

th 2000 - 501  
Numéro d'ordre : 2462

Année : 1998

# THESE

présentée à

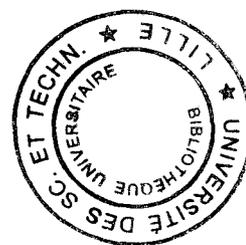
L'Université des Sciences et Technologies de Lille

pour obtenir le titre de

**Docteur en Informatique**

par

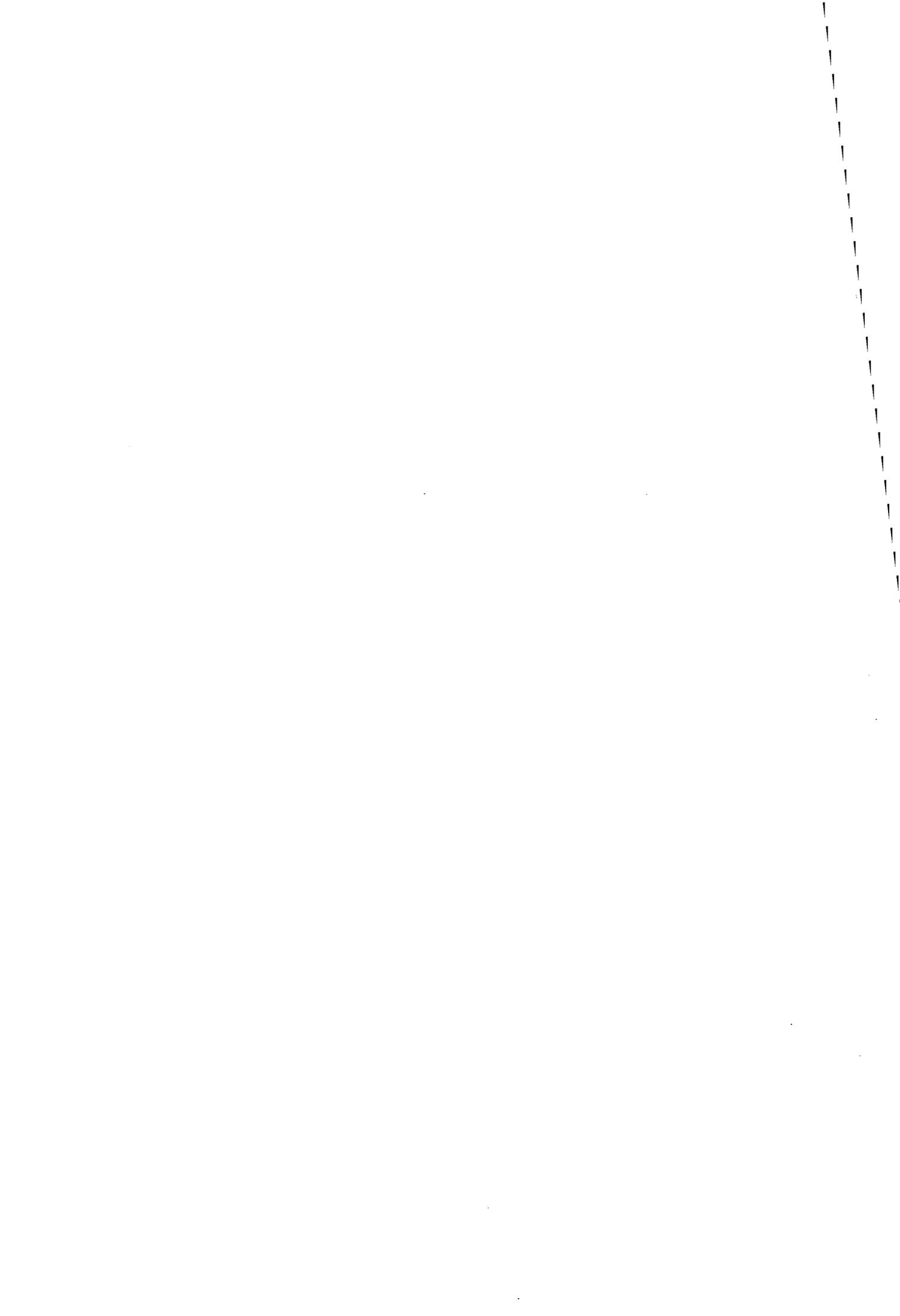
**Raouf Boulkroune**



## **Gestion de la Présentation Vidéo pour les Applications Coopératives dans un Milieu Hétérogène.**

Soutenue le 18 Décembre 1998, devant la commission d'examen :

|                             |                        |                                  |
|-----------------------------|------------------------|----------------------------------|
| <i>Président du Jury :</i>  | M. Christophe CHAILLOU | Université de Lille 1            |
| <i>Rapporteurs :</i>        | M. Christophe KOLSKI   | Université de Valenciennes       |
|                             | M. Michel TREHEL       | Université de Franche-Comté      |
| <i>Directeur de Thèse :</i> | M. Alain DERYCKE       | Université de Lille 1            |
| <i>Examineurs :</i>         | M. Jean VANDERDONCKT   | Université de Louvain (Belgique) |
|                             | M. Pierre VINCENT      | ENIC                             |



## Remerciements

Je tiens à remercier les membres de mon jury :

Monsieur Christophe KOLSKI, Professeur à l'Université de Valenciennes, et Monsieur Michel TREHEL, Professeur à l'Université de Franche-Comté d'avoir accepté de rapporter cette thèse.

Monsieur Christophe CHAILLOU, Professeur à l'Université de Lille 1, d'avoir accepté d'être président de mon jury de soutenance.

Monsieur Alain DERYCKE, Professeur à l'Université de Lille 1, directeur de ma thèse, de m'avoir accueilli au sein de l'équipe NOCE au laboratoire TRIGONE, ainsi que pour ses précieux conseils lors de nos nombreuses discussions.

Monsieur Pierre VINCENT, directeur d'étude à l'ENIC, de m'avoir accepté au sein du Département Informatique de l'ENIC, d'avoir dirigé mon travail de thèse et pour ses précieux conseils durant le long de la thèse.

Monsieur Jean VANDERDONCKT, Professeur adjoint à l'Université Catholique de Louvain (Belgique) d'avoir accepté d'examiner mon travail.

Je tiens également à remercier les nombreuses personnes qui m'ont beaucoup aidé durant ma thèse à l'ENIC et à TRIGONE. Je citerai particulièrement Sabine LEROY, Stanislas MATUSIAK et Jean-Philippe VANDEBORRE à l'ENIC. Isabelle LOGEZ, Jean Claude TARBY et Nadège NEAU à TRIGONE.

Je n'oublierai pas de remercier ici mon ami de longue date Nabil BELHOUANE pour l'aide inestimable qu'il m'a donné afin d'achever mon travail de thèse.

Enfin, je remercie les membres de ma famille pour leurs encouragements. Ce travail leur est dédié.



## SOMMAIRE

|   |           |
|---|-----------|
| <b>1. INTRODUCTION GÉNÉRALE.....</b>  | <b>11</b> |
| <b>2. LE TRAVAIL COOPÉRATIF SYNCHRONE.....</b>  | <b>15</b> |
| 2.1. INTRODUCTION.....  | 16        |
| 2.2. LE TRAVAIL COOPÉRATIF.....   | 17        |
| 2.2.1. <i>Les dimensions Espace et Temps du CSCW</i> .....                            | 17        |
| 2.3. LES SYSTÈMES DE VIDÉOCONFÉRENCE.....   | 19        |
| 2.3.1. <i>Définitions</i> .....   | 20        |
| 2.3.2. <i>La gestion de conférence</i> .....  | 21        |
| 2.3.2.1. Création d'une conférence.....   | 21        |
| 2.3.2.2. Pour participer : se joindre ou être invité.....                             | 22        |
| 2.3.2.3. Comment participer ?.....  | 23        |
| 2.3.2.4. Pour sortir : quitter ou être éjecté.....                                    | 24        |
| 2.3.2.5. Terminer une conférence.....   | 24        |
| 2.3.3. <i>L'espace partagé</i> .....  | 24        |
| 2.3.3.1. Le tableau blanc.....  | 25        |
| 2.3.3.2. Les applications partagées.....  | 26        |
| 2.3.3.3. Les télépointeurs.....   | 26        |
| 2.3.4. <i>T.120, une norme pour la visioconférence et le travail coopératif</i> ..... | 28        |
| 2.3.4.1. L'infrastructure de communication de T.120.....                              | 29        |
| 2.3.4.2. Les protocoles d'applications.....   | 31        |
| 2.3.5. <i>Conclusion</i> .....  | 31        |
| 2.4. LE MEDIA SPACE.....  | 32        |
| 2.4.1. <i>Définitions</i> .....   | 32        |
| 2.4.2. <i>Les services d'un Media Space</i> .....                                     | 33        |
| 2.4.3. <i>Les expériences majeures du Media Space</i> .....                           | 35        |
| 2.4.4. <i>Conclusion</i> .....  | 36        |
| 2.5. LA VIDÉO, SOURCE D'INFORMATIONS.....   | 36        |
| 2.5.1. <i>Le comportement visible des participants</i> .....                          | 37        |
| 2.5.1.1. Le regard.....   | 37        |
| 2.5.1.2. L'expression du visage et les mouvements de la tête.....                     | 38        |
| 2.5.1.3. La gestuelle.....  | 39        |
| 2.5.1.4. La posture.....  | 39        |
| 2.5.2. <i>Discussion</i> .....  | 39        |
| 2.6. CONCLUSION.....  | 40        |
| <b>3. HÉTÉROGÉNÉITÉ DES SYSTÈMES ET DIVERSITÉ DES ACTIVITÉS COOPÉRATIVES .....</b>    | <b>43</b> |

|           |  |           |
|-----------|--|-----------|
| 3.1.      | INTRODUCTION.....  | 44        |
| 3.2.      | DIVERSITÉ DES ACTIVITÉS COOPÉRATIVES .....                                   | 44        |
| 3.2.1.    | <i>L'enseignement à distance</i> .....                                       | 45        |
| 3.2.2.    | <i>La télé-médecine</i> .....  | 46        |
| 3.2.3.    | <i>La télé-réunion</i> .....   | 47        |
| 3.2.4.    | <i>Conclusion</i> .....  | 48        |
| 3.3.      | DIVERSITÉ DES ÉQUIPEMENTS UTILISATEURS .....                                 | 48        |
| 3.3.1.    | <i>Les systèmes de type poste de travail</i> .....                           | 49        |
| 3.3.2.    | <i>Les systèmes de type Salle de Conférence</i> .....                        | 49        |
| 3.3.3.    | <i>Conclusion</i> .....  | 50        |
| 3.4.      | HÉTÉROGÉNÉITÉ DES RÉSEAUX .....  | 51        |
| 3.4.1.    | <i>Caractéristiques des réseaux pour le multimédia</i> .....                 | 51        |
| 3.4.1.1.  | La bande passante réseau .....   | 51        |
| 3.4.1.2.  | Partage efficace de ressources réseaux .....                                 | 52        |
| 3.4.1.3.  | Garantie de performance .....  | 53        |
| 3.4.1.4.  | Scalabilité du réseau .....  | 54        |
| 3.4.1.5.  | Le Multicast .....   | 54        |
| 3.4.2.    | <i>Les réseaux</i> .....   | 56        |
| 3.4.2.1.  | Le réseau RNIS.....  | 56        |
| 3.4.2.2.  | Les réseaux ATM.....   | 57        |
| 3.4.2.3.  | Le réseau Internet.....  | 58        |
| 3.4.3.    | <i>Conclusion</i> .....  | 59        |
| 3.5.      | LES SOLUTIONS ACTUELLES POUR LE PROBLÈME DE L'HÉTÉROGÉNÉITÉ .....            | 60        |
| 3.5.1.    | <i>Passerelles de transcodage</i> .....                                      | 60        |
| 3.5.2.    | <i>Codage vidéo en couches</i> .....   | 62        |
| 3.6.      | UN EXEMPLE PRATIQUE.....   | 63        |
| 3.7.      | CONCLUSION.....  | 65        |
| <b>4.</b> | <b>PROPOSITION D'UN MODÈLE POUR LA GESTION DE LA PRÉSENTATION VIDÉO.....</b> | <b>69</b> |
| 4.1.      | INTRODUCTION.....  | 70        |
| 4.2.      | POURQUOI ASSISTER L'UTILISATEUR DANS LA PRÉSENTATION VIDÉO ? .....           | 71        |
| 4.3.      | LES MODES TRADITIONNELS POUR DIRIGER LA PRÉSENTATION VIDÉO.....              | 72        |
| 4.3.1.    | <i>Le mode sonore</i> .....  | 72        |
| 4.3.2.    | <i>Le mode directeur</i> .....   | 74        |
| 4.4.      | PRÉSENTATION VIDÉO PERSONNALISÉE.....  | 75        |
| 4.5.      | ANALYSE DE L'ACTIVITÉ COOPÉRATIVE.....                                       | 76        |
| 4.6.      | LES CARACTÉRISTIQUES DE L'ACTIVITÉ POUR LA GESTION VIDÉO.....                | 79        |
| 4.6.1.    | <i>Rôles des participants</i> .....  | 79        |
| 4.6.2.    | <i>Etats des participants</i> .....  | 80        |
| 4.6.2.1.  | Gestion des outils avec les rôles .....                                      | 80        |
| 4.6.2.2.  | Outils pour la détermination de l'état .....                                 | 81        |
| 4.6.2.3.  | Etats dépendant de plusieurs outils.....                                     | 83        |

|  |            |
|--|------------|
| 4.6.3. Les règles .....  | 85         |
| 4.7. ETUDE DE CAS : UN COURS À DISTANCE .....  | 87         |
| 4.7.1. Les états .....   | 87         |
| 4.7.2. Les règles .....  | 88         |
| 4.7.3. Scénario .....  | 89         |
| 4.8. PLUS D'INTELLIGENCE DANS LES RÈGLES.....  | 91         |
| 4.9. CONCLUSION .....  | 91         |
| <b>5. SOLUTIONS POUR LA QUALITÉ DE SERVICE VIDÉO.....</b>  | <b>93</b>  |
| 5.1. INTRODUCTION.....   | 94         |
| 5.2. LA QUALITÉ DE SERVICE : QoS.....  | 95         |
| 5.2.1. Définitions.....  | 96         |
| 5.2.2. La spécification de QoS.....  | 96         |
| 5.2.3. La qualité de service à différents niveaux du système.....                                | 98         |
| 5.2.3.1. La QoS au niveau utilisateur.....   | 99         |
| 5.2.3.2. La QoS au niveau application.....   | 100        |
| 5.2.3.3. La QoS au niveau système d'exploitation.....  | 100        |
| 5.2.3.4. La QoS au niveau réseau.....  | 100        |
| 5.3. LA QUALITÉ VIDÉO.....   | 100        |
| 5.4. LA QoS DU POINT DE VUE RÉCEPTEUR.....   | 103        |
| 5.4.1. Les mécanismes pour la QoS.....   | 104        |
| 5.4.1.1. Les mécanismes pour la phase d'établissement .....                                      | 104        |
| 5.4.1.2. Les mécanismes pour la phase d'opération.....   | 106        |
| 5.5. QoS DU POINT DE VUE ÉMETTEUR .....  | 107        |
| 5.5.1. Différence entre les intérêts des participants.....                                       | 108        |
| 5.5.2. Différence entre les ressources des participants.....                                     | 109        |
| 5.5.3. Détermination des vecteurs QoS optimaux.....  | 110        |
| 5.5.4. Solution du point de vue activité coopérative.....  | 113        |
| 5.5.4.1. Degré d'importance .....  | 115        |
| 5.5.4.2. Degré d'intérêt .....   | 116        |
| 5.5.4.3. Détermination des facteurs de pondération.....  | 116        |
| 5.5.5. Remarque.....   | 116        |
| 5.6. CONCLUSION.....   | 117        |
| <b>6. CONCEPTION ET MISE EN ŒUVRE D'UNE ARCHITECTURE ADAPTABLE À L'ACTIVITÉ COOPÉRATIVE.....</b> | <b>119</b> |
| 6.1. INTRODUCTION.....   | 120        |
| 6.2. LES FRAMEWORKS .....  | 121        |
| 6.3. ARCHITECTURE OUVERTE .....  | 122        |
| 6.4. INTERFACE ENTRE COMPOSANT ACTIVITÉ ET LE FRAMEWORK D'APPLICATION.....                       | 123        |
| 6.5. LES COMPOSANTS DU FRAMEWORK .....   | 125        |
| 6.5.1. Le récepteur .....  | 125        |

|            |  |            |
|------------|--|------------|
| 6.5.2.     | <i>L'émetteur</i> .....                                      | 127        |
| 6.6.       | DÉPLOIEMENT DU SYSTÈME .....                                 | 127        |
| 6.7.       | LA TRANSMISSION VIDÉO .....                                  | 129        |
| 6.7.1.     | <i>Emission vidéo</i> .....                                  | 130        |
| 6.7.2.     | <i>Réception vidéo</i> .....                                 | 131        |
| 6.8.       | CHOIX DU LANGAGE D'IMPLÉMENTATION .....                      | 131        |
| 6.8.1.     | <i>Le langage retenu : le C++</i> .....                      | 131        |
| 6.8.2.     | <i>Le langage Java, l'autre candidat</i> .....               | 132        |
| 6.9.       | EXEMPLE DE PRÉSENTATION VIDÉO .....                          | 135        |
| 6.10.      | VALIDATION FONCTIONNELLE.....                                | 141        |
| 6.10.1.    | <i>Gestion de la QoS Vidéo</i> .....                         | 141        |
| 6.10.2.    | <i>Adaptation à différentes activités coopératives</i> ..... | 142        |
| 6.10.2.1.  | <i>Conversation de groupe</i> .....                          | 143        |
| 6.10.2.2.  | <i>Cours à distance</i> .....                                | 143        |
| 6.10.3.    | <i>Résultats</i> .....                                       | 144        |
| 6.11.      | CONCLUSION .....   | 145        |
| <b>7.</b>  | <b>CONCLUSION ET PERSPECTIVES.....</b>                       | <b>147</b> |
| <b>8.</b>  | <b>ANNEXE I : LA COMPRESSION VIDÉO .....</b>                 | <b>153</b> |
| 8.1.       | LES TECHNIQUES DE COMPRESSION.....                           | 154        |
| 8.1.1.     | <i>Le codage RLE</i> .....                                   | 154        |
| 8.1.2.     | <i>Le codage de Huffman</i> .....                            | 155        |
| 8.1.3.     | <i>La transformée DCT</i> .....                              | 155        |
| 8.1.4.     | <i>La quantification</i> .....                               | 156        |
| 8.1.5.     | <i>Compensation et estimation de mouvement</i> .....         | 157        |
| 8.2.       | LES STANDARDS DE COMPRESSION.....                            | 157        |
| 8.2.1.     | <i>La compression JPEG</i> .....                             | 158        |
| 8.2.2.     | <i>La compression H.261</i> .....                            | 160        |
| 8.2.3.     | <i>La compression MPEG</i> .....                             | 162        |
| <b>9.</b>  | <b>ANNEXE II : CODAGE EN COUCHES .....</b>                   | <b>165</b> |
| 9.1.       | LE CODAGE PROGRESSIF.....                                    | 165        |
| 9.1.1.     | <i>Par sélection spectral</i> .....                          | 165        |
| 9.1.2.     | <i>Par approximation successive</i> .....                    | 166        |
| 9.2.       | LE CODAGE HIÉRARCHIQUE.....                                  | 166        |
| <b>10.</b> | <b>ANNEXE III : CLASSIFICATION .....</b>                     | <b>169</b> |
| 10.1.      | AGRÉGATION AUTOUR DE CENTRES MOBILES .....                   | 170        |
|            | <b>BIBLIOGRAPHIE.....</b>                                    | <b>173</b> |

