

Numéro d'ordre : 1731

Année : 1996

THESE

présentée en vue de l'obtention du

Doctorat de l'Université des Sciences et technologies de Lille
Spécialité : Informatique

par

Laurent BARME



Le Son dans les Systèmes Informatiques pour le Travail Coopératif

Présentée le 30 avril 1996

Jury :

Président

M. Jean-Marc GEIB

Directeur

M. Alain DERYCKE

Rapporteurs

M. Michel BEAUDOUIN-LAFON

Mme Joëlle COUTAZ

Examineurs

Mme Marie-France BARTHET

M. Claude KINTZIG

à Marie-Françoise,

pour Louis.

Remerciements

Je remercie mon directeur de thèse, tout particulièrement pour les conditions de travail, les moyens exceptionnels dont il m'a permis de disposer, les documents auxquels il m'a donné accès et les personnes qu'il m'a fait rencontrer.

Je souhaite exprimer ma plus profonde reconnaissance à Joëlle Coutaz pour son soutien décisif lors des rencontres doctorales d'IHM'95, ses nombreux et très utiles commentaires sur la première ébauche de mon mémoire et pour avoir accepté d'être rapporteur de cette thèse.

Je remercie également Michel Beaudouin-Lafon, Jean-Marc Geib, Marie-France Barthet et Claude Kintzig qui m'ont fait l'honneur d'accepter de faire partie de mon jury, respectivement en tant que rapporteur, président et examinateurs. Je remercie Marie-France Barthet aussi pour ses conseils et les informations qui m'ont permis d'améliorer mon approche du problème de l'évaluation.

Je remercie Danièle Clément, Maître de Conférence en Psychologie, avec qui le travail coopératif passe du concept à la réalité, Patricia Plénacoste qui a consacré beaucoup de temps à me conseiller pour la rédaction de cette thèse, Isabelle Logez ainsi que tous les autres membres du laboratoire Trigone qui m'ont aidé plus qu'ils ne le pensent.

Il y a encore bien d'autres personnes que j'ai eu le plaisir de rencontrer à l'occasion de cette thèse. Qu'ils me pardonnent de ne pas citer leurs noms, mais qu'ils soient remerciés pour leur aide.

Sans Marie-Françoise, je n'aurais jamais entrepris cette thèse. Sans elle, je n'aurais jamais pu en venir à bout. Les remerciements que je lui dois ne peuvent pas s'exprimer ici.

Cette thèse n'aurait pas non plus abouti sans le soutien, les conseils tant sur le plan professionnel qu'humain et les encouragements de mes parents et amis. Qu'ils en soient remerciés ici.

Enfin, je réserve mon plus grand Merci pour le plus petit. A partir du 20 février 1994, Louis Barme a fait plus pour cette thèse que n'importe qui : il est la source d'une énergie inestimable ...

Table des Matières

Introduction	13
Chapitre 1 : Présentation des travaux antérieurs exploités dans cette thèse	19
1 Notions de base concernant le son	19
1.1 Caractéristiques physiques et psychophysiques	19
1.1.1 Définitions de base	19
1.1.2 Description analytique d'un son	20
1.1.3 L'intensité et le timbre d'un son	23
1.1.4 Qualité sonore de la manipulation du son	25
1.1.5 Localisation de l'origine spatiale d'un son	26
1.2 Manipulation technique du son	27
1.2.1 Prise de son, restitution et transport	27
1.2.2 Le problème de la compression du son	31
1.3 Manipulation informatique du son	35
1.3.1 Les possibilités sonores des machines actuelles	35
1.3.2 Le partage des ressources sonores dans un système informatique	37
2 Exploitation du son dans les systèmes informatiques	40
2.1 Notification, rétroaction et sonification : un premier niveau d'exploitation du son	40
2.2 Annotations sonores : exploitation élémentaire de la voix	44
2.3 Détection de prise de parole	47
2.4 Spatialisation du son	48
3 Les travaux relatifs aux aspects sonores des collecticiels temps réel et multimédia	51
3.1 Les projets orientés communication multiparties	53
3.1.1 Une approche liée aux problèmes d'acheminement et de routage dans les réseaux	54
3.1.2 Une approche plus abstraite par rapport au réseau de transport	56
3.2 L'intégration du son dans une architecture logicielle	58
3.3 Approche centrée sur l'utilisateur	60
Chapitre 2 : Services proposés pour l'aspect sonore des collecticiels synchrones distribués	63
1 Services offerts par les systèmes actuels : analyse d'un cas concret	63
1.1 Description de la pré-expérimentation de CoLearn	64
1.1.1 Configuration matérielle et géographique	64
1.1.2 Fonctionnement de l'audioconférence	65
1.2 Observations effectuées en première analyse	66
1.2.1 Critiques de l'aspect technique du système	66

Table des matières

1.2.2 Les canaux sonores d'une audioconférence : une chaîne complexe	68
1.2.3 Apports et limites du canal sonore	69
1.3 Améliorations suggérées à partir de cette première expérience	71
2 Améliorations proposées de l'aspect sonore des collecticiels synchrones distribués	72
2.1 Synthèse sonore : spatialisation du son	75
2.2 Gestion et contrôle de la diffusion du son	77
2.2.1 Extension de la notion de participant	78
2.2.2 Structure de conférence dynamique	79
2.2.3 Synthèse des propositions concernant la gestion et le contrôle du son	81
2.3 Données associées aux flux sonores	82
2.3.1 Informations "statiques"	82
2.3.2 Informations dynamiques	83
2.4 Synthèse des services proposés	85
3 Analyse des services proposés par rapport aux modèles conceptuels des collecticiels	85
3.1 Par rapport au trèfle fonctionnel du collecticiel	86
3.2 Par rapport au modèle conceptuel d'Ellis	88
Chapitre 3 : Modélisation, intégration et implémentation	93
1 Modélisation d'ACAO	93
1.1 Formalisme	94
1.1.1 Les composants atomiques d'une audioconférence	94
1.1.2 Définition d'une audioconférence	94
1.1.3 Exemples	95
1.2 Un langage de commande	97
1.3 Synthèse de la modélisation d'ACAO	99
2 Intégration dans un collecticiel	99
2.1 Le contexte d'un collecticiel synchrone distant	99
2.2 Le modèle du système multi-agents	100
2.2.1 Définition d'un agent	100
2.2.2 Organisation des agents d'un collecticiel synchrone distribué	102
3 Réalisation d'un système expérimental	104
3.1 Architecture matérielle	105
3.2 Un serveur d'ACAO	108
3.2.1 L'agent d'ACAO : un assemblage d'objets se partageant les traitements	110
3.2.2 Les liens logiques entre les objets principaux de l'agent d'ACAO	111
3.2.3 La Détection Automatique des Locuteurs	112
3.2.4 L'application sous Windows NT	112
3.3 Des interfaces clients	113
3.3.1 Architecture de base	114
3.3.2 Une interface de suivi et de commande	116
3.3.3 Une interface de démonstration	117
4 Points clés de l'agent d'ACAO en vue d'une généralisation pour des développements ultérieurs	118

Table des matières

4.1 Implémentation de l'agent d'audioconférence	120
4.1.1 Les objets de base représentant les composants atomiques d'une audioconférence	122
4.1.2 Les variables d'instance définies dans la classe Controler	126
4.1.3 Le contrôle des flux sonores par l'objet Ctrl	128
4.1.4 La gestion du langage de commande par l'objet Ctrl	131
4.2 Les principales modifications nécessaires en vue d'une implémentation plus généraliste	132
4.2.1 Adaptation de la gestion des commandes	133
4.2.2 Adaptation de la gestion des flux sonores	134
4.3 Conclusion sur la généralisation de l'implémentation	135
Chapitre 4 : Evaluations des services proposés	137
1 Introduction	137
1.1 La problématique de l'évaluation	138
1.2 Les différents niveaux d'une évaluation	139
1.3 Les différentes méthodes d'évaluation	141
1.4 Présentation des évaluations réalisées	141
2 Potentiel d'un système stéréophonique	143
2.1 Présentation	143
2.1.1 Objectif de l'expérimentation	143
2.1.2 Méthode de "positionnement"	144
2.1.3 Précision sur la méthode de positionnement	145
2.1.4 Procédé 1	145
2.1.5 Procédé 2	146
2.1.6 Procédé 3	147
2.2 Expérience	147
2.2.1 Sujets	147
2.2.2 Matériel	147
2.2.3 Logiciels	148
2.2.4 Procédure	149
2.3 Résultats - Discussion	151
2.4 Conclusion sur le potentiel d'un système stéréophonique	154
3 Apport de la spatialisation aux audioconférences	155
3.1 Présentation	155
3.2 Mise au point du protocole expérimental	157
3.2.1 Contexte général	157
3.2.2 Choix d'un protocole expérimental	158
3.2.3 Choix d'un ensemble d'agendas	159
3.3 Méthode de l'expérience réalisée	161
3.3.1 Précisions pour l'expérimentation	161
3.3.2 Matériel	162
3.3.3 Sujets	165
3.3.4 Procédure	165
3.4 Résultats	166
3.4.1 Mesures effectuées	167
3.4.2 Questionnaire	168
3.5 Discussion	169
3.5.1 Analyse statistique et interprétation	169

Table des matières

3.5.2 Propositions d'explication des résultats	171
3.6 Conclusion sur l'apport de la spatialisation du son	173
4 Conclusion sur l'évaluation des services proposés	174
Conclusion	177
Bibliographie	181
Annexes	193
1 Description du langage de commande proposé pour la gestion des audioconférences	193
1.1 Liste des commandes	193
1.2 Liste des messages retournés	196
2 Première évaluation : résultats bruts	198
2.1 Exemple d'un fichier de mesures	198
2.2 Bilan des mesures	201
3 Deuxième évaluation	202
3.1 Documents utilisés	202
3.2 Résultats bruts	211
3.2.1 Mesures	211
3.2.2 Questionnaires	214

Table des illustrations

<i>figure 1 : place du son dans les collecticiels</i>	14
<i>figure 2 : un son périodique</i>	20
<i>figure 3 : spectre de Fourier</i>	21
<i>figure 4 : évolution d'un spectre de Fourier au cours du temps</i>	22
<i>figure 5 : définition d'une enveloppe</i>	23
<i>figure 6 : conversion analogique numérique</i>	29
<i>figure 7 : un CLAUD "répondeur" formé de trois LAUDs</i>	40
<i>figure 8 : Architecture de MERMAID</i>	52
<i>figure 9 : différentes organisations de réseau</i>	56
<i>figure 10 : exemple d'assemblage d'objets de base pour une application de téléconférence</i>	60
<i>figure 11 : les maillons de la chaîne sonore d'une audioconférence</i>	68
<i>figure 12 : spatialisation des sons pour un participant</i>	76
<i>figure 13 : configuration réelle ou virtuelle</i>	77
<i>figure 14 : le trèfle fonctionnel du collecticiel</i>	86
<i>figure 15 : définition d'un agent</i>	101
<i>figure 16 : organisation d'un agent de téléconférence</i>	103
<i>figure 17 : notre pont audio expérimental</i>	106
<i>figure 18 : schéma de notre pont audio expérimental</i>	107
<i>figure 19 : implémentation de l'agent d'audioconférence</i>	110
<i>figure 20 : interface de l'agent d'ACAO</i>	113
<i>figure 21 : liens logiques des objets d'un agent d'interface</i>	114
<i>figure 22 : une interface de suivi et de commande</i>	116
<i>figure 23 : interface de gestion et de suivi de commande sous Unix</i>	117
<i>figure 24 : une interface de démonstration</i>	118
<i>figure 25 : définition de la classe Controler</i>	121
<i>figure 26 : exemple de configuration des instances gérées par l'objet Ctrl</i>	123
<i>figure 27 : définition de la classe Canal</i>	123
<i>figure 28 : ce que gère un canal</i>	124
<i>figure 29 : fichier de configuration</i>	124
<i>figure 30 : définition de la classe Chaise</i>	125
<i>figure 31 : définition de la classe Conf</i>	126
<i>figure 32 : données propres de la classe Controler pour la gestion des audioconférences</i>	126

Table des matières

<i>figure 33 : exemple de réponse</i>	127
<i>figure 34: définition de la classe Rep</i>	127
<i>figure 35 : principe du contrôle des flux sonores</i>	129
<i>figure 36 : définitions des classes SndTuning et SndState</i>	130
<i>figure 37 : méthodes de traitement des commandes</i>	131
<i>figure 38 : les variables utilisées pour l'interprétation des commandes</i>	131
<i>figure 39 : traitement des commandes</i>	132
<i>figure 40 : positionnement en stéréo</i>	144
<i>figure 41 : procédé de "positionnement" linéaire</i>	145
<i>figure 42 : procédé de "positionnement" parabolique</i>	146
<i>figure 43 : procédé de "positionnement" cubique</i>	147
<i>figure 44 : interface de configuration de l'expérience</i>	148
<i>figure 45 : état initial de l'interface de l'expérience</i>	149
<i>figure 46 : bilan de la première évaluation</i>	153
<i>figure 47 : un des quatre postes de travail lors de l'expérience</i>	162
<i>figure 48 : interface de suivi et de contrôle de la deuxième évaluation</i>	163
<i>figure 49 : enregistrement d'une solution trouvée</i>	164

Introduction

Deux événements concomitants sont à l'origine des recherches présentées dans cette thèse : l'émergence du Travail Coopératif Assisté par Ordinateur et la maturité des systèmes informatiques pour la manipulation du son. Nous avons cherché à les faire converger en exploitant le second dans les applications du premier. Cet objectif, qui relève d'une démarche pluridisciplinaire, s'appuie sur un vaste champ d'investigation composé de trois domaines de recherche à part entière : le son, les systèmes informatiques et le travail coopératif. La finalité des travaux présentés dans cette étude n'est donc pas d'apporter une contribution à chacun de ces domaines pris séparément. Dans notre travail, ces trois domaines interviennent de différentes façons :

- les résultats des recherches dans le domaine du son sont exploités,
- les systèmes informatiques constituent à la fois un contexte général et le domaine d'application,
- le Travail Coopératif Assisté par Ordinateur (TCAO), plus connu sous son vocable anglais CSCW (*Computer Supported Cooperative Work*), donne un cadre plus spécifique, permettant de restreindre un peu le domaine des systèmes informatiques, trop vaste en lui même.

Nous nous intéressons donc à l'exploitation du son dans les logiciels appliqués au CSCW : les collecticiels (*groupware*). Les collecticiels sont définis par Ellis et al [ELLIS 91] comme des "*systèmes informatiques qui assistent un groupe de personnes engagé dans une tâche commune*". Vilers [VILERS 92], Croisy [CROISY 95] et Hoogstoel [HOOGSTOEL 95] en fournissent une synthèse détaillée.

Pour mieux cerner la dimension sonore dans les collecticiels, nous reprenons plus particulièrement la classification spatio-temporelle des collecticiels. On distingue ainsi les applications qui sont utilisées par des utilisateurs qui travaillent d'une part au même endroit (local) ou non (distant) et d'autre part en même temps (synchrone) ou non (asynchrone). Johansen [JOHANSEN 88] est à l'origine de cette classification, inspirée des travaux de DeSanctis [DESANCTIS 87]. Grudin [GRUDIN 93] l'a récemment étendue en introduisant la notion de prévisibilité. Cette notion n'intervient pas dans notre problématique. Dans [BARME 92] nous adaptons, à la dimension sonore (figure 1), la classification spatio-temporelle de Rodden [RODDEN 91a].

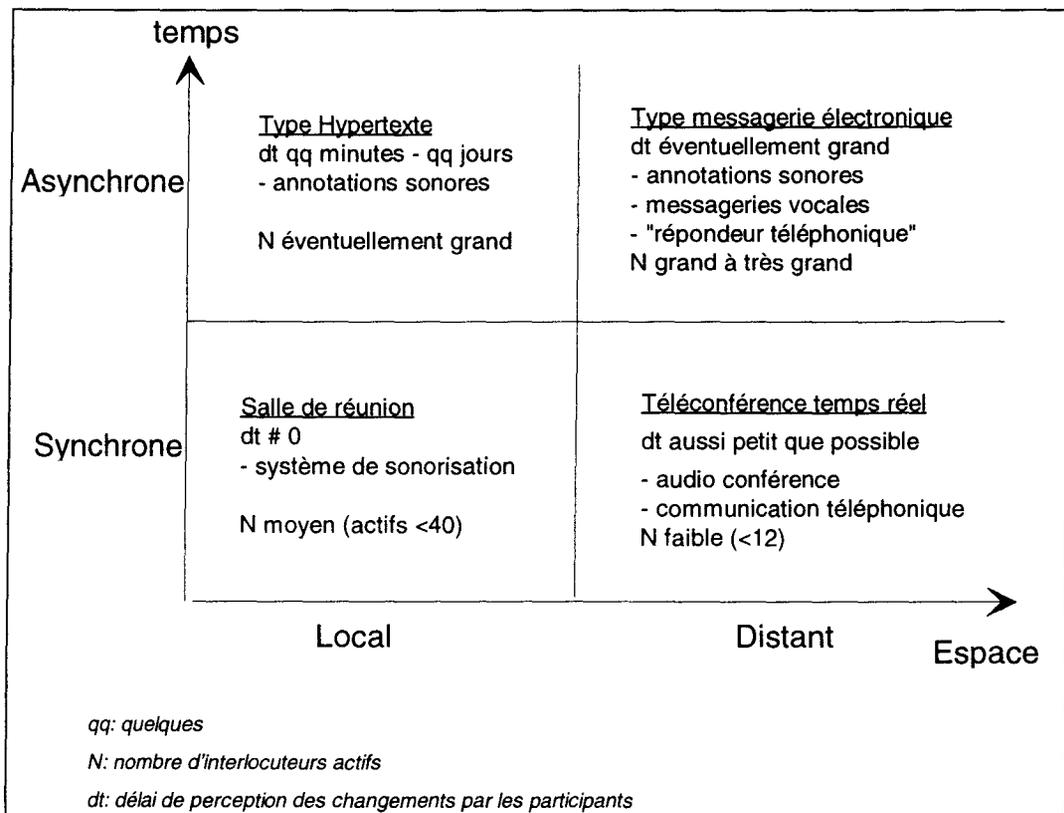


figure 1 : place du son dans les collecticiels

Indépendamment de la dimension sonore, Rodden [RODDEN 91b] affine la distinction local ou distant en quatre catégories suivant le type d'usage présumé et les moyens utilisés :

- local (*co-located*) : les utilisateurs sont réunis dans une même pièce où ils exploitent un système informatique en commun,
- virtuellement local (*virtually co-located*) : similaire à la catégorie local mais à distance avec des moyens de communications audiovisuels,
- localement distant (*locally remote*) : suppose l'utilisation de collecticiels n'ayant pas d'équivalent en mode local,
- distant (*remote*) : suppose des moyens d'accès minimaux entre les utilisateurs tels que des systèmes de messages ou des systèmes de conférences rudimentaires.

Les systèmes de conférences s'étendent alors entre virtuellement local et distant. Selon une perspective sonore, en prolongeant la réflexion exposée dans [RODDEN 91b], nous soulignons un domaine d'applications appartenant partiellement à deux classes de collecticiels ; les conférences multimédias se situent en partie dans chacune des classes de collecticiels synchrones ou asynchrones [DERYCKE 94].

Par ailleurs, l'adaptation à la dimension sonore de la classification spatio-temporelle de Rodden est confrontée au fait que les composantes spatio-temporelles sont intrinsèques aux sons : un son est, à la base, une variation de pression en fonction du temps qui se propage en trois dimensions dans un milieu porteur (air, eau, objets ...) [HANDEL 89]. En mode asynchrone il en résulte que la problématique du son n'est pas spécifique aux collecticiels. La résoudre revient au même pour un utilisateur indépendant ou isolé, soit parce qu'il utilise une application mono-utilisateur, soit parce qu'il utilise un collecticiel asynchrone. L'usage du son dans les systèmes informatiques pour un utilisateur indépendant a fait l'objet d'une recherche importante par ailleurs. Dans cette thèse, nous

n'approfondirons pas le mode asynchrone, mais nous exploiterons les travaux qui s'y appliquent.

En mode synchrone (salle de réunion, téléconférence temps réel), le canal sonore est principalement utilisé pour la communication verbale. Si les utilisateurs sont réunis au même endroit, cette communication verbale est réalisée par l'air ambiant qui transporte les ondes sonores et échappe ainsi en grande partie au système informatique. Nous avons donc concentré nos efforts sur l'aspect sonore des collecticiels synchrones distribués. Nous nous intéressons ainsi plus particulièrement à l'espace de communication, tout en ayant soin de maintenir ses interactions avec l'espace de coordination et l'espace de production¹.

Ce document est organisé en quatre chapitres. Le premier chapitre présente les travaux antérieurs exploités dans cette thèse. Le deuxième chapitre énonce les services proposés pour les collecticiels synchrones distribués. Le troisième chapitre décrit l'implémentation de ces services. Le quatrième chapitre rend compte des évaluations effectuées.

Le premier chapitre est divisé en trois parties. Les résultats des recherches concernant le son dans des domaines variés sont présentés dans un ordre croissant de spécificité par rapport au sujet de cette thèse. La première partie expose des notions de base utilisées dans cette thèse concernant le son. La deuxième partie présente un aperçu de l'exploitation du son dans les systèmes informatiques. La troisième partie rassemble les quelques informations les plus proches du sujet de cette thèse.

Le deuxième chapitre décrit la réflexion menée à partir de l'analyse d'un cas concret pour proposer de nouveaux services en vue d'améliorer la gestion du son dans les collecticiels synchrones distribués. L'objectif de cette réflexion a été de fournir un composant indépendant gérant l'aspect sonore et s'intégrant harmonieusement au sein d'un collecticiel. Cet objectif dépasse le cadre de

¹ les espaces de communication, de coordination et de production, qui forment le trèfle fonctionnel

l'amélioration des audioconférences et inscrit la démarche suivie vers son exploitation dans les systèmes informatiques. Deux idées ont prévalu pour la formulation de ces nouveaux services : tenter de se rapprocher de ce que l'on a déjà sans l'apport de l'informatique, et profiter du fort potentiel de contrôle et de traitement automatique de l'informatique pour apporter des services inaccessibles dans la réalité. A la fin de ce chapitre nous situons notre réflexion par rapport aux modèles conceptuels des collecticiels.

Par ailleurs, il a été pressenti dès le départ la nécessité de réaliser des expérimentations pour valider (ou invalider) les services proposés. Le troisième chapitre décrit l'effort entrepris pour concevoir et réaliser un système expérimental offrant concrètement ces services. Cette réalisation, bien que forcément spécifique pour la technologie actuellement exploitable, a été conduite avec le souci permanent d'une généralisation future. Une fois le matériel adéquat choisi, une première modélisation des audioconférences a été formulée pour en permettre le contrôle. La définition d'un langage de manipulation s'est ensuite avérée nécessaire pour en permettre l'exploitation. Celle-ci a été prévue dans le contexte d'un collecticiel synchrone distribué en utilisant le modèle des systèmes multi-agents. Un agent d'audioconférences assistées par ordinateur a donc été développé ainsi que des agents d'interface de test et de démonstration.

Enfin, l'évaluation des services proposés a été abordée. Le quatrième chapitre détaille les problèmes rencontrés et les expérimentations que nous avons réalisées. Deux expériences sont présentées. La première nous a permis de vérifier le potentiel discriminatoire d'un simple système stéréophonique. Le système expérimental d'audioconférences assistées par ordinateur a été exploité pour la seconde afin de vérifier l'apport de la stéréo aux audioconférences.

La conclusion dresse un bilan du travail accompli et présente les perspectives envisagées.



Chapitre 1 : Présentation des travaux antérieurs exploités dans cette thèse

1 Notions de base concernant le son

Dans cette partie, nous résumons les aspects essentiels des connaissances que nous avons dû acquérir sur le son. Après quelques précisions sur les caractéristiques physiques et psychologiques élémentaires du son, nous donnons un aperçu des techniques et problèmes de manipulation du son. Ensuite, nous présentons l'état de l'art dans l'exploitation des bruits et de la parole par les systèmes informatiques. Enfin, nous décrivons les travaux sur les problèmes de la détection de prise de parole et de la spatialisation du son que nous avons mis à profit dans cette thèse.

1.1 Caractéristiques physiques et psychophysiques

1.1.1 Définitions de base

Un *son* est, à la base, un phénomène vibratoire propagé (et transformé) par un milieu conducteur, altéré par les composants de ce milieu et perçu comme une variation de pression, principalement au niveau des tympans.

Cette définition d'un son est résumée de l'introduction de livre de Handel [HANDEL 89] qui est, avec [BOTTE 89], la principale source des informations qui suivent sur les caractéristiques physiques d'un son.

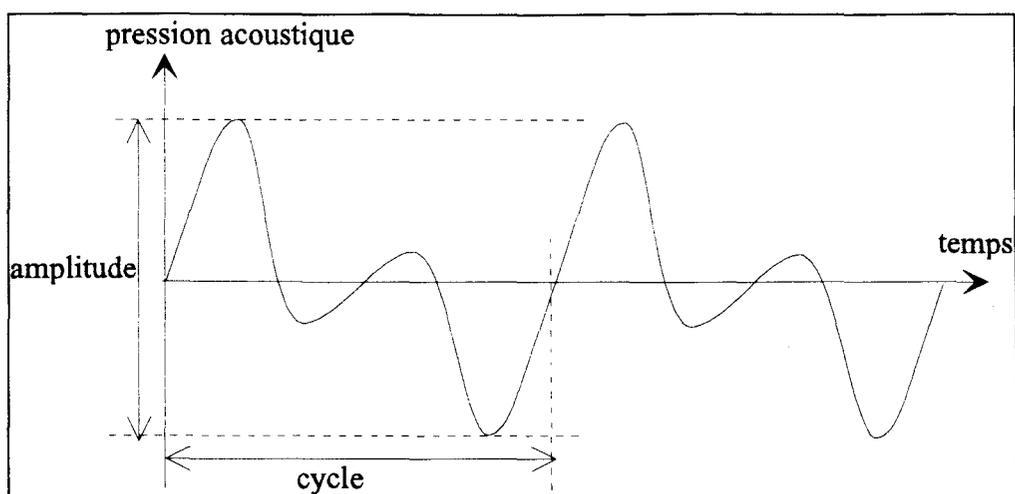


figure 2 : un son périodique

On utilise les grandeurs suivantes pour caractériser un son :

- *amplitude* : distance maximale atteinte autour de la position d'équilibre au cours du déplacement de l'élément oscillant (figure 2).
- *cycle* : motif d'oscillation répété dans un mouvement périodique (figure 2).
- *période* (en secondes) : durée d'un cycle, notée T .
- *fréquence* (en Hertz) : nombre de cycles par seconde, noté F . On a $F = \frac{1}{T}$
- *phase* (en degrés) : progression de l'onde au cours d'un cycle; est principalement utilisée pour exprimer le décalage entre deux ondes.
- *longueur d'onde* (en mètres) : liée à la vitesse de propagation de l'onde dans le milieu conducteur : le produit de la longueur d'onde par la fréquence est égal à la vitesse de propagation du son. La vitesse du son varie beaucoup d'un milieu à l'autre; plus la densité du milieu est grande, plus elle sera élevée : 340 m/s dans l'air, 1420 m/s dans l'eau, près de 6000 m/s dans l'acier [BOTTE 89].

1.1.2 Description analytique d'un son

Les grandeurs physiques précédentes ne s'appliquent qu'aux sons périodiques, appelés sons élémentaires. Un son quelconque peut être décrit comme la composition de sons élémentaires par l'analyse de Fourier. Un son élémentaire

correspond à une vibration sinusoïdale appelée mouvement harmonique. D'après le théorème de Fourier, les sons plus complexes se décomposent en somme de vibrations sinusoïdales.

Un son est dit complexe s'il faut plusieurs fréquences (sons purs) pour le décrire. On distingue deux catégories de sons complexes : les sons complexes périodiques et les sons complexes non périodiques.

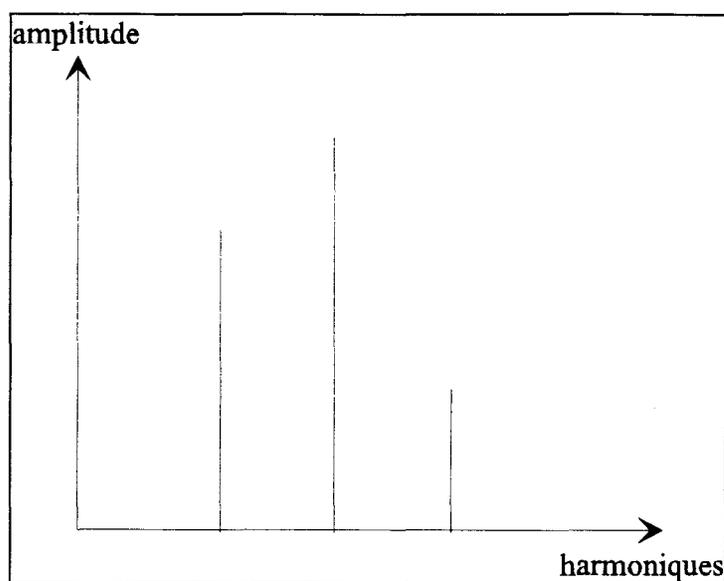


figure 3 : spectre de Fourier

Dans le cas de sons complexes périodiques (et non sinusoïdaux), les éléments sinusoïdaux de la décomposition du son sont appelés des harmoniques. Ils ont la propriété d'être un multiple entier de l'harmonique correspondant à la fréquence fondamentale du son, appelé premier harmonique. Le déphasage entre les différents harmoniques n'intervient pas dans la perception du son [HANDEL 89]. On représente donc seulement les amplitudes des différents harmoniques par des traits verticaux proportionnels à l'amplitude et espacés selon la fréquence des harmoniques, ce qui donne un spectre de Fourier. Nous donnons dans la figure 3 un exemple de spectre de Fourier tiré de [HANDEL 89]. Il correspond à celui du son représenté dans la figure 2.

A propos de l'importance de la phase dans la perception du son, si nous sommes peu sensibles au déphasage inter harmoniques, nous sommes par contre sensibles au déphasage d'une oreille à l'autre. Botte [BOTTE 89] précise que nous utilisons cette information comme un indice pour la localisation de la source.

Pour les sons non périodiques, le nombre d'harmoniques est infini. On représente toujours les amplitudes en fonction de la fréquence par une courbe correspondant à une fonction continue appelée "transformée de Fourier", ce qui permet de définir précisément des bruits particuliers tels que des bruits impulsionnels très brefs ou des "bruits blancs" (par analogie avec le spectre de la lumière blanche) dont la transformée de Fourier est une droite horizontale (une infinité d'harmoniques dont l'amplitude est constante). Handel précise cependant que l'analyse de Fourier prend mal en compte l'évolution d'un son réel au cours du temps.

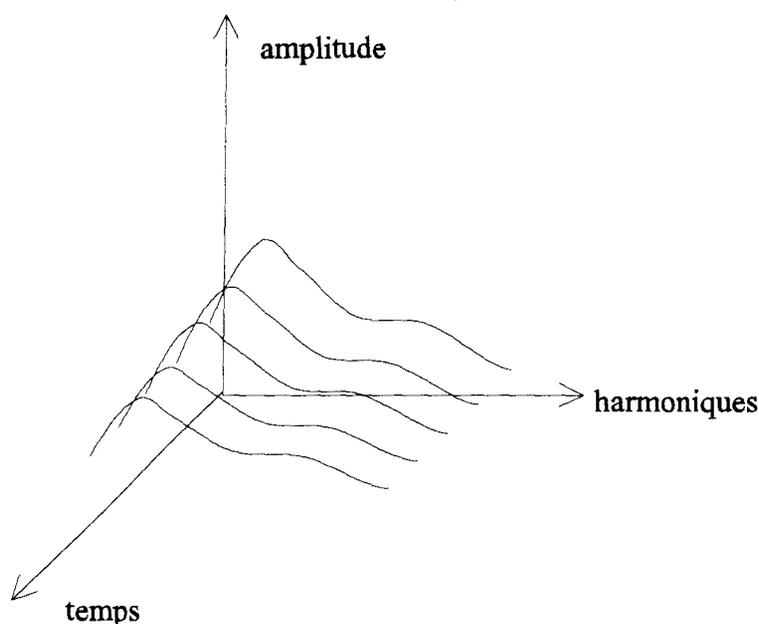


figure 4 : évolution d'un spectre de Fourier au cours du temps

On peut alors découper le temps en petits segments et appliquer à tous ces segments une analyse de Fourier. On pourrait aussi représenter par une perspective d'une surface en trois dimensions la variation du spectre de Fourier au cours du temps (figure 4).

L'évolution d'un son au cours du temps peut aussi être représentée en considérant soit la variation de fréquence, soit la variation d'amplitude. Cette dernière est plus utilisée et décompose l'enveloppe en trois [HANDEL 89] ou quatre [STOLZ 92] phases avec une première phase de chute en plus (figure 5):

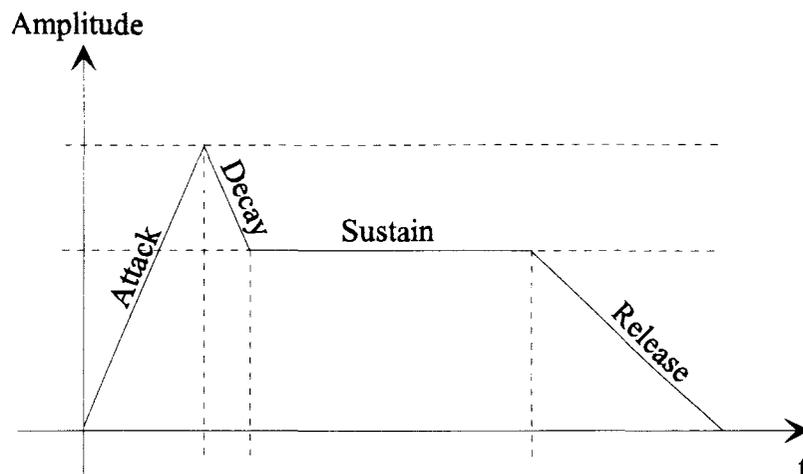


figure 5 : définition d'une enveloppe

- le temps de montée (*Attack*) : durée nécessaire pour atteindre le volume maximal,
- la première phase de chute (*Decay*) : durée du passage du volume maximal au volume de phase de stabilité,
- le niveau maintenu (*Sustain*) : durée de la partie stable de l'enveloppe,
- la deuxième phase de chute (*Release*) : durée de la partie terminale du son, après la phase de stabilité.

1.1.3 L'intensité et le timbre d'un son

La mesure de l'intensité physique des sons et leur perception subjective sont précisément décrites dans les travaux de Marie-Claire Botte [BOTTE 89]. On retiendra que l'intensité acoustique d'un son est exprimée en dB (décibels, du nom de Graham Bell, l'inventeur du téléphone) qui mesure en fait 20 fois le logarithme du quotient de la pression acoustique d'un son sur la pression acoustique d'un son

de référence : $20 \log (p / p_0)$. En prenant un son de référence standard, on mesure des dB SPL dont Botte donne la définition suivante :

"Lorsqu'un niveau, en dB, est exprimé par référence à la valeur conventionnelle de $20 \mu\text{Pa}$, il porte le nom de niveau de pression acoustique et correspond à la valeur dite SPL en anglais (pour Sound Pressure Level)."

Handel précise que ce niveau de référence correspond au seuil de l'audition humaine d'un son sinusoïdal d'une fréquence de 1000 Hz. Handel donne d'autres repères d'un point de vue perceptif, parmi lesquels : 20 dB correspond à un chuchotement, 80 dB à un trafic routier et 160 dB est le seuil de douleur. Cependant la perception d'un son varie aussi en fonction de sa fréquence. L'échelle logarithmique pour l'intensité des sons est adaptée à notre perception mais réserve des surprises comme le souligne Botte :

"... l'échelle des décibels permet de représenter dans des limites acceptables, car elle la comprime énormément, la marge immense des pressions (ou des intensités) acoustiques auxquelles l'oreille est sensible : le seuil d'audition est 10^{-12} W/m^2 , tandis que le seuil de la douleur est environ 1 W/m^2 .

... l'échelle des décibels, si pratique pour l'acousticien, réserve des surprises si l'on oublie ses bases; ainsi, par exemple, lorsque deux machines émettent, l'une un bruit de 100 dB, et l'autre un bruit à 90 dB, pour connaître le niveau total, on ne peut en aucun cas additionner les décibels! Ce sont les valeurs des puissances acoustiques correspondantes, en Watts, qui doivent être sommées, puis on transforme le résultat en dB."

Par ailleurs, les études en psychoacoustique ont révélé des effets de masquage d'un son par un autre. Ainsi, en présence de plusieurs sons, on peut n'entendre que le plus fort alors que si les sons étaient isolés, ils seraient tous audibles. Cette caractéristique a été judicieusement exploitée par le système de compression de

l'information sonore, voir le paragraphe 1.2.2 *Le problème de la compression du son*, page 31.

La perception du timbre, si elle est subjectivement précise pour l'homme, ne correspond cependant pas à une grandeur physique bien définie. Elle est liée à la répartition de l'énergie sonore dans l'ensemble des harmoniques ainsi qu'à l'enveloppe [BOTTE 89]. Ainsi, le timbre d'un son nous permet de distinguer le son d'un saxophone de celui d'un violon. Dans le cadre de cette thèse, on retiendra que le timbre est une caractéristique qui permet d'identifier une voix et donc le locuteur [ARONS 92a].

1.1.4 Qualité sonore de la manipulation du son

La qualité d'un dispositif manipulant du son (par exemple l'amplificateur d'une chaîne Hi-fi, un magnétophone ou un microphone) se mesure en fonction de la fidélité avec laquelle il permettra de restituer le son. La qualité d'un dispositif est évaluée à partir de trois grandeurs physiques mesurables :

- la *bande passante*, en Hz : c'est la plage des fréquences transmises avec une atténuation inférieure à 3 dB en puissance. Les sons trop graves ou trop aigus ne sont en général pas transmis ou conservés par un dispositif manipulant du son. Pour mettre les chiffres en rapport avec nos capacités auditives, l'oreille humaine peut détecter des sons dont la fréquence est comprise entre 16 Hz et 20 000 Hz. La voix humaine émet des sons entre 70 et 2 000 Hz. [STOLZ 92]. La norme AFNOR prévoit une bande passante de 20 Hz à 20 kHz pour la hi-fi.
- la *dynamique*, en dB : c'est la plage des intensités sonores possibles, c'est-à-dire le volume sonore (exprimé en dB) maximal moins le volume minimal (en dB). L'oreille humaine est sensible sur une plage d'environ 100 dB.
- le *rapport signal / bruit*, en dB : un dispositif restituera, en plus du signal entré, un bruit parasite. Plus ce bruit est faible par rapport au signal traité (c'est-à-dire plus le rapport signal / bruit est grand), plus la qualité est grande.

Il y a cependant d'autres caractéristiques qui échappent à ces mesures. Elles concernent par exemple la qualité de restitution des attaques et des transitions qui interviennent dans le timbre d'un son. C'est ce qui participe aux caractéristiques des différents instruments de musique et des voix humaines.

1.1.5 Localisation de l'origine spatiale d'un son

Il y a longtemps que l'on s'intéresse à ce sujet, notamment en psychologie expérimentale. Dès 1953, Cherry présentait ses travaux sur le maintenant célèbre "cocktail party effect" où il démontrait la possibilité que nous avons de discerner un locuteur particulier dans un brouhaha ambiant en le localisant dans l'espace grâce à notre écoute binaurale [CHERRY 53].

Depuis, on a analysé et décomposé les mécanismes de base intervenant dans la perception spatiale des sons; une synthèse se trouve dans [MULDER 94]. Il y a 8 principaux paramètres identifiés :

- l'*Interaural Time Delay* (IDT) qui correspond au décalage temporel avec lequel un son atteint chacune des oreilles. C'est le cas pour les hautes fréquences (petites longueurs d'onde). Pour les basses fréquences, on a une longueur d'onde qui dépasse la largeur de la tête : on perçoit alors un déphasage du son d'une oreille à l'autre. Pour donner une idée de la finesse avec laquelle nous exploitons ces paramètres, l'IDT ne dépasse pas 0.6 ms [HANDEL 89],
- l'*Head Shadow* ou l'*Interaural Intensity Difference* (IID) correspond à l'altération du son qui traverse la tête, un corps creux (sinus, fosses nasales, ...) et complexe (os, cartilages, tissus de consistances diverses, ...),
- la *Pinna response* qui est due à la modification du son par les pavillons externes,
- les *Shoulder echoes* : les épaules et le haut du corps créent des échos dépendant de l'élévation de la source sonore,
- les *mouvements de la tête* nous permettent d'avoir plusieurs "points de vue" pour lever un doute quant à la localisation d'un son,

- la *vision* qui est prépondérante : on croira plus volontiers ses yeux que ses oreilles si l'on identifie l'objet qui est à l'origine du son,
- l'*intensité sonore* qui diminue avec le carré de la distance; mais dans une pièce avec une forte réverbération, cette information est un peu faussée; et l'importance du volume du son réverbéré par rapport au son direct augmente avec la distance jusqu'à être totalement prépondérant,
- les *premiers échos* ont aussi été identifiés comme jouant un rôle.

1.2 Manipulation technique du son

Les définitions des principaux termes employés pour le son que nous venons de rappeler vont nous permettre d'aborder le problème de la manipulation technique du son. Nous les présenterons en deux sous-parties. Dans la première nous réunissons les considérations sur la prise de son, le transport et la restitution du son. Nous traitons à part le problème de la compression du son.

1.2.1 *Prise de son, restitution et transport*

La prise de son passe par un microphone dont le principe de fonctionnement est standard : une membrane capte les vibrations de l'air et les transforme en oscillations d'une tension électrique, soit parce que la membrane est couplée à un électro-aimant (microphone électrodynamique), soit parce que la membrane est partie intégrante d'un condensateur (microphone électrostatique). A partir de ces deux techniques se décline une grande variété de mises en œuvre avec des effets très divers². Il y a ainsi une grande marge de manœuvre, mais des précautions à prendre au niveau de la prise de son. Par exemple, si l'on veut capter une ambiance sonore ou plusieurs personnes dispersées dans une pièce, on utilisera un microphone omnidirectionnel. Si l'on veut être plus sélectif et ne capter que le son d'une seule personne dans un environnement bruyant, on utilisera un microphone

² nous tenons ces informations d'une photocopie du cours de M. Michel Boudinet, professeur à l'Institut National de l'Audiovisuel, qui nous a été donnée par Monsieur Jean-Paul Gevaert, Ingénieur du son.

"hyper-cardioïde". Il est possible de le faire à distance avec un micro "canon" mais pas dans une ambiance à forte réverbération. Il convient aussi de tenir compte de la directivité des sons et des phénomènes de diffraction. Les aigus se propagent dans une direction précise et sont plus facilement arrêtés par les obstacles que les graves qui diffusent. Ceci découle des différences de longueur d'onde et des phénomènes d'ondes stationnaires en espace fermé. Le choix du bon microphone est crucial et dépend du contexte.

La restitution du son se fait de toute façon au moyen de transducteurs électroacoustiques (haut-parleurs). Leur principe est équivalent à celui des microphones mais dans l'autre sens : un électro-aimant active une membrane pour recréer les vibrations de l'air correspondant au son à reproduire. D'autres techniques existent à l'instar des microphones : on peut aussi utiliser des condensateurs. Pour reproduire des sons graves (basses fréquences), il est préférable d'avoir une grande surface de membrane car la longueur d'onde est grande; pour les sons plus aigus (hautes fréquences), il vaut mieux une petite membrane car elle aura une moins grande inertie. C'est pourquoi les enceintes acoustiques comportent plusieurs haut-parleurs de tailles différentes et des filtres passe-haut et passe-bas pour répartir la restitution du son sur les différents haut-parleurs. Leur qualité ne doit pas être négligée au risque de perdre le bénéfice d'une grande qualité sonore en amont. Pourtant, dans le domaine de l'informatique, les petits haut-parleurs qui équipent actuellement les stations de travail sont loin d'être de qualité suffisante. Même les petites enceintes amplifiées généralement fournies avec des cartes sonores de grande qualité ne permettent pas une bonne restitution du son.

Entre microphone et haut-parleur, l'information sonore est représentée par un signal électrique ou l'équivalent codé de ce signal. Pour un son manipulé en mode analogique, la variation de pression acoustique est représentée par la variation d'une grandeur physique proportionnelle, le plus souvent sous la forme d'une variation de tension d'un signal électrique pour la manipulation et le transport, et sous la forme d'une variation de l'intensité d'une empreinte magnétique pour le stockage. L'information se dégrade si elle transite à travers plusieurs dispositifs ou

quand elle est transportée sur une longue distance; les deux détériorations s'ajoutent. Par contre, la manipulation du son en mode analogique est quasi instantanée (temps de propagation \leq au 2/3 de la vitesse de la lumière + retard dans les fonctions électroniques) et, comme on conserve le mode de représentation (la variation d'une grandeur physique), il n'y a pas d'altération de la qualité du son du fait des conversions. L'expérience existante de ce mode de manipulation du son permet en outre de disposer d'outils finement mis au point.

Un son manipulé en mode numérique est représenté par une suite de valeurs qui codent la variation de pression acoustique. Pour passer en mode numérique, on utilise des circuits électroniques spécialisés, appelés convertisseurs analogique/numérique. Ils effectuent un échantillonnage du signal analogique. Une valeur moyenne est substituée à chaque petit segment de la courbe de variation du signal en fonction du temps (figure 6). Si l'on faisait tendre la taille des segments vers 0, et si l'on codait les valeurs avec une précision infinie, on retrouverait exactement le signal analogique, au prix d'une puissance de traitement infinie ...

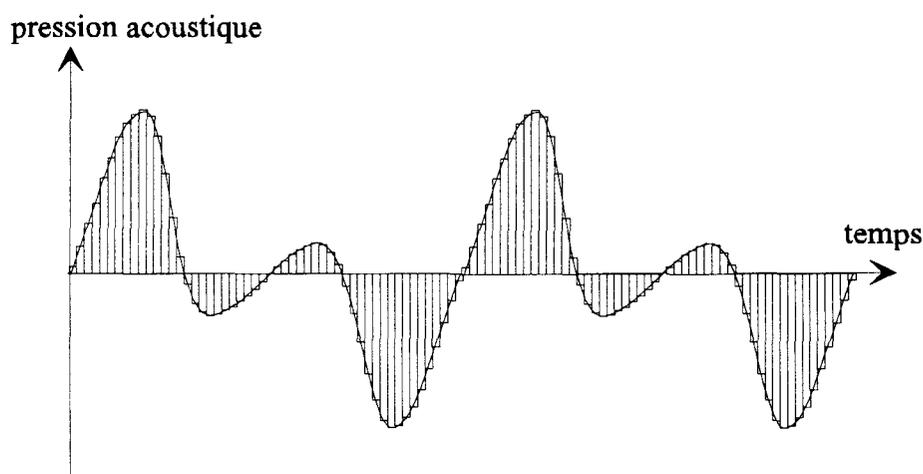


figure 6 : conversion analogique numérique

L'information est dégradée lors de la conversion [FIELDER 89] mais sa qualité sera ensuite maintenue constante par des dispositifs de correction d'erreur pour le

transport, le stockage, la reproduction et la diffusion. De ce fait, le mode analogique est progressivement abandonné au profit du numérique même s'il présente quelques inconvénients :

- il nécessite une très grande puissance de traitement pour être d'une qualité comparable à l'analogique,
- son traitement n'est pas instantané,
- pour le transport, le mode numérique souffre des limites actuelles des débits des réseaux informatiques [PASIEKA 91].

La numérisation du son est basée sur trois paramètres : la fréquence de l'échantillonnage, la taille et le mode de codage des échantillons. La fréquence d'échantillonnage, exprimée en hertz, est le nombre d'échantillons que l'on calcule par seconde. Plus ce nombre est élevé, plus la bande passante sera grande. En fait, Nyquist [NYQUIST 24] a montré qu'en théorie il faut une fréquence d'échantillonnage double de la fréquence que l'on veut numériser [SCHMANDT 94]. Celle de la technologie des disques lasers : 44 100 Hz permet le codage de sons d'une fréquence allant jusqu'à 22 000 Hz et correspond bien aux capacités de notre système auditif.

De la taille et du mode de codage des échantillons dépend le nombre de valeurs différentes que l'on peut coder. On code ces valeurs en binaire; la taille des échantillons correspond au nombre de bits utilisés. En fonction de cette taille, on aura une dynamique plus ou moins importante ou plus exactement, pour une dynamique donnée, plus de finesse dans son expression. Cela intervient directement sur le rapport signal / bruit. Prenons par exemple la norme des CD (Compact Disk) qui prévoit un échantillonnage codé sur 16 bits. Si la conversion était parfaite, seul le dernier bit pourrait être erroné (car il arrondit la valeur réelle). On a donc, en théorie, un bruit minimum de quantification de $1/2^{16}$ soit un rapport signal / bruit de $20\log(2^{16} / 1) \approx 96$ dB [MOUGEY 93].

Une grande fréquence d'échantillonnage et des échantillons de grande taille sont nécessaires pour avoir un son numérique de grande qualité. Ceci introduit les problèmes de stockage et de transport évoqués plus haut. Reprenons la référence de 44 100 Hz sur 16 bits (qualité CD). Une minute d'enregistrement en stéréo demande plus de 10 Mo, et transporter un tel son nécessite un débit de plus 1.4 Mbits/s (1.4112 Mbits/s exactement)! Pour donner un ordre de grandeur, un CD classique enregistre 60 ou 72 minutes pour un volume de 600 Mo. Si l'on veut faire une audioconférence à six en transportant le son sur un réseau numérique, il est nécessaire d'avoir au minimum un débit constant supérieur à 4.2 Mbits/s; ce débit correspond au transport simultané de six flux sonores échantillonnés sur 16 bits à 44 100 Hz (on suppose ici une émission en broadcast, chaque destinataire ayant à charge de mixer les sons qui lui arrivent).

