenteux antipsychotiques notamment) ne permettent pas toujours une rémission complète de la symptomatologie hallucinatoire et l'on considère que 25 à 30% des hallucinations auditives sont « pharmaco-résistantes ». C'est à partir de ce constat que, ces dernières années, ont émergé, pour le traitement des hallucinations auditives, des techniques de neuromodulation comme la stimulation magnétique transcrânienne répétée ou la stimulation électrique transcrânienne par courant continu. Toutefois, les résultats de ces nouvelles thérapies sur les hallucinations auditives résistantes restent modérés et le développement de stratégies alternatives demeure un enjeu de recherche majeur.

Actuellement, les travaux en imagerie fonctionnelle permettent d'affiner les modèles physiopathologiques des hallucinations auditives, mais leur intérêt pourrait aller au-delà de la recherche fondamentale, avec possiblement des applications cliniques telles que l'assistance thérapeutique. Ce travail de thèse s'inscrit précisément dans le développement de l'imagerie cérébrale de « capture » des hallucinations auditives, c'est-à-dire l'identification des patterns d'activation fonctionnels associés à la survenue des hallucinations auditives.

La première partie de ce travail est consacrée à la détection automatisée des hallucinations auditives en IRM fonctionnelle. L'identification des périodes hallucinatoires survenues au cours d'une session d'IRM fonctionnelle est actuellement possible par une méthode de capture semi-automatisée validée. Celle-ci permet une labellisation des données acquises au cours d'une session de repos en périodes « hallucinatoires » et « non-hallucinatoires ». Toutefois, le caractère long et fastidieux de cette méthode limite largement son emploi. Nous avons donc souhaité montrer comment les stratégies d'apprentissage machine (*support vector machine* ou SVM, notamment) permettent l'automatisation de cette technique par le développement de classificateurs performants, généralisables et associés à un faible coût de calcul (indispensable en vue d'une utilisation en temps réel). Nous proposons également le développement d'algorithmes de reconnaissance de la période « pré-hallucinatoire », en mettant en évidence que ce type de classificateur présente aussi des performances largement significatives. Enfin, nous avons pu montrer que l'utilisation de stratégies d'apprentissage-machine alternatives au SVM (*e.g,* le *TV-Elastic-net*), obtient des performances significativement supérieures au SVM.

La deuxième partie de cette thèse propose une réflexion théorique sur les perspectives thérapeutiques offertes par le développement de ces stratégies de capture de l'hallucination auditive. Le neurofeedback est une méthode thérapeutique non-invasive consistant à mesurer l'activité d'une ou de plusieurs régions cérébrales chez un sujet et à lui présenter en temps réel l'enregistrement de cette activité. Nous présentons les avancées récentes de cette technique dans le champ de la médecine, mais rappelons également les polémiques qu'elle suscite de par le faible nombre d'essais contrôlés randomisés actuellement disponibles. A partir d'un travail de revue systématique de la littérature sur l'utilisation du neurofeedback guidé par IRM fonctionnelle pour le traitement des troubles psychiatriques, nous explorons les différentes stratégies envisageables pour mettre en place un protocole visant spécifiquement la prise en charge des hallucinations auditives.

Enfin, nous proposons à partir de cette réflexion théorique, un protocole de neurofeedback guidé par IRM fonctionnelle basé sur une interface multi-classificateurs permettant l'identification des différentes périodes de l'hallucination auditive (phase dite d'entrée dans l'hallucination, phase hallucinatoire, phase de sortie de l'hallucination, période sans hallucination) et nous présentons l'étude que nous souhaitons mettre en place afin de tester son efficacité dans un essai contrôlé randomisé.

COMPOSITION DU JURY

Rapporteurs :	Pr. Arnaud CACHIA, Université Sorbonne-Paris-Cité
	Dr. Emilie OLIE, CHU & Université de Montpellier
Assesseurs :	Dr. Ingrid DE CHAZERON, CHU de Clermont-Ferrand
	Pr. Pierre THOMAS, Université & CHU de Lille
	Dr. Delphine PINS, CNRS & Université de Lille
Directeur de thèse :	Pr. Renaud JARDRI, Université & CHU de Lille