## Table des figures

|   |     |
|---|-----|
| <i>Figure 2-1 : Classification Temps / Espace du travail Coopératif.</i>  | 18  |
| <i>Figure 2-2 : Conférence Multipartie.</i>   | 20  |
| <i>Figure 2-3 : Exemple de création d'une nouvelle conférence.</i>  | 22  |
| <i>Figure 2-4 : Exemple d'une liste de conférences.</i>   | 23  |
| <i>Figure 2-5 : Capture d'écran de GroupDraw montrant différentes formes de pointeurs.</i>  | 27  |
| <i>Figure 2-6 : Modèle général d'un système T.120 ( source [ITU 96]).</i>   | 29  |
| <i>Figure 3-1 : Site élève de l'ENIC.</i>   | 46  |
| <i>Figure 3-2 : Système poste de travail.</i>   | 49  |
| <i>Figure 3-3 : Salle de vidéoconférence.</i>   | 50  |
| <i>Figure 3-4 : Comparaison des modes de transmission de données entre une source S et des récepteurs R.</i>                          | 55  |
| <i>Figure 3-5 : une passerelle entre deux sessions vidéo.</i>   | 60  |
| <i>Figure 3-6 : un transcodeur JPEG / H.261 optimisé.</i>   | 61  |
| <i>Figure 3-7 : Priorité entre flux vidéo.</i>  | 64  |
| <i>Figure 3-8 : Cas où la VOD est sur un autre site.</i>  | 65  |
| <i>Figure 4-1 : Visibilité des participants dans une session de coopération.</i>  | 70  |
| <i>Figure 4-2 : Modèle pour la déduction des intérêts des participants.</i>   | 77  |
| <i>Figure 4-3 : Liaison entre Rôles et Participants</i>   | 80  |
| <i>Figure 4-4 : Gestion de ressources partagées avec les rôles.</i>   | 81  |
| <i>Figure 4-5 : Diagramme d'état d'un participant par rapport à un outil.</i>   | 82  |
| <i>Figure 4-6 : Prise en compte de l'état Demandeur.</i>  | 82  |
| <i>Figure 4-7 : Prise en compte de l'historique de la détention d'un outil.</i>   | 83  |
| <i>Figure 4-8 : Agrégation d'états en fonction du nombre d'outils.</i>  | 83  |
| <i>Figure 4-9 : Etat dépendant de deux outils.</i>  | 84  |
| <i>Figure 4-10 : Règles pour la déduction des intérêts.</i>   | 86  |
| <i>Figure 4-11 : Modèle de classe pour l'activité cours à distance.</i>   | 87  |
| <i>Figure 4-12 : Diagramme d'état d'un participant.</i>   | 88  |
| <i>Figure 5-1 : Modèle en couche pour la QoS.</i>   | 99  |
| <i>Figure 5-2 : Différentes qualité vidéo.</i>  | 102 |
| <i>Figure 5-3 : Négociation de QoS.</i>   | 105 |
| <i>Figure 5-4 : Requêtes QoS différentes</i>  | 108 |
| <i>Figure 5-5 : Présentation vidéo attendues par les participants.</i>  | 109 |
| <i>Figure 5-6 : Représentation graphique montrant la dispersion des formats vidéo standards.</i>                                      | 111 |
| <i>Figure 5-7 : Correspondance entre Vecteurs QoS Demandés et Vecteurs QoS optimaux.</i>  | 112 |
| <i>Figure 5-8 : Répartition des requêtes QoS dans un espace à deux dimensions. Exemple de requêtes pour la dimension de la vidéo.</i> | 114 |
| <i>Figure 5-9 : Degré d'importance.</i>   | 115 |

|  |     |
|--|-----|
| <i>Figure 5-10 : Intervalles de variation de la taille de la vidéo.</i>                            | 117 |
| <i>Figure 6-1 : Composant Activité connectable.</i>  | 123 |
| <i>Figure 6-2 : Interface entre le composant activité et le Framework d'application.</i>           | 125 |
| <i>Figure 6-3 : Architecture côté récepteur.</i>   | 126 |
| <i>Figure 6-4 : Architecture côté émetteur.</i>  | 127 |
| <i>Figure 6-5 : déploiement du système.</i>  | 128 |
| <i>Figure 6-6 : Chaîne d'émission vidéo.</i>   | 130 |
| <i>Figure 6-7 : Chaîne de réception vidéo.</i>   | 131 |
| <i>Figure 6-8 : Principe de la machine virtuelle Java.</i>   | 133 |
| <i>Figure 6-9 : Codage en couche temporel.</i>   | 141 |
| <i>Figure 6-10 : Combinaison Spatio-temporelle pour la détermination d'une QoS Vidéo.</i>          | 142 |
| <i>Figure 6-11 : Couplage avec TutelCast.</i>  | 144 |
| <i>Figure 8-1 : Diagramme bloc de l'encodeur JPEG.</i>   | 159 |
| <i>Figure 8-2 : Diagramme bloc du décodeur JPEG.</i>   | 159 |
| <i>Figure 8-3 : Schéma synoptique d'un codeur H.261.</i>   | 161 |
| <i>Figure 8-4 : Structure hiérarchique d'une image H.261.</i>                                      | 161 |
| <i>Figure 8-5 : Exemple d'une séquence MPEG montrant la dépendance entre les images I, P et B.</i> | 163 |

---

# **1. Introduction Générale**

---

Aujourd'hui, les ordinateurs individuels possèdent la puissance et la capacité de manipuler des données multimédia comme l'audio et la vidéo. Ces données, aussi appelées médias continus, ont des caractéristiques spécifiques :

- Ils nécessitent un traitement en temps réel pour respecter leurs contraintes temporelles ;
- Le volume de données nécessaire pour leur encodage est considérable.

Les nouvelles technologies réseaux peuvent supporter ces caractéristiques. Ils possèdent de larges bandes passantes et peuvent garantir l'isochronisme nécessaire pour la transmission des médias continus.

La combinaison des nouvelles capacités réseaux avec l'accroissement de puissance des processeurs a facilité le développement des applications coopératives multimédia distribuées. La conception de ces applications reste cependant difficile, surtout lorsqu'on essaye de respecter les habitudes des utilisateurs pendant leur activité en groupe.

Pour permettre aux utilisateurs de travailler via des systèmes de travail coopératif, de la manière la plus naturelle possible, comme s'ils étaient en présentiel, les participants doivent se voir et s'entendre, s'échanger des documents, ainsi que partager des espaces et des ressources [Karmouch 93]. Nous pouvons ainsi distinguer deux espaces, un espace de coopération et un espace de communication.

- L'espace de coopération contient les applications pour le travail en groupe tels que l'éditeur partagé et le tableau blanc, et les outils comme le télépointeur pour la désignation des objets partagés.
- L'espace de communication contient l'ensemble des flux audio et vidéo des participants utilisés dans la conférence [Bly 93, Gaver 92a].

---

Pour éviter que les utilisateurs se trouvent confrontés aux problèmes d'incompatibilité entre les systèmes provenant de constructeurs différents, des normes comme H.320 [ITU 90] et T.120 [ITU 96] pour l'échange de données audio-vidéo et d'interopérabilité ont été établies.

Les normes établies permettent l'interopérabilité entre systèmes provenant de différents constructeurs. Cependant, ces normes ne résolvent pas les problèmes de l'hétérogénéité des capacités des systèmes utilisateurs d'une part, et de la diversité des activités coopératives d'autre part.

L'hétérogénéité des systèmes utilisateurs peut se situer au niveau des équipements terminaux et des accès réseaux. Lorsque plusieurs sites ayant des capacités différentes participent à une conférence, c'est le site le moins puissant qui impose la qualité de service utilisée. En particulier, la vidéo qui est une source consommatrice de ressources, aura la qualité permise par les ressources du site le moins puissant. Ceci pénalise les autres participants ayant des capacités plus grandes.

De plus, chaque activité coopérative possède des caractéristiques spécifiques comme les rôles que doivent prendre les participants, les outils utilisés, et les objectifs à atteindre. Il en résulte que chaque activité peut avoir des besoins vidéo spécifiques.

L'hétérogénéité des systèmes et la diversité des activités coopératives montre que dans les applications de travail coopératif, assister l'utilisateur dans la gestion de sa présentation vidéo devient une nécessité, surtout au sein des grands groupes. Les ressources étant limitées, le système utilisateur ne peut pas recevoir toutes les vidéos des sites distants. Le système doit sélectionner les vidéos des sites pour lesquels le participant a de l'intérêt. Cette gestion doit être dynamique avec l'évolution des intérêts du participant. Elle doit tenir compte des ressources disponibles, et de leur hétérogénéité chez les participants.

Les travaux présentés dans ce mémoire ont pour objectif d'assister la composition de la présentation vidéo des participants dans les sessions d'activités coopératives, en tenant compte des spécificités de l'activité, des ressources des participants, et de leur hétérogénéité au sein du groupe.

Ce mémoire est organisé comme suit. Dans le chapitre deux, nous donnons un aperçu général du travail coopératif synchrone, et nous présentons ses deux types, la vidéoconférence et le Media Space. Puis nous décrivons les apports de la vidéo pour le travail coopératif.

Dans le troisième chapitre nous discuterons de l'effet de l'hétérogénéité des systèmes utilisateurs et de la diversité des activités coopératives sur l'utilisation de la vidéo dans les systèmes de travail coopératif. Nous présentons les solutions actuelles pour l'hétérogénéité des systèmes, puis à l'aide d'un exemple, nous montrerons les limites de ces solutions

Le quatrième chapitre décrit notre proposition de modèle pour assister le participant dans la composition et la personnalisation de sa présentation vidéo. A partir de la connaissance des rôles des participants, ainsi que de leurs états, nous déduisons les intérêts du participant à voir ses interlocuteurs. A partir des intérêts du participant, nous établissons des priorités entre ses interlocuteurs afin de lui composer la meilleure présentation vidéo possible.

Nous faisons le lien entre les intérêts du participant et la qualité de service permise par les ressources disponibles dans le cinquième chapitre. Nous présentons nos solutions qui permettent de remonter les décisions de choix de la qualité de service vidéo à fournir aux participants du niveau système au niveau utilisateur. Nous verrons les points de vues de l'émetteur et du récepteur pour la détermination de cette qualité vidéo.

Enfin, dans le chapitre six nous présentons notre architecture ouverte permettant de s'adapter à différents types d'activités coopératives pour assister la présentation vidéo de l'utilisateur. Cette architecture utilise la notion de Framework, qui permet de personnaliser une application par l'utilisation des composants connectables. Nous décrivons également des détails de mise en œuvre ainsi que les expérimentations réalisées avec notre prototype de gestion vidéo.

---

## **2. Le Travail Coopératif**

### **Synchrone**

---

## **2.1. Introduction**

Le terme CSCW (Computer Supported Cooperative Work), traduit par TCAO (Travail Coopératif Assisté par Ordinateur), a été utilisé pour la première fois par Irene Greif et Paul Gashman en 1984, comme prélude à la première conférence CSCW à Austin Texas en 1986.

Le CSCW est né pour résoudre certains problèmes liés au besoin de personnes désirant travailler en groupe.

L'émergence du CSCW a été favorisée par la forte pénétration des stations de travail personnelles dans les bureaux, ainsi que le déploiement des réseaux locaux et distants qui permettent de relier ces stations entre elles. Une fois les stations de travail reliées, leurs utilisateurs peuvent alors être mis en relation à l'aide d'applications de travail coopératif.

Le CSCW est un domaine de recherche pluridisciplinaire. Il utilise les développements accomplis dans les technologies de l'information. De ce fait, l'informatique avec ses diverses branches, constitue un domaine prédominant. Cependant, le CSCW est indépendant de la technologie. La technologie est un moyen pour réaliser des systèmes de travail coopératif, mais n'est pas la force motrice principale. Le CSCW dépend de l'aspect social. Il étudie la manière dont les personnes interagissent et collaborent afin de développer les systèmes qui assistent les personnes à travailler ensemble. De ce fait, les disciplines des sciences humaines comme la sociologie et la psychologie doivent alors contribuer à la conception des systèmes de travail coopératif, afin d'éviter des échecs lors de l'utilisation des produits conçus [Grudin 88]. Ces disciplines permettent d'étudier le comportement humain pendant les activités de groupe, pour en tenir compte pendant la conception des systèmes de travail coopératif. Le développement d'un système de travail coopératif nécessite donc la coopération entre les chercheurs des sciences humaines, des sciences de l'informatique, et des utilisateurs finaux [Karsenty 94].

Le lecteur trouvera plus de détails sur le CSCW dans les travaux du laboratoire Trigone [Vilers 92, Croisy 95, Hoogstoel 95, Carlier 98].

Dans la suite de ce chapitre, nous allons décrire les systèmes de travail coopératif. Nous nous intéresserons plus particulièrement aux systèmes de travail coopératif synchrone, dans lequel nous présenterons les deux systèmes types : les systèmes de vidéoconférence et les systèmes de Media Space. Nous détaillerons les principes, les concepts, les normes et les

techniques inhérents à ces systèmes. Ensuite, nous discuterons des apports de l'utilisation de la vidéo comme média de communication dans ces systèmes de travail coopératif synchrone.

## 2.2. Le travail Coopératif

Ellis [Ellis 91] définit un système CSCW par :

*C'est un système informatique qui supporte un groupe de personne engagé dans une tâche commune (ou but commun), et qui fournit une interface à un environnement partagé.*

Dans cette définition apparaissent trois notions importantes. **Le groupe**, qui représente un ensemble de personnes qui s'engagent à travailler ensemble. **La tâche commune**, qui signifie que chaque membre du groupe aura à effectuer un travail qui est soit concurrent ou complémentaire à ceux des autres membres afin d'atteindre le but commun fixé au départ. Et **l'environnement partagé**, qui représente les documents et les outils permettant aux membres d'un groupe de travailler ensemble.

Cette définition est un bon point de départ de discussion sur le CSCW, cependant elle ne met pas en évidence l'aspect communication entre les personnes constituant le groupe, bien que cet aspect puisse être considéré comme faisant partie de l'environnement partagé.

Dans un système de travail coopératif, nous pouvons distinguer deux espaces, l'espace de coopération et l'espace de communication.

L'espace de coopération comprend tous les outils et les objets partagés entre les membres du groupe pour accomplir la tâche commune. Alors que l'espace de communication, permet aux personnes de discuter ou de se voir aux moyens de techniques audio-vidéo.

### 2.2.1. Les dimensions Espace et Temps du CSCW

En fonction des différentes applications coopératives, Rodden [Rodden 91] a établi une classification espace/temps (Figure 2-1). Cette classification divise les applications coopératives en quatre groupes.

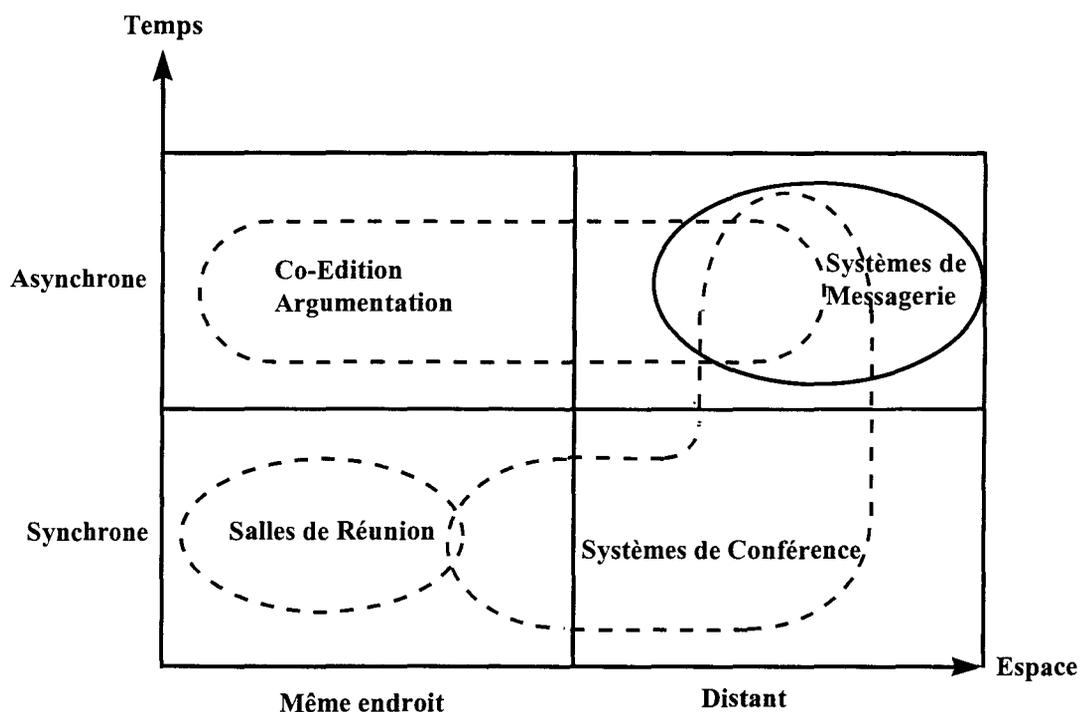


Figure 2-1 : Classification Temps / Espace du travail Coopératif.

Cette classification permet de caractériser le travail coopératif par deux dimensions, la dimension temporelle qui représente la forme d'interaction, et la dimension spatiale qui représente la dispersion géographique des membres du groupe.

- **La dimension temporelle**

Le travail coopératif peut être synchrone ou asynchrone. Dans le cas d'un travail synchrone, les membres du groupe interagissent entre eux directement en temps réel. Des exemples de travail coopératif synchrone sont les réunions de travail face à face, et les systèmes de téléconférence [Schooler 96].