Les débits des signaux numériques relatifs aux informations sonores paraissent a priori compatibles avec ceux des réseaux locaux informatiques : 10 Mbits/s pour un réseau Ethernet (débit instantané) à comparer avec les 1.4 Mbits/s pour un son stéréo haute fidélité ou 64 kbits/s pour le téléphone (un canal sur RNIS). En réalité, les protocoles mis en oeuvre dans les réseaux locaux sont dits non déterministes, c'est-à-dire que le débit moyen et le délai d'acheminement sont une fonction de la charge du réseau et peuvent évoluer dans de très grandes proportions. Il faut ajouter à cela que les réseaux informatiques fonctionnent le plus souvent en mode paquet (exemple : le protocole TCP/IP). Ceci ajoute des retards de transport supplémentaires. En réalité, les délais peuvent atteindre des valeurs supérieures à 100 ms.

Tous ces chiffres ne tiennent pas compte des possibilités de compression.

1.2.2 Le problème de la compression du son

Le son est difficile à compresser sans dégradation. La compression de données est un thème de recherche à part entière [NELSON 93], tout comme celle appliquée spécialement au son [GERSHO 94]. C'est un problème qui a donc été très approfondi, bien qu'il soit ardu, car l'enjeu est grand; un son numérique de

bonne qualité exige un débit trop grand par rapport aux techniques actuelles et face à la demande en moyens de communication.

Pour la compression d'un son numérique, on peut jouer sur les trois paramètres donnés plus haut, mais il faut tenir compte de deux contraintes : la qualité subjective du son compressé et les délais de compression qui sont rédhibitoires dans le cadre d'une communication en temps réel. Le tableau 1 résume les principaux procédés de codage classés par ordre de complexité. Ces procédés sont expliqués ensuite.

Dénomination	Type de codage	Remarques	Utilisation type
PCM	codage linéaire	pas de compression	fichier ' .WAV '
exponentiel	codage sous la forme d'une mantisse et d'un exposant	perte de qualité par troncature	NICAM (TV)
μ -Law, A-law	codage logarithmique	précision plus importante pour les sons faibles que les sons forts	Téléphone standard
ADPCM	codage différentiel	s'appuie sur la caractéristique de continuité d'un signal sonore	PC Multimédia Editeur de son
MPEG Audio	exploite les phénomènes de masquage	problème de délais	Hi-fi CD-ROM etc.
LPC	codage sémantique	adapté pour la parole	téléphonie mobile
MIDI	codage sémantique	adapté pour la musique	instrument musique
compression temporelle	accélère le débit d'un message verbal	uniquement en asynchrone	Messagerie vocale

tableau 1 : quelques procédés de compression du son

Le format de codage PCM³ correspond au niveau zéro de la compression. C'est celui qui est actuellement utilisé dans les fichiers '.WAV' de Windows. Pour diminuer le débit ou le volume nécessaire au stockage, il n'y a pas d'autre possibilité que de diminuer la fréquence d'échantillonnage et la taille des échantillons avec une dégradation importante de la qualité du signal par une répercussion directe sur la bande passante et la dynamique.

Pour une fréquence d'échantillonnage donnée, on cherche donc à intervenir sur le mode de codage des échantillons pour diminuer leur taille en conservant la dynamique. Un codage non linéaire des échantillons est un premier niveau de compression. Un codage sous la forme d'une mantisse et d'un exposant étend la plage de codage au prix d'une imprécision due aux erreurs de troncatures. Un codage logarithmique permet d'avoir une plus grande précision pour les sons faibles au prix d'une moins grande précision pour les sons forts. C'est ce type de codage qui a été retenu pour le téléphone. Deux standards équivalents existent : le codage A-law et codage μ -law qui est aussi utilisé pour les fichiers '.au' des stations de travail sous Unix. Ils permettent d'avoir sur 8 bits une dynamique équivalente à celle obtenue avec 12 bits en linéaire ou, en d'autres termes, d'avoir pour 8 bits une dynamique de 72 dB au lieu de 48 dB [SCHMANDT 94].

En s'appuyant un peu sur les caractéristiques d'un son, on obtient des systèmes de codage plus performants. Un signal sonore est une fonction continue, d'où l'intérêt d'un codage différentiel. Schmandt [SCHMANDT 94] en décrit plusieurs parmi lesquels on retient l'ADPCM⁴ qui est prévu pour remplacer prochainement le PCM sous Windows [MS 93].

D'autres systèmes efficaces exploitent le phénomène de masquage évoqué plus haut : le MUSICAM⁵ utilisé pour le DAB⁶ [A. V. T. 92; CUSSO 92], ainsi que le

³ Pulse Code Modulation

⁴ Adaptative Differential Pulse Code Modulation

⁵ Masking pattern adapted Universal Sub-band Integrated Coding And Multiplexing

⁶ Digital Audio Broadcasting

système de codage PASC⁷ utilisé pour la DCC⁸ [MOUGEY 93]. Le système MUSICAM est celui choisi pour former la base de la norme MPEG⁹. A qualité d'audition équivalente, il permet de ramener à 256 kbits/s (environ 1/5 du débit sans compression) le débit nécessaire pour un son stéréo de qualité CD [CUSSO 92]. Des taux de compression supérieurs sont possibles mais avec une dégradation de la qualité. L'article précise que cette conversion est faite "presque en temps réel" (> 375ms). Or, le temps réel est très important lors d'une audioconférence : un délai de 250ms est perceptible. Pour donner un autre ordre de grandeur, nous percevons les décalages inférieurs à la milliseconde d'une oreille à l'autre (voir le paragraphe 2.4 *Spatialisation du son*).

En s'appuyant sur les caractéristiques sémantiques d'un son d'un type particulier, on peut encore améliorer les procédés de codage. C'est ce qui est fait par le GSM pour la parole, qui ramène le débit à 4.8 kbits/s pour une qualité sonore correspondant à la norme téléphonique.

Signalons au passage une autre forme de représentation du son. Il s'agit ici uniquement des sons musicaux dont les conventions de représentation remontent au XI^{ème} siècle lorsque Guy d'Arezzo prit les premières syllabes d'un hymne à Saint Jean-Baptiste pour symboliser les notes de la gamme [ELMER 94]. Depuis 1980, la norme MIDI¹⁰ (dont on trouve des explications sommaires mais suffisantes dans [STOLZ 92]) définit, entre autres, un système de codage des sons musicaux. Chaque note est codée ainsi que la manière dont elle est jouée et sur quel instrument elle est jouée, ce qui donne un moyen très efficace de représenter une partition. Le son sera ensuite produit par un synthétiseur qui recrée le son à partir de sons de base préenregistrés et par la mise en série et en parallèle d'opérateurs fonctionnels qui génèrent des fonctions périodiques de base. Un opérateur est composé d'un générateur de phase en série avec un générateur

⁷ Precision Adaptive Sub-band Coding

⁸ Digital Compact Cassette

⁹ Motion Picture Expert Group

¹⁰ Musical Instrument Digital Interface

d'enveloppe. Le premier règle la fréquence et le deuxième l'évolution de l'amplitude. L'intérêt de l'utilisation du codage MIDI pour un son musical ne se borne pas à la compression. Il permet aussi, lors de la restitution, de modifier le rythme, le timbre et la hauteur du son. Ces altérations sont exploitables dans les interfaces sonores (voir le paragraphe 2.1 *Notification, rétroaction et sonification : un premier niveau d'exploitation du son*).

Une toute autre approche pour diminuer la charge des réseaux de communication vocale consiste à restituer un message sonore dans un temps plus court sans en altérer l'intelligibilité : il s'agit de la compression temporelle. ARONS [ARONS 92b] présente une étude sur ce sujet. Une solution tirée de la théorie des ondelettes, la décomposition en grains de Gabor donne une idée de la complexité des techniques requises. Elle est présentée dans un article de vulgarisation par Guillemain [GUILLEMAIN 91].

1.3 Manipulation informatique du son

1.3.1 Les possibilités sonores des machines actuelles

Les recherches dans le domaine du son sont contraintes par les possibilités technologiques qui en ont longtemps freiné les efforts. Voici un rapide aperçu des possibilités actuelles.

C'est sur les PC (*Personal Computers*) que s'est produite la "révolution" sonore motivée par le marché des jeux vidéo qui a fini par se stabiliser autour du standard représenté par la carte Sound Blaster [STOLZ 92]. Les premiers PC étaient équipés d'un générateur de fonction périodique (sinusoïdale, carrée, dent de scie...) dont on pouvait régler la fréquence et l'amplitude. Le signal en sortie de ce générateur de fonction périodique était directement émis sur un petit haut-parleur interne. Certains PC actuels conservent cette extension préhistorique. Les premiers ordinateurs avec un synthétiseur stéréo apparaissent dans les années 80. Le boom des cartes sonores pour PC a commencé en 1987. En 1989 Windows 2.0 se voit doté d'une extension multimédia qui sera intégrée en 1992 dans Windows 3.1. En

1994, un article paru dans la revue "Dr Dobb's" recense 38 cartes sonores différentes uniquement pour des PC. D'autres sont apparues depuis.

Les cartes sonores (standard MPC¹¹) pour PC manipulent trois types de son d'un point de vue technique :

- les sons numérisés : échantillonnés jusqu'à 44 100 Hz sur 2x16 bits en stéréo,
- des sons synthétiques : jusqu'à 32 voix indépendantes stéréo FM avec 4 opérateurs par voie (un opérateur génère un signal de base),
- des sons analogiques avec une entrée ligne en stéréo, une entrée microphone en mono et une entrée pour le lecteur de CD interne.

Ces différents types de son sont mixés avec un contrôle du volume de chaque entrée (des voies de gauche et de droite si l'entrée est stéréo) et du volume de sortie (il n'y a qu'une seule sortie).

Dans le monde Unix, il n'y a pas de standard équivalent à celui des PC. Next a défini l'AIFF qui a été repris par Digital, Silicon Graphics et Apple. Cependant, les stations de travail Unix sont munies au minimum d'un périphérique audio qui permet d'enregistrer et de rejouer des sons numériques échantillonnés à 8 kHz sur 8 bits et comprimés avec le système μ -law. En fait ce périphérique est vu du système comme un fichier utilisable en mode stream (nommé /dev/audio). Utilisé en lecture, il permet d'enregistrer un son (numérisé); utilisé en écriture, il permet de rejouer un fichier son (d'extension '.au').

Sous la pression du marché sonore des PC, les stations offrent de plus en plus de services comparables au standard de la Sound Blaster. Elles n'étaient déjà disponibles que sur les stations haut de gamme comme les Silicon Graphics.

Les MacIntoshs (Apple) ont, dès le début, prévu de produire des sons numérisés mais en sortie seulement. C'est sur les Macs qu'a été inaugurée la

¹¹ Multimedia Personal Computer

rétroaction sonore. Pendant très longtemps, les capacités sonores de Macs n'ont pas évolué. Puis il y a eu un bond en avant et les Mac moins anciens disposent d'un DSP sur la carte mère et offrent une numérisation de qualité CD, simultanément en entrée et en sortie, les deux en stéréo !

1.3.2 Le partage des ressources sonores dans un système informatique

L'amélioration de la facilité d'accès aux ressources sonores d'un système informatique a suivi l'évolution des capacités sonores disponibles. Au début, l'accès aux ressources sonores a été quasiment réservé à la programmation en assembleur, sous réserve que l'on ait une documentation très précise des composants sonores de l'ordinateur (Mac mis à part). Aucune fonction d'un langage de plus haut niveau n'y donne accès. Une première étape a été franchie lorsque l'on est passé de l'utilisation des ressources sonores pour une seule application (MSDOS, Windows) ou processus (Unix) au partage de ces ressources entre plusieurs applications. Il y a plusieurs niveaux de partage :

- les applications peuvent utiliser les ressources sonores chacune à leur tour,
- les applications peuvent utiliser les ressources sonores en même temps,
- les ressources sonores sont partagées sur le réseau avec l'un des deux niveaux de partage précédent.

Nous détaillerons deux systèmes d'exploitation permettant un tel partage des ressources sonores : Windows NT et Unix.

1.3.2.1 Les services sonores sous Windows NT

Les services sonores sous Windows NT s'utilisent au plus bas niveau par l'intermédiaire de l'API¹² Win32 qui est le standard actuel de Windows, pour Windows 3.1, Windows NT et Windows 95. Il existe en plus un protocole qui permet d'en contrôler l'essentiel à un niveau macroscopique (c'est-à-dire

¹² Application Programming Interface

l'équivalent de la manipulation d'appareils audio/vidéo classiques), le MCI¹³. Le MIDI est lui aussi complètement intégré à l'environnement de programmation [MS 93]. Il y a quand même un décalage entre l'intégration des autres éléments de l'interface (fenêtres, graphiques, contrôles, etc.) et celle du son. Les premiers sont modélisés dans les MFC¹⁴, des hiérarchies de classes d'objets en C++, mais pas les sons que l'on doit toujours manipuler avec des fonctions de la bibliothèque de base.

Contrairement au monde Unix, le son sur PC n'est pas intégré harmonieusement au système d'exploitation. On doit rester très proche du matériel : allouer et stabiliser une zone de la mémoire, définir les paramètres précis du format d'échantillonnage, préparer la zone de mémoire allouée. Ensuite on passe le tout au contrôleur de la carte. Cette dernière déclenche une interruption matérielle quand la zone mémoire qui lui a été attribuée est remplie. Il faut intercepter cette interruption et relancer la numérisation.

D'un côté, on a une intégration macroscopique des fonctions multimédias, qui n'offre pas une souplesse suffisante pour sortir du cadre très conventionnel de ce qu'elle a prévu. Et de l'autre, on a une manipulation très technique de bas niveau. Cette dernière motive une extension du niveau d'abstraction qui sera entravée par la première, imparfaite, mais solidement intégrée dans le système d'exploitation.

En résumé, Windows NT offre la possibilité de partager les ressources sonores (gérées au niveau du système d'exploitation) entre les applications. Ce partage est exclusif à un moment donné : si une application utilise un son numérique pour une rétroaction, une autre ne pourra pas le faire pour une notification simultanément. Les deux ensembles pourraient pourtant être utiles car il est possible d'être attentif à plusieurs types de son. D'autres combinaisons sont impossibles, telles que l'enregistrement et la restitution simultanée d'un son numérique, ou la numérisation et la lecture sur le port MIDI. La première s'explique par les

¹³ Multimedia Control Interface

¹⁴ Microsoft Foundation Classes

possibilités de la carte sonore, la deuxième par celle du pilote logiciel. Par contre, l'organisation est bien modulaire : entre une carte sonore et une application, il y a deux couches de pilotes de périphérique, l'une spécifique à la carte, l'autre spécifique à l'API. Il devrait donc être possible d'améliorer le traitement du son sans altérer l'organisation du système.

1.3.2.2 Les services sonores sous Unix

L'absence totale de ce type de considération et la simplicité d'utilisation des ressources sonores sur les stations de travail du monde Unix ont suscité de nombreuses recherches visant à fournir un partage des ressources sonores entre les applications distribuées sur le réseau, et permettant éventuellement un mixage de plusieurs sons. Comme on pouvait s'y attendre sur la plate-forme type pour ce genre de mécanisme, le partage des ressources sonores se fait au moyen d'un serveur audio dont les clients sont les applications. On reprend pour le son le principe du serveur graphique d'X-Windows qui sert d'intermédiaire entre les applications et l'affichage.

Le *VOX audio server* décrit dans [BINDING 89] a été développé selon ce principe pour le compte d'Olivetti. Il encapsule les dispositifs sonores disponibles dans des entités logiques appelées LAUD (Logical Audio Devices). Ces entités logiques isolent les applications des spécificités du matériel, assurent le partage entre plusieurs applications et rendent leurs services accessibles de manière transparente sur le réseau. Elle peuvent être assemblées en CLAUD (Composite LAUDs) pour remplir une fonction particulière. L'exemple donné est celui du répondeur (figure 7).

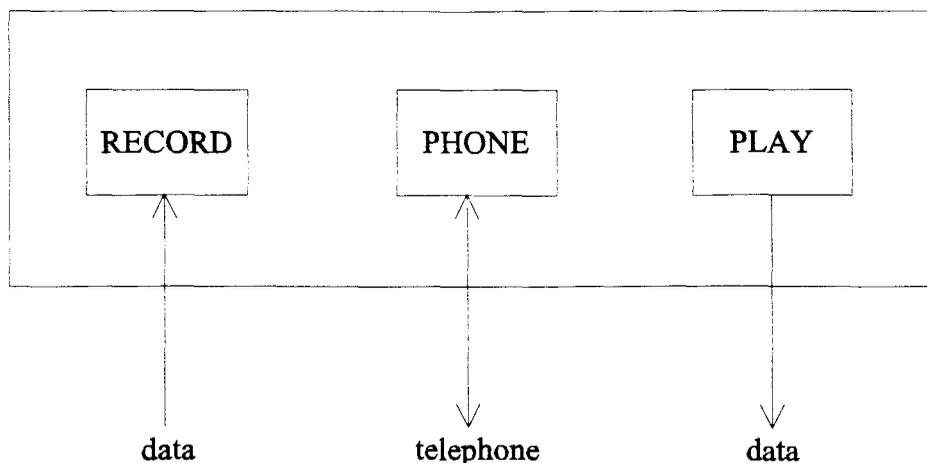


figure 7 : un CLAUD "répondeur" formé de trois LAUDs

Ce serveur audio a été développé en collaboration avec Arons et Schmandt qui a agi en tant que consultant. Schmandt a par ailleurs dirigé le développement d'un travail équivalent, pour le compte de Digital cette fois [ANGEBRANNDT 91]. On retrouve l'organisation du *VOX audio server* à deux détails près. Les LAUD sont devenus des LOUD qui peuvent être assemblés en LOUD composites. Le même exemple de répondeur est repris pour illustrer l'architecture.

Arons [ARONS 92c] a pris du recul par rapport au principe de ces hiérarchies de composants logiques et proposé une approche plus simple où le serveur audio n'a qu'un seul composant qui centralise les requêtes sonores, l'accent étant davantage porté sur les services et sur les mécanismes de communication.

2 Exploitation du son dans les systèmes informatiques

2.1 Notification, rétroaction et sonification : un premier niveau d'exploitation du son

Un premier niveau de l'exploitation du son dans les systèmes informatiques consiste à manipuler des sons synthétisés ou enregistrés à l'avance. Plusieurs taxinomies existent pour organiser ce domaine d'applications. Elles sont basées sur différents types de sons ou différents usages d'un même type de son. Dans [BARME 92] nous distinguons trois types de sons : les bruits, la musique et la voix.

Beaudouin-Lafon [BEAUDOUIN-LAFON 94b] propose une taxinomie basée sur l'usage du son dans les systèmes interactifs. Outre les usages musicaux et ceux de la parole et de la voix, il apporte en plus une distinction pour les bruits :

- *feed-back sonore* : utilisation du son comme moyen de rétroaction, principalement des bruitages, sans exclure l'usage de messages vocaux. La rétroaction désigne une information fournie par le système informatique à l'utilisateur comme réponse à son action.
- *notification* : utilisation du son pour signaler des événements. On n'exclut pas l'utilisation de messages vocaux qui peuvent jouer le même rôle qu'un bruitage. On remarque que dans le cadre d'un collecticiel en mode synchrone, cela peut présenter un inconvénient en se superposant aux échanges verbaux entre les participants.

La communauté scientifique réunie à ICAD'92¹⁵ a dégagé une troisième distinction pour les bruits appelée *sonification*, voir la préface de [KRAMER 94a]. La sonification désigne l'utilisation du son pour remplacer ou compléter l'affichage de données par le canal visuel [KRAMER 94b]. La sonification possède un potentiel insoupçonné, comme par exemple le débogage de programmes à l'oreille, avec un langage de programmation classique [JAMESON 94] ou parallèle [JACKSON 94], la représentation sonore de données multivariées [MADHYASTHA 94; BLY 94], l'étude de sismographes [HAYWARD 94].

Quoique Jonathan Cohen [COHEN 94] rende hommage aux travaux, qui remontent à 1952, de John Cage sur la représentation d'informations au moyen du son, l'utilisation des bruitages dans les systèmes informatiques est récente. Gaver [GAVER 89] en a été l'un des précurseurs. Il met bien en avant la complémentarité du son et de l'image qui découle de leurs relations spatiales et temporelles respectives. Gaver résume cette complémentarité en croisant les sons et la vision avec le temps et l'espace. Quatre phrases condensent cette réflexion en jouant sur les significations de *in* et *over* : *Sound exists in time. Sound exists over*

space. Visual objects exist over time. Visual objects exist in space. Le tableau 2 extrait de [GAVER 89] illustre la complémentarité des sons et des objets visuels.

	TEMPS	ESPACE
SON	<p><i>Sound exists <u>in</u> time.</i></p> <ul style="list-style-type: none"> • Convient bien pour rendre compte d'événements changeants. • Disponible pour un temps limité. 	<p><i>Sound exists <u>over</u> space.</i></p> <ul style="list-style-type: none"> • Ne nécessite pas d'être face à la source. • Un nombre limité de messages peuvent être rendus simultanément.
VISION	<p><i>Visual objects exist <u>over</u> time.</i></p> <ul style="list-style-type: none"> • Convient bien pour l'affichage d'objets statiques. • Peut être échantillonné dans le temps (image par image). 	<p><i>Visual objects exist <u>in</u> space.</i></p> <ul style="list-style-type: none"> • Nécessite d'être face à la source. • Les messages peuvent être répartis dans l'espace.

tableau 2 : complémentarité du son et de l'image

Brown et al [BROWN 89] ont vérifié la complémentarité son/vision par une expérience. De même, une interface tactile (braille) et une interface sonore peuvent se compléter dans le cas des non voyants [MYNATT 94a]. Des recherches ont été entreprises pour affiner l'utilisation des messages sonores dans les Interfaces Homme - Machine [MYNATT 94b; BLATTNER 94; BREWSTER 94].

La rétroaction, la notification et la sonification exploitent toute une gamme d'informations qui peuvent être transmises par l'intermédiaire du son, telle que la spatialisation du son, ce qu'a fait [WENZEL 94]. Michael Cohen a approfondi la manipulation du son : il a étudié les effets que l'on peut obtenir en jouant sur les

¹⁵ International Conference on Auditory Display

différents paramètres d'un son. Dans [COHEN 91], il propose d'utiliser la spatialisation du son et l'altération des sons. Dans [COHEN 93], il présente une synthèse sur le potentiel de l'altération des sons comme moyen de rétroaction, notification et sonification. La musique, à première vue, serait perçue comme un document qu'il faut être capable de présenter dans une interface multimédia, tout comme on le fait avec des photos, des images ou des vidéos. Sa forte structuration et le codage précis des sons musicaux en a fait aussi un moyen de rétroaction privilégié, décrit par exemple dans [BLATTNER 89; KENDALL 91; KARSHMER 92].

La plupart des travaux dans ces domaines étudient le potentiel de ce premier niveau d'exploitation du son pour une interface mono utilisateur. Cependant, leurs principes s'étendent facilement aux collecticiels et certains l'ont fait très tôt avec succès [GAVER 90; KARSENTY 91; COHEN 91].

Avec l'usine virtuelle ARKola, conçue pour une expérience visant à démontrer l'intérêt de la notification sonore, Gaver et al [GAVER 91a] ont, en fait, mis en évidence le potentiel de l'aspect sonore dans une interface de collecticiel synchrone. Dans cette expérience, deux sujets ont en charge la surveillance d'une usine virtuelle de conditionnement de boissons gazeuses. Ils disposent chacun pour cela d'un écran de contrôle qui ne leur permet de visualiser qu'une partie de l'usine à la fois. Les machines de cette usine peuvent tomber en panne ou nécessiter une intervention. L'expérience est menée sous deux conditions : avec ou sans bruitage sonore. Les bruitages donnent des indications quant au bon fonctionnement des machines. L'expérience a révélé que non seulement les sujets exploitent les bruitages pour suivre le bon fonctionnement des machines qu'ils ne voient pas sur leur écran de contrôle, mais encore qu'ils mettent à profit cette possibilité pour s'entraider à résoudre les problèmes. Ainsi, sans bruitage, chacun s'occupe d'une partie de l'usine et il y a peu d'interaction entre les sujets, alors qu'avec les bruitages, ils coopèrent davantage, s'appuyant sur les sons qu'ils perçoivent pour surveiller le bon fonctionnement des autres machines.

Pourtant, ce premier article ne parle que de *group work* dans les mots clés, l'expérience ayant surtout été menée pour prouver l'intérêt des bruitages sonores. Cependant, l'expérience a aussi démontré l'intérêt des bruitages sonores pour le travail coopératif. Dans un autre article publié à la fin de la même année, Gaver [GAVER 91b] met bien en évidence l'application au travail coopératif et précise qu'il s'agit d'un travail à peine entamé.

2.2 Annotations sonores : exploitation élémentaire de la voix

La première image qui vient à l'esprit quand on associe voix et informatique est sans doute celle de *Hal*, l'ordinateur intelligent de "2001: A Space Odyssey", [CLARKE 68]. L'idée de pouvoir dialoguer avec un ordinateur comme avec un être humain a été reprise dans d'autres ouvrages de science-fiction. Le monde de la recherche s'est aussi très sérieusement penché sur le problème selon deux axes : la reconnaissance de la parole et la synthèse vocale. Ces aspects sont chacun des domaines de recherche à part entière, introduits dans [CALLIOPE 89]. Ils ne sont pas traités dans cette thèse. Dans le domaine des IHM, le traitement de la parole s'inscrit dans le cadre des recherches sur les interfaces multimodales qui exploitent conjointement les différents modes d'interaction mis à la disposition de l'utilisateur : clavier, souris, stylo, voix et capteurs de positions (essentiellement des mains).

Il est vrai que la reconnaissance de la parole est à la base d'un grand nombre d'applications. Cependant dans le cadre du Travail Coopératif en mode synchrone avec un canal sonore, la parole est déjà fortement exploitée entre les utilisateurs. L'exigence en termes de qualité et de performance est telle dans ce contexte, que les systèmes actuels de reconnaissance de la parole paraissent pour le moment difficilement exploitables, ce que souligne Hindus [HINDUS 92] : la reconnaissance de la parole dans le cas d'une expression en langage naturel fluide et sans contrainte est encore loin de la réalité.

Cependant, Schmandt [SCHMANDT 94] exploite au mieux un premier niveau d'analyse de la parole. Par exemple, il tire parti simplement de la prosodie de la

parole avec efficacité. Cette technique a été utilisée pour consulter son courrier électronique à distance dans le cadre d'un système mono-utilisateur [SCHMANDT 88].

D'autres travaux témoignent des efforts entrepris pour exploiter la parole dans les systèmes informatiques, tel que la manipulation de messages vocaux appelés annotations sonores quand ils sont insérés dans un document. Elle convient surtout pour de brefs messages, car si elle offre un moyen très commode à celui qui veut commenter un document, elle est moins pratique pour ceux à qui sont destinés les commentaires. En effet, un message est plus vite enregistré sous forme sonore qu'écrite, mais moins commodément écouté que lu. Grudin [GRUDIN 88] en parle dans son étude sur les causes d'échec potentiel des collecticiels. Il souligne que la voix numérisée par rapport à une saisie au clavier avantage presque uniquement celui qui enregistre le message. L'expression orale est plus rapide, transmet facilement de l'émotion et des nuances, et est accessible indépendamment d'un terminal d'ordinateur. Les inconvénients d'un message enregistré sont accablants pour celui qui l'écoute. C'est plus difficile à comprendre. Il faut plus de temps pour en prendre connaissance. C'est difficile à parcourir rapidement ou à vérifier. Un message enregistré a plus de chances de contenir des erreurs et est difficile à manipuler.

En effet, un message est fortement structuré par l'information qu'il contient. Cette structure peut être visible s'il est écrit et perdu s'il est oral. Cela ne gêne pas notre compréhension mais rend la manipulation des messages beaucoup plus difficile. Arons [ARONS 93] attribue à la nature temporelle des sons ces problèmes de manipulation en comparaison avec celle de documents affichés visuellement.

Muller et Daniel [MULLER 90] ont travaillé sur la structure des annotations sonores qu'ils appellent un document vocal. Ils le codent sous la forme d'un hypermedia qui permet de rompre la caractéristique séquentielle inhérente à une

annotation sonore. Un document vocal est défini comme une liste chaînée de *PVS*¹⁶. A chaque noeud, il est possible de commander vocalement une *excursion*, ce qui correspond à l'exécution d'un lien dans un hypertexte, mais dynamiquement. Une telle structuration offre un fort potentiel de manipulation. Le problème de la création d'un tel document vocal est apparu aux auteurs qui proposent des applications d'aide à leur création.

Degen et al [DEGEN 92] ont abordé le problème à l'envers. Plutôt que de chercher à imposer une structure précise à un message sonore en vue d'une manipulation ultérieure, ils ont cherché à capter au vol des indices qui serviront de points de repère lors de l'écoute. Ils ont adapté pour cela un magnétophone portatif pour insérer dynamiquement ces repères sur un message enregistré. Ces repères sont exploités lors de l'écoute du message par un logiciel adapté. Ils permettent des points d'entrée directement accessibles. Ces appareils ont été testés et appréciés par des utilisateurs réels qui ont souhaité cependant plus de services d'annotation des repères eux-mêmes.

Le problème de la manipulation d'annotation sonore a été un important thème de recherche de Schmandt [SCHMANDT 94]. Schmandt [SCHMANDT 93] a, en partie, contourné le problème en prenant un contexte favorable, c'est-à-dire en l'absence d'autres moyens d'interaction tels que l'écran et le clavier. Le domaine de prédilection de Schmandt inclut l'utilisation du téléphone comme terminal d'accès à distance ainsi que toutes les nouveautés en matière de dispositifs informatisés portables.

Arons, qui a travaillé avec Schmandt dans un premier temps, a approfondi le problème des annotations sonores. Pour traiter le problème de la manipulation des messages, il a combiné plusieurs techniques qu'il expose dans [ARONS 93] : la compression temporelle (détaillée par ailleurs dans [ARONS 92b]) pour accélérer l'écoute des messages, la détection automatique des silences pour leur suppression

¹⁶ Parsed Voice String

ou comme point d'entrée et même l'écoute dichotique¹⁷ (détaillée par ailleurs dans [ARONS 94b]). Son travail s'étend à la conception et réalisation d'un clavier spécial pour la manipulation d'annotations sonores. L'intérêt de l'approche d'Arons par rapport à celles de Muller ou Degen est le souci d'automatisation de la structuration du message vocal par l'exploitation de techniques très variées telle que la détection de prise de parole.

2.3 Détection de prise de parole

L'utilisation de sons enregistrés ou synthétisés pour la rétroaction, notification et sonification constitue un premier niveau d'exploitation du son dans les systèmes informatiques. L'utilisation, plus complexe, de messages sonores est une deuxième étape. Or cette utilisation induit, comme on l'a vu, la nécessité d'une compréhension même très sommaire du contenu sémantique des messages sonores. L'étape suivante consiste donc à aborder l'analyse des communications verbales. La détection automatique de prise de parole en est le niveau le plus élémentaire.

La détection automatique de prise de parole est couramment utilisée dans les systèmes de prise de son, que ce soit pour l'enregistrement, la sonorisation ou la diffusion à la radio ou la télévision. Elle est réalisée par des appareils à détecteur de seuil (*noise gates*) directement derrière les microphones. Ils permettent de ne transmettre le signal que lorsque quelqu'un parle, éliminant ainsi les bruits de fond qui seront masqués par la parole ; rien n'est transmis pendant les silences.

C'est à partir de tels *noise gates* que Sellen [SELLEN 92] a conçu un système de détection de prise de parole (et explicitement, non plus de silence cette fois). Les *noise gates* commandent le déclenchement d'un générateur de sons purs. Un détecteur de son les convertit en messages de contrôle MIDI envoyés à un ordinateur qui les estampille avec un codage SMPTE¹⁸ et les enregistre. Ce système a été utilisé dans le cadre d'une expérience pour évaluer différentes

¹⁷ L'écoute dichotique consiste à entendre deux messages différents au niveau de chaque oreille

situations de discussion : dans une même pièce et à distance avec deux systèmes vidéos. Les enregistrements des détections des prises de parole effectués en temps réel par ce système pour le moins complexe ont, en plus, fait l'objet d'un traitement supplémentaire avant leur exploitation pour en affiner la validité.

La détection de silence (équivalente à celle de la prise de parole) est aussi utilisée dans les systèmes de téléconférence sur les réseaux pour compenser les délais et diminuer la bande passante nécessaire [BOURGUIN 95]. Elle est réalisée par différentes techniques à partir du son numérisé. Les principales techniques de détection des silences sont résumées et critiquées dans [ARONS 93], qui ne mentionne cependant pas celle utilisée par Sellen. La technique finalement retenue par Arons n'est guère plus simple que celle de Sellen. Elle n'est pas plus utilisable en temps réel puisqu'elle nécessite plusieurs étapes, chacune avec un examen de l'ensemble des échantillons.

Comme on peut le voir à travers ces quelques exemples, la détection automatique du locuteur (le problème dual de la détection automatique des silences) est d'abord un problème technique complexe. Cependant, ses applications sont vastes et variées. A mi-chemin entre ce domaine et celui des annotations sonores, on trouve aussi le travail d'Hindus et Schmandt [HINDUS 92]. Ils proposent d'enregistrer et d'analyser automatiquement les conversations d'un groupe de personnes pour aider l'exploitation ultérieure d'une discussion.

2.4 Spatialisation du son

La spatialisation du son est transversale à l'exploitation des bruitages, la manipulation des annotations sonores et la détection de prise de parole. Elle peut être appliquée au niveau de la prise de son avec des antennes acoustiques, lors de la restitution de n'importe quel type de son, bruitage ou message sonore, et aussi en temps réel, lorsque le son ne fait que transiter dans le système informatique.

¹⁸ Single Motion Picture Time Encoding : système de codage temporel utilisé pour la vidéo.

La spatialisation du son en transit dans le cadre d'une audioconférence assistée par ordinateur est l'une des trois principales propositions de cette thèse et sera exposée plus loin. La spatialisation du son concerne bien entendu tous les types de sons et sera bénéfique dans toutes les situations de collecticiels. Une étude précise et approfondie de ce domaine est présentée dans [WENZEL 92].

On ne connaît pas bien l'importance respective de chacun des paramètres (cf. *Localisation de l'origine spatiale d'un son* page 26) qui entre en jeu dans notre perception de la position d'une source sonore. Une manière de contourner le problème est d'avoir une approche globale pour les quatre premiers paramètres et négliger les quatre derniers.

Pour exploiter cela dans les interfaces des systèmes informatiques, l'approche la plus courante est celle des HRTF : *Head Related Transfer Filters* expliquée dans la plupart des articles, notamment [WENZEL 92]. Cette technique consiste à enregistrer la perception sonore d'un sujet passif. De petits microphones sont placés au fond des deux canaux auditifs du sujet. Plusieurs sons de fréquences variées sont produits à différentes positions tout autour du sujet et enregistrés par ces petits microphones qui subissent ainsi les quatre premiers paramètres donnés plus haut. On effectue ainsi un ensemble de mesures qui seront interpolées lors de la synthèse sonore pour spatialiser un son.

La technique des HRTF a été appliquée dans bien des domaines, depuis les systèmes d'aide au pilotage d'avion (dès 1946) jusqu'à la simulation de salle de concert [JOT 93]. Comme la technique des HRTF demande une grande puissance de traitement, Burgess [BURGESS 92] propose une approche plus économique en simplifiant et en optimisant le processus de calcul des interpolations.

La recherche sur la réalité virtuelle a redonné un surcroît d'intérêt à la spatialisation du son. Cela a été rendu possible par l'augmentation significative des possibilités de traitement du son par des processeurs spécialisés dans le traitement du signal, les DSP. Cette fois, on synthétise un par un chacun des paramètres intervenant dans la spatialisation du son [MULDER 94; POPE 93]. La technique

semble particulièrement efficace puisque les adeptes de l'HRTF commencent eux aussi à adjoindre avec succès des indices synthétisés dans leurs systèmes [JOT 93; BURGESS 94].

L'HRTF est loin d'être parfait (problème de localisation avant/arrière par exemple) comme le reconnaissent ses utilisateurs et présente des inconvénients [RICARD 94]. En effet, les filtres mesurés ne sont bien adaptés que pour la personne avec qui on les a mesurés, ce qui n'est pas une surprise puisque personne n'a la même tête. D'autre part, l'HRTF ne met en oeuvre que quatre des paramètres de la spatialisation du son alors que les techniques de synthèse les traitent tous, pouvant même les amplifier artificiellement. Enfin, l'HRTF suppose l'utilisation d'un casque, ce qui n'est pas adapté à toutes les situations, notamment pas celle du travail coopératif.

En effet, même le port d'un simple casque auditif peut entraver la liberté de mouvement et de discussion, sans parler de l'inconfort quand il doit être porté longtemps. Heureusement, le contexte du Travail Coopératif supporte une relaxation du degré de qualité de la spatialisation. A priori, il n'est pas nécessaire de pouvoir discuter avec quelqu'un derrière, dessous ou au-dessus de soi.

Il reste donc une troisième approche, déjà étudiée pour les bruitages [BEAUDOUIN-LAFON 94a] qui consiste à utiliser simplement un dispositif stéréophonique avec une paire d'enceintes. Dans le compte rendu du *CHI'94 Workshop* [ARONS 94a] cette approche est déjà pressentie. On y note qu'il doit être possible d'aller loin sans recourir à un système complet de spatialisation du son, par exemple jouer sur la balance du volume entre les enceintes droites et gauches peut être suffisant dans certaines applications.

Enfin, Arons [ARONS 92a] a récemment remis en question l'aspect purement spatial du "cocktail party effect" pour mettre en avant d'autres facteurs tels que la sémantique des échanges (cohérence, prédictivité du discours), le timbre des voix, les accents, etc. Il relate aussi l'utilisation de modifications artificielles des voix pour apporter des indices utiles.

3 Les travaux relatifs aux aspects sonores des collecticiels temps réel et multimédia

Au travers d'un certain nombre de travaux significatifs, mais non exhaustifs, il est possible de mettre en évidence les problèmes posés par les aspects sonores dans les collecticiels temps réel et multimédia, et bien souvent l'absence de réelles solutions.

L'étude des travaux antérieurs a donc confirmé le domaine choisi comme étant une zone à défricher. L'exploration des travaux sur les collecticiels synchrones distribués est en effet décevante quand on s'intéresse à l'aspect sonore qui se trouve davantage mentionné qu'étudié : il apparaît plus comme une préoccupation secondaire. Un son monophonique de mauvaise qualité (celle du téléphone) est mis en œuvre avec un service minimum : des échanges verbaux entre les participants. L'importance de l'aspect sonore est pourtant reconnu comme le dit Schooler [SCHOOLER 91a] dans une synthèse des travaux sur les téléconférences où elle affirme que les systèmes de téléconférences sont impraticables sans communication vocale en temps réel. Schooler insiste aussi sur l'importance de la qualité du son qui passe avant celle de l'image car la plupart des informations transitent par le canal audio.

Avant d'aborder plusieurs classes de travaux de recherche, nous pouvons étudier plus précisément un projet MERMAID [MAENO 91], développé dans la société NEC Corporation. Ce projet vise à valoriser les réseaux numériques à intégration de services tels qu'ils ont commencé à émerger au début des années 1990. Le but est la réalisation d'un système de téléconférence multimédia, utilisé en mode synchrone et réparti dans des postes de travail individuels.

MERMAID est le sigle correspondant à *Multiple Environment Remote Multiple Attendee Interactive Decision-making*. Chaque poste de travail est équipé d'un PC sous OS/2, d'une caméra, d'un microphone, d'une tablette à digitaliser pour saisir l'écriture manuscrite. Il permet à un groupe de personnes installées à distance de discuter ensemble en visualisant simultanément des documents multimédias qu'ils

peuvent annoter. MERMAID utilise deux types de réseaux : Ethernet pour les données, ISDN pour le son et la vidéo (compressée par CODEC). On voit apparaître la notion de serveur audio. Les voix des participants sont mixées par le serveur audio et transmises à tous les autres participants simultanément. Ce serveur audio est en fait réalisé par le système d'audioconférence d'un autocommutateur téléphonique.

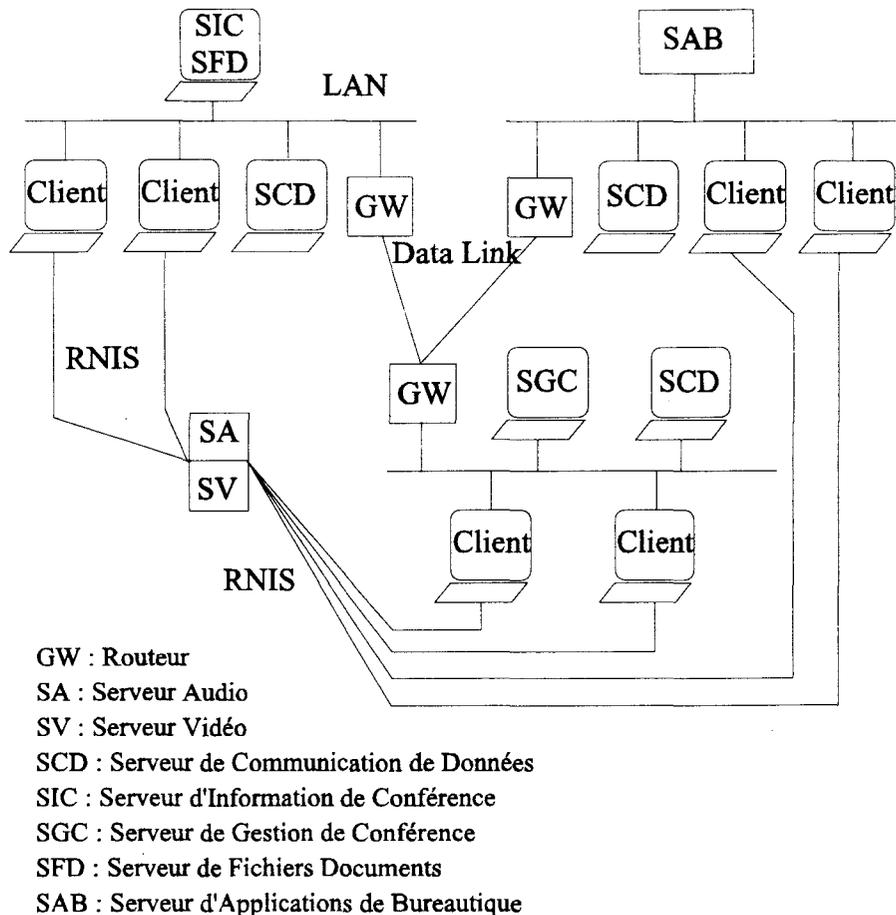


figure 8 : Architecture de MERMAID

Au niveau de l'architecture logicielle, MERMAID s'appuie sur un modèle fonctionnel dégageant des groupes logiques : interface utilisateur, contrôle de la communication, application, base de données pour la gestion des éléments du système et l'interface de communication appelé *Group Communication Architecture* (GCA). L'implémentation révèle une répartition des services en plusieurs serveurs qui reprennent le découpage fonctionnel prévu. Les auteurs

assurent le bon fonctionnement du système et annoncent le développement d'applications coopératives telles que des éditeurs partagés exploitant la GCA.

Si MERMAID a permis une utilisation réelle d'un système de téléconférence assisté par ordinateur, celle-ci a en fait révélé de nombreux problèmes. Watabe et al [WATABE 90] rapportent les problèmes de l'intégration du canal audio.

"La voix a été le médium le plus simple et le plus communément utilisé. Cependant, quand plus de quatre personnes sont en conférence et que les participants ne connaissent pas encore leurs voix, il y a des difficultés pour déterminer qui parle. Comme les microphones incorporés dans le haut parleur ne captent pas les mots prononcés à voix basse suffisamment clairement, les utilisateurs trouvent qu'il est parfois difficile de distinguer ce qui est dit et par qui."

3.1 Les projets orientés communication multiparties

Il existe de nombreux projets dans le domaine des collecticiels dont le point d'entrée est constitué par des services de communication multiparties ($1 \rightarrow N$ et $M \rightarrow N$). Alors que dans les réseaux de données traditionnels, les mécanismes de sessions sont en général orientés pair à pair ($1 \leftrightarrow 1$).

Il existe bien sûr des mécanismes de très bas niveau permettant l'acheminement de paquets d'informations vers de multiples destinations appartenant à un même groupe et ceci simultanément. C'est le multicast.

A partir de cela, des chercheurs ont exploré le potentiel des réseaux numériques pour réellement créer des interactions multiparties. Dans les premières réalisations de téléconférence sur les réseaux informatiques, les problèmes abordés sont purement techniques. Le transport a été le premier problème auquel on a été confronté pour le son, suivi ensuite de près par celui de la gestion des connexions. Les deux ensembles ont été étudiés du point de vue de l'architecture logicielle et matérielle.

3.1.1 Une approche liée aux problèmes d'acheminement et de routage dans les réseaux

Dès 1990, Leung et al [LEUNG 90] proposent une solution qui semble résoudre tous les problèmes, y compris l'intégration des applications existantes sous Unix. Cependant, ils abordent essentiellement les problèmes d'acheminement et de routage en mettant l'accent sur les fonctions d'éclatement et de regroupement nécessaires aux communications multiparties. Ils proposent le concept de *connector* pour l'interconnexion des composants logiciels et matériels. Un *connector* assure le contrôle dynamique de la diffusion des données entre les composants. Le son, la vidéo et les données sont bien sûr séparés d'un point de vue logique mais transportés sur un seul réseau. Ils disent utiliser pour cela un réseau à commutation de paquet expérimental offrant un débit supérieur à 100 Mbits/s. Pourtant, ils prévoient sans hésitation l'utilisation de l'ISDN (équivalent anglais du RNIS¹⁹), dont le débit est de 64 kbits/s pour deux canaux de données. Les autres composants sont eux aussi plus ou moins ad hoc. En ce qui concerne le son, on relève que la communication verbale sera réalisée par un terminal téléphonique. On déplore l'absence de commentaires sur l'utilisation de l'ensemble du système.

Le projet *Rapport* décrit par Ensor et al [ENSOR 91] propose de même une vue d'ensemble depuis le multimédia jusqu'aux applications de partage de documents. Le son, la vidéo et les données sont transportés sur des réseaux différents et variés, depuis des réseaux câblés dédiés jusqu'à des réseaux informatiques propriétaires en passant par l'ISDN et Ethernet. Ils ont utilisé le réseau téléphonique, l'ISDN et leur propre pont audio pour transmettre la voix, des lignes analogiques pour la vidéo, un réseau Ethernet et *HPC/VORX* pour les données et les signaux de contrôle. Il n'y a malheureusement pas de commentaires comparatifs de ces différentes solutions.

Tout comme Schooler, Ensor note au passage l'importance du canal audio. L'accent est cependant mis sur le problème du contrôle des connexions, de

l'échange d'informations et de la coordination des différents médias. Le système de contrôle est simple et logique, son implémentation a permis aux auteurs d'avoir une première expérience du fonctionnement d'un tel collecticiel. Quant à l'aspect sonore, l'article précise que le pont audio réalise un mixage de type n-1 : chaque participant reçoit toutes les autres voix hormis la sienne; un mixage spécifique est donc utilisé par participant.

Avec une approche plus spécialisée, Ramanathan et al [RAMANATHAN 92] ont analysé le potentiel d'un mixage hiérarchisé du son. Une fois posé le principe du mixage de plusieurs sources audio réparties sur une arborescence d'ordinateurs, ils en expriment mathématiquement les délais engendrés, en prenant en compte les temps de mixage et les temps de transport entre les différents noeuds au sein de l'arbre de mixage. On relève un problème potentiel : un seul mixage commun pour tous les participants est envisagé. Ceci va à l'encontre des techniques employées dans les systèmes réels (comme *Rapport* par exemple) où chaque participant doit recevoir les sons de tous les autres participants mais pas le sien pour éviter l'effet Larsen (l'effet Larsen est un sifflement strident à la fréquence de résonance du dispositif d'amplification du son, dû à un bouclage du son amplifié). Pourtant, les auteurs en sont conscients puisqu'ils prévoient que les participants pourraient avoir à extraire leur propre son du mixage restitué. Que deviennent les résultats si l'on souhaite fournir à chaque participant un mixage et des effets qui lui sont spécifiques?

Toujours selon une approche centrée sur les problèmes d'acheminement et de routage dans les réseaux, mais à un autre niveau, Clark [CLARK 92] décrit les projets *MIAC*²⁰ puis *MIAS*²¹ où l'effort est porté sur une meilleure intégration du flux multimédia. A l'opposé du projet *Rapport*, un seul réseau (ISDN) est utilisé pour le son, la vidéo et les données. Clark donne plusieurs exemples de répartition des différents médias sur les deux canaux B d'une liaison ISDN. Au niveau de la

¹⁹ RNIS : Réseaux Numérique à Intégration de Services

²⁰ Multipoint Interactive Audio-visual Communication

²¹ Multipoint Interactive Audiovisual System

configuration du réseau, Clark écarte les réseaux maillés du fait du trop grand nombre de connexions qu'ils impliquent et propose une extension du réseau étoilé par l'intermédiaire de "concentrateur", des *MCU*²², ce qui rejoint le concept du *connector* de Leung (figure 9).

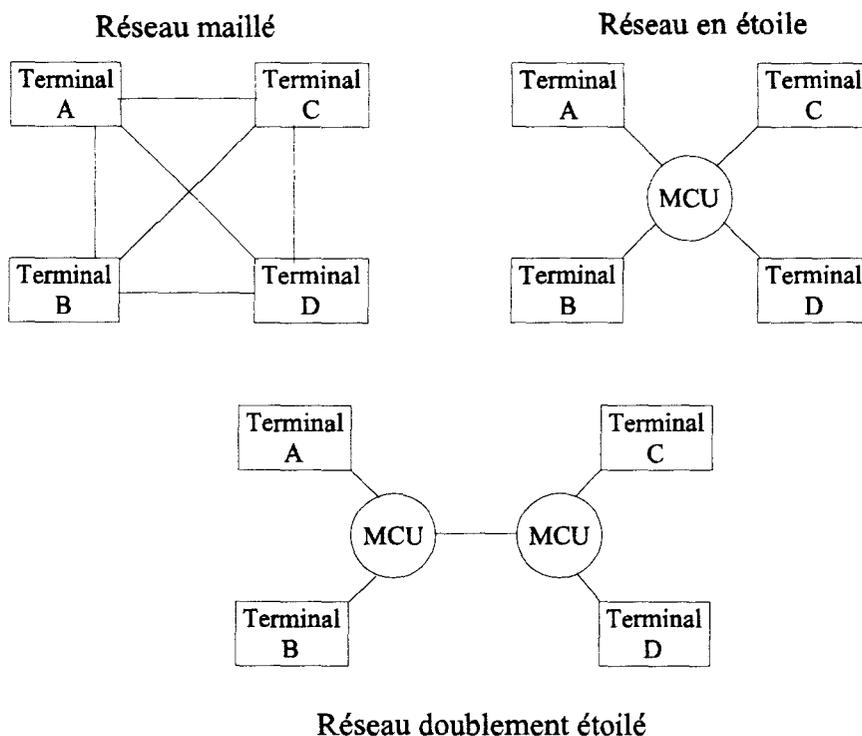


figure 9 : différentes organisations de réseau

3.1.2 Une approche plus abstraite par rapport au réseau de transport

A l'opposé de ces considérations très pragmatiques, Rangan et Vin [RANGAN 91] ont poussé la réflexion sur le problème du contrôle. De leur rapide survol de l'état de l'art, ils mettent en exergue le niveau élémentaire auquel les travaux ont abordé le problème du son et l'absence de réflexion sur la problématique de la conférence à part entière.

Ils proposent donc un modèle formel qui part d'une réflexion sur les conférences, élevées au rang de paradigme, et poussent l'analyse jusqu'à la

²² Multipoint Control Unit

définition de commandes, dans le but de permettre des interactions multimédias dans un collecticiel synchrone distribué. Rangan et Vin font les propositions suivantes :

- deux types de participants, un simple utilisateur et un *super* participant qui est une conférence entière; les conférences et les participants sont munis d'attributs qui précisent leurs caractéristiques tant d'un point de vue instantané que durable,
- une taxinomie des conférences avec les distinctions suivantes :
 - statique ou dynamique,
 - simple ou super (une super conférence contient au moins un super participant),
 - sans points communs ou hiérarchiques,
 - fugitive ou persistante,
 - symétrique ou asymétrique (dans une conférence asymétrique, les participants peuvent utiliser des moyens de communication différents),
 - homogène ou hétérogène (les conférences hétérogènes comportent des moyens de conversions d'un moyen de communication à l'autre),
 - séquentielles ou concurrentes,
- un modèle de connexion pour les conférences multimédias,
- des opérations sur les conférences :
 - création d'une conférence,
 - ajout d'un participant dans une conférence,
 - retrait d'un participant d'une conférence,
 - terminaison d'une conférence,
 - fusion de conférences,
 - scission d'une conférence,

- copie et déplacement de conférence,
- contrôle du mode d'une conférence (active ou inactive)
- modification des attributs d'une conférence.

Les travaux de Rangan et Vin sont purement théoriques bien qu'ils mentionnent une implémentation en cours. Dans cette thèse, nous présentons un travail qui :

- affine l'approche de Rangan et Vin en se concentrant sur l'aspect sonore,
- prolonge la réflexion de Rangan et Vin en confrontant l'approche théorique à une réalisation pratique : nous avons ainsi été amené à adapter certaines des propositions théoriques de Rangan et Vin (détaillé au Chapitre 2).

Avec un point de vue plus macroscopique, Schooler présente une architecture complète : MMCC²³ et pose des problèmes génériques aux collecticiels. [SCHOOLER 91b, 93b]. Dans MMCC, on s'intéresse au problème de l'extension du nombre de participants et de l'étendue du réseau. Schooler focalise alors son attention sur les problèmes d'échelle dans ce type de système [SCHOOLER 93a].

3.2 L'intégration du son dans une architecture logicielle

Si la plupart des travaux sur les collecticiels synchrones temps réel n'ont fait qu'évoquer l'aspect sonore, ceux présentés par Anderson se concentrent sur l'aspect sonore et ne font qu'évoquer la dimension du collecticiel. L'objectif de ses travaux est d'exploiter le potentiel d'abstraction de la programmation à objets pour la gestion de composants concrets (microphone, haut-parleur, mixeur...). Anderson a par ailleurs proposé MOOD [ANDERSON 91a], une hiérarchie de classes écrites en C++ pour les applications manipulant les sons musicaux.

Anderson [ANDERSON 91b, 92] propose aussi une boîte à outils sous la forme d'une hiérarchie de classes écrite en C++ pour les applications multi-utilisateurs manipulant du son et de l'image vidéo. Un peu comme dans [LEUNG 90],

²³ MultiMedia Conference Control program

COMET permet de résoudre tous les problèmes. Aucun exemple de réalisation n'est indiqué. Anderson précise que son prototype présente une lacune importante : il ne prend en compte que l'aspect sonore. D'après l'auteur, la solution fournie pour le son par COMET devrait s'étendre sans difficulté à l'image.

Nom	Description	Dynamique
<i>CM Connexion</i>	Connexion pour des données de type média continu (CM) au niveau de transport	oui
<i>Strand</i>	un flux de média continu	non
<i>Rope</i>	une combinaison de flux de médias continus	non
<i>PDev</i>	dispositif physique d'E/S pour des données de type média continu	non
<i>LDev</i>	dispositif logique	oui
<i>CLDev</i>	Ensemble de LDev formant une source ou une destination pour une combinaison de flux de médias continus	oui
<i>LTS</i>	horloge de synchronisation de flux d'E/S	oui

tableau 3 : les abstractions de bases ACME d'Anderson

Les travaux d'Anderson présentent cependant plusieurs aspects intéressants. Ils mettent en avant le concept de media continus (*Continuous Media* : CM), la notion de flux sonores (ou visuels) (*streams*) organisés en cordes (*ropes*) comme une composition de sous-flux élémentaires appelés brins (*strands*). Ces concepts sont présentés précisément dans [ANDERSON 91c] ainsi qu'un ensemble d'abstractions (ACME) permettant de les manipuler au moyen de dispositifs physiques, logiques ou composites (tableau 3). Anderson donne un exemple d'un assemblage d'objets pour une application de téléconférence (figure 10). Cela constitue un premier pas au niveau logiciel vers une utilisation du son numérique complètement intégré dans les réseaux informatiques. Il reste à le concrétiser par une application réelle.

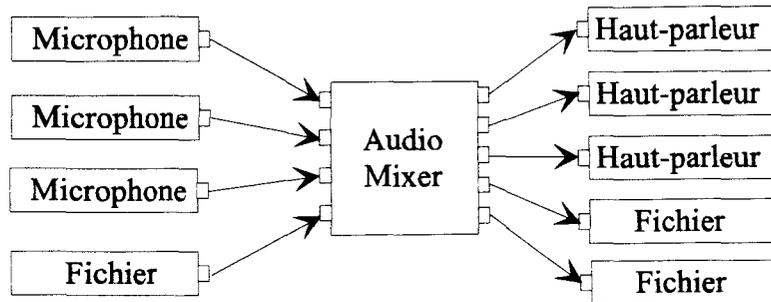


figure 10 : exemple d'assemblage d'objets de base pour une application de téléconférence

3.3 Approche centrée sur l'utilisateur

Les travaux présentés ci-dessus, qui s'attachent à résoudre des problèmes techniques, négligent un aspect essentiel : la dimension humaine. Gaver est l'un des premiers à avoir exploré cette dimension. Il a eu une démarche plus pragmatique tirant parti de l'apport potentiel des Sciences Humaines. Comme Anderson, il aborde le problème par l'aspect sonore. Ses premiers travaux ont eu pour objectif d'explorer le potentiel du son dans des interfaces homme-machine mono-utilisateur avant d'aborder la dimension multi-utilisateurs (voir le paragraphe 2.2 *Notification, rétroaction et sonification : un premier niveau d'exploitation du son*).

Gaver [GAVER 92] apporte ensuite une vision plus globale et une autre façon d'aborder les problèmes posés par MERMAID. Il insiste sur l'importance de l'écologie de la perception. Il s'intéresse plus particulièrement aux possibilités offertes par les systèmes de télécollaboration. Gaver définit les *affordances* comme les propriétés de l'environnement qui offrent des actions aux organismes appropriés :

"Affordances are properties of the environment that offer actions to appropriate organisms."

Cette notion est sans doute plus facile à comprendre sur des exemples. Gaver donne pour cela le cas de la vidéo dans un système de téléconférence comparée à la vision dans une conférence réelle. L'une n'offre qu'un champ de vision étroit et

fixe alors que l'autre offre une grande souplesse d'utilisation, permettant une sélection active et une recherche dynamique. Pour le son, il relève la qualité monophonique du son des systèmes actuels qui empêche la localisation, et donne un extrait partiel de l'information sonore distante. Ainsi, son analyse démontre que les systèmes actuels ne permettent pas d'exploiter toutes les capacités de communication que nous avons en présentiel.

Sa conclusion souligne que les systèmes de télécollaboration ont encore un fort potentiel inexploré de services. C'est une des idées à la base du travail entrepris dans cette thèse, avec en plus le souci d'une approche pluridisciplinaire encouragée par l'expérience du projet Indigo décrit par Harrison et al [HARRISON 94]. Le nom *Indigo* est un pseudonyme pour masquer l'identité réelle de la société qui a utilisé le mediaspace²⁴ étudié. Le projet a réuni des concepteurs d'interface, ingénieurs, psychologues et sociologues. L'article n'aborde aucun problème technique. Il disserte plutôt sur les interactions entre les différentes disciplines au cours du projet. Nous en retenons la nécessité d'une coopération en amont du développement, comme l'affirme aussi Karsenty [KARSENTY 94] qui conclut une synthèse des travaux sur les collecticiels par la nécessité d'une approche pluridisciplinaire, ou plus précisément d'une coopération entre plusieurs disciplines :

"La recherche dans ce domaine est une synergie entre les recherches dans le domaine informatique qui sont créatrices de nouveaux problèmes pour les sciences humaines et les recherches en sciences humaines qui éclairent les chercheurs en informatique sur les raisons de l'échec d'un collecticiel et les aiguillent sur de nouvelles pistes."

²⁴ mediaspace ~ espace médiatique; la forme la plus récente des systèmes de télécollaboration

Chapitre 2 : Services proposés pour l'aspect sonore des collecticiels synchrones distribués

1 Services offerts par les systèmes actuels : analyse d'un cas concret

La plupart des collecticiels synchrones distribués actuellement opérationnels sont des prototypes expérimentaux. La communication verbale est assurée par une audioconférence faiblement coordonnée au reste du système. Techniquement, elle est soit réalisée par le système lui-même, soit par une audioconférence du réseau téléphonique. La première solution offre des possibilités de contrôle plus importantes, mais guère exploitées, avec une qualité sonore très médiocre, voire insuffisante pour un travail réel. La deuxième solution offre une qualité sonore permettant une communication fonctionnelle, mais sans presque aucun contrôle possible par le système. Comme il était primordial d'avoir une communication effective, c'est cette deuxième solution qui a été choisie pour CoLearn, un système conçu essentiellement pour l'enseignement à distance avec un souci de présenter une homogénéité entre les modes synchrone et asynchrone [DERYCKE 92, 93b; VIEVILLE 95; CROISY 95; HOOGSTOEL 95]. Une première expérimentation de CoLearn nous a permis de mieux appréhender les possibilités et les limites de telles audioconférences.

1.1 Description de la pré-expérimentation de CoLearn

Il est important de préciser que cette expérimentation ne visait pas seulement à vérifier que le système CoLearn était opérationnel d'un point de vue technique. Il s'agissait plutôt d'avoir une première idée de l'utilisabilité et de l'utilité du système. Le système CoLearn est un collecticiel conçu essentiellement pour l'enseignement à distance. L'expérience réunissait donc un enseignant et des élèves. L'enseignement portait sur la résolution graphique d'un système de contraintes. La fenêtre d'espace public fut le principal outil testé de CoLearn. Cependant, d'autres outils ont aussi été utilisés. Cette expérience n'a pas donné lieu à un rapport précis et complet. Seuls des enregistrements audio et vidéo ont été conservés. Nous détaillons ici le compte rendu de l'aspect sonore que nous en avons fait.

1.1.1 Configuration matérielle et géographique

Pour pouvoir apprécier l'aspect sonore, nous avons relevé un minimum de détails permettant de présenter le contexte. La configuration du système mis en oeuvre comportait quatre stations de travail (des PC sous Windows 3.1) outre les autres éléments informatiques nécessaires à son fonctionnement (serveur, routeur...). L'accès au réseau téléphonique depuis les différents postes était réalisé soit par un téléphone (mains libres), soit par une carte RNIS utilisée en partie pour une connexion au réseau informatique et en partie pour une connexion au réseau téléphonique. Cette liaison est réalisée par l'intermédiaire d'une carte sonore elle-même utilisée avec un microphone et des petits haut-parleurs amplifiés.

Une de ces stations, installée sur le campus de l'Université à Villeneuve d'Ascq, a été utilisée en tant que poste tuteur, les trois autres comme postes élèves. Le poste tuteur a été utilisé par l'enseignant aidé par un informaticien pour les problèmes purement techniques. Les trois postes élèves étaient installés respectivement l'un à Dunkerque, l'autre à Lens et le troisième à Villeneuve d'Ascq, mais dans un lieu distinct du poste tuteur. Les postes élèves ont été utilisés par un nombre variable d'élèves, de un à trois, tous aidés par un informaticien (sauf celui qui se trouvait seul, étant lui même informaticien). Les informaticiens

et un psychologue chargé de l'évaluation ne sont pas intervenus, ou très peu, sur le canal sonore. L'évaluation de l'aspect sonore a été faite depuis Paris, en utilisant uniquement un téléphone dans le but :

- d'évaluer le potentiel d'un poste CoLearn n'ayant accès qu'au canal sonore,
- de tester la communication sur une plus grande distance (inter-départements),
- de se concentrer sur l'aspect sonore.

Le tableau 4 résume l'organisation matérielle et la configuration géographique de l'expérimentation.

Lieu	Type de poste	Type de connexion	Utilisateur(s) et rôles
Villeneuve d'Ascq	poste tuteur	téléphone	1 enseignant 1 informaticien
Villeneuve d'Ascq	poste stagiaire	téléphone	2 élèves 1 informaticien
Dunkerque	poste stagiaire	carte RNIS et micro casque	1 stagiaire informaticien
Lens	poste stagiaire	carte RNIS, microphone et haut-parleurs	3 élèves 1 informaticien 1 psychologue
Paris	téléphone	téléphone	évaluation de l'aspect sonore

tableau 4 : configuration matérielle et géographique de la pré-expérimentation de CoLearn

1.1.2 Fonctionnement de l'audioconférence

La partie sonore de cette expérimentation a été réalisée par une audioconférence prise en charge par France Télécom. Elle a été utilisée pendant 3h30 depuis l'organisation initiale de l'expérimentation jusqu'à son bilan à chaud.

Son utilisation est très simple. La connexion à cette audioconférence se fait en appelant un numéro classique (préfixé par le 16 depuis Paris). Un message d'accueil prévient celui qui se connecte qu'il entre en audio-réunion et que la fin de

celle-ci sera indiquée par un bip sonore 10' et 5' avant la fin. Pour ceux qui sont déjà connectés à l'audio-réunion, un bip sonore leur indique qu'une nouvelle connexion est établie. Toutes les personnes connectées se retrouvent dans un unique espace sonore "brut" : tout son capté par l'un des microphones de n'importe quel participant est restitué tel quel vers tous les autres participants.

1.2 Observations effectuées en première analyse

Au cours de cette expérimentation, qui s'est déroulée en juin 93, nous avons fait un certain nombre d'observations que nous rapportons ici telles qu'elles nous sont apparues alors, en l'absence du recul que nous avons maintenant. Nous les présentons selon l'organisation suivante :

- des critiques quant au fonctionnement technique du système,
- des remarques concernant le problème des connexions sonores dans une audioconférence,
- un premier bilan concernant le potentiel de communication offert par ce système.

1.2.1 Critiques de l'aspect technique du système

La première catégorie de critiques concerne les problèmes de connexions :

- Rien n'indique à la personne qui entre dans l'audio réunion le début de sa connexion.
- Le bip sonore qui indique la présence d'une connexion supplémentaire aux personnes déjà connectées est insuffisant : on ignore qui s'est connecté et d'où.
- Rien n'indique la déconnexion d'un participant.
- La connexion depuis Paris a été systématiquement coupée par France Télécom très régulièrement toutes les demi-heures. Le service organisant les audio réunions n'a pas pu donner d'autre explication que la confirmation que ce n'est pas normal ...

Outre le problème anecdotique de la rupture périodique de la liaison depuis Paris, qui soulève cependant la question de la gestion des défaillances possibles du réseau, il s'agit essentiellement d'un problème d'absence de rétroaction et de notification des événements de connexion et déconnexion. Cette notification n'est pas réalisée par le canal audio, sans doute par souci de discrétion. Cependant, elle n'est pas non plus réalisable par le système informatique du fait de l'absence de couplage avec l'audioconférence.

La deuxième catégorie de critiques concerne la qualité du son :

- La retransmission brute du son émis engendre des effets Larsen, une ambiance "cathédrale" ou des échos.
- Si la qualité sonore s'avère suffisante pour reconnaître les interlocuteurs aux caractéristiques propres de leurs voix, il est cependant difficile de distinguer certains mots ou noms. Il y a eu en l'occurrence une confusion entre Laurence et Laurent.

Enfin, la dernière catégorie de critiques concerne un problème plus subtil qui découle de l'absence du canal visuel et des informations paraverbales qu'il permet :

- Quand quelqu'un est seul dans l'audioconférence ou que personne ne parle, le silence est "angoissant" ; il est impossible de savoir s'il y a un problème dans le canal sonore ou si ce silence est normal.
- De même, si l'on n'entend pas un des participants, il n'est pas possible de savoir s'il ne dit rien ou si sa connexion est coupée.

1.2.2 Les canaux sonores d'une audioconférence : une chaîne complexe

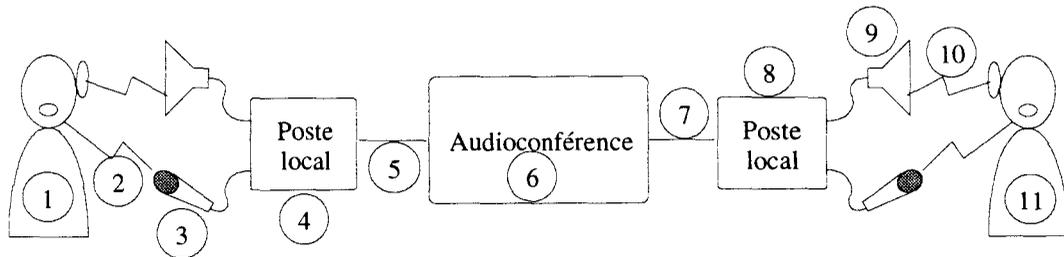


figure 11 : les maillons de la chaîne sonore d'une audioconférence

Souvent, au cours de la réunion, un des participants s'est plaint de ne pas entendre un autre suffisamment bien. Or, ce dernier était parfaitement audible par tous les autres participants. De même, la qualité d'audition et de perception des différents participants était très variable. En fait, il y a de multiples sources de problèmes dont certains sont purement techniques, voire très élémentaires, mais n'en restent pas moins rédhibitoires.

On ne compte pas moins de 11 maillons dans la chaîne du canal sonore quand une personne parle à une autre (figure 11). Chacun de ces maillons pouvant être la cause du dysfonctionnement de toute la chaîne, il y a :

- la personne qui parle (1); elle doit le faire distinctement, de manière à être entendue.
- sa position relative au micro (2); un microphone sélectif est plus que souhaitable pour éviter de transmettre les bruits ambiants, mais il doit alors être précisément positionné par rapport à l'utilisateur.
- le micro lui même (3); il doit être correctement connecté, mis sur "On" ...
- le poste de réception/*transmission* local (4); il doit être correctement configuré (réglage de la sensibilité de la prise de son). Ceci est d'autant plus important que cette configuration peut être complexe à mettre en oeuvre.
- la connexion du poste "émetteur" à l'audioconférence (5); un connecteur mal enfoncé n'est pas évident à repérer, bien qu'il soit une cause potentielle de panne.

- le système d'audioconférence lui même (6); cf. le problème des déconnexions systématiques toutes les demi-heures.
- la connexion du poste "récepteur" à l'audioconférence (7).
- le poste de *réception/transmission* local (8).
- le système d'amplification du son (9); le haut parleur doit être branché, l'ampli aussi.
- la position relative du haut-parleur à la personne qui écoute (10). Et réciproquement ...
- la personne qui écoute (11); elle doit être dans de bonnes dispositions d'écoute.

Cette décomposition de la chaîne du canal sonore peut servir de base au test du bon fonctionnement de celui-ci. Tous les moyens objectifs de contrôle seraient les bienvenus (par exemple des mesures physiques accessibles aux utilisateurs, vumètre, boucle de contrôle etc. ...).

1.2.3 Apports et limites du canal sonore

Les apports sont :

- L'expérimentation concernait plusieurs sujets (système pédagogique, système informatique, audioconférence ...). Le canal sonore a été massivement utilisé pour soutenir les autres canaux. Nous avons relevé par exemple :
 - le bon fonctionnement du système informatique,
 - les directives pédagogiques,
 - le contrôle de l'avancement des travaux dans l'espace privé.
- A un moment, le système informatique est tombé en panne, sa remise en route a été organisée par le canal sonore.
- Le canal sonore est le principal média de communication; un simple téléphone permettait de suivre et de participer pleinement à l'expérimentation.

Les limites sont :

- Le canal sonore ne se suffit pas à lui-même. Quand tout ce que l'on peut faire est "parler" alors que des documents sont nécessaires à la discussion, on n'a pas grand chose à dire.
- En absence de support visuel, on ressent vivement le besoin de se confectionner une image mentale de la situation des autres participants : un schéma indiquant *qui est où* est bien utile.
- Il est difficile de savoir qui parle. Une rétroaction visuelle sur les postes informatisés (sur le schéma indiquant *qui est où* par exemple) serait la bienvenue.
- Même avec seulement 5 participants, une régulation des prises de parole est nécessaire. Un "arbitre" d'audio-réunion s'impose. Automatique, semi-automatique ou "manuel" ?
- Suivant la qualité de la prise de son locale, les bruits ou les conversations privées peuvent envahir l'audio-réunion même si les participants peuvent veiller à tenir leur conversation locale loin des microphones.
- Si plusieurs conversations interviennent simultanément, elles s'embrouillent. Il serait souhaitable de pouvoir dégager des participants dans des sous audio-réunions temporaires, avec un mécanisme de rétroaction adapté ...
- L'absence de rétroaction visuelle se fait ressentir quand une question est posée : est-ce que personne ne répond parce que personne ne connaît la réponse ou parce que personne n'ose prendre la parole ?
- Dans le cadre de cette expérimentation, il y avait d'autres moyens de communication. Quand on n'a que le téléphone (situation prévue dans CoLearn pour un utilisateur "de classe E"), il n'est pas possible de savoir ce qui se dit à travers le "chatter" (outil d'envoi de message écrits) ... ni de voir la désignation d'un point particulier du document partagé dans la fenêtre publique.

1.3 Améliorations suggérées à partir de cette première expérience

Dès lors que l'on ne voit pas les personnes avec qui l'on parle, il n'est plus possible d'interpréter les silences. Même si l'on avait une expérience très importante des systèmes d'audioconférence, on ne pourrait s'empêcher de se demander si un silence prolongé ne trahit pas une défaillance du système. En effet, cela fait des décennies que la radio existe, les postes de réception sont d'une fiabilité quasi totale et tout le monde y est habitué. Cependant les émissions radio se caractérisent toujours par une absence totale de silence. De même, les standards téléphoniques vous envoient une musique pour vous faire patienter quand ils recherchent votre correspondant.

Dans le cas d'une audio réunion, il y a deux formes de silence qui sont "inquiétants" : le silence total (dans le cas où personne ne parle) et le silence local (un des participants n'émet aucun son pendant une longue durée).

Dans le premier cas, un fond musical sonore est très appréciable. Bien sûr, quand les échanges à travers l'audio-réunion sont importants, il devient rapidement nuisible. Il serait peut-être alors judicieux d'avoir un fond sonore adapté automatiquement. Par exemple, dans le cas d'un silence dépassant une durée fixée, un fond sonore pourrait alors apparaître progressivement, et de même s'atténuer totalement quand les conversations reprennent.

Dans le deuxième cas, il serait peut-être envisageable de remplacer un silence prolongé d'un poste par un fond sonore propre (dépendant donc du poste et différent pour chaque poste). Un tel service peut être envisagé au niveau du serveur de conférences ou localement à chaque poste. Dans le premier cas, il sera plus simple à implémenter (un même mécanisme appliqué à toutes les connexions), dans le deuxième cas, cela permettrait en plus de vérifier le bon fonctionnement des maillons 4 et 5 de la chaîne du canal sonore. Ce mécanisme de remplacement des silences locaux par un fond sonore propre rend de plus inutile le mécanisme de remplacement du silence total, puisqu'il en assure alors le service.

Les fonds sonores de remplacement ne doivent pas forcément se limiter à de la musique; le fond sonore du tuteur pourrait être un bruit de craie sur le tableau, ceux des élèves un bruit de crayon, de cahier, de raclement de siège, de bavardage etc.

Il y a plusieurs types de sorties auxquelles il faudrait associer des bruitages différents :

- involontaire : c'est le système qui "casse",
- volontaire mais discrète : le participant désire se retirer sans gêner les autres,
- normale : le participant qui se retire veut en avertir les autres,
- ... (ex. une sortie colérique ...)

Ces problèmes renvoient aux travaux sur la notification, rétroaction et sonification cités au paragraphe 2.1 *Notification, rétroaction et sonification : un premier niveau d'exploitation du son*, page 40.

C'est donc cette première expérience qui nous a guidé pour la réflexion qui suit, bien que nous l'ayons abordée à un niveau plus général.

2 Améliorations proposées de l'aspect sonore des collecticiels synchrones distribués

Le travail coopératif assisté par ordinateur est prévu pour mettre la puissance de l'informatique au service du travail d'un ensemble de personnes. Une des applications phares est de le faire à distance; cela représente en effet un fort potentiel d'application : téléconférence, télé-réunion, télé-assistance, télétravail, télé-enseignement ...

Cependant ce potentiel est compromis par des inconvénients engendrés par la distance. Les principaux facteurs négatifs se cristallisent autour de :

- la différence de contexte,

- les déficiences du canal visuel, voire son absence,
- les déficiences du canal sonore,
- les performances et l'utilisabilité des outils de partage de documents.

En effet, le fait d'être à distance implique un contexte différent pour chacun, ce qui représente une surcharge cognitive liée à la représentation des environnements des autres participants.

Cette difficulté supplémentaire est accrue par les limitations visuelles qui rendent encore plus difficile cette représentation. Aux deux difficultés précédentes s'ajoute celle d'une amputation du potentiel de communication orale par la piètre qualité du canal sonore.

D'un point de vue sonore, le problème se résume à deux questions : "*Qui parle ?*" et "*Qui entend quoi ?*". En effet, pour comprendre ce qui se dit, il faut savoir de quoi on parle. Pour savoir de quoi on parle, il faut connaître le contexte. Et comme ici les contextes sont différents, il faut donc savoir qui parle pour bien saisir le contexte (à supposer que l'on ait pu se faire une représentation suffisante du contexte de chaque participant). Réciproquement, pour bien se faire comprendre, il faut pouvoir s'adapter au contexte de ceux qui écoutent et savoir qui entend quoi.

Quand tous les participants sont réunis au même endroit, la question du contexte ne se pose pas. La vision fournit une grande quantité d'informations. L'audition vient en plus très efficacement compléter la vision. En effet, nous ne nous contentons pas de décoder les messages verbaux qui nous sont adressés, nous identifions aussi la personne qui parle, par le timbre de sa voix bien sûr, mais aussi par notre perception spatiale des sons [CHERRY 53; ARONS 92a].

Il n'est pas étonnant de constater qu'un système d'audioconférence fourni par le réseau téléphonique, en mono et de mauvaise qualité, est loin d'être adapté aux ambitions du travail coopératif [HINDUS 92]. Déjà, au téléphone, certains sons se distinguent difficilement (le "s" et le "f" par exemple), certains mots peuvent se

confondre. Le timbre des voix est difficile à identifier, surtout sur des interventions courtes (par exemple "oui", "non", "humm"). Quant à notre capacité à localiser un son, elle nous permet de savoir où se trouve le combiné téléphonique... Pour étayer cette impression, il suffit de se référer à une expérience que l'on peut faire quasi quotidiennement : écouter une discussion faisant intervenir plusieurs personnes à la radio (bande passante 5 fois supérieure à celle du téléphone, en FM) sur un poste monophonique. Si plusieurs participants parlent en même temps, on n'y comprend plus rien; et pourtant ils se comprennent, eux, qui se trouvent au même endroit, puisqu'ils continuent de discuter sans paraître incommodés, à l'exception du responsable de l'émission qui les prie de parler chacun à leur tour.

Pour remédier à ce problème, il y a deux approches complémentaires, qui relèvent de celles de la réalité virtuelle :

- essayer de reproduire au mieux les conditions sonores du présentiel, c'est-à-dire au moyen d'un son de grande qualité et d'une reconstitution des informations sonores nous permettant de localiser les sons,
- tirer parti de ce que le son est transmis sous contrôle du système informatique pour fournir des services qui ne sont pas possibles dans la réalité afin de compenser, au moins, les restrictions qui en découlent.

Il ne faut pas se focaliser sur l'aspect "audioconférence" uniquement. Nous utilisons notre système auditif en collaboration avec nos autres sens. La composante "audioconférence" doit donc fonctionner harmonieusement avec les autres composantes du collecticiel. C'est en fait une nécessité. Par exemple, si l'on veut établir un contact sonore avec quelqu'un d'autre, en particulier lors d'une audioconférence (c'est le mécanisme d'aparté que nous détaillerons plus loin), on ne peut pas le faire par le canal sonore lui même, il faut pouvoir le négocier à travers les autres médias de communication.

Il est cependant souhaitable que l'audioconférence soit traitée à part pour qu'elle soit exploitable à tous les niveaux et qu'elle ne soit pas rigidement liée au reste du

système dont les composants évolueront sans doute indépendamment. Dans le même ordre d'idées, à un niveau interne cette fois, les services de l'audioconférence doivent, eux aussi, être organisés de manière à pouvoir évoluer.

Notre démarche s'organise en trois parties :

- la restitution du son avec une synthèse spatiale de la scène sonore,
- le contrôle de la diffusion du son,
- l'ajout d'informations au niveau de la prise de son,

2.1 Synthèse sonore : spatialisation du son

La technique que l'on se propose d'employer, évoquée dans l'état de l'art, est une version très simplifiée de la synthèse des paramètres intervenant dans notre perception spatiale des sons. Peu importe la technique pourvu qu'elle convienne aux services envisagés.

La spatialisation du son est prévue pour aider à répondre à la question "*Qui parle ?*" en référence à une position dans l'espace. Notre capacité à localiser l'origine d'un son est plus fine dans une pièce que dans un champ libre : perception des échos, réverbération, à condition de ne pas être saturé par ces informations complémentaires. Le contexte de base est donc une pièce, ce qui correspond bien au contexte habituel de travail. Ce choix est supposé suivi au niveau de l'interface visuelle. Les deux devant être en correspondance pour maintenir une cohérence nécessaire à la bonne utilisation du collecticiel. On suppose donc une métaphore d'interface de type *ROOM* maintenant classique [DERYCKE 93a; CARD 91; HAMMAINEN 91].

L'usage présumé du système est la discussion, la coopération, ou l'apprentissage. Pour la discussion, le mode de fonctionnement est le face à face; pour la coopération, c'est le côte à côte et pour l'apprentissage l'un des deux précédents suivant que l'on est dans une situation de transfert du savoir ou d'aide à l'appropriation du savoir. Il n'y a pas lieu, au moins dans un premier temps, de

prévoir des échanges verbaux avec des personnes derrière, au dessus ou en dessous de soi.

Un des principes acquis d'une bonne interface suppose qu'elle soit "centrée utilisateur". On considère donc chacun des participants à une conférence indépendamment et on se propose de lui restituer chacun des sons qu'il est susceptible d'entendre selon deux paramètres, en coordonnées polaires : la distance et le gisement (figure 12). S'il y a n "participants" (cf. plus loin la définition de "participants"), il y aura $n-1$ positions spatiales à rendre pour chacun des participants, soit $n(n-1)$ sons à spatialiser au total.

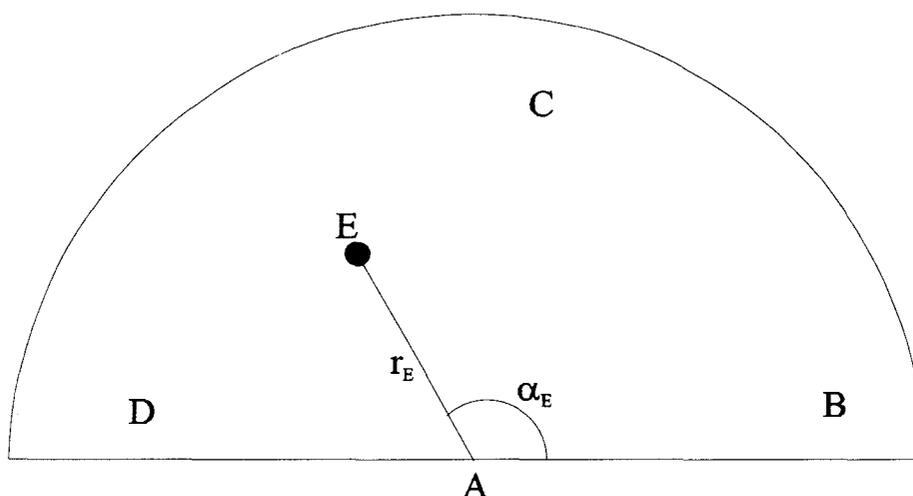


figure 12 : spatialisation des sons pour un participant

Un pas décisif est franchi de l'une à l'autre des deux approches évoquées plus haut (reproduire au mieux la réalité, exploiter le potentiel de la virtualité). D'une part, on renonce à une spatialisation complète en se limitant à une aire dans un plan horizontal en face de chaque participant. D'autre part, on s'affranchit de la contrainte du réel qui veut que les positions soient relatives. La figure 13 illustre ce principe. Dans la réalité, si A entend B à sa droite, alors B entendra A à sa gauche. Alors que l'on peut faire en sorte que A entende B à sa droite et B entende A à sa droite aussi. Sur cet exemple, il est raisonnable de se demander si ce n'est pas une nuisance plutôt qu'un bienfait. Cependant cette approche n'empêche en rien de faire correspondre la virtualité à la réalité et, par contre, elle ouvre la porte

à de nouveaux services que l'on devra bien sûr évaluer, mais il n'y a pas de raison de les rejeter d'emblée et arbitrairement.

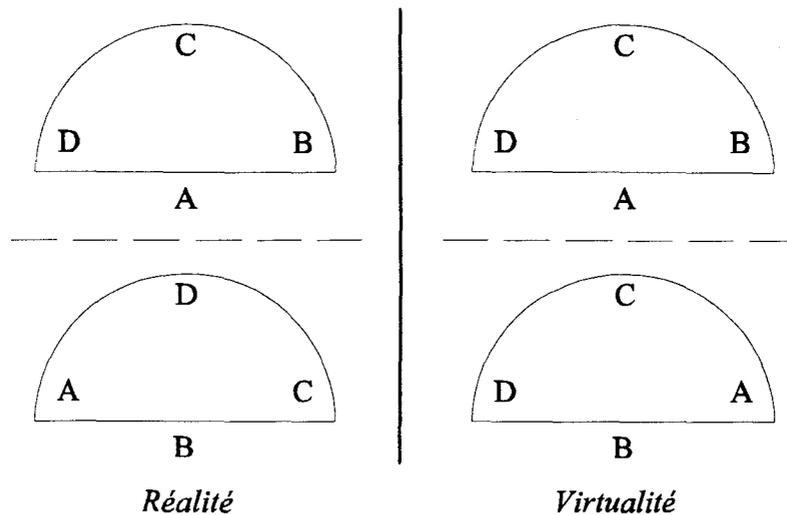


figure 13 : configuration réelle ou virtuelle

2.2 Gestion et contrôle de la diffusion du son

Comme pour la spatialisation du son, on propose un traitement par utilisateur pour la diffusion du son. C'est-à-dire que l'on pourra définir une perception sonore particulière à chaque participant. C'est un choix techniquement imposé et stratégiquement choisi.

Il est techniquement nécessaire de faire un traitement spécifique. En effet, il faut éviter de restituer le son émis par un participant là où on le capte avec un microphone. Sinon, on risque un accrochage entre le microphone et les haut-parleurs et un effet Larsen immédiat. Cette restitution est aussi inutile : la personne s'entend déjà parler toute seule. Par contre, avec un casque, il est plus confortable pour l'utilisateur d'avoir un léger retour dans ses oreilles du son qu'il émet, surtout si le casque isole bien des bruits extérieurs (et l'extérieur, des sons qu'il restitue). La plupart des systèmes de conférences appliquent la règle dite du n-1 qui consiste à restituer à un endroit donné, tous les sons sauf un, celui qui en provient. Cette règle doit donc être assouplie si l'on utilise des casques.

Il est stratégiquement intéressant de faire un traitement spécifique pour chaque participant. Par exemple, dans une réunion, les participants parlent plus ou moins fort, sont plus ou moins loin. Si l'utilisateur a un degré de contrôle à travers le système d'audioconférence sur le volume des sons qu'il reçoit, il pourra à loisir reconfigurer la scène sonore à son gré. Comme pour la spatialisation, cette configuration n'a pas obligatoirement à respecter une cohérence par rapport à la réalité.

2.2.1 Extension de la notion de participant

Cette rupture de cohérence avec le monde réel ne se limite pas à la restitution des sons. On propose aussi, comme Rangan [RANGAN 91], une extension de la notion de participant, mais pas dans la même optique.

Rangan prévoit une vision récursive d'un participant : une conférence rassemblant plusieurs participants pourrait être vue comme un "super" participant. Nous avons aussi envisagé ce principe dans un premier temps. Arrivé au niveau de l'implémentation, il s'est avéré sans intérêt. D'abord parce qu'il faut aplatir les super conférences pour pouvoir calculer la perception du son pour chaque participant. Ensuite parce que ce procédé n'apporte rien au niveau de l'utilisateur. Contrairement à ce qu'argumente Rangan dans son article, nous préférons une vision plus centrée utilisateur où ce n'est pas une conférence qui, d'un bloc, devient un participant d'une conférence, mais chaque utilisateur qui entre dans une autre conférence. Cette dernière solution paraît aussi plus facile à gérer au niveau de l'utilisabilité du système. En effet, si une conférence est vue comme un super participant, comment résoudre sa manipulation ? Qui décide : un membre de cette conférence, un vote de tous les participants ?

Notre extension de la notion de participant est plus pragmatique. Dans une conférence classique, un participant est une personne physique; dans une ACAO, on aura par exemple :

- une personne isolée disposant d'un système de prise de son et de restitution du son (c'est la notion classique d'un participant),

- tout dispositif générant du son (document sonore retransmis ou enregistré par exemple),
- tout dispositif enregistrant du son (magnétophone ou magnétoscope par exemple).
- un groupe de personnes partageant un dispositif de restitution du son. Si chaque personne dispose d'un microphone individuel et s'il est possible de repérer automatiquement et en temps réel qui prend la parole, on prévoit de les assimiler à des personnes isolées, recevant cependant le même flux sonore.

2.2.2 Structure de conférence dynamique

Le deuxième volet des services rendus possibles par le contrôle de la diffusion du son requiert l'apport de l'informatique. On propose d'offrir la possibilité de modifier en temps réel l'organisation des conférences et de permettre à un participant "d'assister" à plusieurs conférences simultanément.

2.2.2.1 Réorganisation dynamique de l'ensemble des conférences

La structure d'une conférence classique est statique : elle est définie par le groupe des personnes présentes dans le lieu de la conférence. Modifier cette structure impose un déplacement physique des personnes dans d'autres lieux, ce qui ne peut se faire ni instantanément ni sans perturbations importantes. Dans une ACAO, on veut avoir par exemple, sans délai ni perturbation, la possibilité de réorganiser à loisir la ou les conférences en cours, c'est-à-dire de :

- scinder une conférence en sous-conférences,
- fusionner plusieurs sous-conférences.

Ainsi, si au cours d'une réunion, une discussion intervient pour un sous ensemble des participants, ils pourront "s'isoler" dans une nouvelle conférence et rejoindre la précédente au terme de leur discussion. Dans une conférence classique, ce mécanisme est rendu possible, dans une certaine mesure, par notre faculté à nous concentrer sur une discussion parmi plusieurs (grâce à notre

perception binaurale). Et nous continuons à percevoir en arrière-plan l'évolution des autres discussions, en particulier le moment où elles sont terminées. Dans une ACAO, si ceux qui ne sont pas partis en sous-réunion veulent avoir une idée de l'évolution de la discussion, ils pourront y "assister", en plus. C'est là qu'intervient le don d'ubiquité que nous proposons dans une ACAO.

2.2.2.2 Don d'ubiquité des participants

Nous proposons donc qu'un même participant puisse assister à plusieurs conférences simultanément. Selon toute vraisemblance, cela devrait se limiter naturellement à 2, voire 3 conférences tout au plus et être transitoire. Cela nous paraît utile dans les cas suivants :

- pour permettre l'équivalent des "super-conférences" de Rangan,
- pour suivre, éventuellement à un niveau sonore moindre, une sous-conférence,
- pour permettre des apartés entre deux participants.

Ainsi, alors que Rangan propose d'inclure globalement une conférence comme participant d'une autre, nous proposons, pour obtenir le même effet, que les participants d'une conférence puissent aller simultanément dans une autre. Ce don d'ubiquité peut aussi être exploité, par exemple, pour permettre aux participants d'une conférence de discuter d'une autre conférence; cet exemple sera détaillé plus loin au paragraphe 1.1.3 *Exemples* du chapitre 3 page 95. Enfin c'est par le don d'ubiquité que nous proposons de réaliser les apartés.

Dans la réalité, on peut parfois se permettre un aparté avec ses voisins immédiats à droite ou à gauche. Dans une ACAO, on pourra le faire avec n'importe qui, qu'il soit un participant de la conférence ou non, et sans gêner personne. Un aparté sera donc réalisé par la création d'une nouvelle conférence où prendront place les personnes désirant avoir une discussion parallèle. Ils seront donc alors dans (au moins) deux ACAO simultanément, avec toute latitude pour définir les paramètres de leur audition indépendamment dans l'une ou l'autre des conférences.

2.2.2.3 Autres services

La liste de ces nouveaux services n'est pas exhaustive. En les combinant, on en obtient facilement d'autres. Par exemple, le président de session pourrait aller virtuellement accueillir un intervenant important. Le scénario est le suivant. Une sous-conférence est utilisée pour établir le contact et éventuellement régler les paramètres de communication. Le président de session n'est plus entendu dans la conférence principale et peut librement parler avec l'intervenant. Il continue cependant à entendre en arrière-plan la conférence principale. Au moment opportun, la sous-conférence est fusionnée avec la conférence principale et l'orateur peut s'exprimer. Dans le même ordre d'idée, le président de session peut utiliser un aparté pour signaler à l'intervenant qu'il ne lui reste plus que le temps de conclure.

Le contrôle de la diffusion du son au sein des participants et le traitement par des filtres permettent d'autres effets intéressants :

- la gestion de la diffusion du son d'une conférence à une autre,
- la gestion du droit à la parole (en collaboration avec le reste du système),
- l'ajout d'effets de présence et de positionnement.

2.2.3 Synthèse des propositions concernant la gestion et le contrôle du son

Dans un souci de généralité plutôt que d'exhaustivité, nous essayons de proposer des mécanismes simples dont la combinaison offre un grand potentiel de services. Le premier principe choisi est de rester centré sur l'utilisateur. Ensuite, nous ne chercherons pas à rester fidèle à la réalité : il pourra y avoir une incohérence globale possible des perceptions sonores par rapport à la réalité, sous réserve qu'elle soit utile aux participants. Enfin, nous permettrons à un même participant d'assister à plusieurs conférences simultanément, ces conférences pouvant être transitoires. A partir de là, nous proposons d'offrir à chaque utilisateur la possibilité de jouer sur les paramètres suivants :

- localisation spatiale indépendante pour chaque source sonore qu'il entend,

- réglage du volume sonore indépendant pour chaque source sonore qu'il entend, et ce pour chaque conférence à laquelle il participe. Si une même source sonore est écoutée dans plusieurs conférences, le niveau sonore rendu sera le niveau sonore de la conférence où il est maximal, tant que celle-ci durera.
- réglage du niveau sonore avec lequel il pourrait être entendu, de manière indépendante dans chaque conférence à laquelle il assiste.

2.3 Données associées aux flux sonores

Il y a d'une part un degré de liberté inhérent aux systèmes d'audioconférence qui n'est pas rendu par les paramètres évoqués ci-dessus et d'autre part un potentiel qui reste inexploité. Pour le premier, il faut prendre en compte que, si le son transite sous le contrôle du système informatique, il passe par des ressources *physiques*. Elles seront bien identifiées, contrairement aux *personnes* qui les utiliseront. Pour le second, il s'agit de faire un premier pas vers l'exploitation de l'analyse du son dont le "saint graal" (sic) [HINDUS 92] est la reconnaissance de la parole pour la rédaction automatique d'un compte rendu. En suivant l'exemple des travaux de Degen [DEGEN 92], nous proposons donc le concept des données associées aux flux sonores, classées en deux parties, les informations statiques et les informations dynamiques.

2.3.1 Informations "statiques"

La première information statique associée à une ressource physique concerne sa description (matériel, localisation géographique, potentiel sonore). C'est à ce niveau que l'on stockera, par exemple :

- les informations de spatialisation du son,
- le mode de restitution, à l'aide d'un casque ou d'un haut parleur, dont dépend la restitution ou non du son capté sur place.

Dans la suite, nous appellerons canal cette ressource physique. Ce canal sera identifié par un *nom physique*.

Ensuite, il convient d'identifier l'utilisation qui est faite de ce canal (le nom de l'utilisateur par exemple). Cette identification sera le *nom logique* du canal.

Il est possible d'imaginer bien d'autres informations qu'il peut être utile d'associer à un canal : les horaires auquel il est utilisable, des droits d'accès, etc.. Ces informations statiques constituent clairement un point d'accueil pour l'extension des services.

2.3.2 Informations dynamiques

Ici, on s'intéresse au contenu du flux sonore transitant par un canal. Le premier niveau que l'on se propose d'exploiter a été évoqué plus haut lors de l'extension de la notion de participant. Il s'agit de la détection de prise de parole. En effet, si les interventions de plusieurs personnes sont regroupées sur un seul canal, si on suppose que la plupart du temps elles ne parleront pas ensemble et si on peut identifier qui parle à un moment donné, il est alors possible, par la puissance du traitement informatique, de simuler plusieurs canaux qui partagent la même restitution sonore.

Il est possible d'aller plus loin dans l'analyse du flux sonore d'un canal. Comme on l'a vu dans le chapitre 1, Schmandt a exploité la prosodie du discours, Arons les silences pour semi-structurer la conversation. Ce niveau d'analyse n'est pas abordé dans cette thèse. Nous nous contenterons de la détection de la prise de parole, appelée par la suite Détection Automatique des Locuteurs (DAL).

La DAL permet de faire bien plus que la séparation d'un canal, elle est exploitable :

- en différé,
- à un niveau instantané,
- d'une manière synthétique.

Ce type d'information a déjà été exploité en différé [SELLEN 92]. Il est à la base d'une analyse des qualités de différents systèmes *visuels* de téléconférence.

Nous pensons qu'il faudrait étudier à part l'aspect sonore et que les résultats de Sellen sont faussés car, dans l'un des cas (l'utilisation du système *hydra*), il y a une spatialisation du son alors qu'avec le PIP (Picture in Picture), il n'y en a pas. Or les travaux de Cherry [CHERRY 53] ont déjà mis en évidence l'effet de la spatialisation du son sur la compréhension du son mis en avant par Sellen. Nous pensons que la légère différence entre le système PIP et le système *hydra* s'est jouée au niveau du son, et c'est ce que nous avons vérifié avec une expérience détaillée plus loin.

Cependant l'information fournie par la détection automatique de prise de parole peut être facilement exploitée au niveau de la rétroaction visuelle, pour signaler à un participant que le son qu'il émet est bien capté. Il peut être exploité au niveau de la notification, pour les autres utilisateurs, afin de leur signaler que tel participant est en train de parler.