Par opposition à la coopération temps réelle, la coopération asynchrone permet aux membres du groupe de travailler à des moments différents. Elle peut prendre un temps indéterminé avant de se terminer, alors que la coopération synchrone est relativement courte dans le temps. Les systèmes de messagerie et de workflow comme les produits Lotus Notes<sup>1</sup> et Exchange<sup>2</sup> sont des exemples d'interaction asynchrone. Dans [Yousfi 96], le principe du

<sup>1</sup> <http://www.lotus.com/>

<sup>2</sup> <http://www.microsoft.com/exchange/>

workflow à été utilisé pour la coordination des activités du personnel médical dans les unités de soins.

- **La dimension spatiale**

Les membres d'un groupe peuvent travailler ensemble au même endroit (co-localisés) ou dispersés géographiquement sur plusieurs sites (distants). Il peut sembler étonnant que le CSCW étudie les applications coopératives entre personnes se trouvant au même endroit au même moment. Cependant, dans une application de co-édition, les participants peuvent se trouver dans la même salle, chacun équipé d'un ordinateur pour rédiger une partie du document partagé.

Généralement, les médias utilisés dans les communications asynchrones imposent moins de contraintes sur les membres du groupe que dans les communications synchrones.

Dans la suite, nous nous intéresserons aux systèmes de travail coopératif synchrone et distribué. Nous présenterons les systèmes de vidéoconférence et les systèmes de Media Space.

## 2.3. Les Systèmes de vidéoconférence

Ces dernières années, une multitude d'outils de vidéoconférence ont vu le jour. Parmi les outils pionniers, citons *ivs* de l'INRIA [Turletti 94], qui fut l'un des premiers outils à utiliser la compression vidéo logicielle. L'outil **CU-SeeMe** de Cornell University [Dorcey 95] fut l'outil de vidéoconférence le plus populaire sur Internet. Aujourd'hui, il est commercialisé par la société White Pine<sup>3</sup>. *vic* est un autre outil qui a connu beaucoup de popularité [McCanne 95]. C'est l'outil le plus utilisé sur le MBone<sup>4</sup> [Eriksson 94, Macedonia 94].

Bien que chacun de ces outils ait ses spécificités, technologiques ou fonctionnelles, ils offrent tous les mêmes services de bases à leurs utilisateurs.

Dans ce paragraphe, après avoir donné un ensemble de définitions pour fixer la terminologie utilisée dans les systèmes de vidéoconférence, nous présenterons les services communs à tous les systèmes de vidéoconférence, à savoir la gestion de conférence, et l'espace partagé permettant aux utilisateurs de coopérer. Nous terminerons ce paragraphe par donner un aperçu sur la norme T.120 qui normalise les services de base d'un système de vidéoconférence.

---

<sup>3</sup> <http://www.wpine.com/Products/CU-SeeMe/>

<sup>4</sup> MBone : Multicast Backbone.

### 2.3.1. Définitions

Une **conférence** est une session de travail coopératif synchrone qui permet à des personnes dispersées géographiquement de travailler ensemble en temps réel. Les temps de début et de fin de conférence sont généralement connus à l'avance.

Une conférence est dite **multimédia** si on associe au moins un média discret avec un média continu. Le média discret peut être un tableau blanc partagé, sur lequel chaque participant peut faire des annotations, alors que le média continu peut être de l'audio permettant aux participants de parler et d'entendre leurs correspondants. Dans ce cas, nous parlons de système d'**audioconférence**. Si de plus un support vidéo est utilisé pour permettre aux participants de se voir, nous parlerons de système de **vidéoconférence** ou de **visioconférence**. Dans la suite, nous utiliserons indifféremment les termes conférence et vidéoconférence pour désigner une session de travail de groupe.

Une conférence est dite **multipoint**, si le nombre de sites participants est supérieur à deux [Clark 92, Boyer 94, Boulkroune 96a]. On utilise le mot **Multipartie** pour désigner des conférences multipoints, dans lesquelles, au moins un site accueille plusieurs participants (Figure 2-2). Un exemple de conférence multipartie est une application de télé-enseignement dans laquelle un site élève peut accueillir plusieurs étudiants.

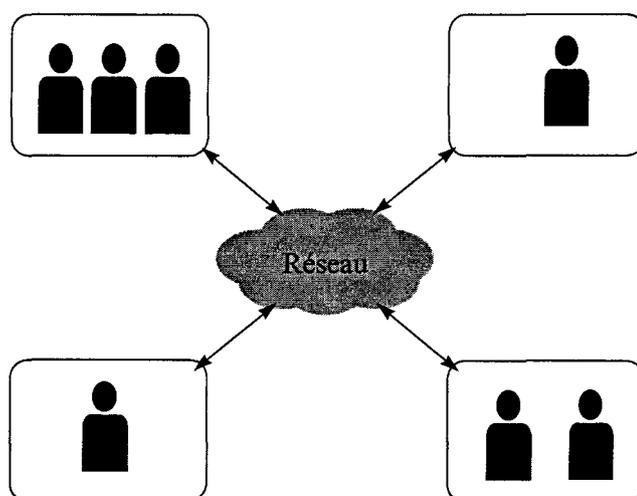


Figure 2-2 : Conférence Multipartie.

L'architecture d'un système de vidéoconférence est généralement divisée en composants : la gestion de conférence, l'espace partagé, et la gestion des flux multimédia audio et vidéo [Handley 93, Gajewska 94, Altenhofen 93, Boulkroune 96b].

### **2.3.2. La gestion de conférence**

Un système de conférence doit fournir aux utilisateurs des fonctions de base qui permettent de gérer une conférence depuis sa création jusqu'à sa terminaison. Pour identifier les plus importants services dans une conférence, suivons les différentes étapes du cycle de vie d'une conférence.

#### **2.3.2.1. Création d'une conférence**

Une conférence peut être créée à la demande ou par réservation. La création à la demande s'effectue par un utilisateur qui se connecte à un serveur de conférence, et lance une requête de création de conférence. Le serveur traite cette requête, et dans le cas d'une réponse favorable, la conférence est créée et l'utilisateur créateur est considéré comme étant le premier participant de la conférence.

La deuxième manière de créer une conférence se fait par réservation. Une personne se connecte à un système de réservation de conférence, pour fixer le jour et l'heure du déroulement de la conférence, ainsi que pour réserver les ressources nécessaires. Dans ce cas, c'est le serveur qui crée la conférence au moment fixé.

Au moment de la création ou de la réservation d'une conférence, certaines de ses caractéristiques sont précisées, comme son nom, son sujet de discussion, et l'aspect sécurité (Figure 2-3). Ce dernier spécifie si l'entrée à la conférence est publique ou privée. Dans ce dernier cas, la conférence est protégée par un mot de passe qui sera demandé aux personnes désirant participer à celle-ci. Un deuxième niveau de protection peut être utilisé, en spécifiant que la conférence ne doit pas figurer parmi la liste des conférences en cours, lorsqu'un utilisateur interroge le serveur.

L'utilisation d'un serveur de conférence s'explique par la nécessité de référencer les conférences à un endroit, afin que les utilisateurs puissent avoir une adresse ou un numéro à contacter pour être au courant des profils des conférences et choisir les conférences dans lesquelles ils peuvent éventuellement participer. Des outils comme SDR sur le MBone, et les réflecteurs pour CU-SeeMe servent à référencer les conférences en cours ou programmées.

Create New Session
[-] [x]

**Session Name:**

**Description:**

Session de travail sur la gestion de la présentation vidéo dans les applications de travail coopératif.

**URI:**

| <b>Security:</b><br><input checked="" type="checkbox"/> Public<br><input checked="" type="checkbox"/> Private | <b>Scope:</b><br><input type="checkbox"/> Site (ttl 15)<br><input type="checkbox"/> Region (ttl 63)<br><input type="checkbox"/> World (ttl 127) | <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">Media:</th> <th style="width: 10%;"></th> <th style="width: 10%;">Format</th> </tr> </thead> <tbody> <tr> <td><input checked="" type="checkbox"/> </td> <td>audio</td> <td>PCM</td> </tr> <tr> <td><input checked="" type="checkbox"/> </td> <td>video</td> <td>H.261</td> </tr> <tr> <td><input checked="" type="checkbox"/> </td> <td>whiteboard</td> <td>wb</td> </tr> <tr> <td><input checked="" type="checkbox"/> </td> <td>text</td> <td></td> </tr> </tbody> </table> | Media: |  | Format | <input checked="" type="checkbox"/> | audio | PCM | <input checked="" type="checkbox"/> | video | H.261 | <input checked="" type="checkbox"/> | whiteboard | wb | <input checked="" type="checkbox"/> | text |  |  |
|---|---|--|--------|--|--------|-------------------------------------|-------|-----|-------------------------------------|-------|-------|-------------------------------------|------------|----|-------------------------------------|------|--|--|
| Media:  |   | Format   |        |  |        |                                     |       |     |                                     |       |       |                                     |            |    |                                     |      |  |  |
| <input checked="" type="checkbox"/>   | audio   | PCM  |        |  |        |                                     |       |     |                                     |       |       |                                     |            |    |                                     |      |  |  |
| <input checked="" type="checkbox"/>   | video   | H.261  |        |  |        |                                     |       |     |                                     |       |       |                                     |            |    |                                     |      |  |  |
| <input checked="" type="checkbox"/>   | whiteboard  | wb   |        |  |        |                                     |       |     |                                     |       |       |                                     |            |    |                                     |      |  |  |
| <input checked="" type="checkbox"/>   | text  |  |        |  |        |                                     |       |     |                                     |       |       |                                     |            |    |                                     |      |  |  |

**Session will be active:**

Once  from:  at  for

Repeat for:

**Contact details:**

Figure 2-3 : Exemple de création d'une nouvelle conférence.

### 2.3.2.2. Pour participer : se joindre ou être invité

Pour participer à une conférence, un utilisateur se connecte au serveur de conférence, et demande à se joindre à une conférence en précisant son nom. S'il n'a pas connaissance du nom de la conférence à laquelle il veut se joindre, il peut d'abord récupérer la liste des conférences référencées par le serveur, puis se joint à la conférence voulue (Figure 2-4).

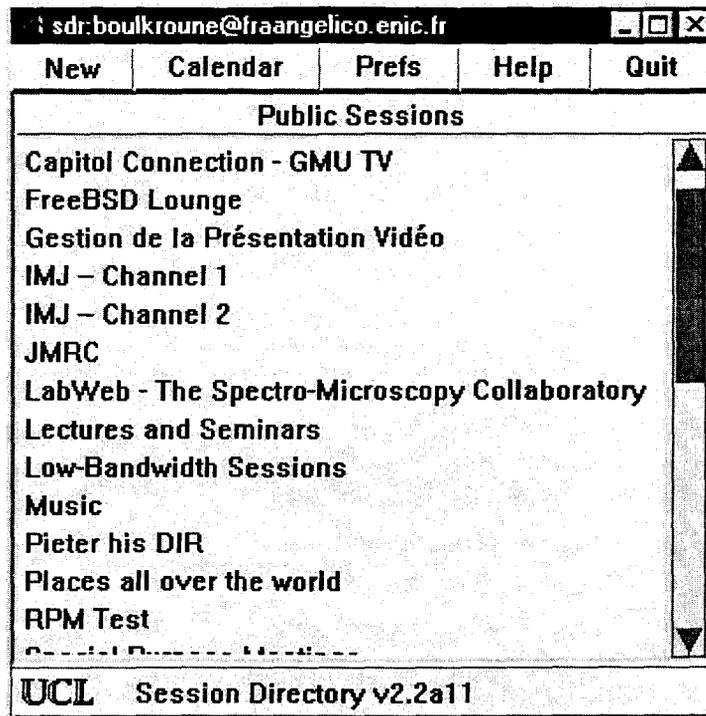


Figure 2-4 : Exemple d'une liste de conférences.

Un utilisateur peut aussi être invité à participer à une conférence soit par un participant, soit par le serveur. Dans ce dernier cas, une liste d'utilisateurs à inviter doit être indiquée au moment de la réservation de la conférence.

### 2.3.2.3. Comment participer ?

En entrant dans une conférence, un utilisateur devient un *participant*. Un participant peut être *conducteur*, *détenteur de jeton*<sup>5</sup> ou *observateur* [Altenhofen 93].

Le conducteur d'une conférence gère le déroulement de la session de travail et possède les privilèges d'attribution des ressources utilisées aux autres participants, comme l'attribution de la parole ou les droits de manipulation des documents partagés pendant la session de travail. Le rôle de conducteur peut passer d'un participant à un autre durant la conférence, mais il ne peut y avoir qu'un seul conducteur à un instant donné.

Un détenteur de jeton est un participant actif, au sens où il a la possibilité d'utiliser les ressources de la conférence et de modifier l'espace et les documents partagés.

<sup>5</sup> token holder.

L'observateur quant à lui, est un participant passif, qui ne peut ni demander une ressource, ni modifier un document. Il n'a droit qu'à la consultation.

#### **2.3.2.4. Pour sortir : quitter ou être éjecté**

Pour sortir d'une conférence, un participant peut quitter en faisant appel à la primitive correspondante. Les autres participants seront aussitôt notifiés de sa sortie par le serveur de conférence, qui effectue une mise à jour de la liste des participants de la conférence.

Aussi, un participant peut être éjecté par un autre participant possédant ce privilège (conducteur). C'est ce qui peut arriver par exemple si le conducteur décide de terminer la conférence alors qu'il y a encore des participants.

#### **2.3.2.5. Terminer une conférence**

Une conférence peut être terminée explicitement par le conducteur, ou implicitement par le serveur dans le cas où il n'y aurait plus de participants, ou parce que le temps alloué à la conférence est écoulé.

### **2.3.3. L'espace partagé**

Les systèmes de vidéoconférence mettent à la disposition des utilisateurs un certain nombre d'outils qui permettent de créer un espace de travail partagé entre les utilisateurs. Parmi les principaux outils généralement fournis, nous pouvons citer le tableau blanc, les applications partagées et les télépointeurs.

Un des concepts clé de l'espace partagé est le WYSIWIS<sup>6</sup> [Stefik 87a, Stefik 87b]. Ce concept permet aux membres du groupe de voir les mêmes informations en même temps. Le WYSIWIS peut être strict ou relâché. Il est strict lorsque les utilisateurs ont exactement la même vue de l'espace partagé. Cependant, les utilisateurs peuvent avoir besoin d'un certain degré de flexibilité. En prenant comme exemple l'édition coopérative d'un document, où à chaque participant est attribué un paragraphe à rédiger, le WYSIWIS strict ne peut pas être utilisé. En effet, le simple défilement de la fenêtre du document partagé par un participant, fera défiler les fenêtres de tous les autres sites participants. Dans ce genre de situation, une autre variante du WYSIWIS est utilisée : c'est le WYSIWIS relâché. Cette variante laisse

---

<sup>6</sup> WYSIWIS : What You See Is What I See.

chaque utilisateur visualiser la partie du document qui est à sa charge, et l'autorise à consulter les autres parties du document en changeant sa vue locale, sans modifier les vues des autres sites participants.

### 2.3.3.1. Le tableau blanc

Le tableau blanc partagé est une application qui permet à plusieurs participants distants de voir la même fenêtre sur leur écran d'ordinateur. Chaque participant peut faire des annotations, dessiner ou entrer un texte [Greenberg 92, Karsenty 93]. Pour identifier la personne qui a dessiné une partie du dessin, il faut utiliser des conventions simples comme l'utilisation d'une couleur pour chaque participant.

Les tableaux blancs partagés émulent des tableaux physiques. Lorsque deux personnes dans un bureau travaillent en utilisant un tableau et une craie, certaines règles sociales gouvernent leur accès à la surface partagée du tableau. Ils n'écrivent pas au même moment, et n'effacent pas ce que l'autre a écrit. Bien sûr, ses comportements sociaux sont difficiles à reproduire à distance. Définir et implémenter les règles qui gèrent l'accès aux objets partagés est un problème important dans l'espace partagé. Ceci s'appelle les politiques du Floor Control [Dommel 97].

Il existe quatre approches de base pour le floor control.

***Pas de contrôle*** : le système laisse chaque personne accéder librement à la surface partagée, faisant « confiance » au comportement social pour éviter et résoudre les conflits. Ceci peut fonctionner raisonnablement bien entre deux personnes, mais n'est pas praticable lorsque le nombre de personnes augmente. Lorsque des personnes travaillent sur un tableau physique, le comportement social fait que l'on laisse une personne s'approcher du tableau pour écrire, afin de ne pas le gêner. S'il y a un changement, la personne qui écrivait recule pour laisser passer la personne qui va écrire, puisqu'elle a besoin d'un espace près du tableau. Ceci est difficilement reproductible à distance, puisque les personnes ne peuvent pas se gêner physiquement.

***Verrouillage implicite*** : chaque fois qu'un participant commence à saisir ou à entrer des données, il prend implicitement le droit d'accès (floor), et les autres participants ne peuvent pas entrer de données. L'accès est automatiquement libéré après un certain temps de la fin de la rentrée de données de la part de celui ou de celle qui a l'accès. Si ce participant garde le

contrôle plus d'un certain temps, il sera notifié qu'un ou plusieurs participants sont en attente. Un temps de quelques secondes est raisonnable.

***Verrouillage explicite*** : cette approche est similaire au cas précédent, sauf que l'utilisateur doit demander ou relâcher explicitement l'accès via une touche ou un clic. Bien sûr les requêtes faites pendant que l'accès est occupé sont mises dans une file d'attente de type FIFO.

***Contrôle centralisé*** : un des participants est désigné pour être le modérateur (chairperson). Le modérateur peut donner ou reprendre l'accès au tableau blanc à tout moment. Il a besoin d'outils pour contrôler la liste des requêtes.

Quelle que soit la politique retenue, le floor control nécessite l'apport de comportements sociaux, qui sont facilités par l'apport de communication directe comme un outil de chat ou un canal audio.

### **2.3.3.2. Les applications partagées**

Elles résultent d'une extension du tableau blanc partagé. Un outil d'application partagé permet aux participants de partager la visualisation et le contrôle d'une application. Le point important est que l'application mise en partage n'a pas été conçue pour supporter plusieurs utilisateurs simultanément.

Un intérêt immédiat de l'utilisation des applications partagées est de remédier aux déficiences du tableau blanc. Nous pouvons ainsi bénéficier des avantages des applications comme des tableurs ou des éditeurs de texte pour éditer respectivement des feuilles de calculs ou des documents textes par plusieurs personnes.

Il existe maintenant des applications multi-utilisateur comme les applications de co-édition (éditeur partagé), qui permettent à plusieurs utilisateurs d'éditer le même document. La différence avec les éditeurs mono-utilisateur mis en partage, est qu'une couche logicielle est rajoutée au-dessus pour gérer l'aspect multi-utilisateur. Les utilisateurs peuvent travailler de manière synchrone ou asynchrone.

### **2.3.3.3. Les télépointeurs**

Un télépointeur est un outil se présentant sous la forme d'une icône que l'utilisateur peut déplacer sur son écran pour montrer (pointer) un objet particulier dans l'espace partagé. Le télépointeur ainsi que ses mouvements sont vus de manière synchrone sur les écrans des autres participants. L'implémentation du télépointeur dépend de son utilisation. Il peut pointer tout l'écran de l'utilisateur ou bien être lié à une fenêtre particulière. La session de coopération peut posséder un télépointeur, unique pour tous les participants, qui est géré par

des droits d'accès ; chaque participant peut aussi posséder son propre télépointeur. Dans ce dernier cas, pour différencier les télépointeurs des différents participants, une couleur ou forme particulière est attribuée à chaque télépointeur. Ceci est illustré par la Figure 2-5 montrant l'interface de GroupDraw [Greenberg 92]. Des identifiants numériques ou textuels (noms) accolés aux télépointeurs peuvent aussi être utilisés.

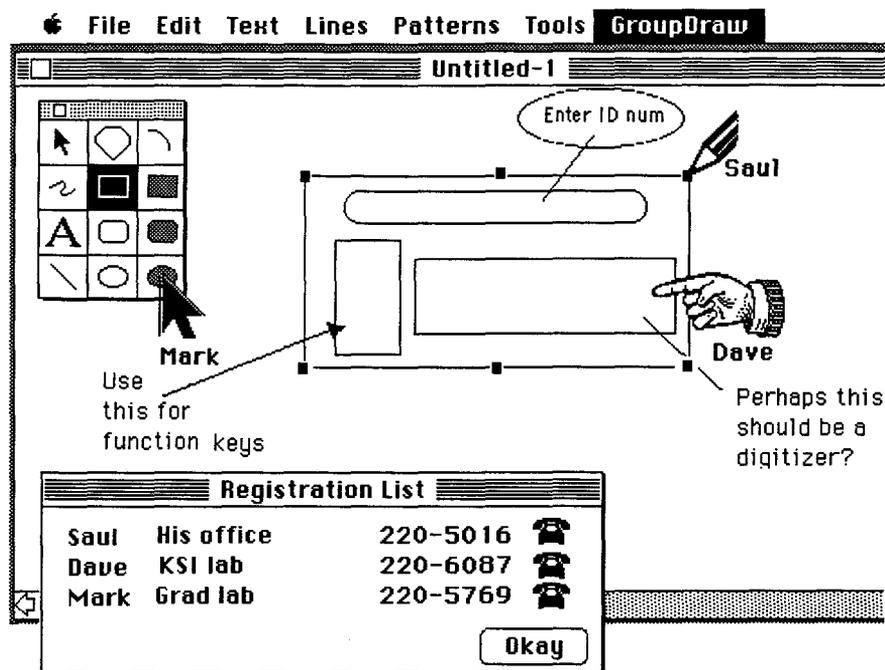


Figure 2-5 : Capture d'écran de GroupDraw montrant différentes formes de pointeurs.

Des difficultés peuvent apparaître lors de l'implémentation ou de l'utilisation des télépointeurs. Lorsque plusieurs télépointeurs sont utilisés au même moment sur l'écran, ceci peut devenir distrayant. Pour résoudre ce problème, nous pouvons restreindre le nombre de télépointeurs qui peuvent être utilisés simultanément. Si un seul télépointeur est utilisé à la fois, une des méthodes de contrôle d'accès vue précédemment pour le tableau blanc peut être utilisée.

Mais le problème le plus difficile à résoudre, est lorsque le WYSIWIS [Stefik 87a, Stefik 87b] relâché est utilisé. Chaque participant place et dimensionne ses fenêtres indépendamment des autres, ce qui fait que les participants n'ont pas exactement la même vue

sur l'espace partagé. Une des raisons du WYSIWIS relâché est la résolution de l'écran qui peut être différente pour chaque participant.

### **2.3.4. T.120, une norme pour la visioconférence et le travail coopératif**

Avec la multiplication des produits commerciaux et expérimentaux de travail coopératif, il était devenu nécessaire d'établir des normes pour permettre une interopérabilité entre les différents produits. L'union internationale de télécommunication (ITU) a récemment développé la série de protocoles T.120 [ITU 96]. T.120 fournit une architecture pour la communication de données en mode multipoint pour les conférences multimédia.

Le protocole T.120 fournit des moyens de communication avec différentes formes de média pour deux ou plusieurs terminaux multimédias, et gère cette communication.

Le modèle général d'un système T.120 est composé d'une infrastructure de communication et de protocoles d'application qui utilisent cette infrastructure (Figure 2-6).

Le contrôleur de nœud (Node Controller), qui ne fait pas partie de la norme T.120, a le rôle de gestion dans un terminal ou un MCU<sup>7</sup> de tout ce qui est en rapport avec la conférence. C'est le contrôleur de nœud qui émet les primitives vers la couche Contrôle Générique de Conférence (GCC) pour le lancement et le contrôle d'une session de communication.

---

<sup>7</sup> MCU : Multipoint Control Unit. C'est un nœud central utilisé pour relier des sites distants lors d'une communication multipoint.

## Model of ITU-T T.120 Series Recommendations

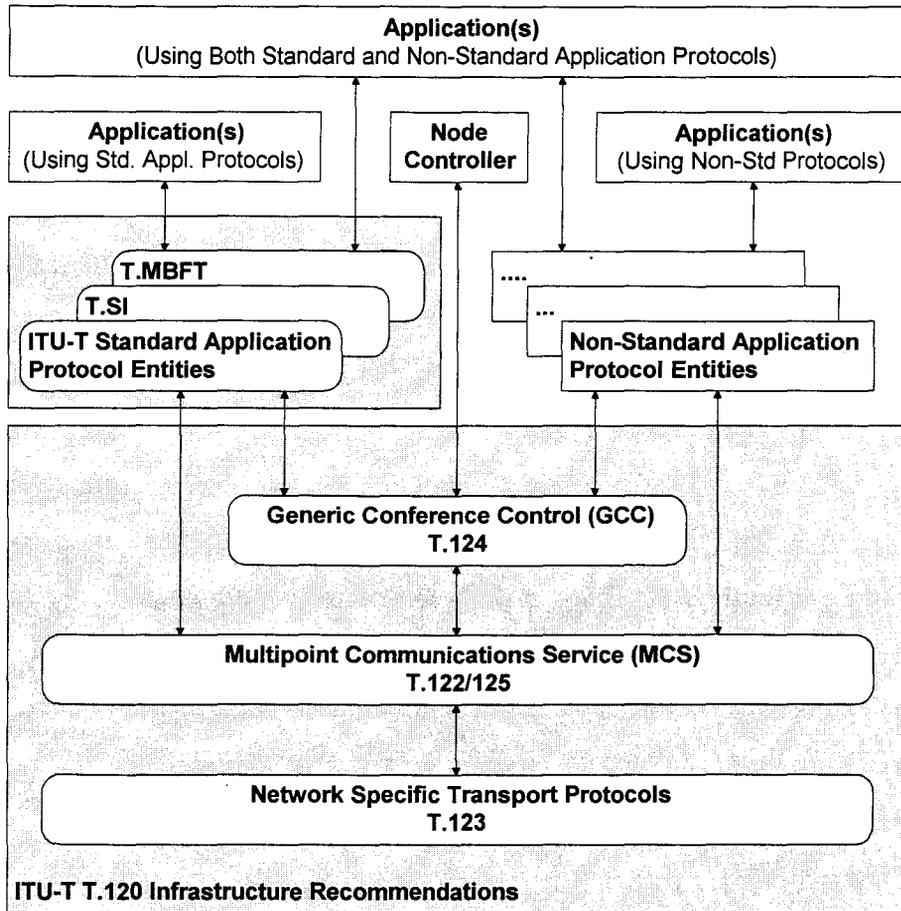


Figure 2-6 : Modèle général d'un système T.120 ( source [ITU 96]).

### 2.3.4.1. L'infrastructure de communication de T.120

C'est la partie obligatoire que tout nœud conforme à la norme T.120 doit posséder. Elle est composée de trois couches, une couche qui s'occupe de l'aspect réseau, une couche MCS qui offre les services de communication multipoint, et une couche GCC pour le contrôle générique de conférence. Ces trois couches sont maintenant successivement décrites.

#### Protocoles de transport spécifique pour chaque réseau

La recommandation T.123 [ITU 93a] fournit un service de livraison fiable des données. Elle définit les piles de transport pour différents réseaux de télécommunication. Elle présente un service transport pour les couches supérieures. Au niveau transport, la conférence est vue

comme un ensemble de connexions point à point (des paires différentes peuvent être sur des réseaux différents).

Parmi les réseaux supportés, nous pouvons citer le RTC (Réseau Téléphonique Commuté), le RNIS (Réseau Numérique à Intégration de Service), et les réseaux X.25. L'intégration du support de réseaux ATM et réseaux locaux est en cours d'étude.

### **Service de Communications Multipoints (MCS)**

La couche MCS fournit un service de connexion multipoint orientée données qui correspond à la couche session dans le modèle OSI.

La couche MCS est définie par deux recommandations, T.122 qui définit les services [ITU 93b] et T.125 les protocoles [ITU 94].

La couche MCS est l'élément principal pour la communication multipoint. Elle prend les connexions point à point fournies par la couche en dessous (T.123) et les combine pour former un domaine multipoint. Un domaine peut correspondre à une conférence ou à une activité particulière d'un groupe. Un nœud peut participer à plusieurs domaines.

Dans un domaine MCS, un nœud est désigné comme étant le fournisseur supérieur du domaine, agissant comme serveur de ressources pour le domaine. Un service de gestion de jetons est alors proposé par la couche MCS, dont seul le fournisseur supérieur du domaine est capable d'y accéder.

### **Contrôle Générique de Conférence (GCC)**

La couche GCC fournit un ensemble de services pour gérer une session de conférence multipoint. GCC fournit des mécanismes qui permettent la création, l'établissement, et la terminaison de conférences. Elle gère aussi la base de données qui contient des informations sur la configuration d'une conférence, comme son nom, le type du nœud participant (terminal, MCU ou terminal), ainsi que des informations sur le site et le participant de chaque nœud. Lorsqu'un nœud joint une conférence, il annonce sa présence. Ceci met à jour la base de conférence et la distribue.

GCC supporte aussi l'enregistrement de protocoles d'applications. Le fournisseur GCC de chaque nœud maintient la liste de protocoles d'applications enregistrés avec leur capacité. Cette liste est envoyée au fournisseur GCC supérieur pour être distribuée à tous les nœuds participant à la conférence. De cette façon, chaque nœud a la connaissance des possibilités de tous les autres nœuds.

Une fois enregistré, un protocole d'application peut utiliser le service multipoint de MCS.

Les services GCC sont spécifiés dans la recommandation T.124 [ITU 95a].

#### **2.3.4.2. Les protocoles d'applications**

Les protocoles d'applications s'appuient sur l'infrastructure de communication pour permettre aux applications utilisateurs de communiquer avec d'autres applications situées sur d'autres nœuds. Ces protocoles d'applications fournissent des fonctionnalités requises pour les applications utilisateurs afin qu'elles puissent interopérer entre elles.

T.127 définit le protocole pour l'échange de fichiers binaires au sein d'une conférence multipoint [ITU 95c]. Ce protocole n'impose aucune restriction sur le contenu des fichiers à transférer. Une application utilisateur peut s'appuyer sur ce protocole pour envoyer un ou plusieurs fichiers en même temps à tous les sites de la conférence, ou à des sites spécifiques.

L'ITU propose également un protocole pour le transfert d'image fixe dans la recommandation T.126 [ITU 95b]. Ce protocole permet de normaliser les applications de tableau blanc partagé, l'échange d'images fixes, le contrôle d'application à distance, l'archivage d'images annotées, ainsi que la définition de type d'outils de dessin.

Les protocoles d'applications sont optionnels. Ils ne sont pas obligatoires pour qu'un nœud soit conforme à la norme T.120. Des protocoles d'applications propriétaires peuvent donc être utilisés dans une conférence, tout en s'appuyant sur les services de l'infrastructure de communication de T.120 (Figure 2-6).

#### **2.3.5. Conclusion**

Dans ce paragraphe, nous avons donné un aperçu sur les systèmes de vidéoconférence. Nous avons vu les services de base pour la gestion de conférence, ainsi que les outils nécessaires aux participants pour coopérer comme les tableaux blancs et les télépointeurs. Nous avons également présenté la norme T.120 qui permet de normaliser ses services et ses outils.

Après l'étude de la norme T.120, il nous est apparu que cette norme est essentiellement basée sur l'utilisation d'un nœud central (MCU) pour établir la communication entre les différents sites. Les solutions utilisant un nœud central ne sont pas extensibles pour un grand nombre de sites désirant coopérer, et les limites du nœud central peuvent être rapidement atteintes si les participants transmettent une grande quantité de données.

De plus, T.120 ne traite pas le problème de l'hétérogénéité des ressources des participants. Généralement, la communication est établie avec les paramètres du site ayant les plus faibles capacités.

Dans nos travaux, nous nous sommes inspirés de la norme T.120 pour l'implémentation des services de base d'une conférence, mais nous avons opté pour une solution distribuée pour les aspects de communication entre les sites pour aboutir à une solution extensible à de grands groupes et adaptable à l'hétérogénéité de leurs systèmes.

## **2.4. Le Media Space**

Imaginez qu'un jour de début de semaine, vous rentrez le matin dans les locaux de votre société à Lille. En passant par des bureaux de vos collègues vous dites bonjour. Puis vous passez à côté d'un écran placé dans le couloir, dans lequel vous apercevez un de vos collègues à Paris qui vient de passer la porte d'entrée des locaux parisiens. Vous vous arrêtez pour discuter de ce que chacun de vous a fait pendant le week-end, puis de vos travaux respectifs. Après cette brève discussion, chacun de vous rejoint son bureau pour commencer une nouvelle journée de travail. Cette façon de pouvoir communiquer de manière informelle et imprévisible avec des sites distants est rendue possible grâce au concept de Media Space.

Nous devons le concept de Media Space aux laboratoires de Xerox qui l'ont utilisé pour la première fois lorsque les membres d'un laboratoire ont été répartis sur deux sites [Bly 93]. Le but du Media Space mis en place était de maintenir un groupe uni pour permettre aux membres des deux sites de continuer à travailler ensemble.

### **2.4.1. Définitions**

Dans [Bly 93], le Media Space est défini comme suit :

*C'est un ensemble de systèmes électroniques dans lequel un groupe de personnes peuvent travailler ensemble, même s'ils ne se trouvent pas au même endroit ou s'ils ne sont pas présents au même moment. Dans le Media Space, une personne peut créer des environnements visuels et acoustiques qui regroupent des sites éloignés. Elle peut aussi contrôler l'enregistrement, accéder et rejouer des images et des sons à partir de ces environnements.*

Dans cette définition, l'accent est mis sur l'utilisation des médias son et vidéo comme moyen de communication pendant le travail en groupe. Un nœud Media Space doit comprendre les moyens nécessaires pour recevoir et envoyer le son et la vidéo, un logiciel pour le contrôle des connexions entre les sites, ainsi que les logiciels et les outils pour faciliter la coordination et la collaboration entre sites.

Cependant, le Media Space ne se limite pas seulement aux sessions de travail formelles, mais accorde une grande importance aux communications informelles. De ce fait, le Media Space est un outil qui met l'accent sur la nature sociale du travail.

### **2.4.2. Les services d'un Media Space**

La différence entre un Media Space et un système de vidéoconférence classique, ne se fait pas par comparaison des dispositifs constituant les deux systèmes, qui peuvent être similaires, mais plutôt par leur utilisation. Un système de vidéoconférence permet à un groupe de personnes de travailler ensemble à un instant généralement fixé à l'avance, et pendant une durée limitée dans le temps. De l'autre côté, en plus des services de conférence, un Media Space offre d'autres services qui permettent d'assister les communications informelles qui ne sont pas prévues dans le temps et qui peuvent durer longtemps. Ces services font en sorte que les utilisateurs retrouvent leur habitude et leur comportement dans le monde réel. Le coup d'œil (glance), la connexion d'arrière plan, le partage de bureau, et la liaison vidéophone en sont des exemples. Ces services sont décrits successivement.

- ***La connexion d'arrière plan***

Une notion à prendre en considération dans les systèmes de travail coopératif est la notion d'awareness [Dourish 92, Gutwin 96, McDaniel 97]. L'awareness, qui peut être traduite par conscience, signifie le fait d'être conscient de la présence des autres.

Dans un Media Space, les utilisateurs peuvent posséder une connexion d'arrière plan pour avoir en continu une liaison audio-vidéo, montrant généralement un endroit public (couloir, cafétéria,...). Ceci augmente la conscience de l'existence et de la présence des autres.

Les vues généralement trouvées dans les bureaux des personnes sont les lieux communs des autres sites. Il y a souvent des sons provenant des autres sites rendant compte des activités en cours. Être conscient de ces activités, donne un aperçu de ce qui se passe et des personnes qui sont autour.

Ce service augmente les chances de la réalisation des rencontres fortuites. Les rencontres entre collègues de travail peuvent généralement se faire par hasard dans les couloirs du lieu de travail, ou dans les cafétérias. Elles ne sont pas intentionnelles et peuvent engendrer l'opportunité d'interaction entre les personnes. Dans ce cas aussi, les lieux communs jouent un rôle important pour les rencontres fortuites. Pour cela, dans la majorité des Media Space développés, des liaisons audio-vidéo relient les lieux communs des sites distants. Lorsqu'une personne passe par un lieu commun, elle jette un coup d'œil sur l'écran du site éloigné. Si dans l'autre site une personne passe aussi par le lieu commun, une rencontre peut se réaliser.

La connexion d'arrière plan permet donc de remplir les principaux buts du Media Space : assister les communications informelles entre ses utilisateurs, et procurer une sensation de groupe permettant de savoir qui est présent et avec qui il est possible de discuter ou travailler.

- ***Le coup d'œil***

C'est une connexion vidéo de courte durée vers un autre site, permettant de savoir si une personne est dans son bureau. Ce service peut être utilisé pour la localisation de collègues. Contrairement aux rencontres fortuites, la localisation d'une personne est intentionnelle. Une personne peut chercher une autre personne particulière ou voir seulement qui se trouve à un endroit bien précis. Cette personne peut même "entrer" dans un bureau pour demander si quelqu'un a vu la personne recherchée.

Ce service peut poser certains problèmes sociaux. Les utilisateurs peuvent trouver extrêmement "indiscrette" la possibilité donnée aux autres utilisateurs de s'immiscer dans leur bureau [Kraut 93, Dourish 93]. Ainsi, afin d'éviter de brusquer les utilisateurs, certains systèmes de travail coopératif, comme Montage [Tang 94], font apparaître les fenêtres vidéo progressivement. Les utilisateurs ne sont pas surpris, et sont préparés à recevoir des "coups d'œil".

- ***Le partage de bureau***

Les collègues qui travaillent en étroite collaboration ont généralement le bureau du collègue sur le moniteur comme vue d'arrière-plan. Cette connexion audio-vidéo est maintenue en continu pendant une longue durée (plusieurs heures ou jours).