En cumulant et en interprétant les informations de la DAL pour chaque participant, on obtient très facilement les temps de parole de chaque personne et le nombre de leurs interventions. En utilisant ces données pour l'ensemble des participants, on a aussi le nombre et les temps de recouvrement des prises de parole qui sont un bon indice de l'animation de la discussion. Ce sont justement ces données que Sellen a utilisées pour son évaluation.

Nous voulons fournir ces informations en temps réel aux utilisateurs et mesurer (et éventuellement valider) les effets par des évaluations. Nous pensons que si les utilisateurs ont, en temps réel, une information objective et impartiale sur les durées de prise de parole respectives de chaque participant, cela pourra peut-être les influencer. Ceux qui parlent le moins seront peut être encouragés à prendre la parole, pour ne pas être en reste. Celui qui parle tout le temps devrait être encouragé à laisser davantage la parole aux autres. Nous n'avons pas d'idée préconçue sur les effets de l'information sur les temps de recouvrement de prise de parole, qui donnent une idée des tentatives d'interruption pour prendre la parole, et donc du caractère "animé" de la discussion.

2.4 Synthèse des services proposés

La spatialisation du son dans une audioconférence doit pouvoir aider les utilisateurs à répondre à la question "*qui parle?*" et améliorer ainsi la qualité de la communication orale. Les données associées à la prise de son, la DAL en particulier, permettent de renforcer cette aide et de réaliser la spatialisation du son, même quand un seul canal regroupe les paroles de plusieurs personnes, ce qui est utile pour interconnecter différents systèmes. L'exploitation de plusieurs systèmes d'audioconférence doit se faire avec un contrôle fin de la diffusion du son. Ce contrôle permet une réorganisation dynamique d'un ensemble d'audioconférences.

Nous avons cherché à fournir des mécanismes génériques ayant un fort potentiel d'exploitation plutôt que de réaliser une liste de services. Une intégration dans un collecticiel complet de tous ces services est cependant nécessaire pour en permettre l'évaluation. La modélisation décrite au chapitre suivant a été conçue en partie en vue de la réalisation d'un système expérimental mais surtout dans le but d'être modulaire et évolutive pour servir de base aux travaux futurs dans ce domaine.

3 Analyse des services proposés par rapport aux modèles conceptuels des collecticiels

Nous retenons pour cette analyse les trois principaux modèles conceptuels ou espaces de classification des collecticiels mentionnés par Salber et al [SALBER 95]. Dans l'introduction, nous avons utilisé la classification spatio-temporelle pour mieux situer notre domaine d'investigation. Nous utilisons le modèle du trèfle fonctionnel du collecticiel pour mieux situer notre contribution par rapport à un collecticiel synchrone distribué complet. Enfin, nous utilisons le modèle conceptuel d'Ellis et al [ELLIS 94] pour décrire cette contribution.

Cette analyse appelle deux remarques préliminaires : l'une quant à l'utilisation des modèles, l'autre quant à l'organisation des informations relevant de cette analyse.

De même que notre domaine d'investigation ne couvre pas l'ensemble des collecticiels décrits par la classification spatio-temporelle, notre contribution concerne un aspect particulier d'un collecticiel synchrone distribué : l'aspect sonore. Aussi, tous les éléments de chaque modèle ne sont pas pris en compte. La précision des éléments non concernés, tout comme celle des éléments concernés, permet de mieux décrire notre contribution.

L'objectif de cette analyse est de donner un cadre conceptuel au chapitre suivant qui décrit plus en détail les services réalisés. Idéalement, ces détails devraient apparaître dans cette analyse, mais alors ils auraient dû être décrits en amont de la précision de ce cadre conceptuel. Nous avons préféré le présenter tout d'abord en donnant les grandes lignes des services réalisés dont les détails pourront alors leur être facilement reliés.

3.1 Par rapport au trèfle fonctionnel du collecticiel

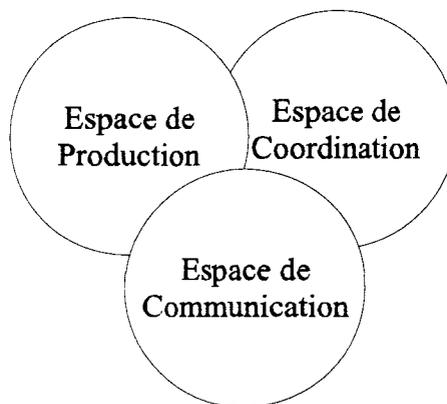


figure 14 : le trèfle fonctionnel du collecticiel

Le trèfle fonctionnel du collecticiel, décrit par Salber et al, décompose le noyau fonctionnel d'un collecticiel en trois parties :

- l'espace de production,
- l'espace de coordination,
- l'espace de communication.

De toute évidence, notre contribution est tout particulièrement centrée sur l'espace de communication. En nous efforçant d'améliorer le potentiel de communication verbale offert par le canal sonore, nous nous situons parfaitement dans la définition de l'espace de communication [SALBER 95] :

"L'espace de communication offre aux acteurs du collecticiel la possibilité d'échanger de l'information. Le contenu sémantique de cette information concerne les acteurs communicants. Il est étranger au système qui se contente de servir de messenger."

Cependant, comme le suggère la figure 14, les trois espaces ne sont pas disjoints. Autrement dit, s'intéresser à l'un d'eux exclusivement n'aurait pas d'intérêt. Notre contribution concerne donc aussi les deux autres espaces.

Elle concerne l'espace de coordination car, si nous souhaitons offrir un composant logiciel gérant complètement les communications verbales, il doit le faire en coordination avec le reste du système et donc être soumis au système de contrôle global du collecticiel. Nous écartons en fait de nos préoccupations les problèmes de coordination au niveau de la communication verbale en les reportant sur un autre composant du système. Cependant, nous mettons à la disposition de ce composant les outils nécessaires à la réalisation de cette coordination. Prenons par exemple le problème de la gestion du droit à la parole. Plusieurs stratégies sont possibles : présidentiel, anarchique, ... Plusieurs tactiques sont possibles : celui qui parle a automatiquement le contrôle du télépointeur, celui qui a le contrôle du télépointeur a automatiquement le droit à la parole... La réalisation de ces stratégies et tactiques n'est donc pas du domaine exclusif de la gestion de l'aspect sonore qui doit cependant les réaliser.

Notre contribution concerne aussi l'espace de production. Il est pris en compte par l'intermédiaire de l'extension de la notion de participant et des données associées aux flux sonores. L'extension de la notion de participant inclut tout dispositif générant ou enregistrant du son. Ainsi, un document sonore exploité dans l'espace de production est accessible aux utilisateurs dans l'espace de communication. De même, le contenu dynamique de l'espace de communication

peut être transféré dans l'espace de production sous la forme d'un enregistrement sonore. Ce dernier sera très utilement complété par les informations associées aux flux sonores qui pourront être mis en correspondance de manière précise avec les éléments de l'enregistrement concerné. Un premier pas peut ainsi être franchi vers la production d'un compte-rendu automatisé de discussion avec un enregistrement enrichi de références au contexte de cette discussion, tant d'un point de vue statique (noms et situations des participants par exemple) que dynamique (DAL).

3.2 Par rapport au modèle conceptuel d'Ellis

Ellis et al [ELLIS 94] proposent une décomposition en trois parties :

- un modèle ontologique,
- un modèle de coordination,
- un modèle d'interface utilisateur.

Le modèle ontologique tel qu'il est défini par Ellis est une description statique des objets manipulables par les utilisateurs. Ellis précise cette définition en ramenant cette description à celle des classes de ces objets et notamment celle de leurs méthodes publiques que le système fournit aux utilisateurs.

Le modèle de coordination coïncide avec l'espace de coordination du trèfle du collecticiel avec un point de vue davantage centré sur les utilisateurs. Il décrit l'organisation des activités des utilisateurs et non l'organisation des processus exécutés par le système. Ellis précise qu'il s'agit de décrire l'activité de chaque participant et comment ces activités sont coordonnées pour le travail du groupe.

Le modèle d'interface utilisateur ne fait pas l'objet d'une définition aussi concise que les précédentes mais se décompose lui-même en trois composants :

- les représentations des objets d'informations,
- les représentations des participants,
- les représentations du contexte.

La description de notre contribution s'étend sur les trois parties du modèle d'Ellis mais plus particulièrement au niveau du modèle ontologique. Par rapport au modèle ontologique, les objets de notre système manipulables par les utilisateurs sont :

- des canaux qui représentent un dispositif sur lequel peut transiter un flux sonore,
- des conférences qui correspondent à un ensemble de canaux et à la description des caractéristiques des flux sonores transitant par ces canaux.

Le modèle ontologique prévoit aussi une description des opérations sur ces objets accessibles aux utilisateurs. Ces opérations sont organisées en quatre classes :

- des opérations de présentation (*view operations*),
- des opérations de création (*create operations*),
- des opérations de modification (*modify operations*),
- des opérations de destruction (*destroy operations*).

Notre système offre aux utilisateurs des opérations permettant de connaître :

- la liste des canaux disponibles,
- les caractéristiques propres à chaque canal,
- la liste des conférences en cours,
- les caractéristiques propres à chaque conférence.

Notre système offre aussi des opérations :

- de création et de suppression de conférence,
- de modification des conférences par insertion ou retrait de canal dans une conférence,
- de modification de l'ensemble des conférences (fusion, scission, aparté),

- de modification des caractéristiques des canaux globalement ou localement à une conférence.

En ce qui concerne le modèle de coordination d'Ellis, notre système est extrêmement souple. Comme nous l'avons dit plus haut, la coordination d'un collecticiel synchrone distribué ne peut pas se restreindre au niveau des communications verbales. De toute évidence, les utilisateurs communiqueront entre eux par l'intermédiaire d'un objet canal qui leur sera associé. Ils pourront prendre part à plusieurs audioconférences et contrôler les paramètres sonores de leur canal globalement et localement à chaque audioconférence. Ils auront aussi la possibilité de réorganiser dynamiquement l'ensemble des conférences en cours, d'instancier des apartés, etc. Ces activités sont cependant marginales par rapport à l'activité essentielle réalisée par l'intermédiaire du canal sonore qui est la communication verbale. Une description de notre système au niveau du modèle de coordination d'Ellis est alors précieuse.

Par rapport au modèle d'interface utilisateur d'Ellis, notre système est incomplet, d'une part parce qu'il ne s'agit que d'un composant particulier d'un collecticiel synchrone distribué, et d'autre part parce qu'il s'agit d'un système novateur et que le problème de la représentation des éléments de ce système est encore un problème ouvert. Une interface de démonstration décrite au chapitre suivant donne un exemple d'une représentation au niveau de l'interface utilisateur des objets manipulables par les utilisateurs. Cette forme de représentation la plus simple (des listes) n'est cependant pas satisfaisante d'un point de vue pratique. Il faudrait y substituer une représentation plus intuitive, moins encombrante pour l'interface visuelle et cohérente avec le reste de l'interface du système. La définition d'une telle interface se heurte au problème de la négociation sous-jacente dans l'utilisation des services proposés qui concernent plusieurs personnes (fusion, scission, aparté ...).

Néanmoins, il y a un aspect relevant du modèle d'interface d'Ellis qui a été abordé. Il se situe au niveau de la représentation du contexte (sonore) : l'utilisation

de la spatialisation du son dans la synthèse sonore. Nous en avons fait la description au paragraphe 2.1 *Synthèse sonore : spatialisation du son*, page 75.

Chapitre 3 : Modélisation, intégration et implémentation

Les informations contenues dans ce chapitre visent les objectifs suivants :

- répondre à la description d'un système telle qu'elle est prévue dans le modèle conceptuel d'Ellis (modèle ontologique),
- préciser le cadre logiciel dans lequel nous avons prévu d'intégrer notre travail,
- donner une description du travail préparatoire nécessaire aux travaux présentés dans le chapitre suivant,
- servir de base à un travail de généralisation qui sera abordé à la fin de ce chapitre.

Les problématiques connexes telles que les mécanismes de mise en place d'une téléconférence, la gestion des droits d'accès, ni même la gestion des droits de prise de parole, n'apparaissent pas dans cette modélisation. Par contre, nous nous sommes efforcé de mettre en place tous les mécanismes permettant de résoudre ces problématiques sans avoir à se soucier de l'aspect sonore. Nous avons aussi cherché à assurer la séparation entre les mécanismes de contrôle et les politiques de gestion des collecticiels, comme le rappelle Rodden [RODDEN 91b].

1 Modélisation d'ACAO

Dans cette modélisation, on se situe au niveau d'un composant interne d'un collecticiel. Il est autonome quant à la manipulation des flux sonores, mais inutilisable isolément. Nous en faisons une description depuis ses entités internes

jusqu'à son interface externe (interface entre composants logiciels). Cette partie reprend et complète [BARME 94].

1.1 Formalisme

Afin d'explicitier la problématique de la gestion des flux sonores pour un ensemble d'audioconférences assistées par ordinateur, nous proposons d'abord une définition plus formelle d'un canal. Puis nous donnons la définition d'une audioconférence et des exemples.

1.1.1 Les composants atomiques d'une audioconférence

L'analyse des services envisageables dans le cadre d'une ACAO amène à définir un canal comme une composition des trois entités suivantes :

- une *source*, par exemple la prise de son d'un microphone ; dans la suite, on note s_i la source du i ème canal,
- une *destination*, par exemple le son restitué par un canal à un participant à la conférence ; notée d_i ,
- des *données associées*, des données au sens informatique, par exemple nom du participant utilisant le canal, détection de prise de parole, information de spatialisation, notées D_i .

1.1.2 Définition d'une audioconférence

Une *audioconférence* se définit alors comme un *ensemble de canaux* d'une part et, d'autre part, la description des *destinations* de chaque *canal* en fonction des sources et *données associées* de l'ensemble des *canaux*.

Cette description peut être formalisée par un système d'équations qui régit pour l'ensemble de ces canaux les destinations en fonction des sources et des données associées. Ce système d'équation décrit en fait les mixages sonores à réaliser. Avant d'en expliquer quelques cas particuliers ci-dessous, nous précisons la notation que nous employons sur la formule générale suivante :

$$d_i = \beta_i \sum_{\substack{k=1 \\ k \neq i}}^n \alpha_k s_k$$

Le signe Σ désigne une sommation sonore, c'est-à-dire un mixage de plusieurs sons, en l'occurrence le mixage des sources sonores notées s_i . d_i désigne ainsi le son tel qu'il doit être restitué. Les α et β sont des coefficients qui correspondent à des gains de mixage du son respectivement pour les sons captés pour chaque source et le son restitué à un participant pour une audioconférence donnée ; ils sont appelés les coefficients d'écoute (β) et de parole (α).

1.1.3 Exemples

Considérons le cas le plus simple, celui d'une audioconférence telle que France Télécom en propose. Chaque canal reçoit le même flux sonore qui est la somme brute (au sens mixage) des sons provenant de tous les canaux sauf le sien, ce qui élimine l'effet Larsen direct. Le canal i est le couple (s_i, d_i) . La conférence est définie par l'ensemble $\{(s_1, d_1), (s_2, d_2), \dots, (s_n, d_n)\}$ et le système d'équations tel que,

$$\text{pour tout } i \text{ de } 1 \text{ à } n, d_i = \sum_{\substack{k=1 \\ k \neq i}}^n s_k$$

On n'élimine ici que la possibilité d'un accrochage direct. Il restera quand même une possibilité d'écho à travers un accrochage entre les microphones et les enceintes des autres participants. Cette dernière éventualité sera résolue par des microphones très sélectifs et un réglage approprié des gains d'amplification.

Considérons un cas légèrement plus complexe : l'interconnexion d'audioconférences. L'articulation *participant* \leftrightarrow *canal* permet d'associer plusieurs

canaux à un seul participant. Comme il a été expliqué plus haut, ce dernier peut participer ainsi à plusieurs audioconférences. Etendu à un groupe de personnes, cela permet par exemple à une conférence d'en écouter une autre, voire d'en discuter et de l'analyser, sans la perturber. Selon notre modèle cela s'exprime comme suit :

La conférence observée $\{(s_1, d_1, D_1), (s_2, d_2, D_2), \dots, (s_n, d_n, D_n)\}$ contient ses participants légitimes et les observateurs, mais ceux-ci ne font qu'écouter, en arrière plan, les discussions qui s'y déroulent. Les données associées à chaque canal distinguent le type de participant (observateur ou non). Simultanément les observateurs sont réunis dans une autre audioconférence $\{(s'_1, d'_1, D'_1), (s'_2, d'_2, D'_2), \dots, (s'_p, d'_p, D'_p)\}$. La conférence observée est décrite par le système d'équation suivant :

$$\text{pour tout } i \text{ de } 1 \text{ à } n, d_i = \beta_i \sum_{\substack{k=1 \\ k \neq i}}^n \alpha_k s_k ,$$

avec $\alpha_k = f(D_k) : \alpha_k = 0$ si le participant k est un observateur et $\alpha_k = 1$ sinon,

et avec $\beta_i = g(D_i) : 0 < \beta_i \leq 1$ si le participant i est un observateur et $\beta_i = 1$ sinon.

Ainsi, les commentaires des observateurs ne sont pas restitués ($\alpha=0$) et les observateurs perçoivent les discussions de la conférence observée en arrière plan ($0 < \beta \leq 1$).

$$\text{Pour la conférence des observateurs on a pour tout } i \text{ de } 1 \text{ à } p : d'_i = \sum_{\substack{k=1 \\ k \neq i}}^p s'_k .$$

Ainsi, les observateurs sont dans une conférence classique, se parlent et s'écoutent normalement. Les d'_i et d_i des observateurs correspondent aux mêmes dispositifs de restitution du son, ce qui donne l'effet escompté : les observateurs écoutent et commentent la conférence principale sans la perturber.

Pour ne pas alourdir l'expression des équations ci-dessus, nous n'avons pas fait figurer la spatialisation du son. Cependant le principe de base reste le même. Une "conférence" est un ensemble de canaux. Pour l'ensemble des canaux écoutés, on définit le son perçu en tenant compte, pour chaque source, des informations contenues dans les données associées. On applique enfin les informations contenues dans les données associées du canal qui écoute.

Ce modèle permet de réaliser les interconnexions sonores correspondant aux services proposés au chapitre précédent. Cependant, telle quelle, la manipulation est difficile. Nous proposons donc un langage de commande.

1.2 Un langage de commande

Chaque commande sera traduite par le système informatique dans les termes du modèle afin de réaliser le service voulu. Nous rejoignons ici la proposition de Rangan et Vin [RANGAN 91] en proposant un jeu d'opérations de base. Notre proposition en diffère cependant avec un noyau plus restreint et des fonctions plus évoluées.

Nous avons retenu quatre commandes de base pour la manipulation dynamique des audioconférences :

- une commande de création de conférence ; cette commande attribuera un numéro d'identification unique à cette conférence, utilisé pour y accéder par la suite,
- une commande de suppression avec, comme paramètre, l'identificateur de la conférence à supprimer,
- une commande d'insertion d'un canal dans une conférence,
- une commande de suppression d'un canal d'une conférence.

En fonction des données associées, il faut rajouter d'autres commandes pour y accéder et les modifier. Il est possible de tout faire avec les quatre commandes de base et les commandes de manipulation des données associées. Cependant, pour

des raisons de commodité et de performance, on rajoutera des commandes spécifiques pour les fusions, scissions et apartés.

Nous donnons ici une description précise du jeu de commandes implémenté dans notre prototype. Il s'agit d'un exemple et non d'un modèle strict. Derrière cette description spécifique qui répond à une motivation d'efficacité ponctuelle, il faut considérer les motivations générales données plus haut. Ce jeu de commandes ne prétend pas être idéal ni répondre à une définition théorique d'un langage.

La syntaxe choisie pour exprimer les commandes est celle des messages DDE²⁵ définis pour Windows. Ce choix découle de l'option prise pour la communication inter-outils dans CoLearn en prévision d'une intégration de ce composant logiciel dans le collecticiel réalisé pour CoLearn.

Les messages sont codés sous formes de chaînes de caractères ASCII. Ce procédé n'est ni économique ni efficace au niveau informatique mais a le mérite d'être parfaitement clair et de faciliter le suivi du bon fonctionnement de l'ensemble. Ce type de codage permet d'éviter aussi l'écueil des différentes formes de codages d'une plate-forme à l'autre. La syntaxe des messages est la suivante :

```
[commande(param1,param2,...,"chaîne de caractère",...)]
```

La liste des paramètres entre parenthèses est optionnelle, une commande peut ne pas avoir de paramètre. Le nombre de paramètres n'est pas défini à l'avance, il peut varier d'une commande à l'autre et même d'un appel à l'autre d'une même commande. Cette syntaxe est appliquée aussi bien pour les commandes envoyées au contrôleur que pour les réponses qu'il renvoie. Chaque message est donc délimité par des crochets. Le premier mot du message correspond au code de la commande. Si la commande amène des paramètres, ceux-ci sont donnés sous la forme d'une liste entre parenthèses. Un paramètre est soit une valeur numérique, soit une chaîne de caractères entre guillemets. Une valeur numérique est aussi codée sous la forme d'une chaîne de caractères ASCII. Toute commande non

²⁵ Dynamic Data Exchange

valide, tant du point de vue syntaxique que logique, est ignorée. La description complète du jeu de commande est donnée en annexe.

1.3 Synthèse de la modélisation d'ACAO

Cette modélisation a pour but de faire le pont entre les services proposés dans le chapitre précédent et leur implémentation dans un système informatique. Elle comporte deux passerelles :

- un langage de commande pour synthétiser les actions,
- une modélisation des audioconférences pour réaliser les connections sonores désirées.

Il n'a pas été question jusqu'ici de la dimension du collecticiel, si ce n'est parce qu'il s'agit d'audioconférence. La partie suivante présente cet aspect.

2 Intégration dans un collecticiel

Nous précisons ici le contexte, d'un point de vue système informatique, dans lequel se place notre recherche. Le modèle du système multi-agents est proposé pour organiser les composants logiciels.

2.1 Le contexte d'un collecticiel synchrone distant

Nous rappelons ici les contraintes liées à notre domaine qui ont influencé nos choix pour l'organisation de nos développements dans une optique plus générale. Le terme *collecticiel* infère qu'il y a plusieurs utilisateurs impliqués dans une activité commune, *synchrone*, que ces utilisateurs sont présents en même temps. Le terme *distant* infère d'une part qu'ils ne sont pas au même endroit mais, surtout, que plusieurs systèmes informatiques, interconnectés par des réseaux, sont mis en oeuvre. Il convient de garder à l'esprit ces contraintes et la vaste dimension qu'elles impliquent. En fait, de nombreux domaines de recherche en informatique à part entière sont impliqués simultanément.

Au niveau des interfaces homme-machine par exemple, non seulement il y a plusieurs utilisateurs, mais encore ils agissent ensemble, en parallèle, sur plusieurs interfaces qui ne sont pas obligatoirement homogènes, et à distance. En outre les modes d'interaction sont multiples : par l'image, par le son, par l'intermédiaire de documents partagés. Pour chaque personne, il y a un contexte privé et, pour l'ensemble, il y a un contexte public dont chacun peut avoir une perception différente.

2.2 Le modèle du système multi-agents

Un atelier de travail sur les systèmes multi-agents a montré que ce concept ne fait pas encore l'objet d'un consensus quant à sa définition [GANYMEDE 95]. Aussi nous précisons ce que nous entendons par "système multi-agents" sans prétendre en imposer une définition dogmatique.

2.2.1 Définition d'un agent

Notre approche pose comme postulat de départ le contexte de la programmation à objet. Notre définition d'un agent peut être vue comme s'inspirant des objets actifs complexes tels qu'ils sont décrits dans [GRANSART 95]. Un agent est pour nous une entité logique autonome qui maintient un état interne complexe et en partie persistant : certaines caractéristiques seront constantes d'une session à l'autre. Un agent reçoit et envoie des messages (figure 15). Les messages reçus peuvent donner lieu à un traitement, une modification de l'état interne et à l'envoi de messages en réponse. Cette définition correspond à une version simplifiée des agents cognitifs de Cardozo [CARDOZO 93] puisqu'elle n'inclut pas de base de connaissances. Elle correspond à une version étendue des agents réactifs d'Abowd [ABOWD 90] puisqu'elle inclut la notion d'un état interne qui influence les réponses de l'agent : suivant la situation courante et les contraintes externes, notre agent peut avoir des réponses différentes à un même message (cf. [CROISY 95] pour les notions d'agents cognitifs et réactifs).

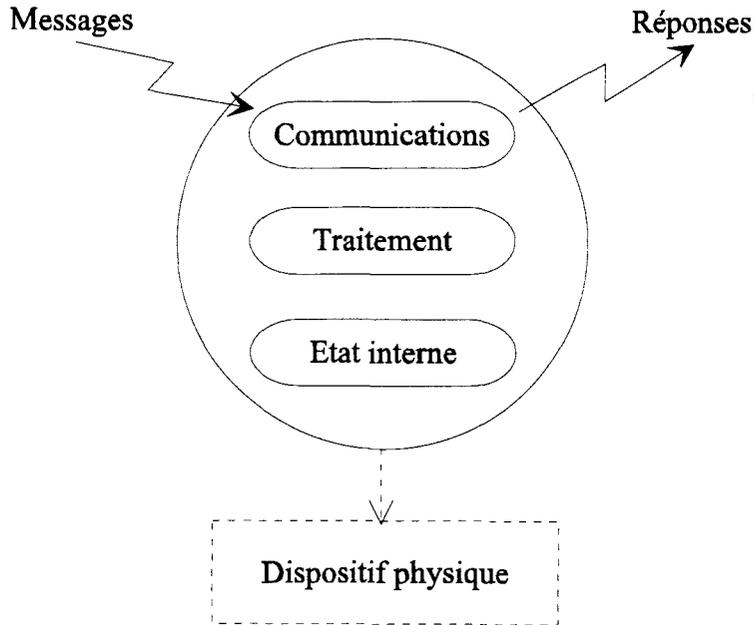


figure 15 : définition d'un agent

Une autre manière de voir un agent est de le considérer comme l'équivalent, au niveau d'un système distribué, d'un objet au niveau d'un logiciel, par rapport aux possibilités d'encapsulation. Aux données (état interne) et méthodes (traitements), nous ajoutons les communications (figure 15). Ces communications ont des spécificités particulières. Elle sont :

- *multiples* : il y a simultanément plusieurs correspondants, chaque correspondant est susceptible d'envoyer et de recevoir des messages,
- *simultanées* : il y aura une concurrence d'accès que l'agent devra gérer,
- *distribuées* : les communications se font entre des systèmes informatiques multi plates-formes distants et interconnectés par des réseaux,
- *contrôlées* : un agent doit maîtriser ses communications, pouvoir envoyer un message à un correspondant unique, à un groupe de correspondants ou à l'ensemble de ses correspondants,
- *ouvertes* : le protocole de communication ne doit pas être propriétaire ; un agent doit pouvoir communiquer avec un autre agent avec un protocole ouvert et standardisé.

Grâce à ce composant "communications", un agent reçoit et (éventuellement simultanément) envoie des messages. La réception et l'émission de messages sont indépendantes. Un agent peut recevoir un message qui provoquera aucune ou toute une série de réponses, à tous ses correspondants, ou à certains d'entre eux. Un agent peut envoyer des messages sans en avoir reçu aucun. Un message peut donner lieu à un traitement, qui amènera éventuellement à modifier l'état interne de l'agent.

Un agent n'est pas obligatoirement une entité abstraite. Un collecticiel utilise des ressources physiques réelles. Pour faire le pont entre les deux, nous proposons qu'un agent puisse piloter un dispositif physique et que les ressources matérielles soient vues du système par l'intermédiaire d'un agent.

2.2.2 Organisation des agents d'un collecticiel synchrone distribué

Nous proposons d'utiliser plusieurs agents pour implémenter un collecticiel synchrone distribué en un système multi-agents, récursivement à plusieurs niveaux. L'ensemble du collecticiel distribué sur les systèmes informatiques sera composé de plusieurs agents, au moins un par utilisateur et un pour le collecticiel de téléconférence. Chacun de ces composants est en fait lui-même composé de plusieurs agents, éventuellement eux-mêmes composés de plusieurs agents.

Nous ne prétendons pas résoudre l'ensemble des problèmes que cette organisation soulève. Nous précisons ce cadre pour que les agents qui nous concernent, et que nous détaillerons par la suite, soient situés dans le contexte pour lequel nous les avons prévus. Cette approche a été utilisée au niveau de la rétroaction visuelle dans les interfaces de collecticiels [CROISY 95].

Comme nous l'avons dit plus haut, par souci de modularité, nous décomposons l'agent de téléconférence d'un collecticiel synchrone distribué en plusieurs entités (figure 16):

- un agent d'interface en relation avec les agents d'interface utilisateurs,
- un agent de visioconférence,

- un agent pour la gestion de l'espace public,
- un agent d'audioconférence,
- un agent de contrôle qui coordonne les trois précédents.

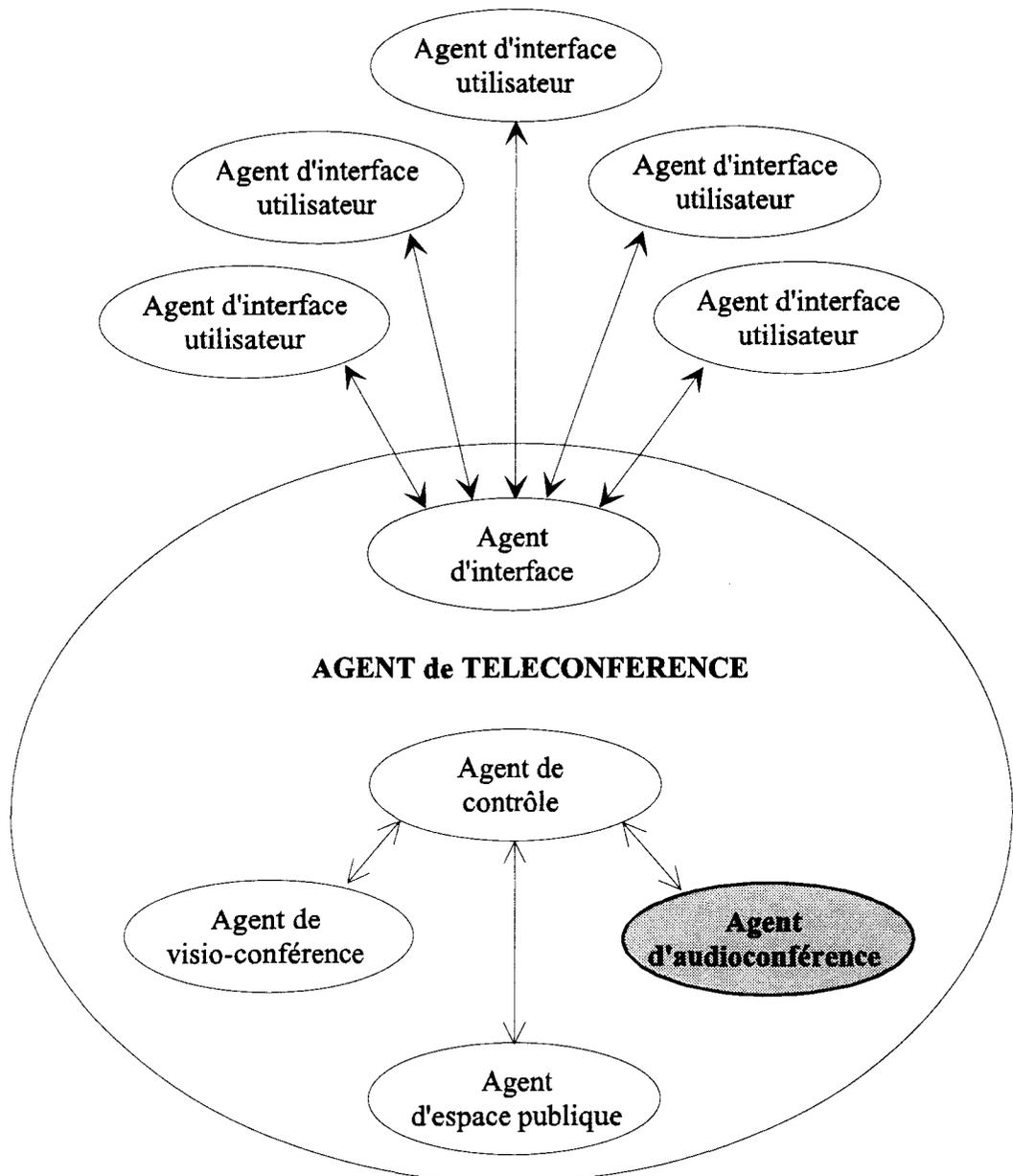


figure 16 : organisation d'un agent de téléconférence

L'idée de décomposer un système de téléconférence en plusieurs agents n'est pas originale. Elle se trouve dans de nombreux travaux, que les composants logiciels soient appelés des agents [SCHOOLER 93b] ou qu'ils soient conçus

plutôt comme des serveurs [MAENO 91]. La figure 16 présente un cadre général dans lequel nous prévoyons l'insertion de notre agent d'audioconférence. Nous faisons remarquer aussi qu'il n'est fait aucun préjugé quant à l'architecture matérielle ni la répartition sur l'ensemble des systèmes informatiques.

De même la gestion et le contrôle sont déportés dans un agent à part. Si la gestion de droit à la parole était gérée par l'agent d'audioconférence, il ne serait plus possible de le synchroniser avec une gestion du partage du contrôle de l'espace publique ou avec l'image vidéo. Le problème du contrôle doit être traité séparément. Par contre, nous cherchons à offrir au niveau de l'agent d'audioconférence tous les outils nécessaires au contrôle du son, et ce, quel que soit le protocole choisi.

De même que l'agent de téléconférence est composite, les agents d'interface utilisateurs sont composites. Un même et unique agent d'audioconférence implémentant les services proposés plus haut peut être utilisé dans des configurations variées. On peut y mettre en correspondance différents types d'agents comme un agent de suivi et de supervision, un agent d'évaluation (destiné à recueillir et exploiter des informations pour évaluer le système d'audioconférence), et des agents d'interface utilisateurs.

On peut relier les agents d'interface utilisateurs directement à l'agent d'audioconférence et réaliser le contrôle au niveau des agents d'interface utilisateurs avec une coordination orchestrée par l'agent de contrôle de téléconférence. On peut aussi passer par l'intermédiaire d'un agent de contrôle situé entre les agents d'interface utilisateurs et l'agent d'audioconférence.

3 Réalisation d'un système expérimental

En partie pour illustrer la modélisation précédente, mais surtout pour mettre en place l'évaluation des services proposés, nous avons réalisé un système expérimental d'audioconférence. Face aux contraintes matérielles et aux limites de la technologie actuelle, il a fallu faire des développements spécifiques et déployer

un assemblage de matériel hétérogène. Cependant cela ne remet pas en cause la généralité du modèle.

3.1 Architecture matérielle

Pour s'affranchir des problèmes de transport du son sur les réseaux et avoir une qualité sonore acceptable, nous avons choisi un réseau analogique pour le canal sonore. Pour effectivement réaliser les interconnexions sonores, tout en gardant le contrôle de l'informatique, nous avons utilisé des régies sonores professionnelles contrôlables par un ordinateur. Le monde du son professionnel a déjà sa norme de contrôle : le MIDI plus ou moins intégré dans le monde informatique. L'informatique a aussi son standard pour les réseaux : Internet, et son standard pour les interactions homme-machine : les stations de travail, avec, dans le monde de la recherche, une préférence pour les stations de travail sous Unix.

Cependant le monde Unix est-il en général difficile à interfacer avec la norme MIDI, alors que celle-ci est complètement intégrée, par des cartes sonores très bon marché, au monde des PC. Cependant les PC sous Windows 3.1 souffrent d'un système d'exploitation obsolète, le MSDOS, poussé au delà des limites de ses possibilités pour offrir une interface convenable. Heureusement, un compromis entre un accès aisé au MIDI et un système d'exploitation convenable est rendu possible avec Windows NT.

En résumé, nous avons utilisé un réseau analogique pour le son. Le pont audio a été réalisé par un assemblage de régies numériques contrôlables par une interface MIDI (figure 17). Le contrôle est fait depuis une station de travail sous Windows NT connectée par un réseau Ethernet à d'autres stations de travail sous Windows NT ou sous Unix.

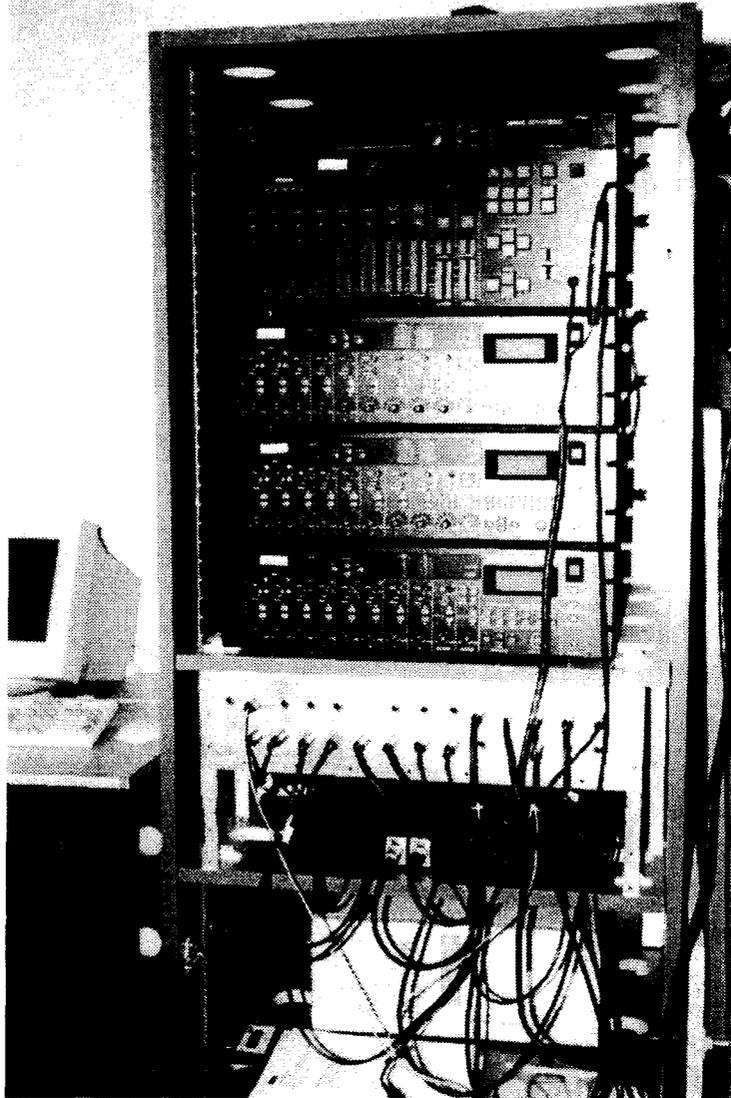


figure 17 : notre pont audio expérimental

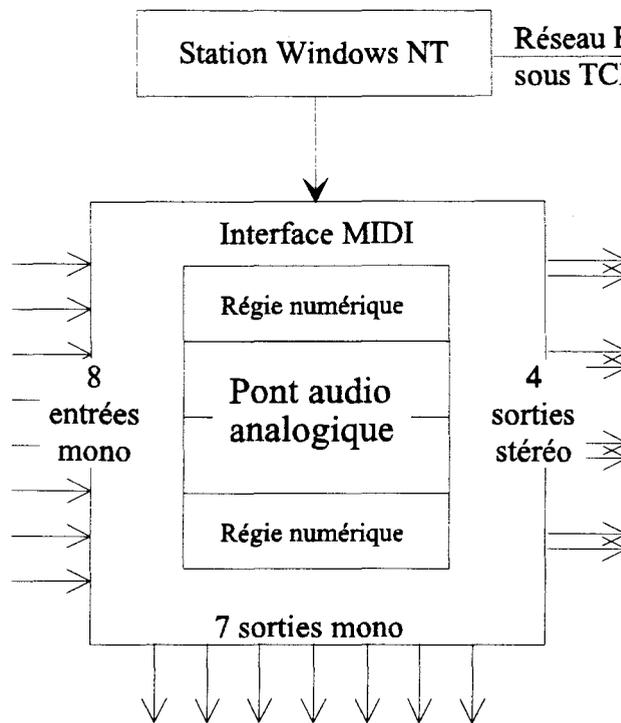


figure 18 : schéma de notre pont audio expérimental

Le pont audio est réalisé à partir d'un assemblage de 4 régies numériques. L'une est un Yamaha DMP²⁶ 11, les trois autres²⁷ sont des Yamaha DMP 9-8. Notre assemblage compose un pont audio analogique entièrement pilotable par ordinateur avec les caractéristiques suivantes (figure 18):

- 8 entrées sonores analogiques,
- 4 sorties sonores analogiques en stéréophonie, qui peuvent être indépendamment découplées chacune en deux sorties monophoniques. Sur chacune de ces sorties, on dispose de deux boucles d'effets indépendantes et dont les paramètres sont réglables,

²⁶ Digital Mixing Processor

²⁷ Le modèle DMP9-8 est le successeur du DMP 11. Cette hétérogénéité est due au laps de temps qui s'est écoulé après la prise en main du DMP11 pour en évaluer le potentiel par rapport à la réalisation d'un pont audio expérimental.

- 7 sorties monophoniques auxiliaires sans effet.

Ce pont audio est actuellement relié par un réseau sonore analogique à 5 pièces différentes équipées de microphones et d'amplificateurs en stéréo :

- le local où se situe le pont audio,
- un poste type équipé d'une paire de haut-parleurs amplifiés de très bonne qualité, d'un microphone hypercardoïde et d'une station de travail Windows NT,
- une salle expérimentale équipée d'un amplificateur permettant de reconstituer n'importe quelle ambiance sonore sur 7 haut-parleurs, de 4 postes de travail incorporés dans un meuble spécialement conçu pour le travail coopératif. Chaque poste est équipé d'un microphone hypercardoïde et d'une station de travail sous Unix (Sparc 5),
- une salle de conférence multimédia,
- un plateau de production audio vidéo.

Les quatre microphones de la salle expérimentale sont regroupés au niveau d'une régie manuelle, si bien qu'ils apparaissent comme une entrée unique au niveau du pont audio. Cependant, on peut les prendre séparément. Les sorties auxiliaires sont utilisées pour la DAL et pour faire des enregistrements sur un magnétophone. Evidemment, le fait que ce soit un réseau analogique entraîne une certaine rigidité dans les configurations possibles, mais un soin tout particulier a été apporté pour conserver un maximum de souplesse.

Ce pont audio est donc le dispositif physique lié à l'agent d'audioconférence qui tourne sur la station de travail Windows NT contrôlant les régies numériques par une interface MIDI.

3.2 Un serveur d'ACAO

Pour l'implémentation de l'agent d'audioconférence, nous aurions aimé pouvoir maintenir la métaphore du système multi-agents, avec un agent de communication,

un agent de contrôle d'ACAO, un agent pour la DAL (Détection Automatique du Locuteur) et d'une manière générale un agent pour chaque composant modulaire. Deux obstacles nous ont gêné : la possibilité d'utiliser le principe du multithread avec les MFC (les classes C++ de base de Microsoft pour développer des applications sous Windows) et le passage obligé des événements par la fenêtre principale d'une application Windows. Ce développement n'étant pas l'aspect principal de cette thèse, nous avons cherché un compromis entre une solution qui fonctionne bien et une solution développée avec un niveau de qualité industrielle.

Nous avons écrit le code du serveur d'ACAO alors que la version des MFC de Windows 32bits ne permettait pas encore le multithread. La nouvelle version des MFC le permet ; il reste à en tirer parti. Nous présentons dans cette thèse une version qui a dû se limiter à des objets passifs avec un seul flux d'instructions commun à toute l'application.

Même avec les MFC 3.0, il reste le passage obligatoire par la fenêtre principale d'une application pour la gestion des événements. Le serveur d'ACAO est donc un assemblage d'objets encapsulant des services de manière modulaire avec une concession obligée par la plate-forme de programmation de Windows qui fait apparaître un objet partagé : celui qui représente la fenêtre principale de l'application. En effet, l'entité de programmation sous Windows NT est l'application (l'équivalent du processus sous Unix). Une application est obligatoirement liée à une fenêtre principale dont le numéro d'identification (le handle) est aux applications Windows ce qu'est le numéro de processus sous Unix. Les événements liés à l'interface utilisateur et les événements liés à la communication par socket ou à l'utilisation de la carte sonore passent tous par la fenêtre principale.

Nous aurions aussi voulu utiliser un mécanisme de communication à la hauteur de la métaphore du multi-agents et intégré au niveau du système. Celui-ci n'existe pas encore sous Windows NT. Nous avons pallié cette lacune par l'encapsulation des communications par sockets dans des objets.

3.2.1 L'agent d'ACAO : un assemblage d'objets se partageant les traitements

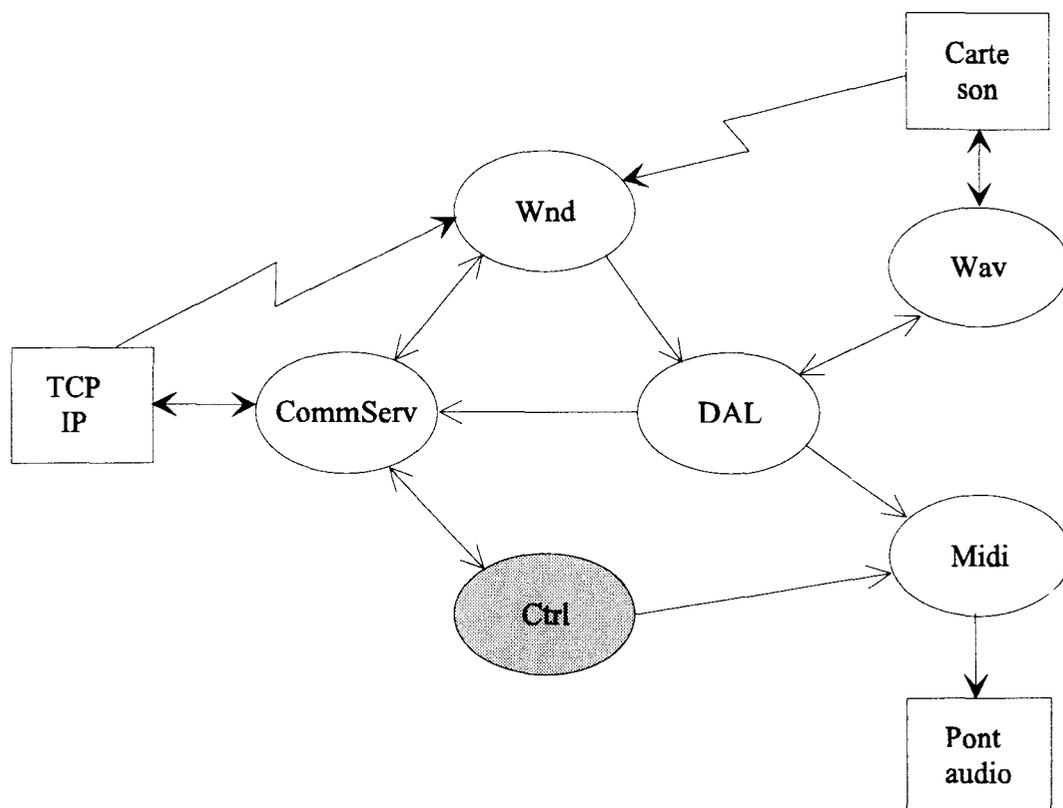


figure 19 : implémentation de l'agent d'audioconférence

Outre les objets de base d'une application utilisant les MFC, le serveur d'ACAO est composé des objets principaux suivants (figure 19):

- un objet Wnd qui intercepte les événements c'est-à-dire tous les messages de Windows,
- un objet CommServ qui encapsule la communication inter-agents pour le serveur : il gère la socket de connexion et les sockets de communication avec tous les agents en correspondance dont il maintient la liste,
- un objet Ctrl qui gère l'état logique du système d'ACAO et traduit cet état logique au niveau du pont audio,
- un objet DAL qui gère la détection des prises de parole sur l'ensemble des huit entrées du pont audio,

- un objet Midi qui permet de piloter le pont audio ; il est partagé par les objets Ctrl et DAL,
- un objet Wav qui pilote la carte sonore pour la numérisation des sons nécessaire à la DAL.

A un autre niveau, le langage de commande implémenté dans l'objet Ctrl est très facilement extensible ou modifiable (cf. les explications données en 4.1.4 *La gestion du langage de commande par l'objet Ctrl*, page131). Ce langage de commande ne devrait pas être difficile à traduire en IDL sur une plate-forme CORBA où l'agent d'ACAO bénéficierait de l'intégration de la métaphore multi-agents au niveau système.

L'objet Ctrl et des détails d'implémentation sont donnés dans la partie suivante :
4 Points clés de l'agent d'ACAO en vue d'une généralisation

3.2.2 Les liens logiques entre les objets principaux de l'agent d'ACAO

La figure 19 explique les liens logiques entre les différents objets composant le serveur d'ACAO. Un agent correspondant communique avec lui par l'intermédiaire de la couche TCP/IP du réseau. C'est d'abord l'objet Wnd qui reçoit l'événement et le communique aussitôt à l'objet CommServ. Celui-ci confie le traitement de la requête à l'objet Ctrl. Suivant les cas, l'objet Ctrl peut simplement répondre (par l'intermédiaire de l'objet CommServ), ou modifier son état interne et éventuellement demander à l'objet Midi de répercuter cette modification sur le pont audio, puis prévenir l'ensemble de ses correspondants de la nouvelle situation. La prise en compte de la DAL pour séparer plusieurs canaux logiques d'un même canal physique n'est pas encore implémentée. Il manque donc le lien entre l'objet DAL et l'objet Ctrl. Pour le moment, l'objet DAL envoie aux correspondants de l'agent d'ACAO des messages les informant sur les canaux actifs. L'explication du mécanisme suit.