- ***Le vidéophone***

Permet d'avoir une communication audio-vidéo courte, analogue aux communications téléphoniques.

### **2.4.3. Les expériences majeures du Media Space**

Le concept de Media Space a vu le jour dans les laboratoires de Xerox PARC<sup>8</sup>, dans le milieu des années 1980 [Bly 93]. Tout a commencé lorsqu'un laboratoire (SCL<sup>9</sup>) du PARC, a été réparti sur deux sites, l'un à Palo Alto (Californie), et le second à Portland (Oregon). Afin de maintenir un groupe uni et un travail d'équipe cohérent, et par la même occasion explorer les technologies pour assister le travail coopératif, des liaisons audio, vidéo, et informatiques ont été établies entre les deux sites.

Initialement, les liaisons reliaient les lieux publics des deux sites. Puis elles ont été étendues aux bureaux, permettant aux utilisateurs d'organiser des réunions, de discuter, ou simplement de chercher une personne. Ce système fut appelé Media Space. Après la fermeture du site de Portland, le système Media Space fut gardé au PARC pour continuer les recherches dans ce nouveau domaine, auquel il donna le nom.

RAVE<sup>10</sup> [Gaver 92b] est le Media Space développé à Rank Xerox EuroPARC. Chaque utilisateur possède un nœud RAVE. L'interface logicielle de RAVE fournit les services de base d'un Media Space, le coup d'œil, le vidéophone, le partage de bureau, et la vue d'arrière plan. Un utilisateur peut contrôler l'accès visuel à son bureau lorsque d'autres utilisateurs veulent y accéder. Par conséquent, il est toujours conscient lorsqu'une autre personne le regarde via les liaisons vidéo.

Le Media Space soulève le problème de protection et de contrôle d'accès à l'espace privé des individus. Godard [Dourish 91] est un composant logiciel de RAVE pour le contrôle d'accès à l'espace privé. Le développement de Godard est basé sur la relation entre les aspects technologique et social du contrôle [Dourish 93].

CAVECAT est le Media Space développé à l'université de Toronto. CAVECAT reprend les travaux de RAVE, et étend le contrôle d'accès à l'espace individuel. En appliquant des

---

<sup>8</sup> PARC : Palo Alto Research Center.

<sup>9</sup> SCL : System Concepts Lab.

<sup>10</sup> RAVE : Ravenscroft Audio-Video Environment.

règles, il est possible de contrôler plus finement une connexion. L'utilisateur peut spécifier pour chaque service, les personnes qui peuvent l'interrompre.

Le "Cruiser" de Bellcore [Root 88] utilise la métaphore du couloir. Il permet à l'utilisateur de se promener dans un couloir virtuel et de "croiser" d'autres utilisateurs. Au niveau du contrôle, Cruiser utilise la réciprocité. Si un utilisateur voit un autre utilisateur, il est en même temps vu par ce dernier.

#### **2.4.4. Conclusion**

Les Media Space sont des systèmes particuliers pour le travail coopératif. La principale différence qui existe entre les systèmes de vidéoconférence et les systèmes de Media Space est leur utilisation dans le temps. Un Media Space s'utilise de façon continue dans le temps, alors que l'utilisation des systèmes de vidéoconférence est ponctuelle. Avec un Media Space, on vise à créer un environnement de travail pour ses utilisateurs se trouvant sur des sites distants, le plus proche possible de l'environnement qu'ils auraient s'ils travaillaient sur le même site. Des services comme la connexion d'arrière plan, le coup d'œil, le partage de bureau et le vidéophone permettent de créer cet environnement.

Nous remarquons que pour fournir chacun des services cités, un Media Space doit établir des liaisons vidéo entre les différents sites. L'utilisation de la vidéo est indispensable pour la réalisation des services d'un Media Space.

La vidéo contribue à améliorer la communication entre les participants en transportant l'information visuelle pendant les sessions de travail coopératif. Dans le paragraphe suivant, nous aborderons les apports de la vidéo dans les systèmes de travail coopératif en présentant les types d'informations qu'elle transporte.

### **2.5. La Vidéo, source d'informations**

Dans les deux paragraphes précédents, nous avons présenté deux types de systèmes de travail coopératif, les systèmes de vidéoconférence et les systèmes de Media Space. Dans chaque type de système, le média vidéo est utilisé pour faciliter la communication entre les utilisateurs. Pour déterminer les apports de la vidéo dans le travail coopératif, Isaacs et Tang [Isaacs 93] ont réalisé une expérience pour comparer les différences entre les interactions d'un groupe de personnes utilisant un système de visioconférence, un système d'audioconférence, et des réunions face à face. L'expérience a montré que comparée à l'audio, la vidéo permettait

de retrouver l'information communiquée pendant une réunion face à face. Elle permet de montrer la compréhension ou l'accord des personnes, de prévoir les réponses, de donner des informations non verbales, de mettre en valeur des descriptions verbales, de gérer les pauses et d'exprimer les attitudes à travers les expressions du visage et les postures.

Ces informations communiquées sont obtenues à partir d'informations visuelles. Elles sont appelées les informations déduites.

Par contre, la vidéo ne permettait pas d'avoir une vision globale de l'environnement des autres participants comme le fait la réunion face à face.

### **2.5.1. Le comportement visible des participants**

L'information véhiculée par la vidéo renseigne sur le comportement visible des participants [Whittaker 97]. Il inclut le regard, l'expression du visage et les mouvements de la tête, la gestuelle et la posture. Ces comportements sont décrits ci-dessous.

#### **2.5.1.1. Le regard**

Le regard permet à une personne d'extraire l'information visible de son environnement. La direction du regard, sa durée, ainsi que la nature du regard sont des paramètres importants pour communiquer de l'information.

Le regard est aussi un indicateur d'attention. En fonction des différents paramètres des regards des participants, le locuteur peut connaître le degré d'attention portée par les autres participants à ce qu'il dit.

Lorsque deux personnes travaillent ensemble via un système de travail coopératif utilisant seulement le canal audio pour la communication, le problème de reconnaître son correspondant ne se pose pas. Mais si plusieurs personnes sont impliquées dans une conférence multipoint, il devient difficile de reconnaître la personne qui parle [Bruce 96]. D'où un intérêt d'utiliser la vidéo. Bien sûr, on peut résoudre le problème de reconnaissance de la personne qui parle ou qui manipule un outil ou une application partagée, en inscrivant son nom dans l'interface de chaque utilisateur. Mais ceci va induire des efforts en plus (surcharge cognitive) pour les participants, qui risquent d'altérer leur concentration.

Il nous est tous arrivé un jour, au téléphone, de demander à notre interlocuteur s'il était toujours à l'écoute, après un certain temps de silence de sa part. Avec la vidéo, un participant

peut facilement interpréter les longues pauses audio de la part de son interlocuteur, qui peut être en train de chercher un objet ou un document, ou en train de lire un mail [Tang 92].

La vidéo permet aussi de rendre compte du comportement général des participants lors d'une conférence. Ceci a pour effet de créer un climat naturel pour les participants. Elle permet de rendre compte de la présence des autres.

Dans une discussion, le fait qu'une personne évoque une autre personne du groupe, provoque chez les autres membres du groupe le besoin de voir la personne en question.

La vidéo apporte une aide considérable pour le conducteur de conférence. Lors d'une conférence, un participant peut être désigné pour conduire la conférence. Son rôle sera de gérer l'attribution des droits à la parole et à l'utilisation des ressources partagées aux participants. En voyant la réaction des participants, leur comportement, leurs expressions et gestes, il peut mieux décider qui sera le prochain à avoir une ressource particulière. Le comportement d'une personne, peut mener un participant à laisser passer son tour, s'il juge que cette personne peut apporter des informations pertinentes pour le groupe.

#### **2.5.1.2. L'expression du visage et les mouvements de la tête**

En regardant une personne, nous pouvons savoir si elle heureuse ou triste, intéressée ou ennuyée, de bonne humeur ou en colère. Cette information est donnée par l'expression de son visage [Ekman 75]. En plus de l'état émotionnel d'une personne, l'expression du visage peut renseigner le locuteur sur le niveau de compréhension des auditeurs. Ainsi, il pourra anticiper s'il doit continuer les explications, ou bien réajuster le contenu de ce qu'il avait à dire, en répétant une partie des explications, ou bien en détaillant encore plus, et parfois même, décider d'abrégé s'il voit que certains détails sont inutiles.

Des interruptions sont ainsi évitées de la part des participants pour dire « est-ce que vous pouvez répéter », ou « est-ce que tout le monde a compris ». Des interruptions fréquentes auraient gêné le bon déroulement de la session de travail.

En plus de l'expression du visage, le hochement de la tête est un autre moyen non intrusif pour véhiculer de l'information. Avec un mouvement de la tête, une personne peut exprimer son niveau de compréhension ou son accord lors d'une discussion.

### **2.5.1.3. La gestuelle**

Les gestes sont des mouvements qu'une personne peut faire avec ses mains et ses bras pendant une communication. En faisant des gestes, on peut améliorer des descriptions verbales, et par suite, augmenter la compréhension des autres sans avoir recourt à un effort supplémentaire. Il est plus simple de décrire des formes, ou des mouvements, non conventionnels, en s'aidant avec des gestes.

Les gestes peuvent aussi être utilisés pour attirer l'attention sur une personne ou un objet en le désignant. Enfin, les gestes peuvent aussi être utilisés pour passer ou demander la parole entre les participants [Goodwin 81].

### **2.5.1.4. La posture**

L'inclinaison et l'orientation du corps des participants est un autre comportement visible permettant de connaître l'intérêt et l'engagement d'une personne dans une discussion [Bull 78, Goodwin 81]. L'intérêt peut être signalé par la position droite d'une personne. D'un autre côté, une personne qui soutient sa tête avec ses mains, révèle son désengagement de la discussion.

La posture d'une personne est moins dynamique que les autres comportements visibles, qui par contre ont une grande fréquence de changement.

## **2.5.2. Discussion**

Bien que la vidéo apporte beaucoup d'information pour les personnes travaillant en groupe, les utilisateurs ne sont pas toujours satisfaits des systèmes de vidéoconférence.

Les effets d'une technologie sur un groupe peuvent mettre des années avant de devenir apparents [Gale 90]. Nous pensons que l'apport de la vidéo dépend de l'application coopérative. Les applications où la négociation est importante, nécessite d'utiliser la vidéo pour pouvoir bénéficier des apports cités plus haut [Chapanis 75]. Alors que d'autres comme la télé-ingénierie nécessite moins de vidéo, puisque les participants sont beaucoup plus focalisés sur l'espace de travail partagé.

La nouveauté de l'outil, nécessite un certain temps d'apprentissage pour les utilisateurs afin qu'ils puissent s'y adapter. Si nous prenons le téléphone, les utilisateurs se sont adaptés à son utilisation, l'ouïe s'est adaptée à la qualité moyenne du son généré, ainsi que les

personnes adaptent leurs discours en sachant que les personnes de l'autre bout du fil ne les voient pas.

La qualité vidéo explique aussi le mécontentement de certains utilisateurs de systèmes de vidéoconférences. Les délais de transport de la vidéo deviennent gênants lorsqu'ils dépassent un certain seuil. La synchronisation entre le son et la vidéo. La qualité de la vidéo est aussi très importante. Nous sommes habitués à travailler ensemble face à face, ou bien à regarder la télévision qui génère une vidéo de bonne qualité. Avec la technologie actuelle, nous arrivons rarement à approcher la qualité d'une vision face à face ou celle de la télévision. La vidéo générée est généralement de la taille d'un quart d'écran, et le rafraîchissement des images est moins fréquent que celui de la télévision (25 images/s). Ce qui donne de la vidéo saccadée dans un quart d'écran.

## **2.6. Conclusion**

Les systèmes de vidéoconférence et de Media Space sont deux types de systèmes permettant à leurs utilisateurs, dispersés géographiquement, de travailler en groupe. Si les configurations matérielles de ces systèmes peuvent être similaires, leurs utilisations sont par contre différentes. Un système de vidéoconférence est utilisé pour les sessions de travail formelles où le temps de début et la durée sont généralement fixés à l'avance, alors qu'un système de Media Space, en plus des services de conférence, offre d'autres services permettant d'assister les communications informelles qui ne sont pas planifiées dans le temps.

L'utilisation de la vidéo dans ces systèmes de travail coopératif permet d'améliorer la communication entre les participants. Elle véhicule l'information visuelle dont bénéficie naturellement les participants dans des sessions de travail en présentiel, brisant ainsi les barrières de distances entre les participants.

L'information visuelle véhiculée par la vidéo inclut les comportements des participants comme le regard, l'expression du visage et les mouvements de la tête, la gestuelle et la posture. Ces comportements peuvent renseigner sur la compréhension des personnes, de donner des descriptions non verbales, ou d'exprimer des attitudes. La vidéo contribue donc de manière significative au bon déroulement de la session de travail coopératif, et nous permet ainsi de retrouver l'information communiquée pendant une réunion de travail face à face.

Cependant, l'utilisation de la vidéo demande beaucoup de ressources : de la puissance de calcul pour son traitement et de la bande passante réseaux pour son transport. Or, les systèmes

des utilisateurs sont généralement hétérogènes. Si les normes établies pour la vidéoconférence permettent à des équipements fabriqués par des constructeurs différents de communiquer, elles ne tiennent pas compte de l'hétérogénéité des ressources des systèmes. Pour communiquer, les différents sites établissent généralement les liaisons en fonction du site ayant les capacités les plus faibles. Tous les sites participant à la session de travail coopératif vont alors recevoir des services vidéo, dont la qualité est celle permise par les ressources du site ayant les capacités les plus faibles. Par conséquent, les sites ayant des ressources plus élevées ne peuvent pas bénéficier de leurs capacités.

Le chapitre suivant traitera de l'hétérogénéité des systèmes utilisateurs et de leurs influences sur l'utilisation de la vidéo. Nous verrons aussi que la diversité des activités coopératives implique des services vidéo différents. Dans certaines activités, la vue des interlocuteurs peut être nécessaire pour le participant afin qu'il puisse accomplir sa tâche convenablement dans le groupe. Par contre, dans d'autres activités, un participant peut très bien se passer de la vue des autres membres du groupe. Dans ce cas, l'utilisation de la vidéo est alors accessoire, et peut même devenir perturbante.



---

### **3. Hétérogénéité des systèmes et diversité des activités coopératives**

---

### **3.1. Introduction**

L'utilisation de la vidéo est nécessaire dans les systèmes de travail coopératif pour permettre aux participants de se voir, améliorant ainsi leur communication. Cependant, la diversité des activités coopératives et l'hétérogénéité des systèmes utilisateurs agissent sur la façon d'utiliser la vidéo.

Comme il existe diverses activités coopératives chacune possédant des caractéristiques propres, des systèmes dédiés sont généralement conçus pour chaque type d'activité. En particulier, la gestion des services vidéo est différente pour chaque activité. Le besoin en vidéo pour les participants varie en fonction du type d'activité coopérative.

De plus, la disponibilité des ressources de temps de calcul et de bande passante réseau sont indispensables pour le codage et la transmission de la vidéo. Si le problème couramment rencontré dans les systèmes distribués est l'hétérogénéité de ceux-ci, il s'accroît pour les applications multimédia distribuées. Dans une activité coopérative, plusieurs participants sont mis en relation pour travailler. Le site de chacun aura un équipement et un accès réseau particulier. Dans un milieu hétérogène, c'est le site qui a le moins de ressources qui impose la qualité de la communication.

Dans ce chapitre, nous passons en revue la diversité des activités coopératives tout en précisant son influence sur l'utilisation de la vidéo ; nous nous intéresserons également à l'hétérogénéité des systèmes terminaux et des réseaux. Puis nous passerons en revue les solutions actuelles pour résoudre le problème de l'hétérogénéité et nous montrerons leurs limites à l'aide d'un exemple.

### **3.2. Diversité des activités coopératives**

Un groupe de personnes peuvent se réunir pour collaborer à différents types d'activités coopératives. En utilisant un système de travail coopératif, l'éloignement géographique entre les participants devient transparent. Cette transparence est en partie rendue possible grâce à l'utilisation de la vidéo pour permettre aux participants de se voir. Dans le chapitre précédent, nous avons vu que l'utilisation de la vidéo contribue à l'amélioration de la communication entre participants. Cependant, l'apport de la vidéo va dépendre du type de l'activité coopérative.

Dans les applications « très conversationnelles » comme dans des sessions de négociation, la vue des personnes avec lesquelles l'utilisateur travaille est indispensable pour déduire leur accord ou leur compréhension à travers leurs gestes et expressions. Par contre, dans les applications où les participants sont surtout focalisés sur le document partagé, comme l'édition en commun d'un document, l'apport de la vidéo n'est pas indispensable. L'écran de visualisation de chaque participant est alors dominé par l'espace de collaboration [Buxton 92]. Dans ce paragraphe, nous allons présenter diverses activités coopératives tout en spécifiant l'utilité de la vidéo pour chacune d'elle.

### **3.2.1. L'enseignement à distance**

En plus de sa contribution à l'élimination de l'éloignement géographique entre personnes, la téléconférence permet une meilleure utilisation des experts de certains domaines spécifiques. En dispensant une conférence à distance, un expert peut faire profiter de ses connaissances un grand nombre de scientifiques. Les systèmes d'enseignement à distance utilisent pleinement cet avantage des systèmes de vidéoconférence.

Le VisioCentre est un système d'enseignement à distance développé à l'ENIC [Lafon 94]. Il permet à un ensemble d'élèves répartis sur plusieurs sites distants de suivre en même temps un cours donné par un professeur.

Pendant le cours, l'image du professeur est diffusée à tous les sites élèves. Chaque site est équipé de pupitres sur lesquels les élèves ont à leur disposition des boutons (Figure 3-1). Pour poser une question au professeur, un élève peut appuyer sur un bouton du pupitre. Le professeur est alors notifié de cette question par l'apparition d'un point d'interrogation sur son écran de contrôle. Le professeur valide la question en appuyant sur le point d'interrogation. L'image de l'élève posant la question est alors diffusée aux autres sites élèves ainsi qu'au site du professeur.



*Figure 3-1 : Site élève de l'ENIC.*

L'exemple du VisioCentre montre que les systèmes de télé-enseignement sont régis par un protocole bien précis. Ce protocole représente le déroulement d'un cours qui consiste en certaines étapes, comme les étapes d'explication du cours, interrompues par les étapes des questions et réponses.

L'utilisation de la vidéo dans les systèmes d'enseignement à distance correspond au protocole de déroulement du cours. Le professeur est vu par tous les autres élèves lorsqu'il donne son cours. Si un élève pose une question, il est alors vu par le professeur et les autres élèves.