3.2.3 La Détection Automatique des Locuteurs

En parallèle avec ces traitements, et de manière continue, l'objet DAL orchestre la détection des locuteurs. Il règle le pont audio pour recevoir sur la carte son deux des huit voies. Il demande à l'objet Wav de procéder à la numérisation de ces voies. Quand celle-ci est finie, la carte son prévient l'objet Wnd qui retransmet aussitôt cet événement à l'objet DAL. Ce dernier demande à l'objet Wav de vérifier si un locuteur est détecté sur ces deux voies. S'il y a une modification de l'ensemble des prises de parole, l'objet DAL prévient les correspondants de l'agent d'ACAO de la nouvelle situation, par l'intermédiaire de l'objet CommServ. Quoiqu'il en soit, l'objet DAL passe aux deux voies suivantes et recommence le processus.

Ce traitement pour la détection automatique des locuteurs peut paraître compliqué. A titre de comparaison, il faut se reporter à celui décrit dans [SELLEN 92] (cf. 2.3 *Détection de prise de parole* page 47). Dans sa version courante, la DAL est réalisée avec une solution centralisée, puisque c'est le seul point où le système informatique peut accéder aux flux sonores. Une solution distribuée où chaque poste réaliserait la DAL pour en diffuser le résultat par le réseau informatique semble cependant plus judicieuse quant à la répartition des traitements.

3.2.4 L'application sous Windows NT

L'agent d'ACAO est donc implémenté comme une application sous Windows NT qui tourne sur le serveur pilotant le pont audio. Comme son rôle ne prévoit aucune interaction directe avec les utilisateurs, son interface ne laisse quasiment rien voir de l'ampleur de la tâche qu'il réalise.

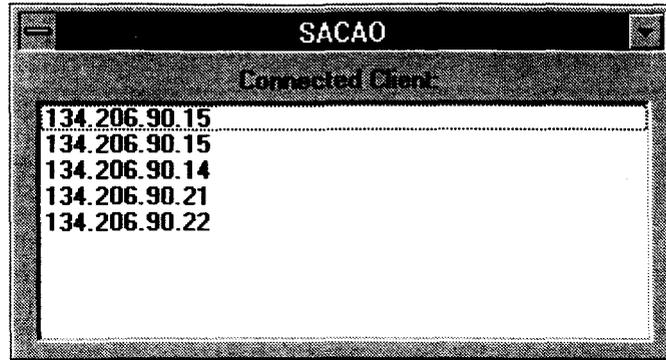


figure 20 : interface de l'agent d'ACAO

La copie d'écran précédente montre cette interface (figure 20). Elle n'affiche que la liste des adresses IP des machines où des agents clients sont connectés. Il peut y avoir autant d'agents clients par machine qu'on le souhaite. Dans cette liste, les trois premières adresses correspondent à des stations sous Windows NT, les deux suivantes à des stations sous Unix.

3.3 Des interfaces clients

L'agent d'ACAO prend en charge la gestion et le contrôle d'audioconférences. Il est prévu pour être utilisé à partir d'agents d'interface. Pour illustrer ce principe, nous avons développé (essentiellement sous Windows NT en Visual C++) des interfaces de test et de démonstration dont la liste suit :

- une interface de suivi et de contrôle des messages envoyés par l'agent d'ACAO ; elle permet de lui envoyer des messages dans le langage de commande et de vérifier les réponses ainsi que de connaître les messages envoyés spontanément,
- une interface identique mais sur station de travail Unix,
- une interface de démonstration des services proposés. La première version fonctionne sous Windows NT. Un portage sous Unix et une spécialisation pour un utilisateur de cette interface ont été entrepris, mais ne sont pas encore achevés.

3.3.1 Architecture de base

Toutes ces interfaces doivent gérer en parallèle les interactions de l'utilisateur et les notifications de l'agent d'ACAO. Suivant le système d'exploitation utilisé, cela suppose une organisation spécifique. Le problème à la base est le même, les interactions de l'utilisateur monopolisent l'attention du flux d'instructions, la communication avec l'agent d'ACAO est gérée par détournement de ce flux sur un traitement d'événements asynchrones. Cependant, sous Windows NT, chaque événement sera intercepté comme un message envoyé à la fenêtre principale de l'application, alors que, sous Unix, il doit être traité comme une interruption logicielle envoyée au processus.

3.3.1.1 Un agent d'interface sous Windows NT

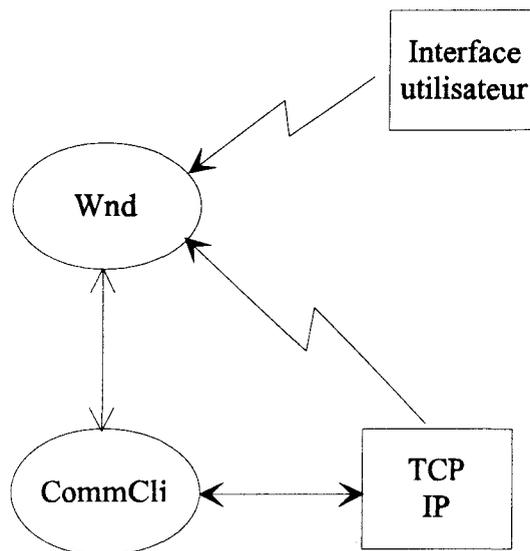


figure 21 : liens logiques des objets d'un agent d'interface

Sous Windows NT, on retrouve une architecture logicielle semblable à celle de l'agent d'ACAO, quoique très simplifiée (figure 21). Pour les interfaces les plus simples, les méthodes de traitement seront directement incorporées à l'objet Wnd. Pour l'interface de démonstration, l'assemblage des objets est très complexe, spécifique à Windows NT et au MFC. Il n'est pas généralisable.

Comme pour l'agent d'ACAO, c'est l'objet Wnd qui intercepte les événements. Quand l'utilisateur effectue une action qui requiert la transmission d'un message à l'agent d'ACAO, celui-ci est envoyé par l'objet CommCli. Quand la réponse arrive sur la socket, un événement est intercepté par l'objet Wnd et retransmis aussitôt à l'objet CommCli qui le récupère et demande à l'objet Wnd d'en répercuter les conséquences au niveau de l'interface utilisateur. Il est à noter que le rôle d'un agent d'interface se borne :

- à traduire les actions de l'utilisateur en messages envoyés à l'agent d'audioconférence,
- à présenter à l'utilisateur les informations fournies sous forme de message par l'agent d'audioconférence.

3.3.1.2 Un agent d'interface sous Unix

Sous Unix, dans l'environnement de programmation que nous avons pu utiliser (Sparcworks 3.0 et SDK 2.3 de Sun), il y a une rupture de cohérence imposée dans la conception à objet. Alors que sous Windows NT, l'application est encapsulée par un objet de la MFC, sous Unix, le processus ne l'est pas. Pour être notifié de la réception d'un message arrivant sur une socket, il faut intercepter le signal SIGIO. Cela ne peut être fait que par une fonction C et pas par une méthode d'un objet instance d'une classe écrite en C++.

3.3.2 Une interface de suivi et de commande

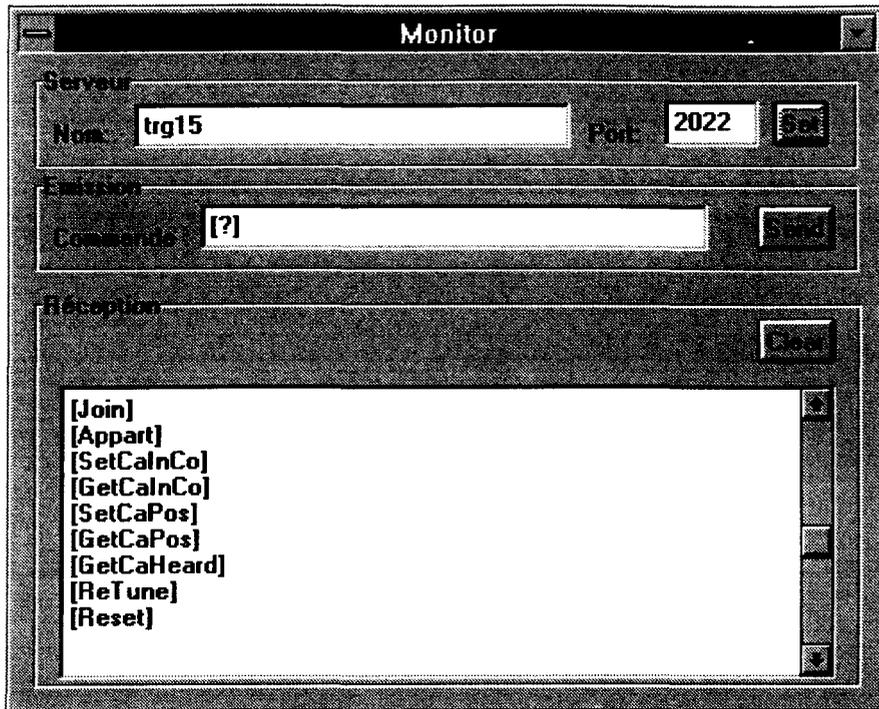


figure 22 : une interface de suivi et de commande

La figure 22 montre une copie d'écran de l'interface de suivi et de commande. Elle permet de désigner l'agent d'ACAO (voire d'en changer dynamiquement) vu comme un serveur s'exécutant sur une machine, avec un numéro de port particulier pour la socket de connexion. Une fenêtre d'édition permet de saisir une commande à envoyer à l'agent d'ACAO. Parallèlement, une fenêtre d'affichage présente les messages reçus. Sous Unix, une interface similaire (figure 23) a été réalisée. Elle nous a permis de démontrer l'indépendance de notre conception par rapport à la plate-forme choisie.

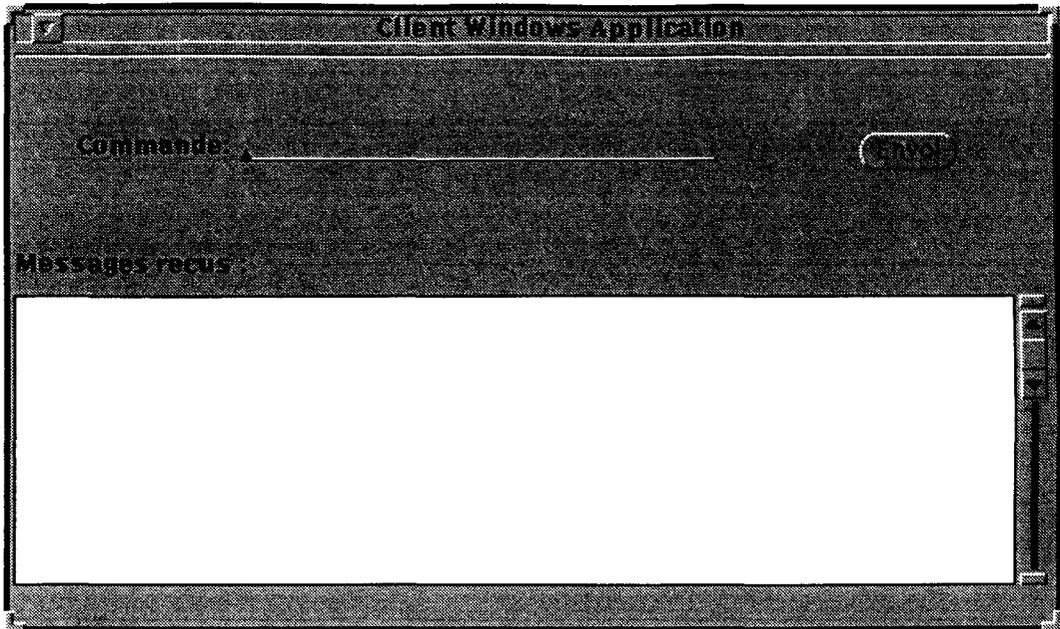


figure 23 : interface de gestion et de suivi de commande sous Unix

3.3.3 Une interface de démonstration

L'interface de contrôle est implémentée comme un agent client de l'agent d'ACAO et manipule trois types de sous-fenêtres (figure 24):

- une fenêtre qui contient la liste des canaux qui est toujours présente,
- une fenêtre par canal, qui peut être ouverte pour en modifier les données associées indépendantes des audioconférences dans lesquelles il pourrait être représenté,
- une fenêtre automatiquement ouverte par audioconférence en cours.

Les services sont directement accessibles par bouton ou par commande de menu avec une sélection des éléments des listes affichées le cas échéant. Les informations affichées sont celles détenues par l'objet contrôleur et non pas par l'interface de contrôle. Ces informations évoluent au rythme des messages de rétroaction envoyés par l'objet contrôleur, de sorte que s'il est piloté par d'autres clients (un autre composant du collecticiel par exemple), il affiche en temps réel l'état du système et des communications sonores en cours entre les participants.

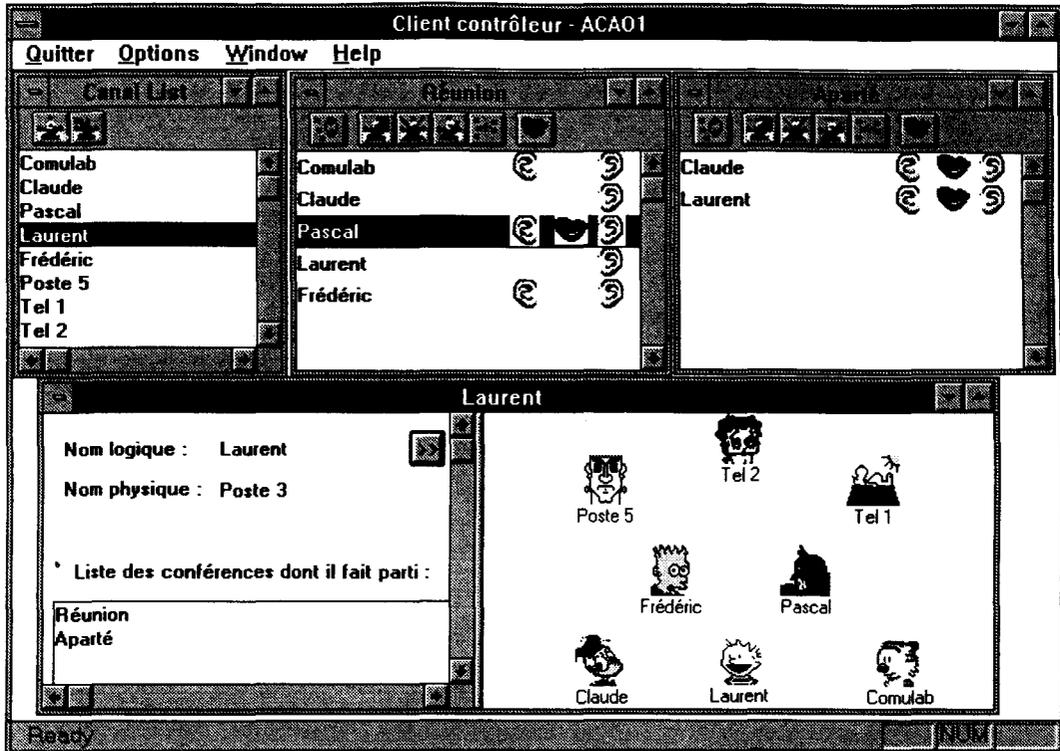


figure 24 : une interface de démonstration

La figure 24 montre un exemple simple d'état du serveur d'ACAO tel qu'il est visualisé par l'interface de contrôle. En haut à gauche figure la liste des canaux potentiels. Les deux fenêtres "Réunion" et "Aparté" présentent deux pièces virtuelles, l'une permettant une audioconférence à cinq participants où "Pascal" est le seul à pouvoir parler tandis que "Claude" et "Laurent" sont en aparté dans une autre pièce virtuelle tout en écoutant en arrière-plan la réunion. La fenêtre du bas affiche et permet de modifier les caractéristiques du canal associé à "Laurent" avec en particulier la position rendue pour le son venant des autres canaux représentée par les petites icônes.

4 Points clés de l'agent d'ACAO en vue d'une généralisation pour des développements ultérieurs

Comme nous l'avons précisé au début de la partie précédente de ce chapitre, l'implémentation réalisée dans le cadre de cette thèse a dû répondre à un souci d'efficacité immédiate en vue d'une première évaluation. Cependant, cette

implémentation comporte de nombreux aspects qui sont indépendants du matériel utilisé ou du contexte logiciel choisi. Nous en précisons ici les caractéristiques en vue de leur réutilisation pour des développements ultérieurs, notamment pour une gestion du son entièrement numérique dans un environnement système qui permet les applications distribuées.

Notre système expérimental comporte deux aspects qui sont particulièrement *ad hoc* :

- la gestion des communications inter-agents,
- le réseau sonore analogique.

Le premier aspect a été résolu par une encapsulation simple des communications par sockets dans des objets C++, ce qui ne mérite pas d'explications détaillées. Le deuxième aspect, particulièrement visible par l'assemblage des régies qu'il suppose, est sans doute la spécificité la plus flagrante de notre système expérimental. Cependant, ces régies sont gérées d'un point de vue logiciel comme de simples mixeurs sonores dont on contrôle les paramètres de mixage. Ces mixeurs physiques pilotés par une interface MIDI peuvent être facilement remplacés par des composants logiciels assurant le mixage de sons numériques encapsulés dans des objets informatiques.

Quant aux développements réalisés, c'est-à-dire la manipulation des composants atomiques d'une audioconférence et la réalisation des services proposés, ils peuvent être largement réutilisés. D'ailleurs, l'agent d'ACAO fonctionne indépendamment du pont audio.

Nous commencerons par une description des points clés de l'objet Ctrl qui comporte l'essentiel de ces développements. Cette description est fidèle à la version courante de la programmation de cet objet telle qu'elle est actuellement. Il s'agit donc d'une solution qui n'est pas optimisée. Nous donnerons ensuite des indications sur les principales modifications nécessaires en vue d'une implémentation plus généraliste.

Nous avons choisi de donner une description en C++ des classes des objets que nous avons implémentées pour la précision et l'exactitude que ce type de description permet. Ces descriptions doivent cependant être considérées comme des documents et peuvent être ignorées. Le texte et les schémas qui les accompagnent en font ressortir les points essentiels et fournissent les explications nécessaires à la compréhension des mécanismes importants sur lesquels nous voulons attirer l'attention.

4.1 Implémentation de l'agent d'audioconférence

L'implémentation de l'agent d'ACAO est centralisée dans l'objet Ctrl. L'objet Ctrl est une instance de la classe Controller dont nous donnons ci-dessous la définition complète en C++ (figure 25), ce qui permet d'en avoir une vue d'ensemble. Des explications structurées sont fournies ensuite.

```

//
class Controller
{
    // Classe encapsulant une réponse du contrôleur
    class Rep : public CObject
    {
    public:
        CString rep;
        RepFlags t;
    };

    // Classe encapsulant les réglages de perception d'un canal pour un autre
    canal
    class SndTuning
    {
    public:
        int state; // audible ou pas <=> présent dans une même conf ou pas
        int vol,pan,effect; // position - coefs d'écoute et de parole
    };

    // Classe encapsulant la perception qu'un canal a d'un autre canal
    class SndState
    {
    public:
        float r,a; // distance et angle
        SndTuning next,current;
    };

protected:    // données propres
    int LastConfNo;
    CObList reponses;
    CObList canaux;
    CObList confs;
    SndState ***CaStatus;
    GestionMidi *devices;
protected:    // méthodes propres
    void loadCanalList();
    Conf* FindConf(int);
    Canal* FindCanal(int);
    Conf* CreateConf();

protected:    // Pilotage du dispositif physique
    void ResetCaStatus();
    void GlobalStatus();
    void Tune();
protected:    // Traitement des commandes
    char buf[128];
    int CmdCount;
}

```

```

Decoder *Decode;
char *CmdName[NBOFCMD];
void(Controler::*CmdAct[NBOFCMD])();
void SetCmd(int,char*,void(Controler::*))();
protected: // Routines de service de traitement des commandes
int NextInt();
void StoreRep(char*,RepFlags);
void SendCanalMsg(Canal*,RepFlags);
void SendConfMsg(Conf*,RepFlags);
void SendCaInCoMsg(Conf*,Canal*,RepFlags);
void SendCaPosMsg(int,int,RepFlags);
protected: // Actions
void VerCmd();
void CmdLstCmd();
void CanalCmd();
void NameCaCmd();
void ConfCmd();
void NewCmd();
void NameCoCmd();
void DelCmd();
void InCmd();
void OutCmd();
void SplitCmd();
void JoinCmd();
void AppartCmd();
void SetCaInCoCmd();
void GetCaInCoCmd();
void SetCaPosCmd();
void GetCaPosCmd();
void GetCaHeardCmd();
void ReTuneCmd();
void ResetCmd();

// _____
public: // Interface
Controler(GestionMidi *);
~Controler();
void Traite(const char*);
int Reponse(char*, int, RepFlags&);
// _____
};

```

figure 25 : définition de la classe Controler

La première remarque concerne les méthodes publiques dont les définitions sont directement visibles sur la figure 25. Il n'y en a que deux (Traite et Reponse), si l'on excepte le constructeur et le destructeur qui gèrent respectivement l'initialisation et la libération des structures dynamiques lors de l'instanciation et de la suppression de l'objet de cette classe. On notera au passage qu'un pointeur sur l'objet Midi, instance de la classe GestionMidi, est passé en paramètre pour permettre à l'objet Ctrl de piloter le pont audio au moyen de l'interface MIDI (cf. la figure 19). Ainsi, il n'y a que deux méthodes publiques, l'une pour passer les messages à traiter, l'autre pour récupérer les éventuelles réponses. Les réponses sont de deux types : pour l'envoyeur uniquement ou pour tout le monde, ce qui est précisé dans une référence de type RepFlags passée en paramètre. Ce type est défini par :

```
enum RepFlags {ToSender,ToAll}; //pour qui la reponse ?
```

La méthode `Reponse` retourne le nombre de réponses à transmettre. La boucle de traitement qui utilise l'objet `Ctrl` est alors très simple. Quand un message est reçu au niveau de l'objet `CommServ`, le message est passé à l'objet `Ctrl` par l'intermédiaire de la méthode `Traite`, puis la méthode `Reponse` est appelée tant qu'elle ne retourne pas la valeur 0. La gestion des audioconférences par l'objet `Ctrl` est donc parfaitement encapsulée et tout passe par le langage de commande.

Avant de détailler davantage les éléments essentiels de l'objet `Ctrl`, nous décrivons les objets de base qu'il manipule.

4.1.1 Les objets de base représentant les composants atomiques d'une audioconférence

Pour gérer une configuration donnée d'audioconférences, l'objet `Ctrl` manipule essentiellement des instances des classes `Canal`, `Chaise` et `Conf`. Il y a une instance de la classe `Canal` pour chacun des canaux du pont audio, une instance de la classe `Conf` pour chaque audioconférence en cours et une instance de la classe `Chaise` pour chaque canal présent dans une conférence. La figure 26 illustre l'organisation de l'ensemble de ces instances pour une configuration possible d'audioconférences. L'objet `Ctrl` gère une liste de canaux et une liste de conférences. Chaque conférence gère une liste de chaises qui sont associées chacune à un canal particulier. Dans l'exemple de la figure 26, il y a deux audioconférences en cours ; l'une contient quatre canaux, l'autre deux canaux parmi ces quatre. La gestion des audioconférences est aussi simple que le suggère la figure 26. Les explications qui suivent donnent les détails de cette organisation.

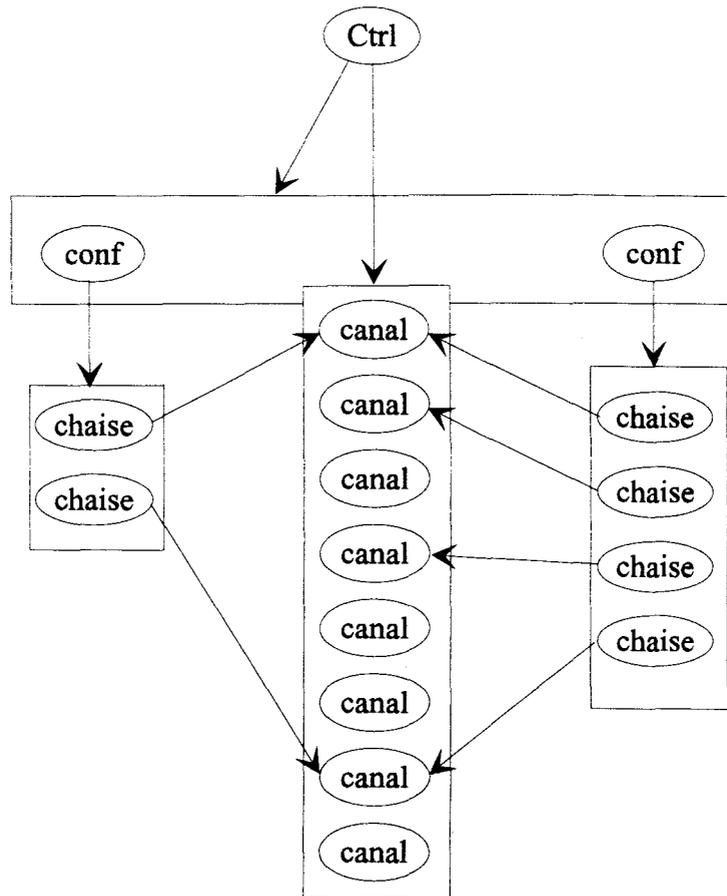


figure 26 : exemple de configuration des instances gérées par l'objet Ctrl

Un canal tel qu'il est défini dans la première partie de ce chapitre est manipulé par l'intermédiaire d'une instance de la classe Canal (figure 27).

```

//
// Classe encapsulant un canal
class Canal : public CObject
{
protected:
    int id;
    char *what,*who;
    int dmp,type;
public:
    Canal(int,char *,int,int);
    ~Canal();
    int GetId();
    char* GetWhat();
    void SetWho(char*);
    char* GetWho();
    int getDMP(){return dmp;};
    int getType(){return type;};
};
    
```

figure 27 : définition de la classe Canal

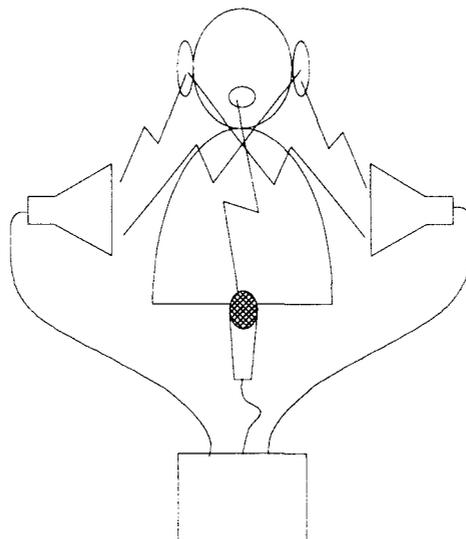


figure 28 : ce que gère un canal

Un canal représente un participant, soit un flux sonore en sortie et en entrée, et les données qui y sont associées (figure 28). Un canal est associé à une entrée et une sortie du pont audio. Cette association est faite lorsque l'agent d'ACAO est créé par la lecture d'un fichier d'initialisation, ce qui permet une reconfiguration de l'utilisation des possibilités du pont audio par simple édition d'un fichier texte (figure 29).

```
# <nom physique> <n° de DMP> <type de sortie>
# sortie n° 1 -> stéréo
# sortie n° 2 -> 1ère sortie d'effet
# sortie n° 3 -> 2ème sortie d'effet si elle existe
#
#
"Server Room" 1 1
"TRG14" 2 1
"COMULAB" 3 1
"CPAV" 4 1
"Conf Room" 2 3
"Tel. 1" 2 2
"Tel. 2" 3 2
"Tel. 3" 4 2
```

figure 29 : fichier de configuration

Les canaux sont numérotés dans l'ordre de leur définition ; ce numéro sert d'identification dans leur manipulation ultérieure et est conservé dans la variable d'instance `id` définie dans la classe `Canal`. La variable `what` conserve la description physique associée à chaque canal. La variable `who` stockera la description logique dès qu'elle sera définie ou redéfinie. La description physique est attribuée une fois pour toutes alors que la description logique sera modifiée

dynamiquement ; c'est pourquoi il n'y a pas de méthode pour modifier la variable `what`. Les variables `dmp` et `type` permettent d'identifier quelles entrées et sorties du pont audio sont associées au canal. L'agent d'ACAO prend ainsi connaissance, lors de son initialisation, du nombre de canaux qu'il aura à gérer, et de leurs caractéristiques qui ne sont pas codées d'une manière figée.

Dans notre modélisation d'audioconférences, un même canal peut faire partie de plusieurs conférences simultanément et avoir dans chacune d'entre elles des caractéristiques différentes. Ceci est réalisé par une classe d'objet intermédiaire appelée `Chaise` (figure 30).

```
//
// Classe encapsulant une chaise qui définit le statut d'un canal dans une
// conférence.
class Chaise : public CObject
{
protected:
    Canal* canal;
    float coef_e, coef_p;
public:
    Chaise(Canal*);
    Canal* GetCanal();
    void SetCoef(float, float);
    void GetCoef(float&, float&);
};
```

figure 30: définition de la classe `Chaise`

La métaphore de conception sous-jacente est qu'un canal prend place sur une chaise dans une audioconférence et que c'est cette chaise qui définit son statut dans cette audioconférence. En l'occurrence, la chaise gère les coefficients d'écoute et de parole (les variables `coef_e` et `coef_p`) qui correspondent aux α et β de la formule synthétique donnée au paragraphe 1.1.3 de ce chapitre.

Une audioconférence est alors encapsulée par une instance de la classe `Conf` (figure 31) qui gère une liste d'instances de la classe `Chaise` qui est elle-même associée à une des instances de la classe `Canal` représentant un canal.

On y retrouve les commandes de base de la gestion des canaux dans une audioconférence réalisée par les méthodes `In` et `Out`. Les commandes de manipulation des données associées à un canal sont réalisées par les méthodes `GetName`, `SetName`, `SetCoef` et `GetCoef`.

```

//
// Classe encapsuant une conférence qui est une liste de chaises
class Conf : public COBject
{
protected:
    int id;
    COBList chaises;
    char *name;
public:
    Conf(int);
    ~Conf();
    int GetId();
    char* GetName();
    void SetName(char*);
    void In(Canal*);
    void Out(Canal*);
    int GetSize();
    POSITION Find(Canal*);
    POSITION First();
    Canal* Next(POSITION&);
    void SetCoef(Canal*, float, float);
    void GetCoef(Canal*, float&, float&);
};

```

figure 31 : définition de la classe Conf

4.1.2 Les variables d'instance définies dans la classe Controller

Chaque conférence est gérée par une instance de la classe Conf. L'ensemble des audioconférences gérées par l'agent d'ACAO l'est par une liste d'instances de la classe Conf, liste stockée dans la variable `confs` définie dans la classe Controller. Cette classe définit peu d'autres données propres pour la gestion des audioconférences (figure 32).

```

... extrait de la définition de la classe Controller ...
protected:    // données propres
    int LastConfNo;
    COBList reponses;
    COBList canaux;
    COBList confs;
    SndState ***CaStatus;
    GestionMidi *devices;
...

```

figure 32 : données propres de la classe Controller pour la gestion des audioconférences

`LastConfNo` permet de retenir le numéro attribué à la dernière audioconférence instanciée, ce qui permet de les identifier avec un numéro unique (dans la limite du nombre maximal stockable dans une variable de type `int`). `canaux` retient la liste des canaux tels qu'ils ont été définis à partir du fichier de configuration lors de l'initialisation de l'agent d'ACAO. `devices` conserve un pointeur sur l'objet qui permet de piloter le pont audio par l'interface MIDI.

reponses contient la liste des réponses éventuelles à une commande. Comme nous l'avons expliqué plus haut, une liste de réponses est générée lorsqu'une commande est passée à l'objet Ctrl par l'appel de la méthode `Traite`. Elle sera accessible par des appels successifs à la méthode `Reponse`. Les réponses comportent deux éléments (figure 33) :

- une chaîne de caractères correspondant au message à transmettre,
- un type précisant si ce message est à envoyer à tous les agents correspondants ou seulement à celui qui a envoyé la commande à laquelle cette réponse correspond.

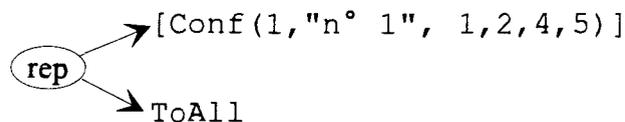


figure 33 : exemple de réponse

Ces deux informations sont réunies dans une classe définie à l'intérieur de la classe `Controler`, la classe `Rep` (figure 34).

```

... extrait de la définition de la classe Controler ...

// Classe encapsulant une réponse du contrôleur
class Rep : public CObject
{
public:
    CString rep;
    RepFlags t;
};
...
  
```

figure 34: définition de la classe Rep

Une structure aurait suffi pour rassembler ces données, mais en définissant une classe héritant de la classe de base `CObject` des MFC, il est possible de gérer facilement une liste des instances de cette classe. Nous avons d'ailleurs utilisé les mécanismes d'héritage du C++ pour réutiliser les mécanismes logiciels déjà programmés dans les MFC, ce qui a permis de se concentrer sur les problèmes spécifiques à notre problématique.

Les explications concernant la variable définie par `SndState ***CaStatus;` concernent un point primordial et beaucoup plus

complexe de l'implémentation de l'agent d'ACAO : le lien entre la gestion des audioconférences et le contrôle des flux sonores correspondants.

4.1.3 Le contrôle des flux sonores par l'objet Ctrl

Les spécifications du contrôle des flux sonores par l'objet Ctrl concernent deux aspects :

- la gestion des audioconférences en cours par l'intermédiaire d'une structure dynamique de liste de listes décrite précédemment,
- la réalisation des flux sonores correspondant à ce que représente cette structure dynamique en termes d'audioconférences.

Ces deux aspects correspondent aux deux éléments de la définition d'une audioconférence donnée au paragraphe 1.1.2. Le premier correspond à la gestion des ensembles des canaux, le second à la réalisation des systèmes d'équations qui synthétisent la description des destinations en fonction des sources. Si le premier se fait indépendamment, conférence par conférence, le deuxième se fait globalement ; il s'agit de définir les flux sonores restitués pour chaque canal en fonction des flux sonores entrant pour l'ensemble des canaux. Cette gestion est organisée par la variable `CaStatus` qui est en fait un tableau à deux dimensions d'instances de la classe `SndState` définie à l'intérieur de la classe `Controller`. `CaStatus` est défini comme un pointeur de pointeurs de pointeurs sur des instances de la classe `SndState`, car c'est un tableau à deux dimensions de pointeurs défini dynamiquement. En effet, comme le nombre de canaux est connu seulement lors de l'initialisation de l'objet Ctrl, il n'est pas possible de définir à l'avance les dimensions de ce tableau. Avant de détailler davantage l'organisation de la gestion des flux sonores par la variable `CaStatus`, nous donnons le principe de l'algorithme sous-jacent.

A chaque fois qu'une commande est susceptible d'altérer l'organisation des flux sonores correspondant à l'état des paramètres des audioconférences en cours, les réglages des mixeurs du pont audio sont recalculés. Les nouvelles valeurs sont

comparées aux valeurs courantes et des ordres sont envoyés au pont audio pour effectuer les ajustements nécessaires. Il est en effet plus facile de recalculer l'ensemble des réglages pour une situation donnée afin d'obtenir les ajustements à effectuer en les comparant aux anciens, que de calculer directement les modifications. Nous évitons aussi de faire un réglage global à chaque fois, pour limiter le contrôle du pont audio au strict minimum.

Le calcul de l'ensemble des réglages pour une configuration donnée d'audioconférences est réalisé par la méthode privée `GlobalStatus` qui modifie le tableau `CaStatus`. Une simple lecture de ce tableau suffit ensuite pour piloter le pont audio (figure 35).

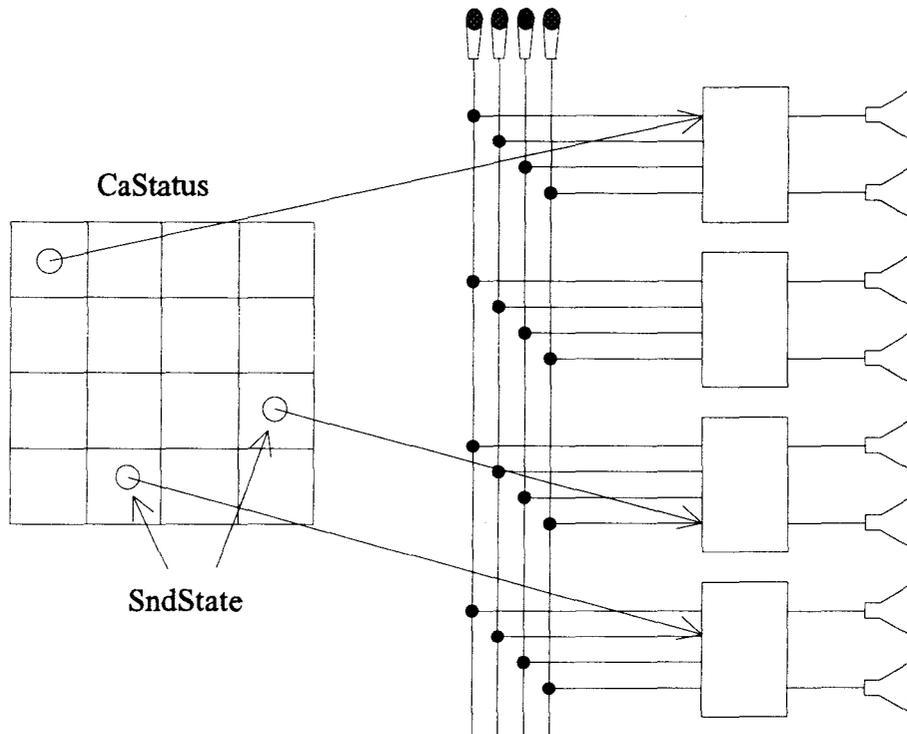


figure 35 : principe du contrôle des flux sonores

Un mixeur est affecté à chaque canal. Pour chacun de ces mixeurs, un réglage particulier est à effectuer pour chaque canal. Ce réglage particulier dépend de la configuration des audioconférences en cours et de la position de spatialisation à réaliser pour chaque canal. Nous rappelons que chaque participant a sa propre vision de la position des autres participants par rapport à lui-même. `CaStatus`

est donc un tableau à double entrée, dont chaque case réunit les paramètres de spatialisation et les deux jeux de réglages correspondant l'un à la situation courante, l'autre à la nouvelle situation lors d'un changement. D'où la définition suivante des classes encapsulant ces données (figure 36).

```

... extrait de la définition de la classe Controller ...
// Classe encapsulant les réglages de perception d'un canal pour un autre
canal
class SndTuning
{
public:
    int state; // audible ou pas <=> présent dans une même conf ou pas
    int vol,pan,effect; // position - coefs d'écoute et de parole
};

// Classe encapsulant la perception qu'un canal a d'un autre canal
class SndState
{
public:
    float r,a; // distance et angle
    SndTuning next,current;
};
...
    
```

figure 36 : définitions des classes SndTuning et SndState

La classe SndTuning reflète les paramètres contrôlés pour chaque voie au niveau des mixeurs :

- l'état on/off de la voie (variable `state`),
- le volume d'entrée de la voie (variable `vol`),
- le réglage de la balance (variable `pan`),
- le volume d'un effet de réverbération utilisé pour restituer une notion de distance (variable `effect`).

Les variables `r` et `a` définies dans la classe SndState précisent la position de la source sonore ; les variables `next` et `current` stockent les jeux de réglages.

Il ne nous reste plus qu'un aspect de la définition de la classe Controller à expliquer : l'interprétation des commandes.

4.1.4 La gestion du langage de commande par l'objet Ctrl

A chaque commande correspond une méthode définie dans la classe `Controler` ainsi qu'une chaîne de caractères ASCII utilisée dans les messages de communication avec l'agent d'ACAO. Les méthodes sont regroupées sous le commentaire `// Actions` dans la définition de la classe `Controler` (figure 37).

```
... extrait de la définition de la classe Controler ...
protected: // Actions
    void VerCmd();
    void CmdLstCmd();
    void CanalCmd();
    void NameCaCmd();
    void ConfCmd();
    void NewCmd();
    void NameCoCmd();
    void DelCmd();
    void InCmd();
    void OutCmd();
    void SplitCmd();
    void JoinCmd();
    void AppartCmd();
    void SetCaInCoCmd();
    void GetCaInCoCmd();
    void SetCaPosCmd();
    void GetCaPosCmd();
    void GetCaHeardCmd();
    void ReTuneCmd();
    void ResetCmd();
...

```

figure 37 : méthodes de traitement des commandes

Le lien entre ces méthodes et les chaînes ASCII correspondantes est réalisé par deux tableaux appariés, l'un des chaînes de caractères (`CmdName`), l'autre des pointeurs sur ces méthodes (`CmdAct`). Ces deux tableaux sont remplis automatiquement lors de l'initialisation au moyen de la méthode `SetCmd`. Voici les définitions de ces éléments de la classe `Controler` (figure 38).

```
... extrait de la définition de la classe Controler ...
    char *CmdName[NBOFCMD];
    void (Controler::*CmdAct[NBOFCMD])();
    void SetCmd(int,char*,void(Controler::*)());
...

```

figure 38 : les variables utilisées pour l'interprétation des commandes

Les conséquences d'une modification ou d'une extension du langage de commande sont ainsi particulièrement limitées, car les modifications impliquées au niveau du programme sont simples et très ponctuelles. Pour modifier une commande, il suffit de modifier la méthode correspondante. Pour rajouter une

nouvelle commande il suffit d'ajouter une nouvelle méthode et un appel à la méthode `SetCmd` dans le constructeur de la classe `Controler`.

Le traitement des commandes est également simple et ne nécessite aucune adaptation si le langage de commande est modifié (figure 39).

```

... extrait de la définition des méthodes la classe Controler ...
//
// Traitement des messages : fonction principale
void Controler::Traite(const char* action)
{
    int c;
    Decode->SetCmdString(action);
    while(c= Decode->NextCmd())
        (this->*CmdAct[c-1]) ();
}
...

```

figure 39 : traitement des commandes

`action` est un pointeur sur la chaîne de caractères ASCII correspondant à une commande (ou plusieurs mises l'une à la suite de l'autre) à traiter selon la syntaxe définie au paragraphe **Erreur! Source du renvoi introuvable.** Cette chaîne est passée à l'objet `Decode` qui en effectue l'analyse. Les appels successifs à la méthode `NextCmd` pour l'objet `Decode` retournent l'index de la première commande décodée ou sinon retourne la valeur 0. Cet index est utilisé pour appeler la méthode appropriée par son pointeur stocké dans le tableau `CmdAct` de l'objet `Ctrl`. La méthode appelée accède aux paramètres éventuels de la commande qu'elle doit traiter par des appels successifs à la méthode `NextParam` pour l'objet `Decode`.

4.2 Les principales modifications nécessaires en vue d'une implémentation plus généraliste

Les principales modifications portent sur les spécificités de notre réalisation : la gestion des commandes et la gestion des flux sonores. Pour donner un cadre plus précis à ces indications, nous supposons une adaptation de l'agent d'ACAO , d'une part sur une plate-forme dont le système gère des objets distribués et les mécanismes de communication entre ces objets, et d'autre part une gestion des flux sonores tout numérique.

4.2.1 Adaptation de la gestion des commandes

Il va de soi que la gestion des communications par une encapsulation artisanale des sockets dans des objets C++ est à remplacer par le mécanisme de communication normalisé du système utilisé. Il suffit pour cela d'intervenir uniquement au niveau de l'objet qui gère la communication dont on conserverait les méthodes publiques d'envoi et de réception des messages.

L'utilisation d'un codage des commandes sous forme de chaîne de caractères ASCII dans le cadre d'une maquette présente certains avantages :

- les commandes sont lisibles directement lors de l'analyse du fonctionnement du programme,
- le langage de commande est directement manipulable au niveau d'une interface en mode texte,
- le langage de commande est facile à adapter,
- il n'y a pas de problème de conversion lors des communications sur le réseau,
- le codage est valable sur toutes les plates-formes.

Cependant, ce type de codage n'est pas dénué d'inconvénients dans la perspective d'une application réelle. Nous en donnons ici une liste non exhaustive :

- il est très peu économique en termes de volume d'informations échangées,
- il nécessite un décodage laborieux en terme de traitement,
- il n'est pas normalisé.

Les modifications nécessaires à l'adaptation d'un langage de commande défini et codé de manière plus rigoureuse se limitent à l'objet Decode. Celui-ci devra simplement continuer à traduire les commandes reçues sous une forme mieux appropriée et à retourner un numéro correspondant à la commande à traiter. Ce numéro doit correspondre à l'indice de la méthode à appeler pour traiter cette commande.

4.2.2 *Adaptation de la gestion des flux sonores*

L'adaptation à une gestion des flux sonores entièrement numériques est un peu moins simple que celle de la gestion du langage de commande : elle nécessite des modifications moins localisées et une interface (inter composant logiciels) convenablement définie.

Nous commencerons à décrire les caractéristiques requises de l'interface du système de gestion de flux sonores numériques. Il est important de continuer à bien séparer la gestion des audioconférences de la manipulation numérique des flux sonores. Dans tous les cas, le système de gestion des flux sonores, à concevoir ou déjà conçu, doit être vu comme un ensemble de mixeurs ayant des caractéristiques homogènes pour être interconnecté avec l'agent d'ACAO.

Chaque mixeur doit accepter en entrée une copie de chacun des flux sonores traités par le système et fournir en sortie un mixage spécifique et contrôlé de ces entrées. Chaque mixeur sera manipulé par l'agent d'ACAO à travers un objet C++ dont les méthodes publiques doivent permettre de régler les paramètres de mixage pour chacune des entrées.

Chaque canal géré par l'agent d'ACAO doit être associé à l'un de ces mixeurs. En ce qui concerne les objets qui gèrent les canaux, il est nécessaire :

- de remplacer les variables `dmp` et `type` définies dans la classe `Canal` qui identifient actuellement quel mixeur est associé à un canal,
- d'adapter le constructeur de la classe `Controller` pour créer les instances de la classe `Canal` à partir du fichier de configuration,
- d'adapter le fichier de configuration pour définir les associations entre canaux et mixeurs.

Au niveau de la classe `Controller`, il faudra :

- éventuellement adapter la définition de la classe `SndTuning` si les mixeurs ont un jeu de paramètres différents,

- modifier la méthode `GlobalStatus` de la classe `Controler` pour le pilotage des mixeurs numériques (simple appel de méthode pour les objets encapsulant les mixeurs),
- supprimer ce qui concerne l'instance de la classe `GestionMidi` permettant le pilotage du pont par l'interface MIDI.

4.3 Conclusion sur la généralisation de l'implémentation

L'agent d'ACAO est fortement structuré, non seulement sous forme d'un ensemble d'objets se partageant les traitements, mais aussi au sein de chacun de ces objets, et en particulier l'objet `Ctrl` dont nous avons détaillé l'organisation. Nous avons principalement exploité les possibilités d'encapsulation du C++ pour bien répartir et cerner les traitements en unités logiques et autonomes. Nous avons aussi exploité les autres outils du génie logiciel tels que le recours à des définitions locales, la répartition de la programmation dans différents modules et l'utilisation de bibliothèques. Le problème de l'implémentation dans son ensemble a été résolu avec le souci constant de bien répartir les entités logiques qui le constituent. Ces entités logiques ont été définies en accord avec la modélisation présentée au début de ce chapitre.

Il en résulte une implémentation largement réutilisable où les modifications sont bien localisées.

Chapitre 4 : Evaluations des services proposés

" ... it also behooves us to do some formal validation studies of the systems we develop in the context of real-world tasks. While we may be personally convinced of their efficacy, we need to start providing concrete evidence of when and how sonification can help users get the job done."

Elisabeth M. Wenzel.

1 Introduction

Dès le début de ce travail, nous avons souhaité confronter à la réalité nos réflexions théoriques et nos propositions pratiques. Nous avons été confortés dans cette idée par l'opinion exprimée dans la citation ci-dessus et la présentation d'une étude des différentes situations de communication dans le cadre d'un travail coopératif²⁸.

La citation précédente tirée de la conclusion de [WENZEL 94] se rapporte à la sonification. Cependant, nous pensons qu'elle doit être appliquée à bien d'autres domaines.

²⁸ présenté par Sylvie Molenda au séminaire Ganymède du 10/3/95, travaux en cours de publication

Nous résumons rapidement ici la présentation de Sylvie Molenda. Quatre situations de communication dans le cadre d'un travail coopératif ont été comparées lors d'une expérience :

- quand les personnes sont ensemble, se voient, et voient l'objet partagé de leur travail,
- quand elles ne voient que leurs visages,
- quand elles ne voient que l'objet sur lequel elles coopèrent,
- quand elles ne peuvent que s'entendre.

L'expérience a montré que la situation qui gêne le plus la tâche à accomplir est celle où les personnes ne voient que leurs visages. Or c'est justement ce mode qui a été choisi dans les collecticiels synchrones distribués, ce qui suggère une réflexion insuffisante sur la pertinence des choix technologiques par rapport aux usages envisagés. Il est vrai que l'utilité de voir les visages des personnes avec qui l'on coopère dépend de l'activité. Elle se fera davantage ressentir pour une discussion pure que lorsqu'il existe un objet central pour la tâche à accomplir. Roos [ROOS 95] critique cependant la pertinence de la possibilité de voir les visages des autres en toutes circonstances et renvoie aux travaux d'Harper [HARPER 94] et Pagani [PAGANI 93] pour étayer son argumentation.

Quoiqu'il en soit, il apparaît très souhaitable de chercher à vérifier l'intérêt des services d'un système par son évaluation. Le domaine de recherche de l'évaluation est si vaste que nous ne pourrions que l'évoquer ici.

1.1 La problématique de l'évaluation

Nous avons abordé la problématique de l'évaluation à l'occasion de l'atelier organisé sur ce thème lors d'IHM'94. Nous avons ainsi un peu tardivement découvert l'ampleur du problème et quelques uns de ses aspects essentiels parmi lesquels nous présenterons ceux qui pourront aider à mieux positionner notre travail.

Une remarque préliminaire et incontournable concerne le champ d'application de l'évaluation qui est essentiellement le domaine des Interfaces Homme-Machine dans un contexte mono-utilisateur. Le problème de l'évaluation d'une interface classique, par lui même très consistant, a suscité une liste impressionnante de méthodes d'évaluation. Zorola [ZOROLA 95] compare pas moins de vingt méthodes d'évaluation par rapport aux attributs d'utilisabilité²⁹. Il n'en va pas de même pour le problème de l'évaluation des interfaces multi-utilisateurs comme le soulignent d'ailleurs aussi bien Zorola [ZOROLA 95] que Roos [ROOS 95], tous deux à l'origine de méthodes d'évaluation d'interfaces de collecticiels. A cela nous rajoutons le point vue particulier de notre centre d'intérêt : l'aspect sonore n'est que partiellement abordé. Pourtant dès 1986, le son est pris en compte dans l'évaluation, aussi bien par Scapin [SCAPIN 86] que par Smith et Mosier [SMITH 86], et fait l'objet de recommandations en tant que procédé de rétroaction, notification ou sonification.

L'aspect de la communication verbale n'est qu'évoqué par Roos. Zorola s'appuie sur la théorie des actes de langage d'Austin et Searle et la théorie cognitive de la communication de Sperber et Wilson pour formaliser les actes de communication. Il en tire deux règles heuristiques traduisant les actes de communication par la parole. Ces règles seront utilisées lors d'une simulation dans le cadre d'une évaluation prédictive de l'utilisabilité d'une interface multi-utilisateurs, pour compléter la base des connaissances et croyances mutuelles qui sont au centre de la méthode MESC proposée par Zorola. Nous renvoyons au travail de Zorola [ZOROLA 95] pour les explications concernant la méthode MESC. Toutefois nous tenterons de préciser ici ce que l'on entend, entre autres, par "utilisabilité" et "évaluation prédictive".

1.2 Les différents niveaux d'une évaluation

Il ne semble pas y avoir une définition exclusive de l'évaluation. Dans un premier temps, Senach [SENACH 90] définit l'évaluation comme une

²⁹ le terme utilisabilité est expliqué plus loin

comparaison de l'objet évalué à un modèle de référence. Puis il la précise en indiquant plusieurs objectifs possibles d'évaluation parmi lesquels il considère plus particulièrement :

- l'analyse de l'utilisation d'un dispositif en situation de travail,
- la sélection d'alternatives de conception,
- la détection et la correction des défauts d'une IHM,
- l'évaluation comparative de différents logiciels.

A la lumière de ces objectifs, il apparaît que l'évaluation consiste d'une part à vérifier si l'ensemble des propriétés et des caractéristiques d'un système donne satisfaction, et d'autre part, à aider le choix d'une autre alternative de conception du système.

Par contre, il est clair que l'évaluation se réalise à deux niveaux. Senach distingue l'évaluation de l'*utilité* de celle de l'*utilisabilité*. L'utilité concerne les aspects de capacité fonctionnelle, de performance du système et de qualité d'assistance, tandis que l'utilisabilité concerne les aspects de facilité d'apprentissage, facilité d'utilisation et qualité de documentation. Ces deux niveaux semblent interdépendants. Senach [SENACH 90] précise ainsi leurs définitions :

"La première dimension [l'utilité] détermine si le produit permet à l'utilisateur d'atteindre ses objectifs de travail.

(...)

L'utilisabilité sert ainsi à poser la frontière entre utilité potentielle et utilité réelle."

Nous en retenons que l'utilité concerne la finalité d'un système et que l'utilisabilité mesure à quel point cette finalité est accessible.

1.3 Les différentes méthodes d'évaluation

Sans entrer dans une description détaillée des nombreuses méthodes d'évaluation, nous nous bornerons ici à en préciser les deux grandes classes qui font l'objet d'un consensus chez les différents auteurs mais sous des noms différents. Zorola distingue les méthodes empiriques ou non empiriques et fait un parallèle d'une part avec la classification de Senach selon une approche empirique ou analytique, d'autre part avec celle de Coutaz qui parle de méthodes expérimentales ou prédictives. Ces deux classes distinguent en fait les évaluations faites avec ou sans utilisateurs réels.

L'intérêt de pratiquer une évaluation avec des utilisateurs réels est clairement souligné, mais les difficultés qu'elle comporte ont fortement motivé les approches prédictives (non empiriques ou analytiques). En effet, le premier type d'évaluation donne des informations plus précises et proches de la réalité, mais au prix d'une réalisation plus lourde et plus coûteuse (au moins celle d'un prototype). Aussi a-t-on cherché à intervenir plus en amont, dès les premières phases de conception. De plus, l'implication d'utilisateurs réels n'est pas dénuée de problème tout comme l'interprétation des résultats ainsi obtenus.

Cependant, le choix d'une approche analytique, qu'elle utilise une méthode d'inspection de l'utilisabilité ou une méthode théorique formelle, sous-entend l'existence d'une théorie cognitive, d'une modélisation de l'interface, ou d'une base d'heuristiques ergonomiques. C'est-à-dire que l'approche analytique nécessite une bonne connaissance du domaine du système évalué. Ceci pose un problème dans le cas des collecticiels, et encore plus pour l'aspect sonore des collecticiels.

1.4 Présentation des évaluations réalisées

Les informations précédentes vont nous permettre de situer notre approche de l'évaluation par rapport à notre problématique. En fait, l'étude de l'évaluation des services que nous avons proposés a été très bénéfique bien au delà de l'évaluation proprement dite.

Tout d'abord, l'étude des deux dimensions de l'évaluation, l'utilité et l'utilisabilité, nous a fait prendre conscience des efforts de recherche qui restent à accomplir avant de pouvoir proposer les services envisagés dans cette thèse dans le cadre d'une utilisation réelle. En effet, si l'interface de démonstration présentée au chapitre précédent nous a permis de vérifier que l'on pouvait aisément définir toutes les configurations et modifications dynamiques envisagées, elle est impropre à une utilisation effective. Nous reprenons par exemple le mécanisme d'aparté. Tout est là pour le mettre en place et pour le configurer, mais rien n'est prévu pour le négocier. Cette négociation ne peut pas, de toute façon, se faire par le canal sonore. Elle dépend donc du reste de l'interface du système. En d'autres termes l'utilisabilité de ce service d'aparté est, pour le moment, inexistante. Il reste à définir au niveau des Interfaces Hommes-Machines, et il y en a plusieurs, car nous sommes dans un contexte multi-utilisateurs : les mécanismes d'accès aux services fournis par l'agent d'audioconférence.

Par ailleurs, nous proposons des mécanismes nouveaux dépendant de techniques dont la mise au point reste encore inachevée tels que la spatialisation du son. Ce qui nous amène à suggérer qu'il existe peut-être une troisième dimension d'évaluation à ajouter à celle de l'utilité et de l'utilisabilité, celle concernant le bon fonctionnement d'un service. En effet, avant d'évaluer l'utilisabilité et l'utilité, par exemple, de la spatialisation du son pour les audioconférences, il convient de s'assurer du bon fonctionnement du service sous-jacent : la spatialisation du son est-elle correctement perçue? Ce type de question correspond, pour le canal visuel, à la qualité de lisibilité d'un texte, de définition d'une image, de la précision et de la variété des couleurs; ces paramètres-là sont déjà bien maîtrisés.

Nous n'avons donc pas pu évaluer l'utilité des services que nous avons proposés, mais nous avons réalisé deux évaluations au sujet de la pertinence de l'utilisation d'un système stéréophonique pour une audioconférence. Nous avons utilisé une méthode d'évaluation empirique dans les deux cas en utilisant des variables dépendantes classiques : le taux d'erreur et la durée d'exécution d'une

tâche comme mesure de performance et le confort d'usage comme variable subjective. Ces précisions se rapportent à la taxinomie décrite dans [SENACH 90].

La première expérimentation a consisté à vérifier qu'un système stéréophonique permettait bien de distinguer plusieurs rendus sonores différents. La deuxième expérimentation, nettement plus complexe, a consisté à évaluer l'utilité d'un tel système pour une audioconférence. A l'occasion de cette deuxième expérience, nous avons été amené à définir une activité spécifique qui pourra être réutilisée pour des expériences similaires. Les deux expérimentations ont fait l'objet d'une étude statistique des résultats.

2 Potentiel d'un système stéréophonique

2.1 Présentation

2.1.1 Objectif de l'expérimentation

Lorsqu'un groupe de personnes se parlent sans se voir (l'audio-réunion en est l'exemple typique), l'absence d'information sur la disposition des interlocuteurs et la difficulté d'identification d'un locuteur se font ressentir. Une solution envisagée ici est de "répartir" les interlocuteurs dans un système audio stéréophonique en jouant simplement sur les volumes sonores des voies de gauche et de droite. Cette expérimentation a pour but d'évaluer le potentiel d'un tel procédé.

Nous avons fait tester cette solution par 20 sujets à qui toute une série de "positionnements" sonores ont été présentés et qui ont précisé d'où leur semblait provenir le son qu'ils percevaient ainsi.

L'ensemble des tests a été informatisé dans la mesure du possible : l'expérimentation par chaque personne est contrôlée par un programme qui déroule la séquence des tests et enregistre automatiquement les réponses dans un fichier. Ces fichiers seront ensuite exploités statistiquement pour sanctionner la validité de ce système de rendu de position.

2.1.2 Méthode de "positionnement"

Le principe est le suivant : un son supposé être le plus à droite possible dans les limites du système est rendu avec l'enceinte de droite uniquement. Un son supposé être à l'extrême gauche est rendu avec l'enceinte de gauche uniquement. Un son supposé provenir d'une position intermédiaire est rendu en partie par l'enceinte de gauche et en partie par l'enceinte de droite.

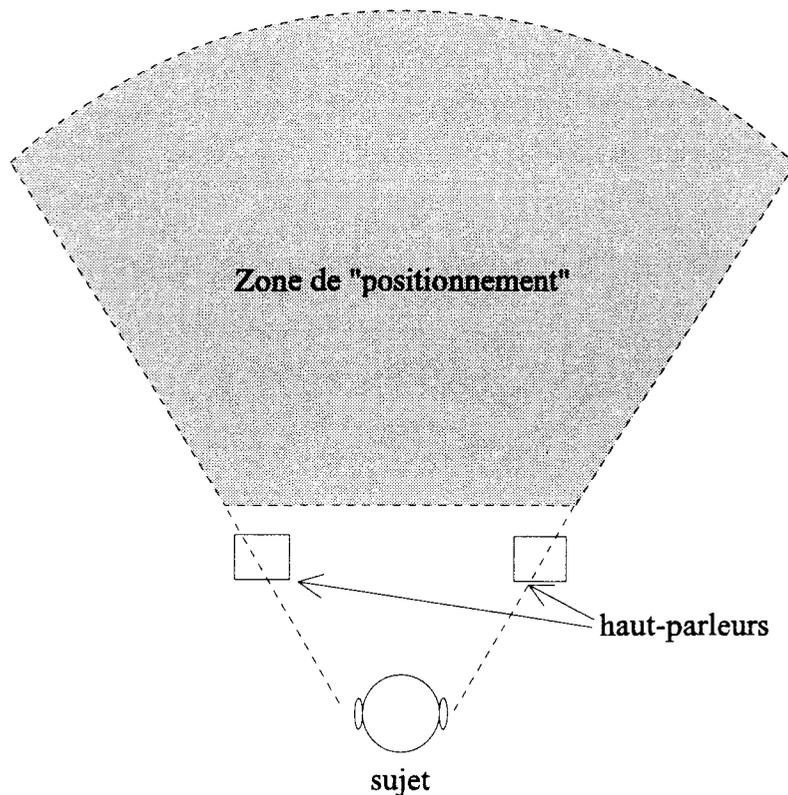


figure 40 : positionnement en stéréo

L'approche proposée est purement empirique. Nous ignorons la relation qui pourrait exister entre cette restitution stéréophonique et une situation sonore réelle. Cette question relève d'une démarche inverse de celles explorées pour le problème de la spatialisation du son (voir au chapitre 1, le paragraphe 2.4) qui partent de la description des positions des sources sonores pour en déduire la restitution sonore correspondante. Ici, nous partons d'une restitution sonore supposée représenter la position spatiale d'une source sonore mais nous ne chercherons pas à établir la relation précise qui existe entre les deux. C'est

pourquoi il conviendrait mieux de parler de rendu sonore différent plutôt que de positionnement. Néanmoins, ce rendu sonore s'appuie sur une métaphore de positionnement spatial en deux dimensions. En jouant sur le volume des deux enceintes, plus ou moins fort et plus ou moins à droite ou à gauche, on croit percevoir le son à une position fictive dans une zone représentée dans la figure 40. L'objet de l'expérience n'est pas de mesurer la corrélation entre la position rendue et perçue mais de vérifier que l'on peut bien distinguer plusieurs rendus sonores.

2.1.3 Précision sur la méthode de positionnement

Le matériel employé (cf. plus loin) nous permet d'automatiser au mieux l'expérience par un contrôle informatisé. Cependant, il induit un flou important sur la maîtrise précise des volumes des sons produits finalement par les enceintes. Les possibilités de contrôle se limitent à un paramètre par voie compris entre 0 et 255. Le volume d'une voie est proportionnel à ce paramètre selon une fonction dont nous ignorons les caractéristiques.

Nous avons donc choisi arbitrairement d'expérimenter trois fonctions simples qui régissent les valeurs de ces paramètres pour les voies droite et gauche représentées par les graphes des figures suivantes dans le cas où 5 positions différentes (A, B, C, D et E) sont rendues.

2.1.4 Procédé 1

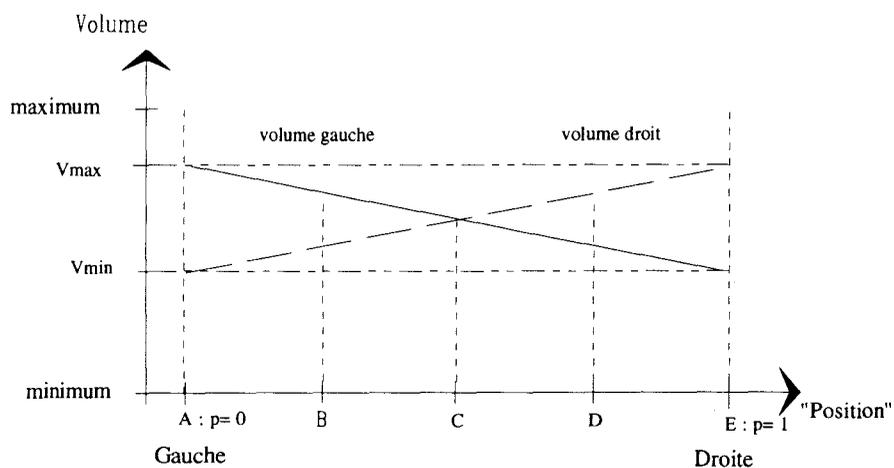


figure 41 : procédé de "positionnement" linéaire

On a, pour le premier procédé de réglage, une répartition des volumes gauche (V_g) et droite (V_d) définie par des fonctions linéaires d'équations :

$$V_g = -(V_{max} - V_{min})p + V_{max},$$

$$V_d = (V_{max} - V_{min})p + V_{min},$$

2.1.5 Procédé 2

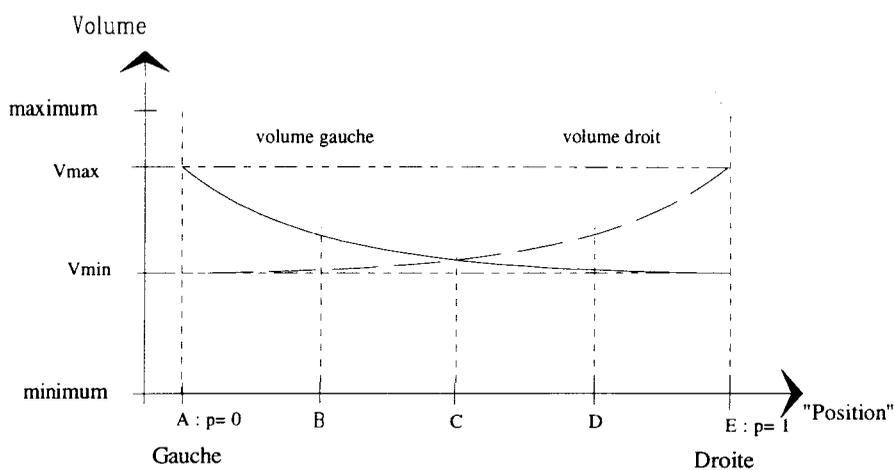


figure 42 : procédé de "positionnement" parabolique

On a, pour le deuxième procédé de réglage, une répartition des volumes définie par des fonctions paraboliques d'équations :

$$V_g = (V_{max} - V_{min})p^2 - 2(V_{max} - V_{min})p + V_{max}$$

$$V_d = (V_{max} - V_{min})p^2 + V_{min}.$$

2.1.6 Procédé 3

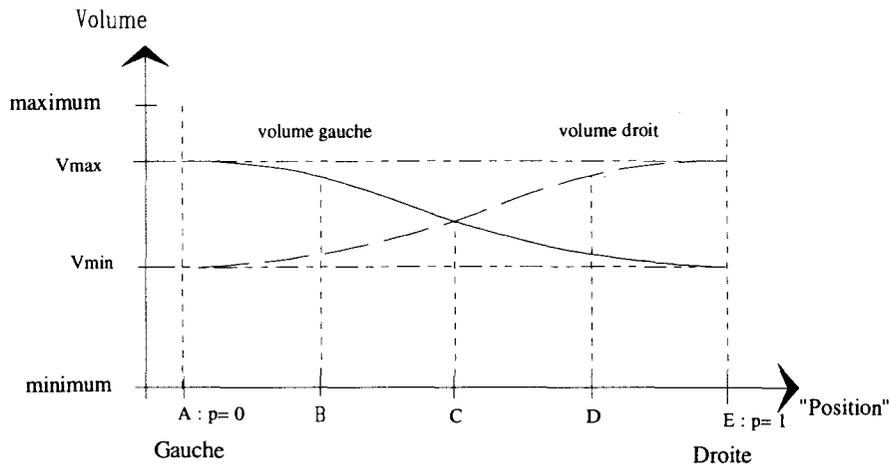


figure 43 : procédé de "positionnement" cubique

On a, pour le troisième procédé de réglage, une répartition des volumes définie par des fonctions cubiques d'équations :

$$Vg = 2(Vmax - Vmin)p^3 - 3(Vmax - Vmin)p^2 + Vmax$$

$$Vd = -2(Vmax - Vmin)p^3 + 3(Vmax - Vmin)p^2 + Vmin.$$

2.2 Expérience

2.2.1 Sujets

L'expérience a été effectuée par 20 adultes, hommes et femmes âgés de 20 à 30 ans pour la plupart, ce qui suppose des capacités auditives moyennes.

2.2.2 Matériel

Chaque sujet a participé à l'expérience sous le contrôle d'un PC sous Windows équipé d'une carte stéréo (une Audio Spectrum 16, mais n'importe quelle autre carte compatible avec Windows aurait pu être utilisée). La source sonore utilisée était produite par une radio captant "France Info" en monophonie, c'est-à-dire essentiellement de la parole. Deux petites enceintes amplifiées restituent ce son. Situées de part et d'autre de l'écran, elles offrent un angle d'écoute d'environ 60°.

2.2.3 Logiciels

Les programmes utilisés ont été développés spécialement pour l'expérience. Une première application (figure 44) permet de fixer les paramètres de déroulement du test. Cette application a été utilisée avant chaque expérience pour faire varier le type de procédé de réglage employé (initialement dénommé modèle), les autres paramètres ne variant pas. Comme on peut le voir sur la figure 44, il avait été prévu de faire des essais avec un casque. Cependant ce type de sortie a été abandonné car il ne correspond pas au confort d'écoute que l'on juge nécessaire dans le cadre d'un collecticiel.

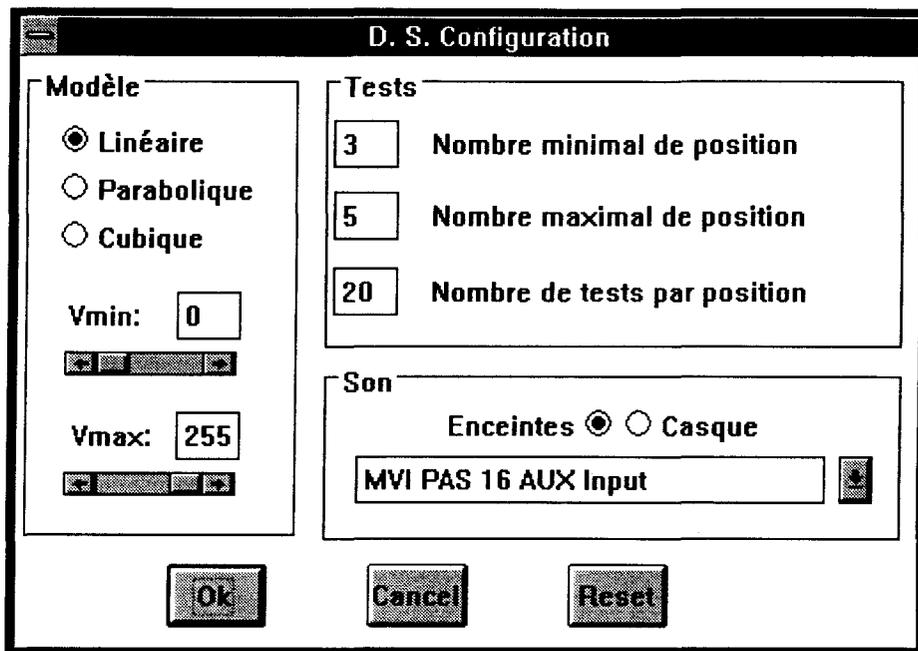


figure 44 : interface de configuration de l'expérience

Cette application enregistre les paramètres dans un fichier d'initialisation. Ce fichier est utilisé par l'application qui contrôle le déroulement de l'expérience. La figure 45 montre l'interface de l'application quand elle est lancée. L'utilisateur est d'abord invité à entrer son nom ; les instructions à suivre sont affichées en dessous. Tout en bas, on aperçoit les 3 boutons qu'il utilisera pour donner ses réponses dans la première phase du test à trois positions (cf. plus loin). Cette interface évolue dynamiquement pour afficher les instructions à suivre au fur et à

mesure, et pour présenter un nombre de boutons correspondant au nombre de "positions" sonores rendues. Les boutons sont centrés et répartis sur toute la largeur de l'interface.

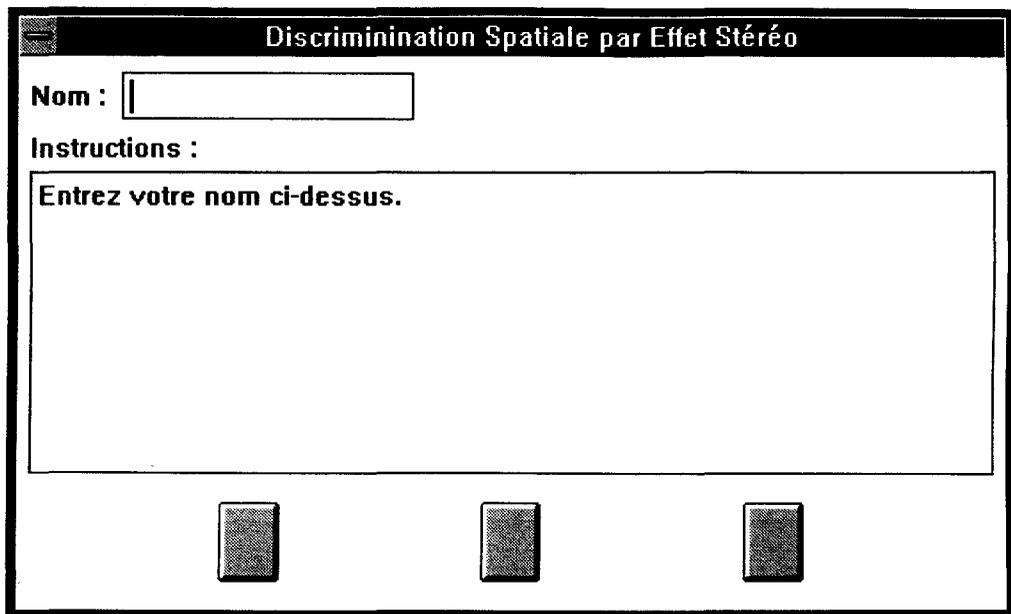


figure 45 : état initial de l'interface de l'expérience

Le nom de l'utilisateur (ou son pseudonyme, suivant ce qu'il décide de donner) est utilisé pour le nom du fichier dans lequel sont enregistrés les paramètres de l'expérience et ses résultats. Bien sûr, un mécanisme assure que tous les fichiers aient un nom différent, même si plusieurs utilisateurs donnent un nom identique. Ce système d'applications simple et robuste a permis d'assurer un déroulement parfaitement homogène de toutes les expériences, et un enregistrement systématique et sans erreur des mesures. Bien entendu nous n'avons pu éliminer les erreurs de mesure dues aux erreurs de manipulation de l'utilisateur.

2.2.4 Procédure

L'expérience s'est déroulée dans des conditions normales de travail (pas de chambre anéchoïque ...) à un bureau devant l'ordinateur pilotant l'expérience. Ces conditions correspondent parfaitement à celles de l'utilisation d'un collecticiel.

Chacun des sujets a effectué l'expérience pour chacun des trois procédés de réglage avec une période de repos entre deux. Les 6 ordres possibles selon lesquels les procédés pouvaient être appliqués ont été répartis uniformément dans l'ensemble des sujets afin que l'effet d'apprentissage n'interfère pas.

Une expérience consiste à tester un procédé de réglage avec successivement 3, 4 et 5 positions possibles. A chaque position correspond un rendu sonore différent, censé permettre d'identifier une origine différente du son rendu. Les tests à 3 et 4 positions ont servi à absorber l'effet d'apprentissage. Le seul test qui compte est celui à 5 positions pour lequel le sujet était bien "entraîné" au déroulement de l'expérience, ce qui permet de bien mesurer son aptitude à distinguer différents rendus sonores et non à effectuer le test. 5 positions correspondent à 6 participants potentiels. Cette limite à 6 n'est pas totalement arbitraire ; Grudin [GRUDIN 93] renvoie à [NUNAMAKER 91] pour expliquer qu'à partir de 7 personnes, les mécanismes d'une réunion s'altèrent.

Pour chaque ensemble de n positions possibles, pour "montrer" dans un premier temps les différents rendus sonores possibles, le son est d'abord positionné de gauche à droite, puis de droite à gauche, le sujet devant à chaque fois appuyer sur le bouton correspondant (qui lui est indiqué dans la fenêtre des instructions) pour passer à la position suivante.

Ensuite $20n$ tests élémentaires sont effectués. Chaque test élémentaire consiste à démarrer un chronomètre au moment où une nouvelle position est rendue et à attendre la réponse de l'utilisateur, ce qui arrête le chronomètre. La position rendue, la réponse de l'utilisateur et son temps de réponse sont alors automatiquement enregistrés alors qu'une nouvelle position est rendue et ainsi de suite. Pour chaque test élémentaire la position est choisie au hasard, mais de telle manière qu'une même position ne soit pas proposée plus de deux fois de suite et que, globalement, chaque position ait été rendue 20 fois. Pour ne pas fausser les résultats, aucune information n'est donnée au sujet quant à l'exactitude de sa réponse, car il s'agit de vérifier si le rendu sonore permet d'identifier une position

particulière et non pas d'apprendre au sujet à associer un rendu sonore particulier à une position donnée.

2.3 Résultats - Discussion

Nous avons ainsi recueilli dans 60 fichiers les résultats de 14 400 tests élémentaires. Les tests à 3 et 4 positions se sont révélés trop faciles. Les résultats retenus ne concernent que les tests à cinq positions. Ce nombre est estimé être la limite pour une discussion efficace. Ces 60 fichiers ont été balayés automatiquement par un programme écrit à cet effet pour fournir les moyennes qui suivent et montrent le bilan de cette expérience.

Nous présentons ici pour chacun des procédés de réglage et pour chaque position rendue les mesures suivantes :

- le pourcentage des réponses rigoureusement exactes, c'est-à-dire quand la position rendue et la position perçue coïncident,
- la moyenne des écarts entre la position perçue et la position rendue,
- le temps moyen de décision des sujets, c'est-à-dire le temps écoulé entre le rendu d'une position et le clic de la souris sur le bouton choisi par le sujet pour la position perçue ; ce temps inclut donc la perception du sujet, son temps de décision, et la coordination de son action pour cliquer sur le bon bouton.

Pour les positions, nous avons pris, de gauche à droite, les abréviations suivantes :

- G : gauche,
- IG : intermédiaire gauche,
- C : centre,
- ID : intermédiaire droite,
- D : droite.

Procédé	G	IG	C	ID	D
1	98.8	97.5	90	94.8	100
2	99.5	98.8	95.3	98.0	99.5
3	98.5	97.0	91.3	94.3	99.8

tableau 5 : pourcentages de réponses exactes

Il s'agit de moyennes, par position et par procédé de réglage pour l'ensemble des sujets. Les positions extrêmes sont mieux identifiées que la position centrale. Le deuxième procédé de réglage donne des résultats légèrement meilleurs que les autres. Cependant le volume sonore des positions centrales est nettement moins fort que pour les positions latérales. S'il facilite l'identification, serait-ce acceptable pour une audioconférence ?

Nous avons choisi le pourcentage de réponses correctes comme résultat significatif de notre expérience. La figure 46 montre les moyennes des pourcentages des réponses exactes, position par position, des trois procédés de réglage, comparées avec le résultat théorique correspondant à des réponses aléatoires des sujets (s'ils étaient incapables de discerner la moindre différence). Dans ce cas, la probabilité maximale d'avoir k% de réponses exactes est de 10% seulement et est atteinte pour k=20. Il est donc indéniable que les sujets ont parfaitement été capables d'identifier les positions rendues : 90% au moins de leurs réponses sont exactes.

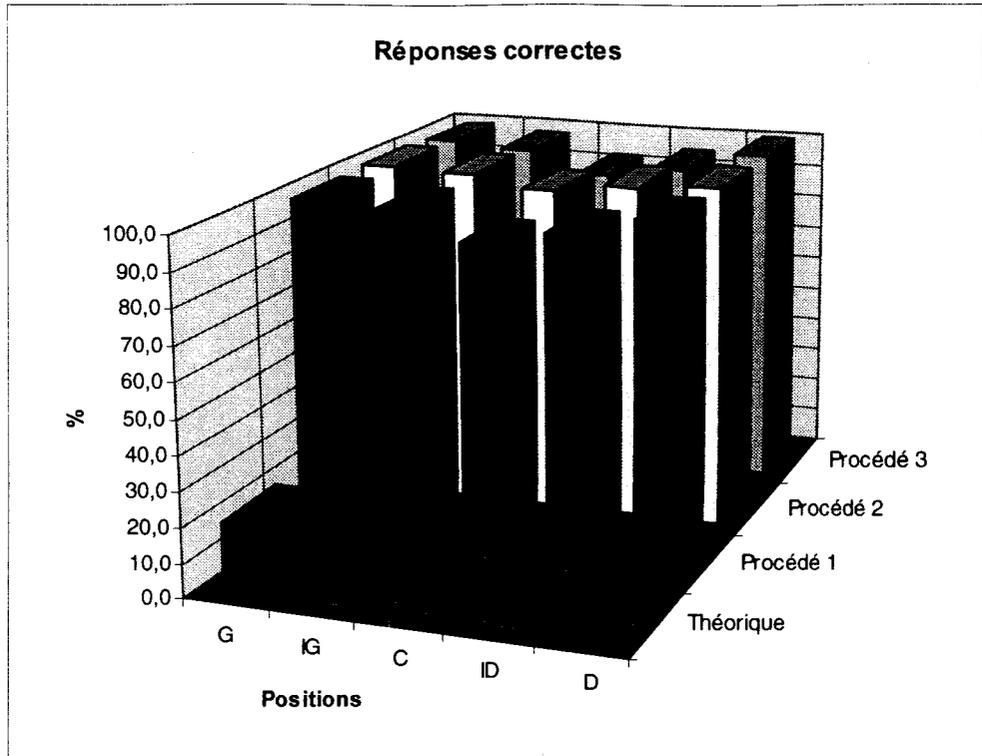


figure 46 : bilan de la première évaluation

Pour préciser ces résultats, considérons l'importance des écarts entre la position rendue et celle perçue :

Procédé	G	IG	C	ID	D
1	0.01	0.03	0.10	0.05	0.00
2	0.01	0.02	0.05	0.02	0.01
3	0.02	0.03	0.09	0.06	0.00

tableau 6 : écarts moyens des réponses

La moyenne des écarts est quasiment nulle. Ceci vient du fait que les erreurs sont exceptionnellement de plus d'une position. Les sujets ont surtout hésité entre le centre et la position intermédiaire à droite. A la vue de cette dissymétrie, un contrôle approfondi des réglages du dispositif a révélé un mauvais centrage de la répartition des volumes dans la partie du dispositif indépendante du contrôle de

l'ordinateur (il y a un réglage de la balance possible au niveau de la radio, de l'entrée et de la sortie de la carte son et, en plus, un autre au niveau des enceintes, le premier et le dernier n'étant pas précis ...). Ce décalage n'a pas eu beaucoup d'influence sur les résultats, pas plus que le fait que le son écouté contenait des informations parfois distrayantes pour les sujets ; mais on compte bien que les conversations dans une audioconférence soient intéressantes.

Les dernières mesures portent sur les temps de réponse des sujets :

Modèle	G	IG	C	ID	D
1	1.6	1.7	1.8	1.7	1.5
2	1.6	1.6	1.6	1.7	1.5
3	1.6	1.7	1.8	1.9	1.6

tableau 7 : temps moyens des réponses

Les sujets ont pris leurs décisions en moins de 2 secondes en moyenne. Cependant, il est à noter qu'ils ignoraient qu'ils étaient chronométrés et donc qu'ils ont pris tout leur temps pour répondre. En outre, l'aisance de manipulation de la souris par les sujets variait considérablement et ce temps de manipulation est certainement dominant dans le temps de réponse. Par contre, globalement, on constate qu'il n'y a pas de différence d'un procédé de réglage à l'autre.

2.4 Conclusion sur le potentiel d'un système stéréophonique

Il est donc prouvé que l'utilisation d'un simple système stéréophonique permet de différencier des "positions" sonores dès lors que l'on règle le volume des voies de droite et de gauche en fonction de la position que l'on veut rendre. De plus, ce système est particulièrement robuste par rapport au problème de réglage.

Il reste à vérifier que ce type d'information purement sonore est bénéfique pour les audioconférences assistées par ordinateur.

3 Apport de la spatialisation aux audioconférences

3.1 Présentation

Cette expérience a pour but de vérifier que l'information fournie à travers un système stéréophonique dans le cadre d'une audioconférence est effectivement utile. Ce travail a fait l'objet d'un premier compte rendu dans [CLEMENT 95].

Nous avons donc utilisé notre système expérimental d'ACAO et conçu une activité de groupe pertinente dans le cadre de l'utilisation d'un système d'audioconférence pour évaluer ce service.

Le principe est de faire participer plusieurs personnes à deux types d'audioconférences : l'une avec une restitution monophonique, et l'autre avec une restitution stéréophonique par laquelle les voix de chaque interlocuteur seront rendues différemment (c'est-à-dire réparties différemment entre les enceintes droite et gauche). Pendant ces audioconférences, les personnes auront une activité bien définie à effectuer impliquant une discussion entre tous les participants et supposée mettre en exergue la nécessité d'identifier les locuteurs, et surtout le fait qu'une aide apportée à cette identification facilite l'utilisation de l'audioconférence. Les résultats de l'expérience seront déduits de la comparaison des grandeurs mesurées pour chacun des deux types d'audioconférence, l'étude comparative étant à la base de toute évaluation [SCHWARTZ 63].

Cette expérimentation a été particulièrement difficile à mettre au point. Nous n'avons pas trouvé d'autres expériences similaires. En Psychologie Cognitive, les expériences d'évaluation de travail coopératif ou d'étude des échanges verbaux entre personnes sont principalement conduites sur des dyades. Quand elles le sont sur des groupes, il y a toujours une composante visuelle prédominante et une asymétrie des activités pour l'ensemble des personnes du groupe. Les méthodologies telles que les cubes de Kohs [GOBET 76], ou le village de Piaget [PIAGET 48] par exemple, ne nous ont pas paru satisfaisantes pour comparer deux modes de communications verbales dans le cadre d'une activité de groupe.

Elles comportent un aspect moteur, non verbal, important. Elle présentent aussi une dissymétrie des rôles par rapport aux besoins de communications.

Contrairement aux expériences les plus proches de notre problématique que nous connaissons, telles que celles de Sellen [SELLEN 92] ou d'Olson [OLSON 95], nous nous sommes focalisé sur l'aspect sonore uniquement. Sellen a comparé trois situations de discussions : face à face ou à distance, selon deux modes de visualisation. Olson a comparé deux situations d'utilisation d'un éditeur partagé : avec ou sans image vidéo. Dans le premier mode de télécollaboration de l'expérience de Sellen, les participants sont visibles sur un même écran de télévision, chacun sur un quart de l'image. Dans le deuxième mode, chaque autre participant est vu et entendu sur un dispositif distinct qui rassemble un petit écran, une caméra et un haut parleur. Sellen fait d'ailleurs remarquer que cela permet d'entendre chaque participant à une position distincte. Olson a utilisé un système d'audioconférence avec un son de très haute qualité sur un réseau sonore analogique dont il dit qu'il est directif en entrée comme en sortie sans donner davantage de précisions. Nous pensons que la dimension sonore a joué un rôle important qui n'a pas été mis clairement en évidence lors de ces deux expériences.

Ces deux expériences présentent à notre avis un autre biais expérimental. Les recherches en Psychologie Cognitive ont montré l'interdépendance entre le type d'activité et la richesse du média utilisé [CLEMENT 95]. Les activités choisies pour ces deux expériences (une activité de conception pour celle d'Olson ou une simple discussion informelle pour celle de Sellen) ne nécessitent pas une très grande richesse de communication. Elle ne sont donc pas à la hauteur d'une évaluation d'un système aussi riche que la visioconférence.

Nous avons donc cherché à concevoir une activité pertinente :

- qui suppose une négociation de conflits d'intérêts,
- qui permette à un groupe de travailler à un but commun bien défini,
- dont chacun détient une partie des contraintes,

- qui implique une répartition équilibrée de la participation au sein du groupe,
- et qui peut se résoudre par le canal sonore uniquement.

Nous avons cherché en plus une activité relativement proche de celle escomptée pour l'usage d'un système d'audioconférence. Nous avons donc choisi comme activité la prise de date pour un rendez-vous. Cependant nous l'avons fortement contrainte pour la définir de manière reproductible.

3.2 Mise au point du protocole expérimental

La mise au point de cette expérience s'est déroulée en plusieurs étapes. Après avoir défini très précisément les contraintes de l'activité, il a fallu définir les objectifs de l'expérimentation, c'est-à-dire éliminer un maximum de paramètres pour pouvoir en mesurer quelques uns.

3.2.1 Contexte général

Chaque sujet dispose d'un agenda (fictif) qui décrit son emploi du temps et uniquement le sien. Cet agenda s'étend sur un mois de 31 jours débutant un lundi et décrivant heure par heure (du lundi au vendredi de 8h00 à 12h00 d'une part et de 14h00 à 18h00 d'autre part) quatre types de contraintes :

1. libre,
2. rendez-vous inamovible,
3. rendez-vous annulable,
4. rendez-vous interne.

Un rendez-vous interne est un rendez-vous pour un sous-ensemble des participants à l'audioconférence et peut être reporté à une autre date dans le mois.

L'objectif de l'activité signifié dans la consigne est de trouver le meilleur rendez-vous d'une heure possible en un minimum de temps. Le classement des solutions suppose que le meilleur rendez-vous est celui qui, dans l'ordre croissant d'importance :

- implique le moins d'annulations possible,
- se situe le plus tôt possible dans le mois,
- implique le moins de reports possible.

Bien entendu, pour donner un peu de consistance au problème, pour l'ensemble des agendas des participants, il n'y aura pas de solution directe (un créneau où tous sont libres). La meilleure solution sera calculée automatiquement par ordinateur (voir plus loin pour le choix d'un ensemble d'agendas) et comporte un ou deux reports. Elle sert de référence pour évaluer les autres solutions. Les discussions pour reporter un rendez-vous interne devraient amener un sous ensemble de sujets à communiquer entre eux et donc mettre en valeur la nécessité pour les participants de s'identifier précisément entre eux.

Les grandeurs mesurées sont la qualité des solutions trouvées (mesurées selon leur écart par rapport à la meilleure solution) et le temps mis par les participants pour y parvenir. Nous aurions bien voulu effectuer des mesures à partir de la DAL pour pouvoir comparer en partie notre expérience avec celle de Sellen et pour enrichir considérablement l'intérêt des résultats, mais notre système n'est pas encore au point et nécessite lui même des expérimentations pour être affiné.

3.2.2 Choix d'un protocole expérimental

Le fait que l'activité à mener pour l'expérience implique un groupe alourdit considérablement son organisation. Pour pouvoir tirer des conclusions, il faut effectuer un minimum de "tests élémentaires". Cependant, pour chacun de ces tests, il faut n sujets. S'il l'on veut faire p tests élémentaires, il faudra np sujets. Or recruter des sujets n'est pas si facile et organiser une expérimentation impliquant un grand nombre de sujets non plus. En l'occurrence, le matériel à mettre en oeuvre est aussi très lourd : une pièce pour isoler chaque sujet est nécessaire. Ces pièces doivent en plus être équipées et connectées à notre système expérimental d'audioconférence.

Les systèmes d'audioconférence étant peu accessibles, nos sujets seraient donc certainement novices. Nous escomptions alors un effet d'apprentissage marqué, non seulement pour l'utilisation d'un système d'audioconférence mais aussi pour l'activité. L'idéal aurait peut-être été d'entraîner nos sujets pour effacer le biais induit par cet effet d'apprentissage. Ceci n'est pas envisageable vu la lourdeur qu'une telle coordination implique, ne serait ce que de pouvoir retenir (et motiver) np sujets, pour l'entraînement et l'expérimentation.

D'autre part, nous voulons aussi faire une évaluation subjective : demander aux sujets de comparer les deux types d'audioconférences et nous dire lequel ils préfèrent, lequel leur semble le plus agréable.

Ces contraintes et ces objectifs nous ont amené à prévoir, pour chaque groupe de n sujets, deux recherches d'un rendez-vous. Une première fois avec un type d'audioconférence, une deuxième fois avec l'autre type. Ainsi, les sujets pourront comparer subjectivement les deux systèmes. Pour ne pas avoir d'effet d'ordre, nous partagerons équitablement nos p groupes de n sujets en deux classes, ceux qui commenceront par le mode monophonique, et ceux qui commenceront par le mode stéréophonique. Il sera alors possible de comparer la prise d'un premier rendez-vous en monophonie avec celle en stéréophonie, puis de même avec le deuxième rendez-vous, donc après un premier apprentissage. Par contre, il ne sera pas possible de faire des comparaisons croisées d'un groupe à l'autre ni au sein d'un même groupe d'un type à l'autre, non seulement à cause de l'effet d'apprentissage mais encore à cause de la difficulté pour comparer deux prises de rendez-vous différentes, comme nous l'avons appris lors de la pré-expérimentation pour le choix d'un ensemble d'agendas.

3.2.3 Choix d'un ensemble d'agendas

Chaque agenda comporte 23 jours (du 1 au 31, sans les week-ends) de 8 heures, soit 184 cases utiles qui peuvent avoir chacune 13 valeurs différentes (libre, inamovible, annulable et les 10 types de rendez-vous internes de 2 à 3 personnes possibles). Il existe donc 13^{184} agendas possibles (un nombre à 205 chiffres ...).

Bien entendu il faut extraire de cet ensemble quasi infini un sous-ensemble cohérent dont les solutions sont d'une difficulté adaptée.

Pour trouver un tel jeu d'agendas, nous avons utilisé une méthode informatisée semi-automatique. Un premier programme génère de façon aléatoire, jour par jour, les n agendas simultanément, avec des contraintes arbitraires de proportion entre les différents types de cases. Un deuxième programme, beaucoup plus complexe, cherche l'ensemble des solutions pour un jeu d'agendas et détermine la meilleure. Donc, pour trouver le "bon" jeu d'agendas, on affine d'abord les paramètres de proportion du premier programme par passage dans le second du jeu d'agendas fourni par le premier. Pour la mise au point finale, on modifie manuellement le jeu d'agendas pour obtenir le résultat désiré. Les jeux d'agendas sont donc codés d'une manière compacte, lisible et éditable. Une fois le jeu d'agendas choisi, une version plus ergonomique est ensuite créée pour les besoins de l'expérimentation. L'avantage de la recherche informatique des solutions d'un jeu d'agendas est que l'on est certain de l'ensemble des solutions et de leur difficulté théorique (nombre d'annulations et de reports minimaux, aux dates les plus tôt dans le mois).

Sans aucune idée de la difficulté réelle de cette activité, nous avons conduit une pré-expérimentation. Quatre de nos collègues ont servi de sujets ; cette précision est importante car elle induit que ces sujets sont loin d'être des novices et qu'ils ont l'habitude de travailler ensemble. Nous avons choisi deux jeux de quatre agendas dont la meilleure solution comportait une seule annulation et un report pour 3 d'entre eux, les dates des deux rendez-vous se trouvant respectivement le 16 et le 18. Le premier jeu comportait 8 autres moins bonnes solutions, le second 5 autres seulement. Nous pensions ces problèmes faciles et à peu près équivalents. Nos collègues ont trouvé en un peu plus de 16 minutes la meilleure solution du premier problème avec un système d'audioconférence monophonique. Nous avons dû les interrompre lors du deuxième problème au bout de 18 minutes alors qu'ils s'acharnaient à trouver la meilleure solution sur laquelle ils étaient déjà passé deux fois sans la voir. L'un d'entre eux a été passablement gêné par l'entrée inopinée du service technique venu pour percer des trous dans le mur pour y fixer un

accessoire, puis par une autre personne venue lui demander un renseignement. Nous mentionnons cet événement anecdotique pour souligner la sélectivité de nos microphones puisque les autres participants ne se sont absolument pas rendu compte de ces interruptions. Non seulement les microphones sont suffisamment sélectifs pour éviter les effets Larsen, mais aussi pour ne pas transmettre des bruits de perceuse (que nous avons perçus à travers le mur seulement) ni un aparté local avec une autre personne.

Ce premier essai nous a cependant apporté beaucoup d'indications. D'abord le système fonctionne bien. Ensuite l'activité était satisfaisante, bien que plus difficile et plus longue à résoudre que prévue. Dans leurs commentaires, nos collègues ont apprécié le mode stéréophonique. L'un nous a dit qu'il avait pu clairement comprendre les interventions simultanées de deux de ses collègues qu'il percevait séparément sur les deux enceintes. Enfin, les deux agendas se sont avérés de difficulté non équivalente.

3.3 Méthode de l'expérience réalisée

En fonction des résultats de cette première expérimentation, nous avons affiné le problème pour l'expérience en vraie grandeur. Pour simplifier, nous avons découpé le mois en deux périodes, avant et après le 15 du mois, plutôt que de proposer deux jeux d'agendas. Ceci permet de plus d'éviter des problèmes de confusion entre les deux jeux. Nous avons aussi choisi un jeu d'agendas de telle sorte que des solutions plus simples à trouver (bien qu'avec un plus grand nombre d'annulations) se trouvent en fin de période ; ceci afin d'être sûr que les sujets trouvent au moins une solution. Nous avons aussi amélioré la présentation des agendas afin de les rendre plus lisibles ainsi que la formulation de la consigne et du questionnaire.

3.3.1 Précisions pour l'expérimentation

Dans le cadre de l'expérimentation, principalement à cause de contraintes matérielles, le nombre de participants est fixé à quatre personnes par audioconférence. A titre de comparaison, le nombre de participants est de trois

pour l'expérience d'Olson et de quatre pour celle de Sellen. Nous n'avons pas connaissance d'expériences d'activités de groupes à plus de quatre participants par groupe.

Pour aider les participants à s'impliquer dans l'activité, la métaphore "réaliste" suivante est proposée. Les quatre personnes représentent différents corps de métiers impliqués dans la construction d'une piscine: un électricien, un maçon, un plombier, un peintre. Ceci donne un cadre général au problème, un rôle équivalent à chacun et explique l'utilité des rendez-vous internes et du rendez-vous commun à trouver. Il s'agit là d'une simple métaphore. Nous n'avons pas l'ambition de simuler ce que serait une discussion réelle entre ces corps de métier, bien que nous ayons défini des contraintes pour notre activité qui sont assez proches d'une situation réelle.