### **3.2.2. La télé-médecine**

Dans [Perednia 95], la télé-médecine est définie comme étant l'utilisation des moyens de télécommunications pour fournir des informations et des services médicaux. Initialement, la télé-médecine trouvait particulièrement son intérêt dans les zones rurales permettant à des populations isolées de bénéficier de services médicaux à distance [Zundel 96]. A l'aide de systèmes de vidéoconférence, les médecins peuvent diagnostiquer à distance des personnes et leur proscrire des traitements à suivre. On parle alors de *télé-diagnostic* et de *télé-consultation*.

Comme pour l'enseignement, les systèmes de travail coopératif permettent une meilleure utilisation des experts dans le domaine médical. Dans certaines spécialités médicales, les

experts sont très sollicités. Pour une urgence, un expert n'ayant pas le temps de se déplacer jusqu'au site où se trouve le patient, peut utiliser des outils de travail coopératif afin d'établir un diagnostic ou diriger une opération chirurgicale.

En *télé-chirurgie*, application permettant à un chirurgien de diriger une opération chirurgicale à distance, l'apport vidéo est indispensable. L'expert aura besoin de disposer d'information vidéo, afin de pouvoir fournir un diagnostic et éventuellement diriger l'opération à distance.

A la différence des autres applications de travail coopératif qui peuvent se contenter d'une qualité vidéo moyenne pour permettre aux participants de voir leurs correspondants, la télé-chirurgie est une application très importante pour sauver des vies humaines. Une haute qualité vidéo est alors exigée pour que l'expert puisse accomplir son travail dans d'excellentes conditions et fournir un bon diagnostic ou opérer avec précision.

### **3.2.3. La télé-réunion**

Les systèmes de vidéoconférence permettent à des personnes dispersées géographiquement de se réunir pour travailler ensemble. Les réunions à distance ont un double intérêt. D'une part, elles permettent la réduction des coûts relatifs aux déplacements des personnes, d'autre part elles permettent un gain de temps. Lorsqu'une personne se déplace pour une réunion de deux heures, elle perd généralement toute une journée, alors qu'avec la réunion à distance, la personne reprend immédiatement ses activités au sein de son organisation, une fois la réunion terminée.

Dans les réunions de conception, plusieurs concepteurs dispersés géographiquement coopèrent dans l'objectif de concevoir un produit. Les concepteurs sont focalisés sur le document de travail partagé. Dans ce type de réunion, l'utilisation de la vidéo n'est pas indispensable pendant toute la durée de la conférence. Dans [Miah 94], les expériences réalisées avec le système de conférence proposé ont montré que les participants se focalisent sur la vidéo au début de la réunion pour échanger des salutations. Une fois le travail commencé, leurs intérêts se déplacent vers les documents partagés pendant toute la durée de la session de travail. Enfin, la phase finale est similaire à la phase initiale. Les participants se focalisent encore une fois sur la vidéo pour échanger des formules d'au revoir.

Contrairement aux réunions de conception, les réunions de négociation sont principalement basées sur la conversation. Chaque participant peut prendre la parole. Il n'y a pas de hiérarchie comme celle existant entre professeur et élèves dans l'enseignement à distance. Le

changement de locuteur peut être fréquent. Il est nécessaire dans ce type de réunions que les participants puissent se voir afin de mieux communiquer en complétant la parole avec les gestes, les hochements de tête, les expressions du visage et tous les autres comportements visibles des participants.

A la différence de la télé-médecine, la télé-réunion n'exige pas une qualité vidéo excellente. Une qualité vidéo permettant de distinguer les participants est suffisante.

### **3.2.4. Conclusion**

D'après les exemples d'activités coopératives que nous avons présentés, il apparaît clairement que l'utilisation de la vidéo doit dépendre de l'activité coopérative. En télé-enseignement, nous avons vu qu'un protocole précis régit le déroulement du cours. Le professeur et les élèves interviennent à des phases déterminées, alors qu'en télé-réunion chaque participant peut intervenir à tout moment de la discussion. La télé-médecine quant à elle, se distingue par le fait de nécessiter une vidéo d'excellente qualité alors qu'une vidéo permettant de reconnaître une personne est suffisante pour la plupart des autres activités coopératives.

Les solutions actuelles pour ce problème de diversité d'activité coopérative, est la conception de systèmes dédiés pour des types d'activités coopératives donnés ou bien la conception de systèmes généralistes.

Pour la première solution, des utilisateurs désirant participer à différentes activités doivent installer autant de systèmes dédiés que d'activités. Les systèmes généralistes quant à eux risquent de ne pas s'adapter à tous les types d'activités.

Dans nos travaux, nous traitons le problème de la diversité d'activité coopérative afin de pouvoir gérer les services vidéo à offrir aux participants en fonction de l'activité coopérative.

## **3.3. Diversité des équipements utilisateurs**

Dans les applications de vidéoconférence, différents types d'équipements sont utilisés par les participants. On distingue les systèmes de type poste de travail (Desktop) et les systèmes de type salles de conférence (Room).

### 3.3.1. Les systèmes de type poste de travail

Les systèmes poste de travail sont généralement utilisés comme systèmes individuels. Ce type de système consiste en une station de travail à laquelle on a ajouté des éléments d'acquisition et de restitution sonore et vidéo, ainsi que les logiciels et outils de coopération comme le tableau blanc et le télépointeur (Figure 3-2).



Figure 3-2 : Système poste de travail.

Les systèmes poste de travail peuvent traiter la vidéo de manière logicielle ou matérielle. Dans le premier cas, la compression vidéo est réalisée par le processeur de la station de travail. Quant à la compression matérielle, elle est possible lorsque l'on rajoute des composants spécifiques dédiés comme des codecs<sup>11</sup> pour la compression vidéo. Cette solution permet de disposer d'une vidéo de meilleure qualité, mais elle a un coût plus élevé que la solution logicielle.

Lorsque les participants d'une même conférence possèdent des systèmes différents, ils disposeront alors de services vidéo différents.

### 3.3.2. Les systèmes de type Salle de Conférence

Les premiers systèmes de vidéoconférence étaient des salles de conférences. Seules les grandes organisations avaient les moyens financiers pour installer de telles salles dans leurs

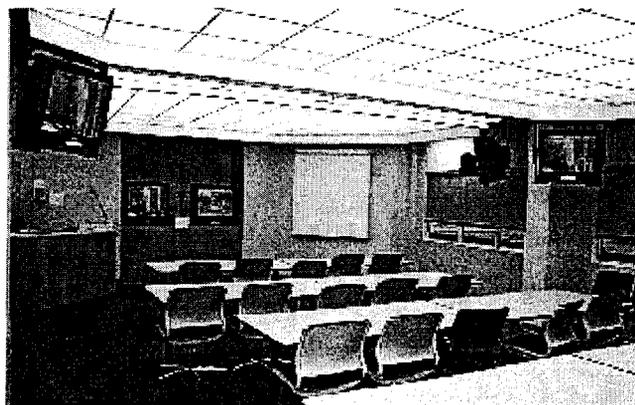
---

<sup>11</sup> Codec : codeur / décodeur.

locaux. A cette époque, les stations de travail ne possédaient pas une puissance suffisante pour gérer les données audio et vidéo.

Aujourd'hui on pourrait supposer que le besoin en de tels systèmes diminue. Au contraire, ceux-ci sont aussi utilisés qu'avant puisqu'ils sont utilisés pour des applications spécifiques comme le télé-enseignement, ou des réunions de travail entre unités d'une même entreprise par exemple.

La configuration matérielle d'une salle de conférence n'est pas la même que celle d'une solution sur station de travail. Généralement, une salle offre des services de meilleure qualité. On peut disposer de plusieurs caméras contrôlables à distance, plusieurs microphones et de plusieurs écrans de visualisation qui sont plus grands que ceux des stations de travail. La disposition des participants dans la salle est étudiée pour que les caméras utilisées puissent avoir tout le monde dans leur champ de vision. On utilise généralement des tables en V ou semi-circulaires, ou des salles de type amphithéâtre (Figure 3-3).



*Figure 3-3 : Salle de vidéoconférence.*

Un système salle de conférence est un système de groupe. Il est équipé pour accueillir plusieurs participants sur un même site pour travailler avec d'autres sites similaires ou d'autres sites poste de travail.

### **3.3.3. Conclusion**

Dans une application de travail coopératif, les participants utilisent des systèmes hétérogènes. Cette hétérogénéité introduit une différence entre les capacités des systèmes

utilisateurs pour traiter les données, particulièrement les données vidéo. Le traitement de la vidéo nécessite beaucoup de puissance de calcul. Plus le participant possède de la puissance, meilleure est la qualité vidéo possible. Lorsque les systèmes sont hétérogènes, les participants peuvent donc disposer de différentes qualités vidéo en fonction de leurs capacités.

Les systèmes de travail coopératif doivent alors fournir des services vidéo adaptables aux ressources des sites utilisateurs.

### **3.4. Hétérogénéité des réseaux**

Le monde des réseaux connaît actuellement une forte évolution au niveau technologique obligeant les sociétés utilisatrices à renouveler leurs infrastructures réseaux. Chaque société effectue un choix sur la technologie à adopter en fonction de ses besoins. Il en résulte que le parc réseaux mondial devient hétérogène. Toutes les technologies réseaux ne sont pas adaptées pour les applications multimédia. L'utilisation des médias continus audio et vidéo en vidéoconférence, exige certaines caractéristiques qui doivent être satisfaites par les réseaux utilisés. Dans ce paragraphe, nous allons étudier ces différentes caractéristiques, puis nous présenterons les réseaux les plus utilisés actuellement [Tanenbaum 96, Pujolle 97].

#### **3.4.1. Caractéristiques des réseaux pour le multimédia**

Pour transporter des flux multimédia, un réseau doit satisfaire certaines caractéristiques. Parmi ces caractéristiques, nous pouvons citer la bande passante qui doit être suffisante pour pouvoir transporter la quantité d'information requise pour les flux multimédia, la gestion efficace de ressources pour permettre à plusieurs applications de les utiliser en même temps, la garantie des performances requises par les applications, la scalabilité<sup>12</sup> du réseau, et sa possibilité à supporter le multicast.

##### **3.4.1.1. La bande passante réseau**

La caractéristique de base d'un réseau pour qu'il supporte les communications multimédia est qu'il ait une large bande passante. On parle aussi de vitesse du réseau pour désigner la largeur de bande passante.

---

<sup>12</sup> Ce terme peut être traduit par extensibilité.

La vitesse d'un réseau dépend du support physique (média) utilisé pour la transmission, de la pile de protocoles, de la distance entre les nœuds intermédiaires (commutateur), et la vitesse de commutation des nœuds intermédiaires.

Entre la paire torsadée, le câble coaxial et la fibre optique, c'est cette dernière qui offre la plus grande bande passante pour la transmission. Théoriquement, une fibre optique peut supporter quelques terabits/s.

Nous devons distinguer la vitesse à deux points différents du réseau. Le premier est l'accès réseau utilisateur, le second est la vitesse entre les nœuds intermédiaires.

En utilisant la compression, un utilisateur a besoin de quelques Mbits/s pour les applications multimédia. Pour supporter les applications conversationnelles et de consultations de données, cet accès doit être disponible dans les deux directions (réception et émission).

La vitesse entre les nœuds intermédiaires doit être de quelques centaines de Mbits/s pour pouvoir supporter plusieurs utilisateurs et plusieurs applications au même moment.

#### **3.4.1.2. Partage efficace de ressources réseaux**

Les données multimédia sont à débit variable (sporadique), surtout après compression. Si chaque utilisateur réserve une bande passante correspondant au débit crête des données, de la bande passante sera perdue lorsque le débit de sortie n'est pas à son maximum. La meilleure approche est d'utiliser le principe de la bande passante à la demande ou le multiplexage statistique : une application peut utiliser autant de bande passante jusqu'au maximum requis. Lorsqu'elle n'utilise pas une partie, les autres applications peuvent l'utiliser.

La technologie de commutation de circuit n'est pas adaptée, parce qu'elle établit un circuit entre la source et la destination avec une bande passante fixe, même si la bande passante n'est pas totalement utilisée.

La technique de multiplexage statistique permet de multiplexer plusieurs flux indépendants à débit variable sur un canal à large bande passante. Lorsque certains flux sont à haut débit, d'autres sont à bas débit. La somme des débits à un instant donné est inférieure à la somme des débits crêtes de ces flux. Ceci est dû à la loi des grands nombres : étant donné un grand nombre de flux de données non corrélés, la bande passante nécessaire reste constante, même si des flux individuels font varier la quantité de données qu'ils émettent. La constante dépend de la caractéristique statistique de chaque flux.

La taille des paquets peut aussi influencer sur l'efficacité d'utilisation de bande passante. Lorsque le paquet est trop long, plusieurs paquets ne sont pas remplis, spécialement le dernier du message, faisant ainsi perdre de la bande passante. Lorsque le paquet est trop court, une bande passante conséquente est perdue pour le transport des données d'en-tête.

Un autre point concernant l'efficacité, est la retransmission des données perdues. Pour les applications temps réel, la retransmission de l'audio et de la vidéo n'est pas souhaitée. Premièrement, la retransmission rajoute un délai supplémentaire qui fait que les données multimédia seront jouées trop tard. Deuxièmement, nous pouvons tolérer des pertes et des erreurs dans les présentations multimédia. En général, la retransmission fait perdre de la bande passante et introduit un délai supplémentaire pour la réception des paquets concernés.

### **3.4.1.3. Garantie de performance**

Pour garantir les performances, un réseau doit garantir qu'un paquet peut accéder au réseau en un temps spécifié, et qu'une fois sur le réseau, le paquet doit être livré en un temps fixé.

Nous avons deux cas où les performances ne sont pas garanties. Le premier cas se situe dans les protocoles de bas niveau pour accéder au réseau, tels que les protocoles MAC<sup>13</sup> qui ne garantissent pas le temps d'accès au réseau. Un paquet peut mettre un temps assez long avant d'être mis sur le réseau. Par exemple, CSMA/CD<sup>14</sup> utilisé par Ethernet [Metcalfe 76] laisse les stations attendre que le réseau ne soit pas occupé avant qu'elles ne puissent émettre. Après la transmission, une collision est encore possible. Si une collision se produit, le même paquet sera retransmis après un certain temps aléatoire.

Le deuxième cas où les performances ne sont pas garanties, se trouve au niveau des commutateurs réseau. Une fois le paquet placé sur le réseau, il peut transiter par un ou plusieurs nœuds intermédiaires, introduisant ainsi un délai supplémentaire avant qu'il ne soit livré. Ceci est dû à la bufferisation des paquets dans les nœuds intermédiaires qui est une opération qui introduit un délai indéterminé. De plus, le buffer du nœud peut être plein provoquant la perte du paquet.

Donc, pour gérer efficacement les ressources réseaux et garantir les performances pour les applications, les points suivants doivent être pris en compte :

---

<sup>13</sup> MAC : Medium Access Control.

<sup>14</sup> CSMA/CD : Carrier Sense Multiple Access / Collision Detection.

- Les caractéristiques d'un trafic doivent être déterminées en termes de débit crête des données, de moyenne du débit de données, intervalle entre les crêtes, délai, et variation de délai. Ces informations sont utilisées pour établir une connexion. Le réseau utilise ces informations pour décider s'il peut accepter l'établissement de cette connexion en fonction des ressources disponibles. Les caractéristiques du trafic peuvent être pensées comme un ensemble d'exigences pour la qualité de service au niveau réseau.
- Le temps d'accès au réseau doit être garanti.
- Les ressources réseau (bande passante et les files d'attente) doivent être gérées efficacement pour pouvoir garantir les performances de plusieurs applications. Cette partie est appelée le dimensionnement du réseau. Elle étudie combien de connexions peuvent avoir leurs performances garanties avec une quantité de ressources réseau donnée. Autrement dit, le but du dimensionnement du réseau est de déterminer la quantité de ressources ou équipement requis pour satisfaire la demande du trafic spécifié.

#### **3.4.1.4. Scalabilité du réseau**

La scalabilité d'un réseau peut s'utiliser pour la distance, la bande passante, ou le nombre d'utilisateurs. En termes de distance, la même architecture réseau et protocole doivent pouvoir opérer en LAN<sup>15</sup> ou WAN<sup>16</sup>. Ceci facilitera l'interconnexion entre ces réseaux. En termes de bande passante, la bande passante du réseau doit pouvoir être étendue avec la demande des utilisateurs sans changer les protocoles réseaux. Et enfin le réseau doit pouvoir supporter un grand nombre d'utilisateurs et la bande passante disponible à chacun ne doit pas être affectée par le nombre de stations connectées sur le réseau. Généralement, les réseaux locaux (LAN) utilisent un média de transmission partagé. La bande passante occupée par chaque station décroît lorsque le nombre de stations actives croît. Ceci n'est pas désiré dans les applications multimédia.

#### **3.4.1.5. Le Multicast**

Dans les applications multimédia, il est souvent utile de distribuer un même flux à plusieurs destinations. Par exemple en vidéoconférence, la vidéo et la voix du speaker doivent

---

<sup>15</sup> LAN : Local Area Network .

<sup>16</sup> WAN : Wide Area Network.

être envoyées à tous les participants de la conférence situés à des endroits différents sur le réseau. Si les données sont transmises en mode unicast, où une copie des données est envoyée à chacun des participants (Figure 3-4-a), il y a introduction d'un long délai et perte de ressources. Ce délai est dû au fait que l'accès au réseau et la transmission mettent un certain temps pour chaque participant. Il y aura aussi perte de ressources parce que les mêmes données seront envoyées plusieurs fois sur les mêmes liens dépendant du nombre de participants et de leur endroit sur le réseau.

La meilleure solution est que la source envoie une seule fois les données et c'est le réseau qui sera responsable de transmettre les données à plusieurs destinations. Cette technique est appelée le Multicast [Deering 91]. En transmission multicast, les mêmes données ne sont jamais transmises plus d'une fois sur le même lien réseau (Figure 3-4-b), permettant ainsi une transmission rapide et efficace des données.