La consigne est formulée de manière à renforcer la métaphore proposée et inciter les sujets à être le plus performants possible.

3.3.2 *Matériel*

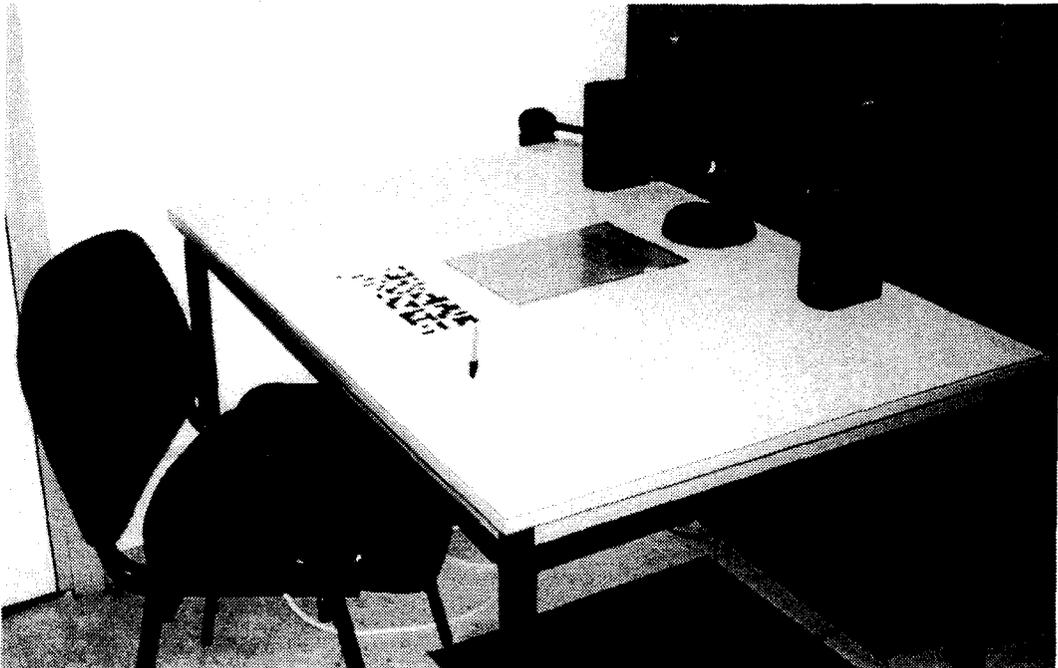


figure 47 : un des quatre postes de travail lors de l'expérience

Chaque sujet est installé à un poste disposant de deux enceintes et d'un microphone (figure 47). Pour les audioconférences de type monophonique, une seule enceinte sera utilisée, l'autre étant présente mais inactive. Chaque sujet dispose d'une feuille de papier décrivant son agenda et d'un crayon pour y porter les annotations qu'il désire. Le même ensemble de quatre agendas est utilisé pour tous les groupes de quatre sujets, mais une copie intacte est utilisée à chaque fois. Cette copie comporte au verso un questionnaire qui est rempli par le sujet dès qu'il a fini l'expérience.

Chaque poste est situé dans une pièce différente d'un même bâtiment ; les participants ne peuvent se voir et ne peuvent s'entendre en dehors du système d'audioconférence. Les postes sont reliés par les liaisons analogiques de haute qualité sonore à travers un système d'audioconférence expérimental permettant les deux types d'audioconférence.

Sur l'ordinateur qui pilote le pont audio expérimental (un PC fonctionnant sous Windows NT 3.5), une application développée spécialement pour l'expérience (figure 48) en contrôle le déroulement. Un bouton permet de choisir le type d'audioconférence utilisé lors de la première phase, une fenêtre texte rappelle les instructions correspondant à l'étape en cours, et un bouton permet de passer à l'étape suivante.

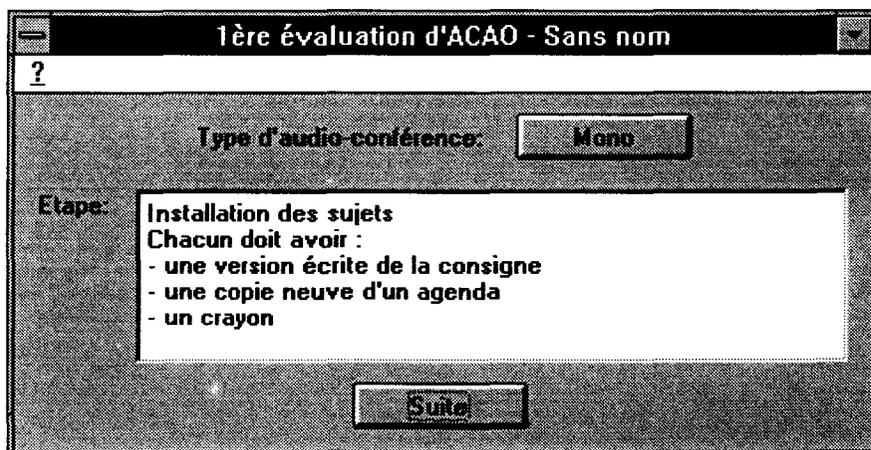


figure 48 : interface de suivi et de contrôle de la deuxième évaluation

Quand une solution est trouvée, une boîte de dialogue (figure 49) permet de la saisir, sans interrompre le déroulement de l'expérience. Bien sûr, c'est aussi cette application qui contrôle le pont audio et qui le bascule automatiquement et instantanément d'un type d'audioconférence à l'autre, tout en enregistrant les paramètres et les résultats de l'expérience (cf. plus loin) dans un fichier identifié avec un nom unique, composé à partir de la date et de l'heure du début de l'expérience. Un tel procédé permet des mesures précises et sûres, ainsi qu'un déroulement homogène de toutes les expériences, ce qui aurait été impossible manuellement.

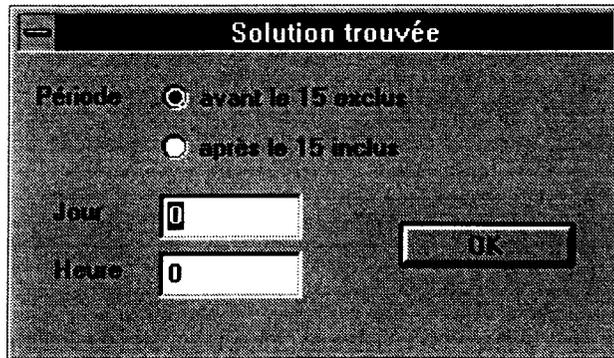


figure 49 : enregistrement d'une solution trouvée

Le pont audio expérimental est aussi configuré pour permettre l'enregistrement sur cassette audio à l'aide d'un magnétophone externe. Nous avons ainsi préféré utiliser un logiciel que nous avons spécialement conçu et réalisé pour gérer cette expérimentation, par souci d'efficacité et de simplicité. Nous aurions pu tout aussi bien utiliser notre agent d'ACAO. Par rapport à notre définition formelle d'une audioconférence, celle utilisée pour cette expérimentation est définie d'une part par un ensemble de six "participants" : les quatre sujets, l'expérimentateur qui lit la consigne et un magnétophone qui enregistre la discussion. D'autre part, la description des destinations sonores en fonction des sources prévoit que les quatre sujets s'entendent et se parlent mutuellement, que l'expérimentateur est entendu par tous et que le magnétophone écoute tout le monde. Les deux modes de fonctionnement correspondent à deux réglages précis et distincts des paramètres

de spatialisation du son pour les participants. Ces configurations s'obtiennent facilement par une série de commandes appropriées envoyées à l'agent d'ACAO.

3.3.3 Sujets

26 groupes de quatre sujets ont participé à l'expérience, soit 104 sujets au total recrutés parmi des étudiants de l'Université. Les expérimentations se sont déroulées entre 12h00 et 14h00 et un repas a été offert aux sujets pour les motiver.

3.3.4 Procédure

Chaque sujet est installé individuellement et séparément à l'un des quatre postes.

Le système d'audioconférence, fonctionnant selon l'un des deux types alternativement monophonique ou stéréophonique, est mis en marche. La consigne est lue au sujet à travers l'audioconférence. Les sujets disposent chacun d'une version écrite personnalisée selon le rôle professionnel qu'ils ont à jouer.

Les sujets cherchent un premier rendez-vous commun avant le 15 (exclu), en mono ou en stéréo selon les groupes.

Dès qu'ils ont trouvé le premier rendez-vous, le système d'audioconférence fonctionnant selon l'autre type est mis en marche. Il y a en fait une commutation discrète et instantanée d'un type à l'autre. Rien n'est fait pour attirer l'attention des sujets sur la modification du fonctionnement. Les sujets cherchent un deuxième rendez-vous commun après le 15 (inclus).

Quand ils ont trouvé le deuxième rendez-vous, le questionnaire figurant au verso de leur agenda est révélé aux sujets. Le système d'audioconférence est arrêté. Les sujets remplissent le questionnaire individuellement.

La durée d'une expérience est limitée arbitrairement à 20 minutes. En fait, ce temps n'a été dépassé par aucun des groupes. La durée de chaque phase est chronométrée et les mesures ainsi que les solutions trouvées sont

automatiquement enregistrées dans un fichier, le déroulement de la procédure étant assisté par ordinateur (séquencement des phases, contrôle du type d'audioconférence et mesures). Les voix des sujets sont enregistrées sur cassette. Tous ces enregistrements sont conservés de manière anonyme : aucun lien entre les mesures enregistrées et l'identité des participants n'est conservé.

Nous n'avons pas prévenu les sujets que le système d'audioconférence fonctionnerait selon deux modes différents pour la recherche des deux rendez-vous, l'un stéréo et l'autre mono, pour ne pas les influencer. Nous ne voulions pas ainsi évaluer leur attention plutôt que leur performance. Nous avons fait cela pour ne pas les influencer, les termes mono et stéréo étant très fortement connotés d'une notion de qualité et de performance.

3.4 Résultats

Nous avons deux ensembles de groupes de quatre sujets (selon le type d'audioconférence utilisé en premier) pour lesquels nous avons respectivement relevé les temps mis pour trouver les rendez-vous et la qualité de la solution trouvée en fonction du type d'audioconférence utilisé. Chaque sujet a également rempli un questionnaire. Nous avons donc fait passer 26 groupes de 4 personnes ; les groupes impairs ont commencé en mono, les groupes pairs en stéréo. Pour chacun des modes nous avons 13 séries de mesures. Nous avons dû écarter le groupe n° 6 pour lequel le système n'a pas basculé d'un mode à l'autre. Nous avons aussi éliminé le groupe n° 18 dans lequel un conflit s'est explicitement déclaré entre les sujets (des insultes ont été échangées). Pour équilibrer le nombre de groupes pour les deux modes nous avons écarté aussi 2 groupes impairs dont les résultats sont notablement différents des autres de la même catégorie, suggérant des mesures erronées par des biais expérimentaux. Ces choix et leurs faibles conséquences sur la validité des résultats sont discutés plus loin. Pour finir, nous avons conservé 11 séries de mesures par mode d'audioconférence.

3.4.1 Mesures effectuées

Les moyennes des mesures obtenues sont résumées dans le tableau suivant, les écart types figurant entre parenthèses :

Rendez-vous	Type	Temps en minutes	dA	dJ
1	mono	5.6 (2.1)	1.64 (0.92)	4.00 (3.22)
2	stéréo	4.5 (2.5)	1.45 (0.68)	6.36 (7.31)
1 et 2		10.1 (2.9)		
1	stéréo	4.4 (1.7)	1.36 (0.92)	3.27 (3.13)
2	mono	3.1 (1.3)	2.00 (1.10)	5.82 (5.88)
1 et 2		7.5 (2.3)		

tableau 8 : bilan de la deuxième évaluation

La colonne nommée dA comporte les moyennes de la différence entre le nombre d'annulations des solutions trouvées et celui de la meilleure solution. Celle nommée dJ concerne la différence de date. Ces deux valeurs représentent la qualité des solutions trouvées : plus elles sont importantes, plus l'écart avec la meilleure solution est grand et donc moins la solution trouvée est optimale. D'après la consigne, le meilleur rendez-vous suppose d'abord le moins d'annulations possibles puis doit être pris le plus tôt possible. La valeur la plus importante est donc celle concernant le nombre d'annulations, puis celle de la différence de date. Nous ne tiendrons pas compte de la différence d'horaire ni de celle du nombre de reports dont l'importance est négligeable par rapport aux deux premiers chiffres.

3.4.2 Questionnaire

Nous n'avons conservé que les questionnaires correspondant aux sujets des groupes retenus pour les mesures données au-dessus, soit un total de 88 questionnaires.

Les réponses aux questionnaires ont permis d'analyser plus finement ces premiers résultats. Aussi les réponses à la première question : "*Avez vous constaté une différence dans la qualité du son entre les deux parties de la tâche ?*" indiquent que 60% des sujets ont repéré un changement dans la qualité d'écoute entre la phase mono et la phase stéréo.

La deuxième question, cherchant à identifier le type de changement perçu indique que la stéréo permet surtout une meilleure identification des partenaires. (parmi ceux qui ont constaté une différence, les autres ne pouvant pas répondre à ces questions) :

- meilleure identification des partenaires.....72%
- confort d'écoute plus important.....60%
- son plus volumétrique53%
- conditions plus proches d'une perception en vis à vis.....51%
- c'est plus agréable.....45%

La troisième partie du questionnaire, visant à vérifier que la qualité du son identifiée comme meilleure correspond bien à la phase stéréo, a montré que 66% des sujets considèrent que la phase stéréo est meilleure que la phase mono. Aucune référence au mode de fonctionnement (mono ou stéréo) n'est présente dans le questionnaire pour ne pas influencer les sujets ; on demande seulement s'ils préfèrent la première ou la deuxième partie. En fait, 74% des sujets ont préféré la deuxième partie. Nous avons alors étudié plus précisément les résultats et constaté que 70% des sujets qui avaient remarqué la différence ont commencé en mono.

3.5 Discussion

3.5.1 Analyse statistique et interprétation

L'étude de la précision des pourcentages³⁰ du sondage par questionnaire donne le chiffre de 60% pour les personnes ayant constaté une différence généralisable avec une précision de 10% au risque de 5%. Le chiffre de 72% pour les personnes ayant trouvé qu'il y avait une meilleure identification des partenaires est généralisable avec une précision de 12% au risque de 5%. On peut donc affirmer que la stéréophonie apporte une amélioration notable qui se traduit par une meilleure identification des participants à l'audioconférence.

Le meilleur temps est réalisé pour la prise du 2ème rendez-vous en écoute monophonique c'est-à-dire quand le 1er rendez-vous s'est réalisé en stéréophonie. Le temps le plus long concerne la prise du 1er rendez-vous en écoute monophonique. Pour entériner la différence de ces moyennes, nous avons procédé à une analyse statistique des données³¹. La différence entre les modes de fonctionnement mono et stéréo n'est cependant pas statistiquement significative ni pour le premier rendez-vous ($t=1.45$, $p<0.16$) ni pour le second ($t=1.70$, $p<0.11$). Une contre-influence des modes stéréo et mono sur les temps de discussion d'un rendez vous et le choix que nous avons fait pour la mesure des temps est peut être en partie à l'origine de cette différence assez peu marquée. En effet, nous mesurons le temps nécessaire à l'obtention d'un consensus entre les participants et non le temps mis pour trouver la solution qui sera finalement retenue. Or l'analyse des enregistrements sonores montre une extension de la discussion en mode stéréo apparemment par simple plaisir d'utilisation du système. En stéréo, le rendez-vous est trouvé très rapidement (en 1 minute environ), mais la discussion continue. Un des groupes a ainsi relancé le débat cinq fois en stéréo, allant même jusqu'à réviser le premier rendez-vous pris en mono ; ils ne se sont mis d'accord pour s'arrêter

³⁰ les calculs de précisions des pourcentages d'un sondage ont été faits d'après les formules données dans [SCHWARTZ 63]

³¹ Les calculs de comparaison des moyennes (test du t ou du Z de Student suivant l'égalité des variances) ont été faits par le logiciel SPSS.

qu'après avoir épuisé toute possibilité de discussion. Au contraire, en mono, la moyenne des 3 minutes de discussion est entièrement consacrée à la recherche de la solution et les sujets semblent être pressés d'en finir, quitte à faire des concessions en annulant des rendez-vous et en évitant les reports.

Ainsi, une première analyse de la qualité des solutions montre que les solutions trouvées en écoute stéréophonique sont toujours de meilleure qualité que celles trouvées dans des conditions d'écoute monophonique. Les meilleures solutions ont été trouvées en moyenne lors de la prise d'un premier rendez-vous en stéréophonie. Par contre, en condition monophonique, l'efficacité est moindre, surtout si la qualité d'écoute se dégrade (passage du système stéréo au mono). Toutefois, la différence entre les modes de fonctionnement mono et stéréo n'est cependant pas statistiquement significative. L'appréciation de la qualité d'un rendez-vous n'est pas évidente. Notre méthode est peut être discutable et, là encore, des effets contraires s'équilibrent. En effet, nous insistons dans la consigne pour que les annulations de rendez-vous soient prioritairement évitées. Cependant, si les agendas fictifs indiquent sans équivoque les rendez-vous annulables, ils ne précisent ni avec qui ni pourquoi ces rendez-vous ont été pris. Les sujets se sentent alors apparemment peu concernés et ne font aucune difficulté pour annuler ces rendez-vous. Par contre, la consigne précise que les rendez-vous internes peuvent être librement reportés, mais les sujets sont tout à fait conscients des efforts de discussion que cela suppose et peu enclins à reporter ces rendez-vous.

En fin de compte, nous avons trouvé que la différence entre les moyennes des temps mis pour trouver les deux rendez-vous selon que les sujets aient commencé en mono ou en stéréo est statistiquement significative ($t=2.34$, $p<0.03$). Nous pouvons donc affirmer que les groupes qui ont commencé en stéréo ont été plus rapides que ceux qui ont commencé en mono et généraliser ce résultat avec un risque inférieur à 3%.

A l'instar de l'expérience d'Olson [OLSON 95] où un groupe a été écarté sur la base de digressions par rapport au problème donné, nous avons essayé d'écarter les mesures susceptibles de biais expérimentaux. Le groupe pour lequel le système n'a

pas changé de mode de fonctionnement doit évidemment être écarté. De même, nous n'avons pas gardé les résultats du groupe pour lequel un conflit sévère, extérieur à l'activité de l'expérience, a resurgi entre deux sujets. Pour avoir un nombre égal de groupes ayant commencé en mono, nous avons écarté aussi deux autres groupes, l'un qui a trouvé les deux rendez-vous en moins de la moitié de la moyenne des autres groupes, l'autre groupe qui a trouvé le premier rendez-vous en deux fois plus de temps que le second et en deux fois plus de temps que la moyenne des autres groupes. Nous aurions pu tout aussi bien le conserver, car les résultats sont alors encore plus significatifs (risque inférieur à 2%). Il en va de même si le groupe conflictuel est conservé, mais pas avec le groupe stressé (risque inférieur à 6%). Nous sommes persuadé que les résultats auraient été encore plus significatifs s'il avait été possible de recruter plus de sujets (c'est-à-dire plus de groupes de *quatre* sujets libres en même temps), et obtenir ainsi davantage de mesures.

3.5.2 Propositions d'explication des résultats

L'impression de confort et d'efficacité du mode stéréophonique pour une audioconférence est telle que nous avons préparé cette expérience en toute confiance. Un de nos sujet nous a même raconté qu'en remplissant le questionnaire, il a regardé autour de lui pour en discuter avec ses collègues et a été surpris de ne voir personne, tellement l'impression de présence avait été forte pendant l'expérience.

Nous nous attendions à avoir des temps de recherche beaucoup plus courts en stéréo qu'en mono et des solutions bien meilleures en stéréo qu'en mono. Il est vrai que le petit nombre de quatre sujets par groupe nous garantit que la spatialisation du son par un simple système stéréophonique permet une discrimination parfaite des positions. Par contre, ce petit nombre a aussi pour conséquence que le problème de la reconnaissance du locuteur (un parmi trois) n'est pas très complexe et met donc moins en valeur l'intérêt de la spatialisation du son. D'autres effets que nous n'avions pas prévus, tel que le prolongement de la discussion par plaisir ou le

peu d'importance accordé aux rendez-vous annulables du fait qu'ils sont anonymes, sont venus certainement tempérer les avantages de la stéréo.

Il n'en reste pas moins que les groupes ayant commencé en stéréo ont été nettement plus rapides que ceux qui ont commencé en mono. Les sujets ont donc été plus efficaces si la prise de contact a été faite en stéréo. Dans ce cas, la recherche du deuxième rendez-vous, bien que réalisée sans spatialisation du son, tire partie des connaissances acquises pendant la première recherche. Au contraire, le plus mauvais temps est réalisé quand le contact est établi en mono et que son effet se prolonge dans la deuxième partie.

Les conditions sonores sont certainement plus importantes pour le premier rendez-vous où il faut commencer par se familiariser avec la situation, établir la communication entre les participants et mettre en place une stratégie de recherche des rendez-vous. Ceci explique pourquoi la deuxième partie est le plus souvent préférée à la première quel que soit le mode de fonctionnement sonore. En outre, les résultats ont montré que la transition mono stéréo avait été plus remarquée que son contraire. Quand le contact est bien établi avec l'aide de la stéréo, les sujets sont absorbés par leur tâche et se rendent moins compte du changement qui intervient dans les conditions sonores.

Une autre explication basée sur l'effet mémoire nous a été suggérée³². L'effet mémoire désigne l'influence de notre mémoire sur notre perception. Ainsi, une orange peinte en bleu et exposée dans la pénombre sera perçue comme étant bien de couleur orange. Alors, une meilleure performance globale réalisée quand la communication est établie en stéréo peut être expliquée par le fait que chaque participant garde en mémoire la position des autres une fois qu'il est habitué à leur timbre de voix. Ensuite, la reconnaissance du timbre suffira pour donner l'impression de la position.

³² Cette suggestion et des explications sur l'effet mémoire viennent de Marie-France Barthelet.

Une autre hypothèse est que la spatialisation du son aide à l'établissement d'une dynamique de groupe et à la distribution des rôles au sein de l'équipe. Nous avons remarqué que les groupes avaient tendance à adopter une démarche similaire : un des quatre sujets finit par diriger la recherche, tandis que les trois autres se bornent à lui fournir les informations à la demande. Il y a d'abord une sorte de compétition pour ce rôle de leader, puis ensuite la recherche est plus efficace. La spatialisation du son permet sans doute de raccourcir ce temps d'organisation et une fois que la dynamique de groupe est installée, les conditions sonores ont moins d'importance. Les groupes qui commencent en mono ne peuvent donc profiter de cet avantage car ils sont déjà organisés quand leur condition d'écoute s'améliore.

Nous tenons aussi à préciser que les résultats ont dû être influencés par le fait que la plupart des sujets se connaissaient au sein d'un même groupe. Ils pouvaient ainsi facilement s'identifier grâce à la connaissance de leur timbre de voix, un autre moyen de reconnaître un locuteur mis en avant par Arons [ARONS 92a]. Nous l'avons découvert pendant l'expérience. Nous ne soupçonnions pas que des étudiants de première année puissent si bien se connaître. En recrutant des étudiants qui se sont inscrits par groupe d'amis à notre insu, nous n'avons pu éviter ce biais expérimental qui cependant renforce nos résultats. Si l'expérience était reconduite avec des groupes de sujets ne se connaissant pas à l'avance entre eux, l'intérêt de la spatialisation du son serait encore plus flagrante.

Pour finir, mener une expérience avec des groupes de sujets ne présente pas que des inconvénients. Quatre personnes impliquées simultanément génèrent une grande variété d'attitudes et de réactions comme le montrent les écart types élevés, mais à notre grand soulagement, aucune solution impossible n'a été trouvée : au moins l'un des quatre sujets détectait l'éventuelle erreur.

3.6 Conclusion sur l'apport de la spatialisation du son

Cette expérimentation nous a permis d'atteindre plusieurs résultats. Notre système expérimental d'audioconférences assistées par ordinateur fonctionne bien et s'est révélé très utile pour mener à bien cette expérience. Notre activité de

recherche d'un rendez-vous est aussi validée. Le problème est bien compris par les sujets, d'une complexité suffisante pour nécessiter un réel travail sans être rébarbatif. Cette tâche précise et fortement contrainte peut être utilisée pour des expériences similaires nécessitant une activité précise et exigeante en termes de richesse de communication. Bien que nous n'ayons pas trouvé exactement ce que nous escomptions, nous avons pu démontrer que la stéréo est préférée par les utilisateurs parce qu'elle permet de mieux identifier les locuteurs. Nous avons aussi pu démontrer qu'une prise de contact en stéréo améliore objectivement l'utilisation d'une audioconférence.

Ce résultat suggère une recommandation quant aux systèmes d'audioconférences. Il est préférable de commencer une audioconférence avec une restitution spatialisée du son aidant l'identification des locuteurs. Ensuite, il est possible de revenir à un mode de fonctionnement monophonique, tant pour réduire la bande passante nécessaire et donc le coût des transmissions que pour raccourcir le temps d'utilisation. Quoiqu'il en soit, la spatialisation du son améliore les systèmes d'audioconférence.

4 Conclusion sur l'évaluation des services proposés

L'évaluation, en confrontant la théorie à la réalité, présente le risque de révéler des déficiences dans un raisonnement par ailleurs irréprochable sur le plan théorique, sans parler de la complexité de ce domaine, de la difficulté à mettre en oeuvre une expérimentation, à en exploiter les résultats, et de la robustesse des réalisations de prototypes expérimentaux que cela suppose. C'est notre conviction qu'une approche incluant l'évaluation est indispensable dans une discipline telle que le travail coopératif où la satisfaction d'utilisateurs réels prime sur la pureté d'un modèle théorique.

Si aborder le problème de l'évaluation pose initialement plus de problèmes que cela n'en résout, cette démarche se révèle très enrichissante à bien des égards. L'évaluation est un problème si vaste qu'il nécessite une approche pluridisciplinaire. Pour un informaticien, il est indispensable de faire appel à des

psychologues et des ergonomes pour préparer et réaliser une évaluation et ce, dès le début de toute conception. Cette coopération, très productive par elle-même, permet d'éviter a priori des erreurs dans les services proposés et de vérifier a posteriori leur validité. Cette coopération n'est pas non plus à sens unique car l'informatique offre de nouveaux outils d'évaluation et permet de réaliser des expérimentations qu'il serait impossible de mener à bien sans elle.

L'approfondissement de la conception jusqu'à un prototype d'expérimentation qu'encourage l'évaluation ainsi que la réflexion qu'elle incite, quant à l'utilisation des services proposés, améliore aussi la qualité de la réflexion théorique. Ainsi, être confronté à l'implémentation d'un modèle d'audioconférence nous a permis d'écarter le concept, pourtant séduisant, d'une structure récursive dans la notion d'un participant. De même, réfléchir aux dimensions d'utilisabilité et d'utilité de l'évaluation, nous a permis de prendre conscience de l'effort de recherche qu'il reste à faire pour mettre nos nouveaux services à la disposition des utilisateurs d'audioconférence au niveau des interfaces homme-machine. Enfin, l'évaluation des services, dont les conséquences sont techniquement loin d'être négligeables, telle que la spatialisation du son et en particulier une spatialisation élémentaire au moyen d'un simple système stéréophonique, permet d'en valider l'intérêt.

Cependant, nous n'avons qu'entamé l'ensemble des évaluations que nécessite le travail présenté dans cette thèse. Il reste de nombreux aspects à évaluer : le mécanisme de la détection automatique du locuteur, l'utilisabilité et l'utilité de son exploitation et bien sûr, une fois le problème de leur mise à disposition au niveau des utilisateurs résolu, l'utilisabilité et l'utilité de l'ensemble des services proposés dans cette thèse.

Conclusion

Le travail présenté dans cette thèse est le fruit d'une réflexion centrée sur les aspects sonores des interfaces des collecticiels synchrones distribués qui sont représentés par leur composante chargée de l'audioconférence. Cette focalisation sur un point précis du sujet étudié au départ, tel qu'il apparaît dans le titre de cette thèse, découle de l'étude de l'état de l'art de ce domaine et a été expliquée dans l'introduction.

Si le point étudié dans cette thèse est assez précis, les domaines abordés pour l'étude de ce point sont en revanche vastes et nombreux. Nous avons ainsi approfondi nos connaissances sur le son en général, étudié l'exploitation qui a été faite du son dans les systèmes informatiques et plus précisément dans les collecticiels.

A partir du constat des insuffisances des systèmes d'audioconférences existants, nous avons proposé un ensemble de nouveaux services pour ces systèmes. Nos propositions concernent trois aspects :

- au niveau de la synthèse sonore, nous proposons l'utilisation de la spatialisation du son pour aider les participants d'une audioconférence à identifier les locuteurs,
- au niveau de la gestion et du contrôle de la diffusion du son, nous proposons une extension de la notion d'un participant et une structure de conférence dynamique qui permet, entre autre, les apartés entre deux participants quelconques, la fusion de deux audioconférences, la scission d'une audioconférence ainsi que d'autres services,
- au niveau de la prise de son, nous proposons d'y associer des informations statiques et dynamiques dans l'optique d'un traitement différé ou instantané.

Dès le début de notre réflexion, nous souhaitions confronter notre réflexion théorique aux contraintes de la réalité. La suite de notre démarche a donc consisté à prévoir l'intégration de ces nouveaux services dans un collecticiel complet et à réaliser un système expérimental permettant d'évaluer nos propositions. Ce travail nous a amené à définir une audioconférence de manière plus formelle, à proposer un langage de commande permettant d'exploiter ses nouveaux services et à adopter la métaphore d'un système multi-agents pour l'intégration dans un collecticiel. Nous avons aussi conçu et réalisé un système d'audioconférences assistées par ordinateur permettant de concrétiser d'une part les services proposés et de réaliser, d'autre part, les communications sonores définies par l'emploi de ces services.

Nous avons alors pu aborder le vaste problème de l'évaluation et réaliser deux expériences qui nous ont permis de confirmer l'intérêt de l'utilisation d'un simple système stéréophonique pour une audioconférence. A l'occasion de la deuxième expérience, nous avons été amené à concevoir une activité adaptée à l'évaluation d'un collecticiel synchrone distribué.

Cette réflexion initiale sur l'aspect sonore des collecticiels synchrones distribués suivie et affinée par une réalisation concrète a en fait généré plus de questions qu'elle n'a apporté de solutions.

Les performances techniques du moment en matière de traitement et surtout de transport du son nous ont contraint à l'étude d'une solution de gestion analogique du son même si nous nous sommes efforcé de prévoir une gestion du son tout numérique. Il reste donc à affiner dans ce sens l'architecture logicielle de notre système expérimental.

En considérant le problème de l'évaluation, nous avons mis en exergue la problématique de l'utilisabilité de nos services. Il reste donc à rendre ces services accessibles au niveau de l'interface utilisateur. Plus généralement, il reste à intégrer notre système expérimental dans un collecticiel complet. Ces étapes

franchies, il sera alors possible de procéder à l'évaluation complète des propositions de cette thèse.

Cependant, certains de ces nouveaux services sont encore à un stade rudimentaire, en particulier ceux qui mettent en jeu la détection automatique des locuteurs. Une fois ces services mieux définis, il suffira de reprendre la même démarche que pour les précédents ...

Bibliographie

[ABOWD 90]

G. Abowd, "Agents: Communicating Interactive Processes", *Proceedings of INTERACT'90*, 1990, pp 143-148.

[ANDERSON 91a]

D. P. Anderson, J. Bilmes, "Concurrent Real-Time Music in C++", *Proceedings of the USENIX C++ Conference*, Washington DC, April 1991, pp 147-161.

[ANDERSON 91b]

D. Anderson, P. Chan, "Toolkit Support for Multiuser Audio/Video Applications", *Proceedings of the Second International Workshop on Network and Operating System Support for Digital Audio and Video*, Heidelberg 1991, in Lecture Notes in Computer Science 614, R; G. Herrtwich (Ed.), pp 230-241.

[ANDERSON 91c]

D. Anderson, R. Govindam, G. Homsy, "Abstractions for Continuous Media in a Network Window System", *Proceedings of the International Conference on Multimedia Information Systems'91*, Japan, 1991, pp 273-298.

[ANDERSON 92]

D. Anderson, P. Chan, "COMET: a Toolkit for Multiuser Audio/Video Applications", *12th International Conference on Distributed Computing Systems*, Yokohama 1992, pp 555-562.

[ANGEBRANNDT 91]

S. Angebrandt, R. Hyde, D. Luong, N. Siravara, C. Schmandt, "Integrating Audio and Telephony in a Distributed Workstation Environment", *Proceedings of USENIX Summer'91*, 1991, pp 419-435.

[ARONS 92a]

B. Arons, "A Review of the Cocktail Party Effect", *Journal of the American Voice I/O Society*, vol. 12, July 1992, pp 35-50.

[ARONS 92b]

B. Arons, "Techniques, Perceptions, and Applications of Time-Compressed Speech", *Proceedings of the 1992 Conference, American Voice I/O Society*, September 1992, pp 169-177.

Bibliographie

[ARONS 92c]

B. Arons, "Tools for Buidings Asynchronous Servers to Support Speech and Audio Applications", *Proceedings of UIST'92*, November 92, pp71-78.

[ARONS 93]

B. Arons, "SpeechSkimmer: Interactively Skimming Recorded Speech", *Proceedings o UIST'93*, November 1993, pp 187-196.

[ARONS 94a]

B. Arons, E. Mynatt, "The Future of Speech and Audio in the Interface", *SIGCHI Bulletin*, October 1994, vol. 26, n° 4, pp 44-48.

[ARONS 94b]

B. Arons, "Efficient listening with two ears: Dichotic time compression and spatialization", *Proceedings of ICAD'94*, November 94,

[A. V. T. 92]

"Le DAB: Radiodiffusion Numérique", *Revue : Audio et Vidéo Tech n°33*, septembre - octobre 1992, pp. 21-31.

[BARME 92]

L. Barne, A. Derycke, "Intégration du son dans les collecticiels", *Actes d'IHM'92*, Novembre 1992 à Paris, pp 157-166.

[BARME 94]

L. Barne, "Audioconférences assistées par ordinateur : gestion et contrôle", *Actes d'IHM'94*, décembre 1994 à Lille, pp 133-138.

[BEAUDOIN-LAFON 94a]

M. Beaudoin-Lafon, W. Gaver, "ENO: Synthesizing Structured Sound Spaces", *Proceedings of UIST'94*, November 1994, pp 49-57.

[BEAUDOIN-LAFON 94b]

M. Beaudoin-Lafon, "L'usage du son dans les systèmes interactifs", *Actes d'IHM'94*, décembre 1994, pp 119-124.

[BINDING 89]

C. Binding, C. Schmandt, K. Lantz, B. Arons, "Workstation Audio and Window-based Graphics: Similarities and Differences", *Engineering for Human-Computer Interaction*, August 1989, pp 233-247.

[BLATTNER 89]

M. Blattner, D. Sumikawa, R. Greenberg, "Earcons and Icons: Their Structure and Common Design Principles", *Human Computer Interaction*, 1989, vol. 4 n° 1, pp 11-44.

Bibliographie

[BLATTNER 94]

M. Blattner, A. Papp, E. Glinert, "Sonic Enhancement of Two-Dimensional Graphics Displays", *Auditory Display*, Kramer ed., Addison Wesley, April 1994, ISBN 0-201-62603-9, pp 447-470.

[BLY 94]

S. Bly, "Multivariate Data Mappings", *Auditory Display*, Kramer ed., Addison Wesley, April 1994, ISBN 0-201-62603-9, pp 405-416.

[BOTTE 89]

M.C. Botte, "Traité de psychologie cognitive 1", *Ed Dunod*, C. Bonnet et al, 1989, ISBN 2-04-018752-9, pages 83-127.

[BOURGUIN 95]

G. Bouguin, "Le transport du son par voie numérique et en temps réel. Modèle d'architecture pour audioconférence multipartie", *Mémoire de DEA*, Université de Lille 1, juin 95.

[BREWSTER 94]

S. Brewster, P. Wright, A. Edwards, "A Detailed Investigation into the Effectiveness of Earcons", *Auditory Display*, Kramer ed., Addison Wesley, April 1994, ISBN 0-201-62603-9, pp 471-498.

[BROWN 89]

M. Brown, S. Newsome, E. Glinet, "An Experiment into the Use of Auditory Cues to Reduce Visual Workload", *Proceedings of CHI'89*, May 1989, pp 339-346.

[BURGESS 92]

D. Burgess, "Techniques for Low Cost Spatial Audio", *Proceedings of UIST'92*, November 1994, pp 49-57.

[BURGESS 94]

D. Burgess, E. Mynatt, "3-D Interactive Percussion: The Virtual Drum Kit", *Conference Companion of UIST'94*, November 1994, pp 45-46.

[CALLIOPE 89]

Calliope, "La parole et son traitement automatique", éditions Masson, 1989.

[CARD 91]

S. Card, G. Robertson, J. Mackinlay, "The information visualizer, an information workspace", *Proceedings of CHI'91*, 1991, pp 181-188.

[CARDOZO 93]

E. Cardozo, J. Sichman, Y. Demazeau, "Using the Active Object Model to Implement Multi-Agents Systems", *Proceedings of the international Conference on Tools with AI*, November 1993, pp 70-77.

Bibliographie

[CHERRY 53]

E. Cherry, "Some experiments on the recognition of speech with one or two ears", *Journal of the Acoustical Society of America*, n° 25, 1953, pp 975-979.

[CLARK 92]

W. Clark, "Multipoint Multimédia Conferencing", *IEEE Communications Magazine*, May 1992, pp 44-50.

[CLARKE 68]

A. Clarke, "2001: A Space Odyssey", July 1968.

[CLEMENT 95]

D. Clément, L. Barne, "Audioconférence spatialisée : évaluation de ses effets sur une activité de groupe", *Actes d'IHM'95*, octobre 1995, pp 35-42.

[COHEN 91]

M. Cohen, L. Ludwig, "Multimédia audio windows management", *Computer Supported Cooperative Work and Groupware*, Greenberg ed., Academic Press, 1991, ISBN 0-12-299220-2, pp 193-210.

[COHEN 93]

M. Cohen, "Throwing, pitching and catching sound: audio windowing models and modes", *International Journal on Man-Machine Studies*, vol. 39, 1993, pp 269-304.

[COHEN 94]

J. Cohen, "Monitoring Background Activities", *Auditory Display*, Kramer ed., Addison Wesley, April 1994, ISBN 0-201-62603-9, pp 499-532.

[CROISY 95]

P. Croisy, "Collecticiel temps réel et apprentissage coopératif : des aspects sociaux et pédagogiques jusqu'au modèle multi-agents de l'interface de groupe", *Thèse de Doctorat*, soutenue à l'Université de Lille 1, 1995, 200 pages.

[CUSSO 92]

M. Cusso, "Des ondes, des nombres et des ombres", *SVM*, novembre 1992, pp 226-234.

[DEGEN 92]

L. Degen, R. Mander, G. Salomon, "Working with Audio: Integrating Personal Tape Recorders and Desktop Computers", *Proceedings of CHI'92*, May 1992, pp 413-418.

[DERYCKE 92]

A. Derycke, "Towards a hypermedium for collaborative learning", *Collaborative learning and computer conferencing*, A. Kaye ed., Springer-Verlag, 1992, pp 211-224.

[DERYCKE 93a]

A. Derycke, C. Chaillou, L. Dumesges, "La troisième dimension dans les interfaces de collecticiels", *Actes d'IHM'93*, Octobre 1993, pp 81-86.

[DERYCKE 93b]

A. Derycke, C. Viéville, "Real time Multimédia Conferencing System and Collaborative Learning", *NATO Advanced Workshop*, Segovia, Spain, April 1993, F. Verdejo, S. Cerri (Ed.), ASI Series, Springer Verlag, 23 pages.

[DERYCKE 94]

A. Derycke, L. Barne, "Multimédia in the Computer Supported Cooperative Work", *Proceedings of BIWIT'94*, February 1994, pp 287-297.

[DESANCTIS 87]

G. DeSanctis, B. Gallupe, "A Foundation for the Study of Group Decision Support Systems", *Management Science*, volume 33 n°5, May 1987, pp 589-609.

[ELLIS 91]

C. A. Ellis, S. J. Gibbs, G. L. Rein, "Groupware: Some issues and experiences", *Communications of the ACM*, vol.34, January 1991, pp 38-58.

[ELLIS 94]

C. A. Ellis, J. Wainer, "A Conceptual Model of Groupware", *Proceedings of CSCW'94*, October 94, pp 79-88.

[ELMER 94]

Elmer, "Le son sur micro-ordinateur", *Ed Dunod Tech*, 1994, ISBN 2-10-0022547.

[ENSOR 91]

J. Ensor, S. Ahuja, R. Connaghan, D. Horn, M. Pack, D. Seligmann, "Control Issues in Multimédia Conferencing", 1991, pp 133-143.

[FIELDER 89]

L. Fielder, "Human Auditory Capabilities and Their Consequences on Digital-Audio Converter Design", *Proceedings of the 7th AES International Conference*, May 1989, pp 45-62.

[GANYMEDE 95]

Ganymède, "Travail coopératif et communication avancée : la notion de Multi-Agents", *Actes des journées de travail sur la notion de Multi-Agents*, février 1995, publication n° 163, 96 pages.

[GAVER 89]

W. Gaver, "The SonicFinder: An Interface That Uses Auditory Icons", *Human-Computer Interaction*, vol. 4, n° 1, 1989, pp 67-94.

[GAVER 90]

W. Gaver, R. Smith, "Auditory Icons in Large Scale Collaborative Environments", *Proceedings of INTERACT'90*, August 1990, pp 735-740.

Bibliographie

[GAVER 91a]

W. Gaver, R. Smith, T. O'Shea, "Effective sounds in complex systems: The ARKola simulation", *Proceedings of CHI'91*, April 1991, pp 85-90

[GAVER 91b]

W. Gaver, "Sound Support for Collaboration", *Proceedings of ECSCW'91*, September 91, pp 293-308.

[GAVER 92]

W. Gaver, "The Affordances of Media Spaces for Collaboration", *Proceedings of CSCW 92*, November 1992, pp 17-24.

[GERSHO 94]

A. Gersho, "Advances in Speech and Audio Compression", *Proceedings of the IEEE*, June 1994, vol. 82 n°6, pp 900-918.

[GOBET 76]

J. Gobet, "Les tests démystifiés", *Editions Aubier Montaigne, Paris*, ISBN 2-7007-0038-4, 1976.

[GRANSART 95]

C. Gransart, "BOX : Un Modèle et un Langage à Objets pour la Programmation Parallèle et Distribuée", *Thèse de doctorat en Informatique*, Université de Lille 1, 1995.

[GRUDIN 88]

J. Grudin, "Why CSCW applications fail: Problems in the Design and Evaluation of Organizational Interfaces", *Proceedings of CHI'88*, September 1988, pp 85-93.

[GRUDIN 93]

J. Grudin, "The CSCW Forum", *Proceedings of the 26th Hawai International Conference on System Sciences*, 1993, pp 51-58.

[GUILLEMAIN 91]

P. Guillemain, "L'intimité des sons", *Pour la Science n°169*, novembre 1991, pp 26-29.

[HAMMAINEN 91]

H. Hammainen, C. Condon, "Form and Room: metaphors for groupware", *Proceedings on Organizational Computing Systems*, SIGOIS Bulletin, vol. 12, n°2-3, November 1991, pp 95-104.

[HANDEL 89]

S. Handel, "Listening : An Introduction to the Perception of Auditory Events", *A Bradford Book, MIT Press*, 1989, 597 pages.

[HARPER 94]

R. Harper, K. Karter, "Keeping people apart", *Journal of Computer Supported Cooperative Work 2 (3)*, pp 199-207.

Bibliographie

[HARRISON 94]

B. Harrison, M. Mantei, G. Beirne, T. Narine, "Communicating about Communicating: Cross-Disciplinary Design of a Media Space Interface", *Proceedings of CHI'94*, April 1994, pp 124-130.

[HAYWARD 94]

C. Hayward, "Listening to the Earth Sing", *Auditory Display*, Kramer ed., Addison Wesley, April 1994, ISBN 0-201-62603-9, pp 369-404.

[HINDUS 92]

D. Hindus, C. Schmandt, "Ubiquitous Audio: Capturing Spontaneous Collaboration", *Proceedings of CSCW 92*, November 1992, pp 210-217.

[HOOGSTOEL 95]

F. Hoogstoel, "Une approche organisationnelle du travail coopératif assisté par ordinateur", *Thèse de Doctorat*, soutenue à l'Université de Lille 1, 1995, 384 pages.

[JACKSON 94]

J. Jackson, J. Francioni, "Synchronization of Visual and Aural Parallel Program Performance Data", *Auditory Display*, Kramer ed., Addison Wesley, April 1994, ISBN 0-201-62603-9, pp 291-306.

[JAMESON 94]

D. Jameson, "Sonnet: Audio-Enhanced Monitoring and Debugging", *Auditory Display*, Kramer ed., Addison Wesley, April 1994, ISBN 0-201-62603-9, pp 253-265.

[JOHANSEN 88]

R. Johansen, "Groupware : Computer Support for Business Teams", *Series in Communication Technology and Society*, chapter 2 p. 12-45, 1988, ISBN 0-02-916491-5, 205 pages.

[JOT 93]

J.M. Jot, O. Warusfel et al, "Binaural concert hall simulation in real time", *Proceedings of the IEEE Monhok workshop*, October 1993.

[KARSENTY 91]

A. Karsenty, M. Beaudoin-Lafon, "Une application de dessin coopératif synchrone", *Actes d'IHM'91*, 1991, pp 141-146.

[KARSENTY 94]

A. Karsenty, "Le collecticiel : de l'interaction homme-machine à la communication homme-machine-homme", *Techniques et sciences informatiques*, vol. 13, n°1, 1994, pp 105-127.

Bibliographie

[KARSHMER 92]

A. Karshmer, "SounStation II: Using Sound and Sound Spaces to Provide High Bandwidth Computer Interface to the Visually Handicapped", *SIGCAPH Newsletter ACM Press*,

[KENDALL 91]

G. Kendall, "Visualization by Ear: Auditory Imagery for Scientific Visualization and Virtual Reality", *Computer Music Journal*, vol. 15, n° 4, Winter 1991, pp 70-73

[KRAMER 94a]

G. Kramer, "Auditory Display", *Ed Addison Wesley*, April 1994, ISBN 0-201-62603-9, 672 pages.

[KRAMER 94b]

G. Kramer, "An Introduction to Auditory Display", *Auditory Display*, Kramer ed., Addison Wesley, April 1994, ISBN 0-201-62603-9, pp 1-78.

[LEUNG 90]

W. Leung, T. Baumgartner, Y. Hwang, M. Morgan, S. Tu, "A software Architecture for Workstations Supporting Multimedia Conferencing in Packet Switching Networks", *IEEE Journal on Selected Areas in Communications*, vol.8, n° 3, April 1990, pp380-390.

[MADHYASTHA 94]

T. Madhyastha, D. Reed, "A framework for Sonification Design", *Auditory Display*, Kramer ed., Addison Wesley, April 1994, ISBN 0-201-62603-9, pp 267-289.

[MAENO 91]

K. Maeno, S. Sakata, T. Ohmori, "Distributed Desktop Conferencing System (MERMAID) Based on Group Communication Architecture", *Proceedings of ICC'91*, 1991, pp 520-525.

[MOUGEY 93]

S. Mougey, "Le Son Numérique", *PIXEL Technique 17*, pp. 19-28.

[MS 93]

Microsoft®, "Win 32 Programmer's Reference", *Microsoft Press*, vol. 2, 1993, 999 pages.

[MULDER 94]

J. Mulder, E. Dooijes, "Spatial Audio in Graphical Applications", *CWI Report CS-R9434*, ISSN 0169-118X, May 1994.

[MULLER 90]

M. Muller, J. Daniel, "Toward a Definition of Voice Documents", *ACM Proceedings of COIS'90*, 1990, pp 174-183.

Bibliographie

[MYNATT 94a]

E. Mynatt, G. Weber, "Nonvisual Presentation of Graphical User Interfaces: Contrasting two Approaches", *Proceedings of CHI'94*, April 1994, pp 166-172.

[MYNATT 94b]

E. Mynatt, "Designing with Auditory Icons: How Well do we Identify Auditory Cues", *Conference Companion of CHI'94*, April 1994, pp 269-270.

[NELSON 93]

M. Nelson, "La compression de données", *Ed Dunod*, 1993, ISBN 2-10-001681-4, 420 pages.

[NUNAMAKER 91]

J. Nunamaker, A. Dennis, J. Valacich, D. Vogel, J. George, "Electronic meeting systems to support group work", *Communications of the ACM*, vol. 34, pp 30-61.

[NYQUIST 24]

H. Nyquist, "Certain Factor Affecting Telegraph Speed", *Bell System Technical Journal* n°3, 1924, pp 324-326.

[OLSON 95]

J. Olson, G. Olson, D. Meader, "What Mix of Video and Audio is Useful for Small Groups Doing Remote Real-Time Design Work?", *Proceedings of CHI'95*, 1995, pp 362-368.

[PAGANI 93]

D. Pagani, W.E. Mackay, "Bringing media spaces into the real world" *Proceedings of ECSCW93*, 1993, pp 341-356.

[PASIEKA 91]

M. Pasioka, P. Crumley, A. Marks, A. Infortuna, "Distributed Multimedia: How Can the Necessary Data Rates be Supported?", *Proceedings of USENIX Summer'91*, 1991, pp 169-182

[PIAGET 48]

J. Piaget, "La représentation de l'espace chez l'enfant", *PUF*, 1948.

[POPE 93]

S. Pope, L. Fahlén, "The use of 3-D Audio in a Synthetic Environment" : An Aural Renderer for a Distributed Virtual Reality System", *Proceedings of IEEE VRAIS'93*, 1993, pp 176-182.

[RAMANATHAN 92]

S. Ramanathan, P. Rangan, H. Vin, T. Kaepfner, "Optimal Communication Architectures for Multimedia Conferencing in Distributed Systems", *Proceedings of the 12th International Conference on Distributed Computing Systems*, 1992, pp 46-53.

Bibliographie

[RANGAN 91]

P. Rangan, H. Vin, "Multimedia Conferencing as a Universal Paradigm for Collaboration", *"Multimedia" Kjell Dahl ed.*, Springer Verlag, April 1995, ISBN 0-387-55201-4, pp 174-185.

[RICARD 94]

G. Ricard, S. Meirs, "Intelligibility and Localisation of Speech from Virtual Directions", *Human Factors*, March 1994, vol. 36, pp 120-128.

[RODDEN 91a]

T. Rodden, "A Survey of CSCW Systems", *Interacting with Computers*, 1991, vol. 3, n° 3, pp 319-353.

[RODDEN 91b]

T. Rodden, G. Blair, "CSCW and Distributed Systems: the Problem of Control", *Proceedings of ECSCW'91*, September 1991, pp 49-63.

[ROOS 95]

S. Roos, M. Ramage, Y. Rogers, "PETRA: Participatory Evaluation Through Redesign and Analysis", CSRP n) 375, March 1995, to appear in *Interacting with Computers*, 20 pages.

[SALBER 95]

D. Salber, J. Coutaz, D. Decouchant, M. Riveill, "De l'observabilité et de l'honnêteté : le cas du contrôle d'accès dans la Communication Homme-Homme Médiatisée", *Actes d'IHM'95*, octobre 1995, pp 27-33.

[SCAPIN 86]

D. Scapin, "Guide Ergonomique de conception des interfaces homme-machine", *Rapport technique de l'INRIA*, n° 77, octobre 86, 91 pages.

[SCHMANDT 88]

C. Schmandt, "Employing Voice Back Channels to Facilitate Audio Document Retrieval", *Proceedings of the ACM Conference on Office Information Systems*, 1988, pp 213-218.

[SCHMANDT 93]

C. Schmandt, "Phoneshell: the Telephone as Computer Terminal", *Proceedings ACM Multimedia 93*, 1993, pp 373-382.

[SCHMANDT 94]

C. Schmandt, "Voice Communication with Computers", *Ed Van Nostrand Reinhold, New York*, ISBN 0-442-23953-1, 1994, 319 pages.

[SCHOOLER 91a]

E. Schooler, S. Casner, J. Postel, "Multimedia Conferencing: has it come of age ?", *Proceedings of the 24th Hawaiï Conference on System Sciences*, January 1991, vol. 3, pp 707-716.

[SCHOOLER 91b]

E. Schooler, "A distributed Architecture for Multimedia Conference Control", *ISI Research Report 91-289*, November 1991. 18 pages.

[SCHOOLER 93a]

E. Schooler, "The impact of scaling on a multimedia connection architecture", *Multimedia Systems*, Springer-Verlag, 1993, pp 2-9.

[SCHOOLER 93b]

E. Schooler, "Case Study: Multimedia Conference Control in a Packet-switched Teleconferencing System", *Internetworking: Research and Experience*, vol. 4, 1993, pp 99-120.

[SCHWARTZ 63]

D. Schwartz, "Méthodes statistiques à l'usage des médecins et des biologistes", *Flammarion Medecine Sciences*, 1963, ISBN 2-257-30326-1.

[SELLEN 92]

A. Sellen, "Speech Patterns in Video-Mediated Conversation", *Proceedings of CHI'92*, May 1992, pp 49-59.

[SENACH 90]

B. Senach, "Evaluation ergonomique des Interfaces Homme-Machine: une revue de la Littérature", *Rapport de recherche de l'INRIA*, n° 1180, février 1990, 74 pages.

[SMITH 86]

S. Smith, J. Mosier, "Guidelines for designing user interface software" *Report ESD-TR-86-278*, Bedford, MA, Mitre Corporation, August 1986, 425 pages.

[STOLZ 92]

A. Stolz, "Le grand livre de la Sound Blaster", *Ed. Micro Application*, ISBN : 2-86899-758-9, 1992, 410 pages.

[VIEVILLE 95]

C. Viéville, "Structuring conversation in asynchronous communication systems to support collaborative learning at a distance", *Proceedings of ED-MEDIA'95*, June 1995, pp 816.

[VILERS 92]

P. Vilers, "Etude d'un système de téléconférence temps réel : application à un jeu de rôle en gestion de production", *Thèse de Doctorat*, soutenue à l'Université de Lille 1, 1992, 221 pages.

[WATABE 90]

K. Watabe, S. Sakata, K. Marno, H. Fukuoka, T. Ohmori, "Distributed multiparty desktop conferencing system : MERMAID", *Proceeding of the ACM CSCW'90 Conference*, October 1990, pp 27-38.

Bibliographie

[WENZEL 92]

E. Wenzel, "Three-Dimensional Virtual Acoustic Displays", *Multimedia Interface Design* edited by M. Blattner and B. Dannenberg, ACM Press ISBN 0-201-54981-6, 1992, pp 257-288.

[WENZEL 94]

E. Wenzel, "Spatial Sound and Sonification", *Auditory Display*, Kramer ed., Addison Wesley, April 1994, ISBN 0-201-62603-9, pp 127-150.

[ZOROLA 95]

R. Zorola Villarreal, "L'évaluation des IHMs Multi-utilisateurs dans le travail coopératif", *Thèse de Doctorat* soutenue à l'Université de Toulouse 1, 1995, 252 pages.

Annexes

1 Description du langage de commande proposé pour la gestion des audioconférences

1.1 Liste des commandes

[Ver]

retourne au demandeur la version du contrôleur d'ACAO utilisée par le serveur sous la forme :

```
[Ver 2.0]
```

[?]

retourne au demandeur la liste des commandes (inutile en usage normal) sous la forme :

```
[ commande1 ]  
[ commande2 ]  
...
```

[Canal]

retourne au demandeur la liste des canaux et leur description. Par exemple :

```
[Canal(1, "Comulab", "nom" )]  
[Canal(2, "Poste 1", "nom" )]  
...
```

où la première chaîne est un descriptif logique détenu par le contrôleur et la deuxième un descriptif logique modifiable par la commande suivante.

[NameCa(ca,"nom")]

définit la description logique d'un canal et diffuse le message :

```
[Canal(ca, "description physique", "nom" )]
```

[Conf]

retourne au demandeur la liste des audioconférences en cours avec la liste des canaux qui y participent. Par exemple :

```
[Conf(1, "n° 1", 1, 2, 4, 5)]
[Conf(2, "Apparté", 2, 4)]
```

qui signifie qu'il y a deux conférences en cours respectivement numérotées 1 et 2. La conférence n°1 concerne les canaux 1, 2, 4 et 5. La conférence n°2, qui a été nommée "Apparté", concerne les canaux 2 et 4.

[NameCo(co,"nom")]

définit la description logique d'une conférence et diffuse le message :

```
[Conf(co, "nom", <liste des canaux dans la conférence>)]
```

[New(ca₁,ca₂,...)]

crée une conférence automatiquement numérotée *i* et retourne d'abord au demandeur le message :

```
[Conf(i)]
```

puis à tout le monde le message :

```
[Conf(i, n° i", ca1, ca2, ...)]
```

[Del(i)]

supprime la conférence n°*i* et diffuse le message [Del(*i*)].

[In(co,ca₁,ca₂,...)]

ajoute à la conférence *co* les canaux *ca₁*, *ca₂*, ... et diffuse le message décrivant la conférence modifiée, par exemple :

```
[conf(co, "n° co", ..., ca1, ca2, ...)]
```

[Out(co,ca₁,ca₂,...)]

retire de la conférence *co* les canaux *ca₁*, *ca₂*, ... et diffuse le message décrivant la conférence modifiée, par exemple :

```
[Conf(co, ...)]
```

[Split(*co*,*ca*₁,*ca*₂,...)]

divise une conférence en deux ; c'est-à-dire : retire de la conférence *co* les canaux *ca*₁, *ca*₂, ... , crée une nouvelle conférence *co'* et y ajoute les canaux *ca*₁, *ca*₂, ... et diffuse les messages décrivant la conférence modifiée, et la nouvelle conférence par exemple :

```
[Conf (co, ... ) ]
[Conf (co', ... , ca1, ca2, ... ) ]
```

[Join(*co*₁,*co*₂)]

réunit deux conférences, c'est-à-dire : ajoute tous les canaux de la conférence *co*₁ à la conférence *co*₂ et supprime la conférence *co*₁ et diffuse les modifications engendrées.

[Appart(*co*,*ca*₁,*ca*₂)]

crée un aparté dans la conférence *co* avec les canaux *ca*₁ et *ca*₂ et diffuse les modifications induites (création d'une nouvelle conférence, modification des coefficients d'écoute et de parole pour les deux canaux. Dans les deux conférence).

[SetCalnCo(*co*,*ca*,*e*,*p*)]

définit le statut d'un canal *ca* dans une conférence *co* où *e* est le coefficient d'écoute et *p* est le coefficient de parole, ces deux coefficients étant compris entre 0 et 1. Tout le monde reçoit alors un message [CaInCo(*co*, *ca*, *e*, *p*)] signalant la modification engendrée. Par défaut un canal est ajouté à une conférence avec *e*=1 et *p*=0.

[GetCalnCo(*co*,*ca*)]

demande le statut d'un canal *ca* dans une conférence *co*, retourne au demandeur un message de type [CaInCo(*co*, *ca*, *e*, *p*)] .

[SetCaPos(*ca*_e,*ca*_p,*ρ*,*θ*)]

définit la position spatiale selon laquelle le canal *ca*_e perçoit le canal *ca*_p où *ρ* et *θ* décrivent les coordonnées polaires de la position de *ca*_p dans un plan horizontal où l'origine est à la position de *ca*_e. Ces deux coefficients étant compris entre 0.0

et 1.0. Pour l'angle, 0.0 correspond à l'extrême droite et 1.0 à l'extrême gauche. Pour la distance, la valeur 1.0 correspond à la position la plus éloignée possible dans la pièce virtuelle perçue par ca_e . Par défaut ρ est fixé pour les autres canaux à 0.5, et θ est fixé de telle manière que les différents canaux soient répartis équitablement. Pour ca_e $\rho=0$ et $\theta=0.5$.

[GetCaPos(ca_e, ca_p)]

demande la position spatiale selon laquelle le canal cae perçoit le canal cap , retourne au demandeur un message de type [CaPos (ca_e, ca_p, ρ, θ)].

[GetCaHeard(ca)]

demande la liste des canaux écoutés par le canal ca , retourne au demandeur un message de type [CaHeard ($ca, ca_i, ca_j, ca_k, \dots$)].

[ReTune]

effectue un re-réglage complet du dispositif physique de sorte qu'il reflète bien l'état logique représenté dans le contrôleur.

[Reset]

réinitialise le contrôleur audioconférence à un état connu : aucune conférence en cours et toutes valeurs par défaut rétablies. Envoie à tout le monde le message : [Reset]

1.2 Liste des messages retournés

[commande]

En réponse à une demande de liste de commande, normalement inutilisable en dehors d'un test manuel.

[Canal($ca, "description physique", "nom"$)]

Précise la description physique et le nom du canal n° ca .

[conf(co,"nom",ca₁,ca₂,...)]

Précise l'existence d'une conférence n° *co* appelée *nom* qui connecte les canaux *ca₁,ca₂, ...*

[Del(i)]

Précise la suppression de la conférence n° *i*.

[CaInCo(co,ca,e,p)]

Précise les coefficients d'écoute *e* et de parole *p* pour un canal *ca* dans une conférence *co*.

[CaPos(ca_e,ca_p,ρ,θ)]

Précise la position (distance ρ et angle θ) selon laquelle un canal *ca_e* écoute un canal *ca_p*.

[CaHeard(ca,ca₁,ca_j,ca_k,...)]

Précise la liste des canaux écoutés par *ca*.

[P(...,ca_i,...)]

Indique la liste des canaux correspondant à ceux qui parlent. Ce message est envoyé automatiquement à chaque fois que cette liste est modifiée.

[Reset]

Signifie que le contrôleur a été réinitialisé.

2 Première évaluation : résultats bruts

2.1 Exemple d'un fichier de mesures

Les résultats bruts de l'ensemble de l'expérience représentent 60 fichiers de mesures tel que l'exemple suivant :

Expérimentation de la discrimination spatiale par simple effet stéréo par MARTIN le 19/10/93 à 15:33:14 Modèle n° 2 Volume minimal: 0 Volume maximal: 255 Type de sortie: enceintes				
Test avec 3 positions:				
N°	rendue	perçue	écart	temps
01	3	3	0	01.70
02	1	1	0	01.49
03	2	2	0	01.04
04	1	1	0	00.99
05	3	3	0	00.99
06	2	2	0	00.99
07	3	3	0	00.99
08	1	1	0	00.98
09	2	2	0	01.05
10	3	3	0	01.70
11	2	2	0	01.26
12	1	1	0	01.05
13	3	3	0	00.99
14	2	2	0	01.59
15	1	1	0	02.37
16	1	1	0	01.04
17	2	2	0	01.10
18	3	3	0	00.83
19	1	1	0	01.21
20	2	2	0	00.76
21	3	3	0	00.99
22	3	3	0	01.10
23	2	2	0	01.92
24	1	1	0	01.93
25	3	3	0	00.93
26	1	1	0	00.71
27	2	2	0	01.15
28	3	3	0	00.83
29	2	2	0	01.70
30	1	1	0	01.26
31	1	1	0	00.93
32	3	3	0	03.41
33	2	2	0	01.92
34	2	2	0	00.82
35	3	3	0	01.21
36	1	1	0	01.15
37	3	3	0	01.04
38	1	1	0	01.38
39	2	2	0	01.37
40	2	2	0	01.10
41	1	1	0	02.25
42	3	3	0	01.26
43	1	1	0	00.94
44	3	3	0	00.99
45	2	2	0	01.15
46	2	2	0	00.88
47	1	1	0	01.65
48	3	3	0	00.93
49	3	3	0	01.26
50	1	1	0	01.65
51	2	2	0	01.21
52	3	3	0	00.82
53	2	2	0	01.15
54	1	1	0	01.37
55	3	3	0	01.38
56	2	2	0	01.15

Annexes

57	1	1	0	01.10
58	2	2	0	03.29
59	3	3	0	01.16
60	1	1	0	00.82
Bilan:				
Position	% réponses correctes		écart moyen	temps moyen
01	100.0		0.00	01.31
02	100.0		0.00	01.33
03	100.0		0.00	01.22

bilan	100.0		0.00	01.28
Test avec 4 positions:				
N°	rendue	perçue	écart	temps
01	1	1	0	01.10
02	2	2	0	01.59
03	4	4	0	01.05
04	3	3	0	00.93
05	2	2	0	01.15
06	3	3	0	01.05
07	4	4	0	01.15
08	1	1	0	01.10
09	1	1	0	00.93
10	4	4	0	02.09
11	2	2	0	01.70
12	3	2	1	01.59
13	4	4	0	02.04
14	2	2	0	01.32
15	3	3	0	01.04
16	1	1	0	01.65
17	4	4	0	01.59
18	3	3	0	01.60
19	2	2	0	00.93
20	1	1	0	01.15
21	3	3	0	01.76
22	2	2	0	02.25
23	1	1	0	01.70
24	4	4	0	01.21
25	2	2	0	01.54
26	1	1	0	02.64
27	4	4	0	02.14
28	3	3	0	01.32
29	4	4	0	01.92
30	1	1	0	01.43
31	3	3	0	01.81
32	2	2	0	01.49
33	2	2	0	01.64
34	1	1	0	01.49
35	3	3	0	02.08
36	4	4	0	01.27
37	4	4	0	01.10
38	1	1	0	01.59
39	3	3	0	02.09
40	2	2	0	01.76
41	4	4	0	01.60
42	3	3	0	01.64
43	1	1	0	01.16
44	2	2	0	01.21
45	3	3	0	02.58
46	1	1	0	01.53
47	4	4	0	04.89
48	2	2	0	07.36
49	2	2	0	01.26
50	3	3	0	01.37
51	1	1	0	02.26
52	4	4	0	03.62
53	2	2	0	04.66
54	4	4	0	01.32
55	3	3	0	01.27
56	1	1	0	00.98
57	3	3	0	02.09
58	4	4	0	01.59
59	1	1	0	01.65
60	2	2	0	02.14
61	3	3	0	01.16
62	4	4	0	01.15
63	1	1	0	01.98
64	2	2	0	01.27
65	2	2	0	02.69
66	1	1	0	01.65
67	3	3	0	01.37

Annexes

68	4	4	0	01.26
69	1	1	0	01.49
70	3	3	0	02.14
71	2	2	0	02.20
72	4	4	0	01.87
73	2	2	0	02.14
74	1	1	0	01.43
75	4	4	0	01.32
76	3	3	0	01.31
77	3	3	0	02.37
78	4	4	0	01.86
79	1	1	0	01.10
80	2	2	0	02.96
Bilan:				
Position	% réponses correctes		écart moyen	temps moyen
01	100.0		0.00	01.50
02	100.0		0.00	02.16
03	95.0		0.05	01.62
04	100.0		0.00	01.80
bilan	98.8		0.01	01.77
Test avec 5 positions:				
N°	rendue	perçue	écart	temps
01	3	3	0	01.43
02	1	1	0	01.48
03	5	5	0	01.43
04	4	4	0	01.05
05	2	2	0	01.15
06	5	5	0	01.49
07	1	1	0	01.98
08	2	2	0	01.15
09	3	3	0	01.10
10	4	4	0	01.48
11	4	4	0	02.03
12	3	3	0	01.98
13	5	5	0	01.04
14	1	1	0	01.54
15	2	2	0	01.26
16	4	4	0	01.76
17	3	3	0	01.15
18	2	2	0	02.09
19	5	5	0	01.15
20	1	1	0	01.98
21	3	2	1	01.59
22	1	1	0	02.03
23	2	2	0	01.81
24	4	4	0	01.38
25	5	5	0	01.37
26	4	4	0	01.76
27	5	5	0	01.43
28	3	3	0	01.15
29	2	2	0	01.65
30	1	1	0	01.65
31	1	1	0	02.69
32	5	5	0	02.03
33	3	3	0	01.98
34	4	4	0	03.02
35	2	2	0	01.32
36	4	4	0	09.83
37	5	5	0	01.26
38	1	1	0	01.26
39	2	2	0	00.83
40	3	3	0	00.99
41	2	2	0	01.48
42	4	4	0	02.03
43	5	5	0	00.99
44	3	3	0	01.59
45	1	1	0	01.32
46	4	4	0	01.70
47	3	3	0	01.48
48	1	1	0	02.31
49	2	2	0	01.43
50	5	5	0	01.42
51	1	1	0	01.65
52	5	5	0	01.05
53	3	3	0	01.92
54	2	2	0	03.19
55	4	4	0	01.10
56	1	1	0	01.26
57	5	5	0	01.15

Annexes

58	4	4	0	01.49
59	2	2	0	01.81
60	3	3	0	00.99
61	2	2	0	01.04
62	4	4	0	01.27
63	5	5	0	01.26
64	3	3	0	02.14
65	1	1	0	01.43
66	4	4	0	01.48
67	1	1	0	01.26
68	3	3	0	01.43
69	5	5	0	01.65
70	2	2	0	05.05
71	5	5	0	01.49
72	4	4	0	02.52
73	2	2	0	01.21
74	1	1	0	00.99
75	3	3	0	05.99
76	2	2	0	00.99
77	1	1	0	00.93
78	4	4	0	01.43
79	3	3	0	02.64
80	5	5	0	00.82
81	5	5	0	00.94
82	3	3	0	03.02
83	2	2	0	06.64
84	4	4	0	01.38
85	1	1	0	01.32
86	1	1	0	01.64
87	4	4	0	02.36
88	2	2	0	01.98
89	5	5	0	01.65
90	3	3	0	02.64
91	3	3	0	02.25
92	4	4	0	01.26
93	1	1	0	01.92
94	2	2	0	03.96
95	5	5	0	01.81
96	4	4	0	01.21
97	1	1	0	01.10
98	3	3	0	01.65
99	5	5	0	01.59
100	2	2	0	01.71
Bilan:				
Position	%	réponses correctes	écart moyen	temps moyen
<hr style="border-top: 1px dashed black;"/>				
01		100.0	0.00	01.58
02		100.0	0.00	02.08
03		95.0	0.05	01.95
04		100.0	0.00	02.07
05		100.0	0.00	01.35
<hr style="border-top: 1px dashed black;"/>				
bilan		99.0	0.01	01.81

Fin des tests à 15:41:26

2.2 Bilan des mesures

Un bilan des mesures réparties dans ces 60 fichiers a été obtenu par un programme spécifique dont le résultat est le suivant :

% de reponses exactes:	G	IG	C	ID	D
Modele 1 (20 sujets):	98,75	97,50	90,00	94,75	100,00
Modele 2 (20 sujets):	99,50	98,75	95,25	98,00	99,50
Modele 3 (20 sujets):	98,50	97,00	91,25	94,25	99,75
ecart moyen:	G	IG	C	ID	D
Modele 1 (20 sujets):	0,01	0,03	0,10	0,05	0,00
Modele 2 (20 sujets):	0,01	0,02	0,05	0,02	0,01
Modele 3 (20 sujets):	0,02	0,03	0,09	0,06	0,00
temps moyen:	G	IG	C	ID	D
Modele 1 (20 sujets):	1,55	1,70	1,77	1,71	1,47
Modele 2 (20 sujets):	1,56	1,62	1,64	1,72	1,46
Modele 3 (20 sujets):	1,55	1,72	1,79	1,93	1,57

3 Deuxième évaluation

3.1 Documents utilisés

Le codage des agendas utilisés pour la recherche d'un ensemble d'agendas convenables, ainsi que la liste des solutions offertes déterminées par programme sont présentés ci-dessous. La version des agendas présentée aux sujets est fournie dans les pages qui suivent.

n 1	1	2	3	4	5	8	9	10	11	12	15	16	17	18	19	22	23	24	25	26	29	30	31
8h00	I	I	13	I	I	I	12	I	124	123	A	14	L	I	I	A	I	I	12	I	I	I	I
9h00	A	A	A	14	I	A	A	I	L	I	12	I	I	I	I	I	I	I	A	I	I	I	A
10h00	I	A	L	I	I	I	A	I	123	I	A	124	I	I	I	A	I	I	A	I	I	I	I
11h00	L	13	I	I	I	A	L	I	I	A	L	I	A	L	I	A	134	L	I	I	14	12	124
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14h00	123	A	L	124	I	14	L	I	134	I	124	L	I	I	I	I	L	I	I	I	I	I	I
15h00	L	123	I	A	A	I	124	I	I	L	I	L	A	I	13	I	L	L	I	L	13	A	12
16h00	I	13	I	A	I	L	I	L	12	A	I	I	123	L	A	I	I	L	I	L	I	A	I
17h00	I	L	123	13	I	L	A	L	I	I	124	I	I	I	L	I	I	I	I	I	L	L	I
n 2	1	2	3	4	5	8	9	10	11	12	15	16	17	18	19	22	23	24	25	26	29	30	31
8h00	A	I	I	A	I	24	12	I	124	123	I	I	I	A	A	24	I	A	12	I	I	L	I
9h00	I	24	L	A	L	L	234	I	I	I	12	L	L	L	A	I	I	234	I	I	I	L	A
10h00	I	L	I	23	I	I	A	I	123	L	A	124	I	A	L	234	I	L	I	L	A	I	I
11h00	I	I	I	I	24	L	I	I	I	I	I	234	I	L	L	I	I	I	I	I	A	12	124
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14h00	123	I	I	124	A	L	I	I	L	I	124	I	I	I	L	I	L	234	A	L	I	I	I
15h00	23	123	I	I	A	124	I	L	L	L	I	L	A	A	A	A	A	24	I	L	24	12	
16h00	I	I	24	I	L	A	I	L	12	I	I	L	123	I	I	I	I	I	A	A	A	L	L
17h00	L	L	123	I	I	A	A	A	A	L	124	L	I	I	24	L	L	I	I	I	A	L	L
n 3	1	2	3	4	5	8	9	10	11	12	15	16	17	18	19	22	23	24	25	26	29	30	31
8h00	A	I	13	I	A	A	L	I	I	123	A	I	I	L	I	L	I	I	I	I	I	I	I
9h00	L	L	I	I	I	A	234	A	I	I	I	I	I	A	I	A	L	234	I	I	I	L	L
10h00	I	I	I	23	A	I	I	I	123	I	A	I	A	I	A	234	I	L	I	I	A	34	
11h00	I	13	A	I	A	L	I	L	L	A	234	L	I	L	L	134	L	L	A	L	I	I	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14h00	123	I	A	I	A	L	I	I	134	A	I	I	L	L	I	A	234	I	I	A	L	I	I
15h00	23	123	A	I	A	I	I	I	I	34	I	I	I	13	L	A	I	L	A	13	I	A	
16h00	A	13	L	A	A	I	I	34	I	A	I	L	123	A	A	A	I	I	A	I	A	I	L
17h00	I	L	123	13	A	I	I	A	I	A	L	A	L	I	A	L	I	L	A	I	A	I	I
n 4	1	2	3	4	5	8	9	10	11	12	15	16	17	18	19	22	23	24	25	26	29	30	31
8h00	I	L	I	I	I	24	I	L	124	A	I	14	I	I	A	24	I	I	A	I	A	L	I
9h00	A	24	A	14	I	I	234	L	L	A	I	I	I	I	I	L	A	234	I	L	A	I	I
10h00	I	I	L	I	I	I	A	I	L	I	L	124	I	I	I	234	I	A	I	L	I	34	
11h00	I	A	I	L	24	A	A	I	I	I	I	234	I	I	L	I	134	I	I	I	14	L	124
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14h00	I	I	L	124	I	14	I	I	134	A	I	124	A	A	L	I	I	234	I	I	I	I	I
15h00	I	L	L	I	I	I	124	I	I	I	34	I	I	A	A	A	I	I	24	I	I	24	L
16h00	A	I	24	I	L	I	I	34	A	I	I	I	I	I	A	L	I	I	A	L	I	I	I
17h00	A	I	I	A	I	I	L	A	I	A	124	I	L	I	24	I	L	I	L	I	L	I	L

Entre le 01 et le 15:
si Rdv interne pour 2 et 4 du 2 à 9h00 reporté au 5 à 16h00,
si Rdv de 1 du 2 à 9h00 annulé,
soit 1 annulation(s) et 1 report(s),
Rdv possible le 2 à 9h00 ...

si Rdv interne pour 1, 2 et 3 du 2 à 15h00 reporté au 2 à 17h00,
soit 0 annulation(s) et 1 report(s),
Rdv possible le 2 à 15h00 ...

si Rdv de 4 du 8 à 11h00 annulé,
si Rdv de 1 du 8 à 11h00 annulé,
soit 2 annulation(s) et 0 report(s),

Annexes

Rdv possible le 8 à 11h00 ...

si Rdv interne pour 1 et 4 du 8 à 14h00 reporté au 3 à 10h00,
soit 0 annulation(s) et 1 report(s),
Rdv possible le 8 à 14h00 ...

si Rdv interne pour 2, 3 et 4 du 9 à 9h00 reporté au 2 à 9h00,
si Rdv interne pour 2 et 4 du 2 à 9h00 reporté au 5 à 16h00,
si Rdv de 1 du 9 à 9h00 annulé,
soit 1 annulation(s) et 2 report(s),
Rdv possible le 9 à 9h00 ...

si Rdv interne pour 3 et 4 du 10 à 16h00 reporté au 2 à 9h00,
si Rdv interne pour 2 et 4 du 2 à 9h00 reporté au 5 à 16h00,
soit 0 annulation(s) et 2 report(s),
Rdv possible le 10 à 16h00 ...

si Rdv de 4 du 10 à 17h00 annulé,
si Rdv de 3 du 10 à 17h00 annulé,
si Rdv de 2 du 10 à 17h00 annulé,
soit 3 annulation(s) et 0 report(s),
Rdv possible le 10 à 17h00 ...

si Rdv interne pour 1, 3 et 4 du 11 à 14h00 reporté au 2 à 15h00,
si Rdv interne pour 1, 2 et 3 du 2 à 15h00 reporté au 2 à 17h00,
soit 0 annulation(s) et 2 report(s),
Rdv possible le 11 à 14h00 ...

si Rdv interne pour 1, 2 et 3 du 12 à 8h00 reporté au 2 à 17h00,
si Rdv de 4 du 12 à 8h00 annulé,
soit 1 annulation(s) et 1 report(s),
Rdv possible le 12 à 8h00 ...

si Rdv interne pour 1, 2 et 3 du 12 à 10h00 reporté au 2 à 17h00,
soit 0 annulation(s) et 1 report(s),
Rdv possible le 12 à 10h00 ...

Il y a 10 solution(s).

La meilleure solution est:

si Rdv interne pour 1, 2 et 3 du 2 à 15h00 reporté au 2 à 17h00,
soit 0 annulation(s) et 1 report(s),
Rdv possible le 2 à 15h00 ...

Entre le 15 et le 32:

si Rdv interne pour 1, 2 et 4 du 15 à 17h00 reporté au 19 à 17h00,
si Rdv interne pour 2 et 4 du 19 à 17h00 reporté au 19 à 11h00,
soit 0 annulation(s) et 2 report(s),
Rdv possible le 15 à 17h00 ...

si Rdv interne pour 1 et 3 du 19 à 15h00 reporté au 24 à 11h00,
si Rdv de 4 du 19 à 15h00 annulé,
si Rdv de 2 du 19 à 15h00 annulé,
soit 2 annulation(s) et 1 report(s),
Rdv possible le 19 à 15h00 ...

si Rdv interne pour 2 et 4 du 19 à 17h00 reporté au 19 à 11h00,
si Rdv de 3 du 19 à 17h00 annulé,
soit 1 annulation(s) et 1 report(s),
Rdv possible le 19 à 17h00 ...

si Rdv interne pour 2 et 4 du 22 à 8h00 reporté au 19 à 11h00,
si Rdv de 1 du 22 à 8h00 annulé,
soit 1 annulation(s) et 1 report(s),
Rdv possible le 22 à 8h00 ...

si Rdv interne pour 2, 3 et 4 du 23 à 10h00 reporté au 19 à 11h00,
si Rdv de 1 du 23 à 10h00 annulé,
soit 1 annulation(s) et 1 report(s),
Rdv possible le 23 à 10h00 ...

si Rdv interne pour 1 et 4 du 29 à 11h00 reporté au 26 à 16h00,
si Rdv de 2 du 29 à 11h00 annulé,
soit 1 annulation(s) et 1 report(s),
Rdv possible le 29 à 11h00 ...

si Rdv de 3 du 29 à 17h00 annulé,
si Rdv de 2 du 29 à 17h00 annulé,
soit 2 annulation(s) et 0 report(s),
Rdv possible le 29 à 17h00 ...

si Rdv interne pour 1 et 2 du 31 à 15h00 reporté au 18 à 11h00,
si Rdv de 3 du 31 à 15h00 annulé,

Annexes

soit 1 annulation(s) et 1 report(s),
Rdv possible le 31 à 15h00 ...

Il y a 8 solution(s).

La meilleure solution est:

si Rdv interne pour 1, 2 et 4 du 15 à 17h00 reporté au 19 à 17h00,

si Rdv interne pour 2 et 4 du 19 à 17h00 reporté au 19 à 11h00,

soit 0 annulation(s) et 2 report(s),

Rdv possible le 15 à 17h00 ...

Peintre

	Lu 1	Ma 2	Me 3	Je 4	Ve 5	we	Lu 8	Ma 9	Me 10	Je 11	Ve 12	we	Lu 15	Ma 16	Me 17	Je 18	Ve 19	we	Lu 22	Ma 23	Me 24	Je 25	Ve 26	we	Lu 29	Ma 30	Me 31	
8h00			Pein Maço					Pein Elec		Pein Elec Plom	Pein Elec Maço		Pein Plom							Pein Elec						
9h00	Pein Plom							Pein Elec								
10h00			Pein Elec Maço			Pein Elec Plom									
11h00		Pein Maço					Pein Maço Plom					Pein Plom	Pein Elec	Pein Elec Plom
14h00	Pein Elec Maço		Pein Elec Plom			Pein Plom			Pein Maço Plom			Pein Elec Plom															
15h00		Pein Elec Maço				Pein Elec Plom								Pein Maço									Pein Maço	Pein Elec
16h00		Pein Maço						Pein Elec				Pein Elec Maço			
17h00			Pein Elec Maço	Pein Maço								Pein Elec Plom															

Légende

	Rendez-vous inamovible
....	Rendez-vous annulable
Elec Plom	Rendez-vous interne reportable (exemple)
	Créneau horaire libre

Electricien

	Lu 1	Ma 2	Me 3	Je 4	Ve 5	we	Lu 8	Ma 9	Me 10	Je 11	Ve 12	we	Lu 15	Ma 16	Me 17	Je 18	Ve 19	we	Lu 22	Ma 23	Me 24	Je 25	Ve 26	we	Lu 29	Ma 30	Me 31	
8h00			Elec Plom	Pein Elec		Pein Elec Plom	Pein Elec Maço						Elec Plom		Pein Elec						
9h00		Elec Plom					Elec Maço Plom					Pein Elec							Elec Maço Plom						
10h00				Elec Maço						Pein Elec Maço			Pein Elec Plom				Elec Maço Plom							
11h00					Elec Plom									Elec Maço Plom												Pein Elec	Pein Elec Plom
14h00	Pein Elec Maço			Pein Elec Plom								Pein Elec Plom								Elec Maço Plom						
15h00	Elec Maço	Pein Elec Maço					Pein Elec Plom									Elec Plom					Elec Plom	Pein Elec
16h00			Elec Plom						Pein Elec					Pein Elec Maço								
17h00			Pein Elec Maço						Pein Elec Plom					Elec Plom									

Légende

	Rendez-vous inamovible
....	Rendez-vous annulable
Elec Plom	Rendez-vous interne reportable (exemple)
	Créneau horaire libre

Maçon

	Lu 1	Ma 2	Me 3	Je 4	Ve 5	we	Lu 8	Ma 9	Me 10	Je 11	Ve 12	we	Lu 15	Ma 16	Me 17	Je 18	Ve 19	we	Lu 22	Ma 23	Me 24	Je 25	Ve 26	we	Lu 29	Ma 30	Me 31	
8h00		Pein Maço							Pein Elec Maço																
9h00							Elec Maço Plom		Elec Maço Plom							
10h00				Elec Maço						Pein Elec Maço					Elec Maço Plom						Maço Plom	
11h00		Pein Maço	Elec Maço Plom							Pein Maço Plom						
14h00	Pein Elec Maço						Pein Maço Plom	Elec Maço Plom						
15h00	Elec Maço	Pein Elec Maço								Maço Plom				Pein Maço				Pein Maço		
16h00	Pein Maço				Maço Plom						Pein Elec Maço			
17h00			Pein Elec Maço	Pein Maço			

Légende

	Rendez-vous inamovible
....	Rendez-vous annulable
Elec Plom	Rendez-vous interne reportable (exemple)
	Créneau horaire libre

Plombier

	Lu 1	Ma 2	Me 3	Je 4	Ve 5	we	Lu 8	Ma 9	Me 10	Je 11	Ve 12	we	Lu 15	Ma 16	Me 17	Je 18	Ve 19	we	Lu 22	Ma 23	Me 24	Je 25	Ve 26	we	Lu 29	Ma 30	Me 31	
8h00							Elec Plom			Pein Elec Plom			Pein Plom				Elec Plom					
9h00	Elec Plom	Pein Plom				Elec Maço Plom			Elec Maço Plom						
10h00														Pein Elec Plom					Elec Maço Plom							Maço Plom
11h00				Elec Plom							Elec Maço Plom						Pein Maço Plom					Pein Plom		Pein Elec Plom	
14h00				Pein Elec Plom			Pein Plom			Pein Maço Plom				Pein Elec Plom					Elec Maço Plom							
15h00								Pein Elec Plom						Maço Plom					Elec Plom				Elec Plom		
16h00		Elec Plom						Maço Plom						
17h00		Pein Elec Plom				Elec Plom											

Légende

	Rendez-vous inamovible
....	Rendez-vous annulable
Elec Plom	Rendez-vous interne reportable (exemple)
	Créneau horaire libre

La consigne distribuée au sujet fut la suivante (version destinée au peintre) :

Consigne

Vous êtes Peintre et vous devez réaliser un chantier (construction d'une piscine) en concertation avec d'autres professionnels (maçon, électricien et plombier). Afin de coordonner vos actions pour une parfaite mise en oeuvre du chantier, il vous est nécessaire de fixer ensemble deux dates pour deux rendez-vous communs respectivement avant le 15 (exclus) et après le 15 (inclus). Etant actuellement chacun sur un autre chantier, vous allez communiquer par un système d'audio-conférence vous reliant par voie sonore aux trois autres personnes.

Pour résoudre ce problème, vous avez chacun sous les yeux un agenda qui décrit votre emploi du temps personnel. Pour fixer un nouveau rendez-vous :

- Vous pouvez être amené à annuler un rendez-vous "annulable" (...) mais vous chercherez à l'éviter.
- Vous devez vous efforcer de trouver ce nouveau rendez-vous le plus tôt possible dans la période considérée (avant ou après le 15).
- Vous pouvez éventuellement déplacer un ou plusieurs rendez-vous "interne" (rendez-vous déjà pris avec un ou deux des trois autres professionnels) mais seulement dans la période considérée et aussi le plus tôt possible.

Vous avez 20 minutes au maximum pour trouver les deux rendez-vous, mais comme l'audio-conférence coûte cher, vous avez intérêt à être le plus rapide possible.

BON COURAGE!

Le questionnaire rempli par les sujets fut le suivant :

Questionnaire

1) Avez-vous constaté une différence dans la qualité du son entre les deux parties de la tâche ?

- oui
- non

2) Cela change-t-il la tâche ? ^(*)

- le confort d'écoute est plus important
- le son est plus volumétrique
- c'est plus agréable
- meilleure identification des partenaires
- conditions plus proche d'une conversation en vis à vis
- on se représente mieux le rôle de chacun
- autre

3) Quelle partie de la tâche vous a semblé meilleure au niveau du son ?

- 1ère partie
- 2ème partie

4) Propositions d'amélioration :

.....
.....

(•) Vous pouvez cocher plusieurs cases

Remarque importante:

Ils vous est expressément demandé de ne pas parler de cette expérimentation autour de vous jusqu'au 22 avril afin de ne pas fausser les résultats.

3.2 Résultats bruts

3.2.1 Mesures

```

Résultats
Groupe 1
Type initial: mono
Phase 1: 1.3 minutes
Phase 2: 4.5 minutes : Rdv le 10 à 17h00 : 3 A
Phase 3: 4.2 minutes : Rdv le 29 à 17h00 : 2 A
Total: 10.0 minutes
Groupe 2
Type initial: stéréo
Phase 1: 1.6 minutes
Phase 2: 6.3 minutes : Rdv le 08 à 14h00 : 1 R
Phase 3: 3.4 minutes : Rdv le 19 à 15h00 : 2 A, 1 R le 15 à 8h00 : 2 A
Total: 11.3 minutes
Groupe 3
Type initial: mono
Phase 1: 2.1 minutes
Phase 2: 7.5 minutes : Rdv le 02 à 15h00 : 1 RA
Phase 3: 2.0 minutes : Rdv le 15 à 17h00 : 1 RA
Total: 11.6 minutes
Groupe 4
Type initial: stéréo
Phase 1: 2.9 minutes
Phase 2: 3.6 minutes : Rdv le 08 à 11h00 : 2A
Phase 3: 3.6 minutes : Rdv le 29 à 17h00 : 2A
Total: 10.0 minutes
Groupe 5
Type initial: mono
Phase 1: 2.0 minutes
Phase 2: 3.2 minutes : Rdv le 02 à 09h00 : 1A, 2 RA
Phase 3: 8.0 minutes : Rdv le 29 à 17h00 : 2A
Total: 13.2 minutes
Groupe 6
Type initial: stéréo
Phase 1: 2.0 minutes
Phase 2: 4.7 minutes : Rdv le 02 à 09h00 : 1A, 1R : 1A
Phase 3: 7.0 minutes : Rdv le 19 à 15h00 : 2A, 1 RA
Total: 13.7 minutes
Groupe 7
Type initial: mono
Phase 1: 1.9 minutes
Phase 2: 6.7 minutes : Rdv le 08 à 11h00 : 2A
Phase 3: 2.4 minutes : Rdv le 15 à 17h00 : 1R : 2A
Total: 10.9 minutes
Groupe 8
Type initial: stéréo
Phase 1: 2.0 minutes
Phase 2: 3.6 minutes : Rdv le 08 à 14h00 : 1RA
Phase 3: 4.8 minutes : Rdv le 19 à 15h00 : 3A 1R
Total: 10.4 minutes
Groupe 9
Type initial: mono
Phase 1: 2.4 minutes
Phase 2: 3.3 minutes : Rdv le 02 à 09h00 : 1A, 1R : 1A
Phase 3: 6.4 minutes : Rdv le 15 à 17h00 : 1R : 1A
Total: 12.1 minutes
Groupe 10
Type initial: stéréo
Phase 1: 2.3 minutes
Phase 2: 3.8 minutes : Rdv le 02 à 09h00 : 1A, 1R : 1 A
Phase 3: 2.1 minutes : Rdv le 22 à 08h00 : 1A, 1R
Total: 8.2 minutes
Groupe 11
Type initial: mono
Phase 1: 2.1 minutes
Phase 2: 3.0 minutes : Rdv le 10 à 16h00 : 1RA
Phase 3: 1.7 minutes : Rdv le 29 à 11h00 : 1A, 1RA
Total: 6.8 minutes
Groupe 12
Type initial: stéréo
Phase 1: 1.9 minutes
Phase 2: 3.8 minutes : Rdv le 02 à 15h00 : 1R
Phase 3: 4.5 minutes : Rdv le 19 à 15h00 : 2A, 1R : 1A
Total: 10.2 minutes
Groupe 13

```

Annexes

Type initial: mono
Phase 1: 1.9 minutes
Phase 2: 6.7 minutes : Rdv le 02 à 15h00 : 1R
Phase 3: 2.8 minutes : Rdv le 15 à 17h00 : 1R : 1A
Total: 11.4 minutes
Groupe 14
Type initial: stéréo
Phase 1: 2.0 minutes
Phase 2: 3.3 minutes : Rdv le 02 à 09h00 : 1A, 1RA
Phase 3: 2.1 minutes : Rdv le 22 à 08h00 : 1A, 1RA
Total: 7.5 minutes
Groupe 15
Type initial: mono
Phase 1: 2.1 minutes
Phase 2: 6.3 minutes : Rdv le 08 à 11h00 : 2A
Phase 3: 4.8 minutes : Rdv le 29 à 17h00 : 2A
Total: 13.1 minutes
Groupe 16
Type initial: stéréo
Phase 1: 2.2 minutes
Phase 2: 3.1 minutes : Rdv le 08 à 11h00 : 2A
Phase 3: 3.1 minutes : Rdv le 29 à 17h00 : 2A
Total: 8.4 minutes
Groupe 17
Type initial: mono
Phase 1: 2.1 minutes
Phase 2: 2.6 minutes : Rdv le 08 à 11h00 : 2A
Phase 3: 5.6 minutes : Rdv le 29 à 17h00 : 2A
Total: 10.3 minutes
Groupe 18
Type initial: stéréo
Phase 1: 2.5 minutes
Phase 2: 7.2 minutes : Rdv le 02 à 09h00 : 3A, 1R
Phase 3: 1.9 minutes : Rdv le 15 à 17h00 : 1RA
Total: 11.6 minutes
Groupe 19
Type initial: mono
Phase 1: 2.3 minutes
Phase 2: 4.6 minutes : Rdv le 08 à 11h00 : 2A
Phase 3: 3.3 minutes : Rdv le 29 à 11h00 : 1A, 1R
Total: 10.2 minutes
Groupe 20
Type initial: stéréo
Phase 1: 2.0 minutes
Phase 2: 4.2 minutes : Rdv le 08 à 11h00 : 2A
Phase 3: 1.1 minutes : Rdv le 15 à 17h00 : 1RA
Total: 7.3 minutes
Groupe 21
Type initial: mono
Phase 1: 2.1 minutes
Phase 2: 3.5 minutes : Rdv le 02 à 15h00 : 1R
Phase 3: 7.8 minutes : Rdv le 15 à 17h00 : 2R
Total: 13.5 minutes
Groupe 22
Type initial: stéréo
Phase 1: 2.1 minutes
Phase 2: 5.5 minutes : Rdv le 08 à 11h00 : 2A
Phase 3: 4.8 minutes : Rdv le 29 à 17h00 : 2A
Total: 12.4 minutes
Groupe 23
Type initial: mono
Phase 1: 2.0 minutes
Phase 2: 9.2 minutes : Rdv le 08 à 11h00 : 2A
Phase 3: 9.0 minutes : Rdv le 19 à 17h00 : 1A, 1R
Total: 20.1 minutes
Groupe 24
Type initial: stéréo
Phase 1: 2.0 minutes
Phase 2: 3.1 minutes : Rdv le 02 à 09h00 : 2A, 1R
Phase 3: 1.6 minutes : Rdv le 15 à 17h00 : 1R : 2A
Total: 6.8 minutes
Groupe 25
Type initial: mono
Phase 1: 2.0 minutes
Phase 2: 7.0 minutes : Rdv le 08 à 11h00 : 2A
Phase 3: 1.5 minutes : Rdv le 15 à 17h00 : 1R : 2A
Total: 10.4 minutes

Annexes

Groupe 26
Type initial: stéréo
Phase 1: 2.0 minutes
Phase 2: 8.6 minutes : Rdv le 02 à 15h00 : 1R
Phase 3: 3.1 minutes : Rdv le 15 à 17h00 : 2R
Total: 13.7 minutes

Annexes

3.2.2 Questionnaires

N°	type	1a	1b	2a	2b	2c	2d	2e	2f	2g	1ère	2ème	mono	stéré o
1	Elec	1		1	1			1				1	0	1
	Maço	1			1		1					1	0	1
	Pein	1		1		1	1				1		1	0
	Plom	1		1	1		1			1	1		1	0
2	Elec		1										0	0
	Maço	1		1	1	1	1				1		0	1
	Pein	1				1	1		1			1	1	0
	Plom		1										0	0
3	Elec	1		1		1	1			1		1	0	1
	Maço	1		1			1	1				1	0	1
	Pein	1		1	1					1		1	0	1
	Plom	1		1	1		1	1				1	0	1
4	Elec		1										0	0
	Maço	1					1	1		1		1	1	0
	Pein	1						1	1			1	1	0
	Plom	1						1				1	1	0
7	Elec		1										0	0
	Maço	1										1	0	1
	Pein	1		1		1	1					1	0	1
	Plom	1		1	1		1					1	0	1
8	Elec	1		1		1		1				1	1	0
	Maço	1					1	1			1		0	1
	Pein		1										0	0
	Plom		1										0	0
9	Elec	1		1	1		1	1				1	0	1
	Maço	1		1	1	1	1	1		1		1	0	1
	Pein	1			1		1	1				1	0	1
	Plom	1		1		1	1					1	0	1
10	Elec	1		1		1	1					1	1	0
	Maço		1										0	0
	Pein	1						1	1			1	1	0
	Plom		1										0	0
12	Elec	1			1						1		0	1
	Maço		1										0	0
	Pein	1			1				1	1			0	1
	Plom	1			1			1				1	1	0
13	Elec	1			1	1						1	0	1
	Maço	1					1	1		1		1	0	1
	Pein	1		1	1			1				1	0	1
	Plom	1		1	1	1	1					1	0	1
14	Elec	1			1	1	1				1		0	1
	Maço		1										0	0
	Pein		1										0	0
	Plom	1		1		1	1					1	1	0
15	Elec	1			1			1				1	0	1
	Maço	1			1		1	1				1	0	1
	Pein	1				1	1				1		1	0
	Plom	1		1		1	1					1	0	1
16	Elec		1										0	0
	Maço	1		1		1						1	1	0
	Pein		1										0	0
	Plom		1										0	0
17	Elec	1			1		1					1	0	1
	Maço	1		1		1	1		1		1		1	0
	Pein	1			1		1	1				1	0	1
	Plom		1										0	0
19	Elec	1					1					1	0	1

Annexes

	Maço		1									0	0
	Pein		1									0	0
	Plom		1									0	0
20	Elec	1		1	1		1				1	1	0
	Maço	1				1	1			1		0	1
	Pein		1									0	0
	Plom	1		1	1	1	1		1		1	1	0
21	Elec		1									0	0
	Maço		1									0	0
	Pein	1			1		1	1			1	0	1
	Plom	1							1	1		1	0
22	Elec	1				1	1			1		0	1
	Maço		1									0	0
	Pein		1									0	0
	Plom	1			1		1				1	1	0
23	Elec		1									0	0
	Maço	1		1		1	1		1		1	0	1
	Pein	1		1		1	1	1		1		1	0
	Plom		1									0	0
24	Elec		1									0	0
	Maço		1									0	0
	Pein		1									0	0
	Plom		1									0	0
25	Elec	1		1	1			1			1	0	1
	Maço		1									0	0
	Pein	1		1				1			1	0	1
	Plom	1		1	1				1		1	0	1
26	Elec		1									0	0
	Maço	1			1					1		0	1
	Pein	1								1		0	1
	Plom		1									0	0