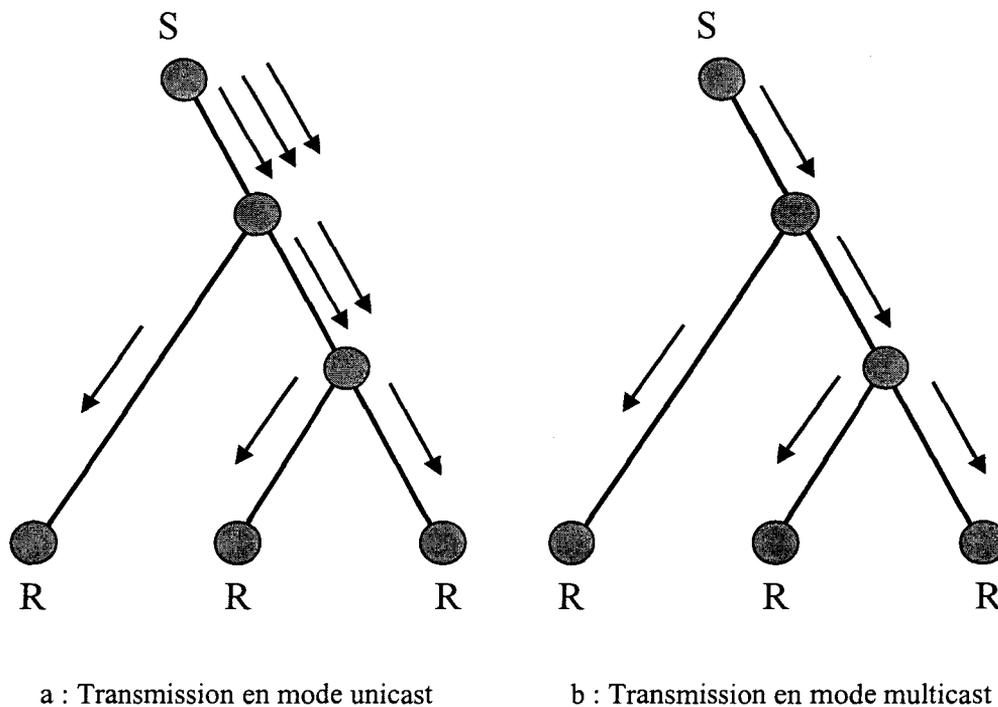


Figure 3-4 : Comparaison des modes de transmission de données entre une source *S* et des récepteurs *R*.

Sur la plupart des LANs, les stations partagent le même médium de transmission. Les données transmises atteignent donc toutes les stations. Dans ce type de réseau, le multicast est simple à implémenter : les stations désirant recevoir les données multicast, peuvent prendre une copie lorsque les données passent.

En WAN, il est difficile pour un commutateur de circuit de connecter dynamiquement un lien en entrée à plusieurs liens en sorties. Mais ceci est facile pour un commutateur paquet. La technique du Multicast a été développée pour les réseaux à commutation de paquet.

Le principe du Multicast est le suivant. Les stations intéressées par recevoir un flux Multicast forme un groupe de Multicast. Ce groupe a une adresse unique. Les commutateurs réseaux seront informés de l'adresse de ce groupe. Lorsqu'une station veut envoyer des données à ce groupe, l'adresse du groupe est utilisée comme adresse de destination. Lorsqu'un commutateur reçoit un paquet avec une adresse de groupe, il envoie ce paquet sur les liens menant aux stations de ce groupe.

Les applications de vidéoconférence représentent un exemple où le Multicast est nécessaire, puisque chaque station doit communiquer avec plusieurs autres stations avec un seul accès réseau. Ceci n'est pas possible avec un réseau à commutation de circuit, par contre c'est possible avec un réseau à commutation de paquet, qui permet d'établir plusieurs circuits virtuels avec d'autres stations à partir d'un seul point d'accès réseau.

### **3.4.2. Les réseaux**

Dans cette partie, nous présentons les réseaux largement déployés dans lesquels des applications de vidéoconférence sont utilisés. Nous décrirons successivement RNIS, ATM, et Internet.

#### **3.4.2.1. Le réseau RNIS**

Le réseau RNIS<sup>17</sup> [Decina 86] est né du besoin d'un réseau de communication unique pour véhiculer indifféremment des données, de la voix, des images fixes ou animées. Il fut adopté par la majorité des constructeurs de produits de vidéoconférence, comme Intel<sup>18</sup> et PictureTel<sup>19</sup> pour respectivement les gammes de produits Proshare et PCS.

RNIS est un réseau à commutation de circuit. Donc, il garantit le transport des médias continus en temps réel. Pour l'utilisateur, deux types de raccordements sont possibles, l'accès de base et l'accès primaire. L'accès de base offre deux canaux B à 64 Kbits/s chacun, alors que l'accès primaire offre jusqu'à 30 canaux B.

---

<sup>17</sup> RNIS : Réseau Numérique à Intégration de Services.

<sup>18</sup> <http://www.intel.com/proshare/conferencing/>

<sup>19</sup> <http://www.picturetel.com/>

Cependant, RNIS est basé sur des technologies anciennes, qui ne lui permettent pas de gérer les ressources efficacement. En établissant une liaison entre deux sites, les ressources ainsi réservées, ne peuvent être utilisées par d'autres sites même s'il n'y a pas de données qui transitent dans cette liaison. L'autre point faible de RNIS, est sa bande passante. Bien qu'il soit possible de choisir différentes bandes passantes pour une liaison en utilisant plusieurs canaux B, nous ne pouvons pas dépasser les 2 Mbits/s qui correspond au débit maximal possible avec l'accès primaire à 30 canaux B.

#### 3.4.2.2. Les réseaux ATM

L'ATM<sup>20</sup> [Minzer 89, De Prycker 93, Onvural 93] est une technologie réseau qui a été « pensée » pour le multimédia, et donc pour combler les lacunes de RNIS au niveau de la bande passante et de la gestion de ressources.

L'ATM offre des bandes passantes nettement supérieures à celles du RNIS. Aujourd'hui, les utilisateurs d'ATM peuvent bénéficier de bandes passantes allant de 25 Mbits/s à 622 Mbits/s. De plus, ATM est conçu de manière à pouvoir véhiculer des données ayant des caractéristiques variées. Ainsi, il est possible de multiplexer de la voix, de la vidéo et des données sur le même lien physique. A l'établissement de la connexion, des paramètres comme la bande passante à allouer et les délais de transfert sont spécifiés pour désigner la qualité de service. Cette qualité de service va dépendre du type d'application et des données à transmettre. Pour des applications de vidéoconférence, le transport des données multimédia doit se faire en temps réel, alors que des applications de transfert de données « statiques » n'ont pas cette contrainte. L'ATM Forum<sup>21</sup>, le principal organisme de normalisation pour ATM a défini différentes classes de qualités de services [ATM Forum 96]. Ces classes de services sont :

**CBR (Constant Bit Rate)**, est un service à réservation de bande passante. A l'établissement de la connexion, un contrat est établi entre le réseau et la station cliente permettant à cette dernière de disposer d'une bande passante fixe pendant la durée de vie de la connexion. Ce service convient particulièrement au transport de données temps réel comme le son et la vidéo avec compression à débit constant.

---

<sup>20</sup> ATM : Asynchronous Transfer Mode.

<sup>21</sup> <http://www.atmforum.com/>

**VBR (Variable Bit Rate)**, est un service qui convient aux applications dont le trafic est à débits variables.

**ABR (Available Bit Rate)**, utilise la bande passante restante. Un débit minimum est garanti.

**UBR (Unspecified Bit Rate)**, aucune réservation de bande passante n'est faite. L'acheminement des données n'est pas garanti. Ce service correspond au service offert par le réseau Internet.

Aujourd'hui, la technologie ATM est principalement utilisée sur les backbones, épines dorsales des réseaux des opérateurs de télécommunication. Elle ne s'est pas encore généralisée sur le poste utilisateur final, à cause du coût qui reste relativement cher et de la jeunesse de la technologie qui ne s'est pas encore stabilisée.

### **3.4.2.3. Le réseau Internet**

Internet [Huitema 95] est l'ensemble des réseaux qui utilisent les protocoles TCP/IP<sup>22</sup> [Socolofsky 91, Comer 95, Feit 96] pour former un unique réseau virtuel. Initialement conçu pour le transfert de données, Internet est sujet actuellement à certaines études pour faire évoluer sa pile de protocoles. En effet, les applications utilisées sur Internet utilisent de plus en plus des données « dynamiques » comme l'audio et la vidéo. Or, la pile de protocole actuelle ne garantit pas les contraintes des médias continus, en l'occurrence :

- Le respect de délai de transport pour garantir la réception en temps réel ;
- Le maintien de la synchronisation des flux continus, pour pouvoir les restituer à la réception, à la même fréquence d'échantillonnage d'émission ;
- La garantie de bande passante lorsque le nombre d'utilisateurs sur le réseau augmente.

Pour pallier ces manques au niveau de la suite TCP/IP, l'organisme de standardisation d'Internet IETF<sup>23</sup>, a établi des protocoles qui seront ajoutés à la pile de protocoles TCP/IP pour la prise en compte des données nécessitant du temps réel comme les médias continus. Parmi ces protocoles, nous pouvons citer RTP<sup>24</sup> [Schulzrinne 96] et RSVP<sup>25</sup> [Zhang 93]. En rajoutant des estampilles temporelles et d'ordres dans les paquets de données audio ou vidéo à l'émission, RTP permet à la réception de présenter ces données ordonnées et à la même

---

<sup>22</sup> TCP/IP : Transmission Control Protocol / Internet Protocol.

<sup>23</sup> IETF : Internet Engineering Task Force. <http://www.ietf.org/>

<sup>24</sup> RTP : Real time Transport Protocol.

<sup>25</sup> RSVP : ReSerVation Protocol.

fréquence de leur production. RSVP est un protocole permettant d'allouer de la bande passante réseau, garantissant ainsi aux applications clientes de bénéficier de cette bande passante pour le transport de leurs données.

### **3.4.3. Conclusion**

Dans ce paragraphe traitant l'hétérogénéité des réseaux, nous avons passé en revue les principales caractéristiques nécessaires qu'un réseau doit satisfaire pour transporter des flux multimédia. Nous avons également présenté les réseaux les plus utilisés pour les applications multimédia, à savoir RNIS, ATM et Internet.

Après cette étude, il apparaît clairement que les utilisateurs de systèmes de travail coopératif peuvent avoir des infrastructures réseaux hétérogènes. Par conséquent, ils bénéficieront de services vidéo différents, puisque la vidéo nécessite une large bande passante pour être transportée sur le réseau et possède des contraintes temps réel. Notamment, ceci va influencer la qualité ainsi que le nombre des vidéos reçues par les participants.

Les applications coopératives sont des applications de groupe. Elles nécessitent alors des communications multipoints [Diot 97]. Une communication de groupe est par définition plus gourmande en bande passante qu'une communication point à point. En vidéoconférence, chaque participant émet sa vidéo à tous ses interlocuteurs, et reçoit leur vidéo.

Le réseau RNIS permet d'effectuer des liaisons multipoints en passant par un nœud central (MCU). Cette solution n'est pas extensible lorsque le nombre de participants augmente dans une conférence. Dans les réseaux ATM, la notion de circuit virtuel point à multipoint est introduite dans les spécifications UNI<sup>26</sup> [ATM Forum 95]. Cependant, comme la technologie ATM est jeune, ces spécifications sont sujettes à modifications.

Notre choix s'est donc naturellement porté sur les protocoles TCP/IP pour l'implémentation de la couche communication de nos travaux. La pile de protocole TCP/IP supporte le multicast qui permet la diffusion des données à un groupe. Cette diffusion se fait sans duplication de paquets sur un même lien réseau. De plus, TCP/IP est indépendant de la couche physique du réseau. Il peut être porté sur ATM afin de bénéficier de la large bande passante de ce dernier [Laubach 98].

---

<sup>26</sup> UNI : User to Network Interface

### 3.5. Les solutions actuelles pour le problème de l'hétérogénéité

Dans les paragraphes précédents de ce chapitre, nous avons vu que les systèmes des utilisateurs sont hétérogènes, et que cette hétérogénéité se répercute sur utilisés pour les applications coopératives présentent une diversité et hétérogénéité des systèmes utilisateur.

Le problème de l'hétérogénéité a seulement été abordé au niveau de l'adaptation de bande passante réseau entre les utilisateurs. Généralement, deux solutions sont envisageables, la première sur le réseau, la seconde à la source.

#### 3.5.1. Passerelles de transcodage

La première méthode pour résoudre les problèmes d'hétérogénéité est de mettre sur le réseau des équipements qui rendent transparent la communication entre les systèmes hétérogènes. La solution la plus utilisée, consiste à mettre en place des passerelles entre des réseaux ayant des bandes passantes différentes. Une passerelle permet de connecter de manière transparente une paire de sessions de conférences, en une seule session logique en manipulant les données et les informations de contrôle des flux vidéo (Figure 3-5).

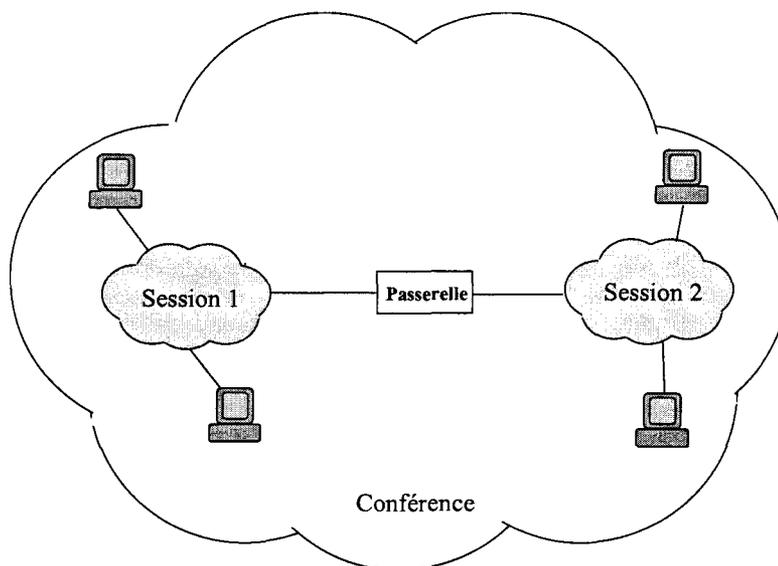


Figure 3-5 : une passerelle entre deux sessions vidéo.

En particulier, la passerelle accomplit les fonctions d'adaptation de bande passante entre les sessions au moyen de transcodage et de contrôle de flux. Des algorithmes de transcodage entre différents types de formats vidéo existent.

Chaque passerelle fera le transcodage entre deux formats vidéo. Ce transcodage peut se faire de deux manières : par conversion directe entre les deux formats vidéo, ou par décodage puis réencodage. La première méthode est plus intéressante au niveau de la performance, mais n'est pas toujours possible si les fonctions de conversion entre deux formats vidéo ne peuvent être déterminées. La deuxième est moins performante puisqu'elle nécessite plus de temps de calcul. Par contre, elle est toujours possible puisque le décodage ramène l'image dans le domaine des pixels, qui est le dénominateur commun de toutes les méthodes de codage.

Dans le cas où les formats source et cible présentent des similitudes, il est possible de faire un décodage/réencodage partiel. C'est le cas par exemple des formats JPEG et H.261 qui sont basés tous les deux sur la DCT<sup>27</sup> (voir Annexe I). Lorsque l'on veut passer des données vidéo au format JPEG d'un réseau vers un autre de moindre bande passante, la passerelle effectue un transcodage vers le format H.261 pour faire baisser le débit. En entrée, la passerelle reçoit le flux JPEG. Chaque image est ramenée vers le domaine DCT, en subissant un décodage d'entropie et une quantification inverse. Les coefficients de la DCT ainsi obtenus, seront réinjectés dans le codeur partiel H.261 pour subir une quantification et un codage d'entropie, pour avoir en sortie de la passerelle un flux H.261 (Figure 3-6).

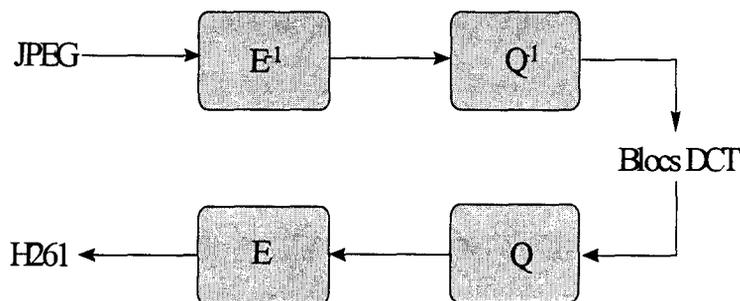


Figure 3-6 : un transcodeur JPEG / H.261 optimisé.

<sup>27</sup> DCT : Discrete Cosine Transform.

IVS-H.261/JPEG [Zhou 95] est une application de vidéoconférence multi-format vidéo. Son architecture est basée sur deux systèmes mono-formats IVS-H.261 [Turletti 94] et IVS-JPEG interconnectés avec une passerelle IVS-SERV permettant l'échange des données et des commandes entre les sous-systèmes mono-formats en transformant leurs formats et en interprétant leur sémantique lorsque cela s'avère nécessaire.

IVS-JPEG utilise un codec matériel (carte XVideo Parallax). Le taux de rafraîchissement d'images peut atteindre 30 images/s ce qui génère un débit d'environ 1500kb/s. IVS-JPEG est adapté pour les participants ayant des stations de travail assez puissantes et des accès réseau à large bande.

Le codec de IVS-H.261 est un codec logiciel. Le débit généré est relativement bas (30kb/s). Ceci est dû d'une part, au codage H.261 qui est un codage inter-trame, et d'autre part à la puissance de calcul des stations de travail qui est limitée (5-10 images/s). IVS-H.261 est donc plus adapté pour les réseaux à bande étroite.

Une passerelle peut aussi faire un transcodage tout en gardant le même format de compression du flux d'entrée, mais avec des paramètres différents pour le flux de sortie. Dans [Amir 95], pour passer d'un réseau à un autre, une passerelle est utilisée. Ainsi, une passerelle est placée entre ATM et Ethernet pour faire baisser les flux vidéo JPEG de 6 Mb/s à 1 Mb/s. Une autre passerelle, qui est en même temps un routeur est utilisée pour transcoder les flux JPEG 1 Mb/s sur Ethernet, en flux H.261 à 128 Kb/s et les envoi sur le Mbone. Une passerelle placée sur Ethernet est utilisée pour faire le transcodage adéquat aux mobiles. Enfin une passerelle est placée entre le Mbone et RNIS pour baisser le débit H.261 à 128 Kb/s sur le Mbone, à 64 Kb/s sur RNIS.

L'utilisation des passerelles rend transparents les problèmes d'hétérogénéité de bande passante, mais en contrepartie, elles peuvent constituer des goulets d'étranglement sur le réseau.

### **3.5.2. Codage vidéo en couches**

La deuxième solution au problème de l'hétérogénéité se situe à la source de l'émission vidéo. Le codage vidéo est constitué de plusieurs couches: une couche principale correspondant au codage de base, et plusieurs autres couches, chacune permettant d'obtenir

une qualité vidéo meilleure lorsqu'elle est ajoutée à la couche de base [Delgrossi 93, McCanne 96, Hoffman 96].

Une fois le codage fait, chaque couche est envoyée sur un canal multicast. A la réception, chaque utilisateur va joindre le nombre de canaux multicast correspondant à sa bande passante.

Le codage en couches peut se faire de plusieurs façons : fréquentiel, spatial, ou temporel. La compression JPEG permet de faire du scaling fréquentiel et spatial avec respectivement le codage progressif et hiérarchique. Le codage progressif permet d'obtenir différentes qualités de la même image en agissant sur les coefficients de la DCT par sélection spectrale ou par approximation successive en agissant sur la précision des coefficients (voir Annexe II).

Dans [Amir 96], le flux en couche est produit en utilisant une DCT en couche (LDCT, layered-DCT). LDCT code les plans de bits des coefficients de la transformée dans l'ordre du plus significatif au moins significatif pour réaliser une structure en couche. Chaque couche est distribuée sur un canal multicast distinct. Chaque récepteur va alors contrôler le nombre de canaux multicast à recevoir en fonction de sa bande passante.

Dans le codage hiérarchique, chaque image est encodée comme une séquence d'images de différentes résolutions. Mais ce codage n'est pas adapté au temps réel puisque qu'il nécessite beaucoup de traitement (un encodage pour chaque résolution d'image). Il est plus utilisé dans les serveurs vidéo à la demande [Chaddha 95a, Chaddha 95b].

L'encodage en couche présente deux inconvénients : il n'est pas normalisé, et il demande un traitement supplémentaire pour le site émetteur.

### **3.6. Un exemple pratique**

De ce qui précède, nous voyons que le problème d'hétérogénéité a été traité seulement au niveau de l'adaptation de bande passante entre sites ayant des accès réseaux différents. Une proposition de gestion de services vidéo dans un milieu hétérogène n'a pas été faite. En effet, un site peut envoyer plusieurs flux vidéo et en recevoir plusieurs en fonction de ses ressources (écran, équipement, réseau) qui sont limitées. Comme dans une session de travail coopératif, nous pouvons être en présence de plusieurs participants, leurs flux vidéo peuvent être de différentes qualités. Aussi chaque flux a une importance particulière pour un utilisateur.

Pour mieux comprendre ce problème, prenant un exemple. Dans une conférence multipartie, chaque site peut avoir un accès réseau différent de celui des autres participants. Un participant ayant un accès ATM veut à un moment donné s'appuyer sur une vidéo enregistrée sur son disque local dans sa présentation (Figure 3-7). Comme son accès réseau lui permet d'envoyer plusieurs flux vidéo, il peut envoyer sa vidéo temps réel capturée par une caméra mise en face de lui, ainsi que la vidéo enregistrée. Les sites récepteurs se trouvant sur ATM peuvent recevoir ces deux flux vidéo. Par contre les sites ayant des accès réseau à bande passante moindre comme Internet ou RNIS ne peuvent pas recevoir tous les flux. Dans ce cas particulier, la solution ne consiste pas à adapter les bandes passantes des différents participants, mais plutôt à effectuer un *choix* du flux à recevoir parmi tous les flux.

Inversement, dans le cas où l'émetteur n'a pas la bande passante nécessaire pour envoyer les deux flux, il faut *privilégier* la vidéo enregistrée sur celle de la personne qui parle.

Donc, en émission ou en réception, il faudra mettre des *priorités* entre les différents flux à émettre ou à recevoir.

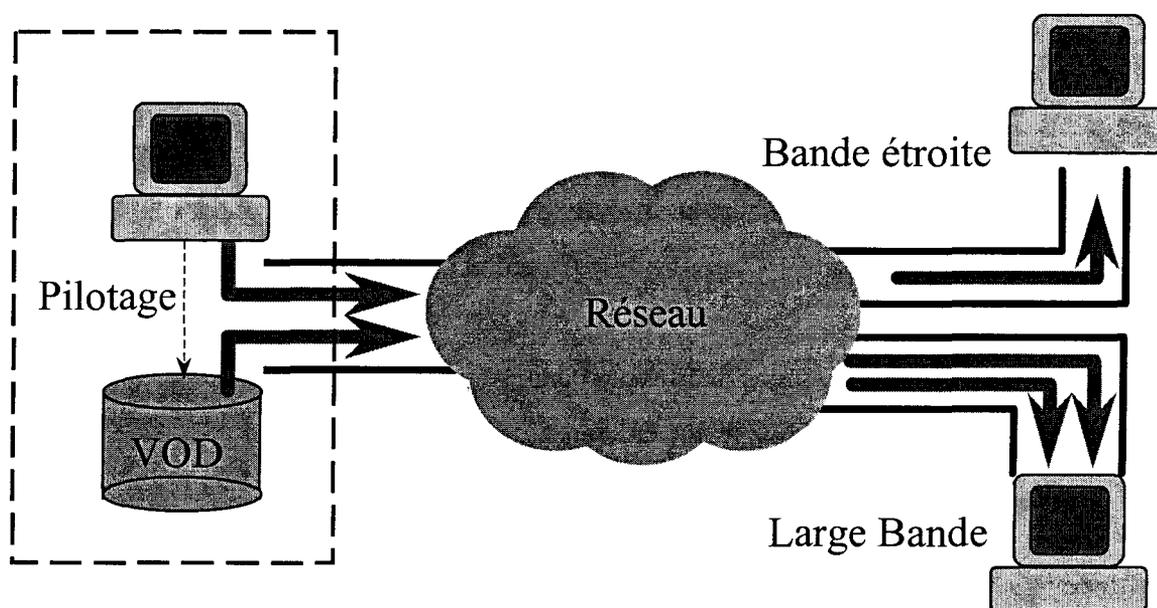


Figure 3-7 : Priorité entre flux vidéo.

Nous pouvons avoir une autre variante de cet exemple en supposant que la vidéo enregistrée peut se trouver sur un serveur à la demande<sup>28</sup> (Figure 3-8). Dans ce cas le site

<sup>28</sup> L'expression anglaise serveur Video On Demand (VOD) est couramment utilisée.

émetteur peut envoyer la vidéo du / des locuteur(s) lorsque la bande passante réseau est étroite (bien sûr en supposant que les bandes d'émission et de réception sont indépendantes).

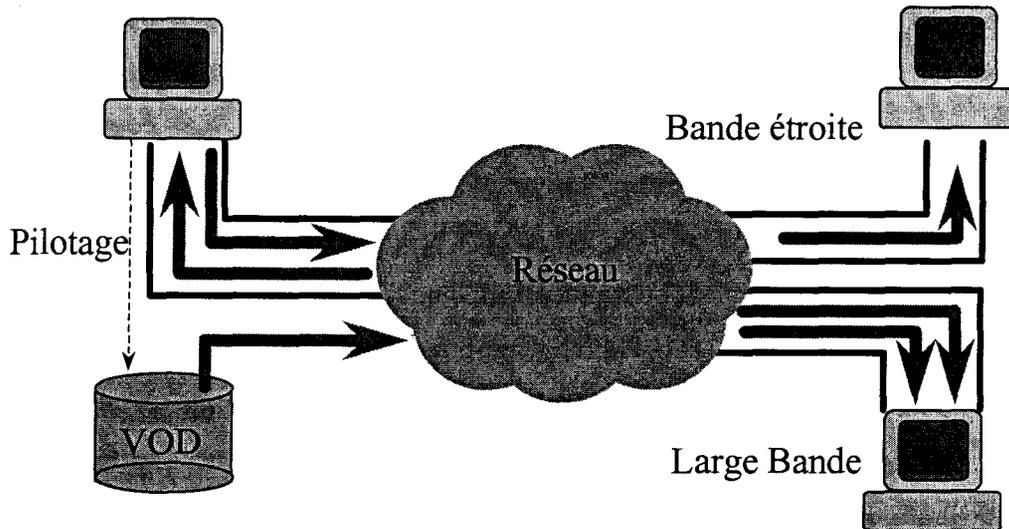


Figure 3-8 : Cas où la VOD est sur un autre site.

Avec cet exemple, nous voyons que l'utilisateur a besoin d'effectuer un choix parmi les vidéos qu'il peut recevoir en fonction de ces intérêts. Les solutions de transcoding vidéo ou de codage en couche agissent au niveau du système. Elles ne tiennent pas compte des intérêts des utilisateurs. Dans l'exemple précédent, une solution utilisant le transcoding vidéo ou le codage en couche aurait pour effet de dégrader les deux flux vidéo au risque de les rendre inexploitable, alors qu'un choix simple aurait été plus judicieux. Nous devons donc donner de l'intelligence au système pour qu'il assiste l'utilisateur dans le choix des vidéos qui l'intéressent.

### 3.7. Conclusion

Dans ce chapitre nous avons mis en évidence la diversité et l'hétérogénéité que l'on rencontre dans les systèmes de travail coopératif et leurs influences sur l'utilisation de la vidéo.

Tout d'abord, nous avons une diversité d'activités coopératives. Chaque activité possède ses propres objectifs à atteindre. Des rôles spécifiques seront affectés aux participants qui accompliront leurs tâches en utilisant des outils particuliers. Il en résulte que chaque activité coopérative nécessitera des besoins particuliers en vidéo. L'apport de la vidéo comme média de communication entre les participants leur permettant de se voir et d'avoir une sensation de groupe est très important dans les activités basées sur la conversation entre participants, alors que dans des activités centrées sur le document, l'utilisation de la vidéo peut être accessoire. Aussi, l'accent peut être plus mis sur la qualité de la vidéo que sur le nombre de vidéo à afficher sur l'écran du participant ; c'est le cas par exemple de la télémedecine où un expert a besoin d'une bonne qualité vidéo pour pouvoir diagnostiquer à distance.

La diversité des activités coopératives implique alors une gestion vidéo spécifique à chacune d'elle. Cette gestion va dépendre des rôles et des outils propres à l'activité coopérative considérée. Ces notions seront abordées dans le chapitre suivant, qui constitue notre contribution pour la gestion de la présentation vidéo des participants dans une activité coopérative.

Nous avons aussi vu que les systèmes des utilisateurs sont hétérogènes. L'hétérogénéité des configurations matérielles des sites utilisateurs implique des différences entre leurs capacités de traitements, alors que l'hétérogénéité de leur accès réseaux implique des différences entre les volumes de données qu'ils peuvent échanger. Comme la vidéo nécessite beaucoup de ressources de calcul et de bande passante réseau, les services vidéo dont peuvent bénéficier les utilisateurs vont dépendre des ressources disponibles de leur site.

Afin d'éviter que les sites les moins puissants n'imposent le niveau de service vidéo à utiliser pendant une session d'activité coopérative, des solutions ont été établies par la communauté de recherche : les passerelles de transcodage et le codage en couches. La solution utilisant les passerelles de transcodage permet d'interconnecter des groupes de sites entre eux. Les sites ayant des capacités similaires seront mis dans le même groupe. La passerelle de transcodage accomplis une transformation des flux vidéo consistant à baisser le débit entre deux groupes de sites ayant des capacités différentes. Ainsi, avec cette solution, chaque site va bénéficier d'une vidéo correspondant à ses capacités, ne contraignant pas tous les sites à s'aligner sur les capacités du plus faible.

La solution utilisant le codage en couches s'effectue sur le site émetteur. Ce dernier va coder la vidéo en plusieurs couches. La première couche va correspondre à la qualité de base de la vidéo, et les couches suivantes vont améliorer successivement la qualité de la vidéo.

Chaque récepteur va choisir le nombre de couches vidéo à recevoir en fonction de ces capacités.

Les passerelles de transcodage et le codage en couches sont des solutions au niveau système. A l'aide d'un exemple, nous avons mis en évidence l'utilité de prendre des décisions au niveau utilisateur. Dans le cas de réception de plusieurs flux vidéo, il est plus judicieux de choisir les flux vidéo à recevoir qui intéressent l'utilisateur, que de laisser le système agir sans intelligence, en baissant tous les flux vidéo, au risque de les rendre toutes inexploitable pour l'utilisateur. Dans le chapitre cinq, nous présentons notre solution permettant de faire un compromis entre les intérêts des participants et la qualité de service offerte par les ressources du système afin de déterminer la qualité vidéo à utiliser. Dans notre solution, nous relèverons la décision du choix de la qualité vidéo au niveau utilisateur, au lieu de laisser les décisions se faire au niveau système comme dans les systèmes existants. En particulier, notre solution tient compte des caractéristiques de l'activité coopérative, afin que cette dernière puisse se dérouler dans de bonnes conditions.

Le chapitre six décrit l'architecture que nous avons adoptée pour notre système. En remarquant que chaque activité possède ses caractéristiques, la gestion de la présentation vidéo sera spécifique à l'activité. Pour cela, nous proposons une architecture ouverte basée sur le principe des Frameworks, permettant ainsi de s'adapter au type d'activité coopérative. La gestion spécifique de la présentation vidéo est encapsulée dans un composant qui se connecte dynamiquement au Framework à l'exécution.



---

## **4. Proposition d'un modèle pour la gestion de la présentation vidéo**

---

## 4.1. Introduction

Dans le mode présentiel, les personnes participant à une réunion, à un cours, ou à une présentation, se trouvent dans le même endroit physique. Chaque personne a un champ de vision lui permettant de voir les autres personnes dans la salle. Ce champ de vision va dépendre de plusieurs facteurs, dont la position de la personne dans la salle, la distance le séparant des autres, ainsi que de son acuité visuelle.

Les salles accueillant les participants sont étudiées pour permettre à chacun d'avoir un champ convenable (Figure 4-1-a). Ainsi, dans une salle de réunion, les participants se mettent généralement autour d'une table, permettant à chaque participant d'être visible à tous les autres membres du groupe. Par conséquent, chaque participant voit tous les autres membres du groupe. Dans un amphithéâtre, les sièges sont mis en escalier de telle sorte que les personnes assises sur les premières rangées ne cachent pas la vue pour ceux qui sont assis derrière.

La correction de l'acuité visuelle quant à elle, reste à l'initiative de chaque personne, en portant des « lunettes », ou bien en s'asseyant sur les premières rangées dans le cas d'un amphithéâtre. Le participant *s'adapte* à son environnement.

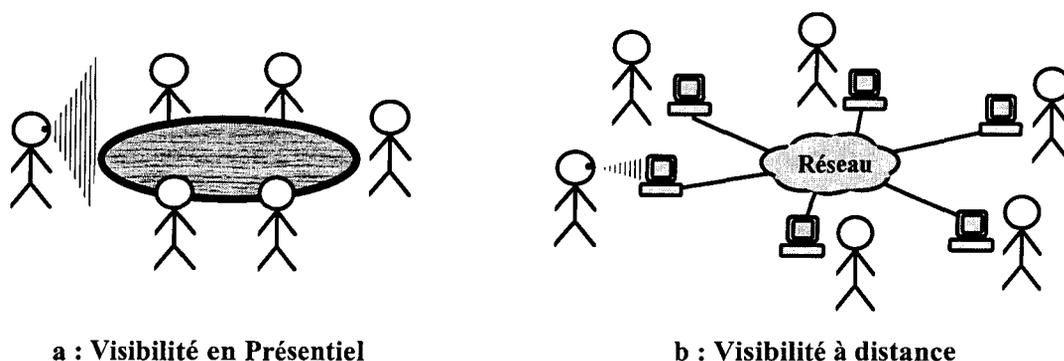


Figure 4-1 : Visibilité des participants dans une session de coopération.

Dans les systèmes de travail coopératif, fournir aux participants la même *visibilité* que celle qui existe en présentiel, ainsi que la *réciprocité* de visibilité entre participants, est une tâche ardue.

Le champ de vision d'un participant est limité par la taille de l'écran de visualisation utilisé (Figure 4-1-b). L'écran de visualisation est utilisé par les applications coopératives pour représenter dans l'interface graphique les espaces de coopération et de communication. Les proportions des surfaces de ces deux espaces sont variables. Elles dépendent à la fois de l'activité coopérative et de son état à un instant précis.

La surface réservée pour l'espace de communication, ne peut contenir qu'un certain nombre de vidéo lorsque le nombre de sites participants augmente. Pour cela, il faut effectuer une sélection d'un nombre réduit parmi les sites participants. Cette sélection se fera en fonction de *l'intérêt* que porte le site récepteur pour les autres sites.

Dans la suite de ce chapitre, nous allons voir pourquoi il faut assister le participant dans la gestion de sa présentation vidéo pendant la session de coopération, puis nous montrerons comment, à partir de l'analyse de l'activité coopérative, nous déterminons les caractéristiques qui nous permettront de déduire les intérêts du participant à voir les autres membres du groupe impliqués dans l'activité coopérative.

## **4.2. Pourquoi assister l'utilisateur dans la présentation vidéo ?**

Le participant dans une application de travail coopératif ne peut pas avoir une visibilité totale lui permettant de voir tous les membres engagés dans l'activité coopérative. Ceci est dû aux limites des ressources des systèmes sous-jacents : la taille de l'écran pour l'affichage, la puissance de calcul pour la manipulation des données et la bande passante réseau pour le transport. Lorsque les ressources du système ne permettent pas de recevoir tous les flux vidéo des autres sites, une sélection de vidéo à recevoir doit être faite parmi les sites participants. Cette sélection se fera en fonction des intérêts que porte l'utilisateur pour les autres membres du groupe. Ce sont ces intérêts qui vont l'amener à vouloir voir certaines personnes et pas d'autres. Cependant, les intérêts d'un participant évoluent pendant l'activité coopérative avec l'évolution de cette dernière. Chaque participant a un rôle qui lui fait accomplir une tâche, qui fait qu'à un instant donné, un participant peut être « plus actif » que les autres, et attirer l'attention des autres sur lui. Cette évolution des intérêts implique un changement dans la sélection des vidéos à recevoir.

Pour cela, deux raisons majeures nous motivent pour assister l'utilisateur dans la présentation vidéo :

- D'une part, assister l'utilisateur lui évite de se charger lui-même de la sélection vidéo, qui peut être une tâche plus ou moins fréquente en fonction du type de l'activité. Ceci diminuera la surcharge cognitive due à la manipulation de l'interface et lui permettra de se focaliser beaucoup plus sur l'accomplissement de sa tâche au sein du groupe.
- D'autre part, nous avons vu que l'utilisateur a une visibilité réduite des membres du groupe. Il ne peut pas avoir une connaissance totale de tous les événements. La perception, que possède l'utilisateur, de son environnement est par conséquent réduite. Si la personne qui prend la parole à un instant donné ne figure pas parmi la sélection faite auparavant, il sera difficile à l'utilisateur de le localiser sans l'assistance du système. L'utilisateur n'est donc pas apte à décider tout seul de la composition de sa présentation vidéo.

Pour que le système assiste l'utilisateur dans sa présentation vidéo, il faudra qu'il ait connaissance des intérêts de l'utilisateur. Cette connaissance n'est acquise qu'en faisant une analyse de l'activité coopérative, des participants et de leurs tâches à accomplir.

### **4.3. Les modes traditionnels pour diriger la présentation vidéo**

Avant d'aborder l'analyse de l'activité coopérative pour la gestion de la présentation vidéo, nous allons d'abord présenter les deux méthodes principalement utilisées dans les systèmes traditionnels, à savoir le mode sonore et le mode directeur, pour en démontrer leurs limites.

#### **4.3.1. Le mode sonore**

Dans ce mode, les participants reçoivent la vidéo de la personne qui parle. Le système détecte la personne qui parle en comparant les niveaux sonores. La vidéo du site dont le niveau sonore est le plus élevé sera diffusée aux autres sites. Le site diffuseur quant à lui, va recevoir la vidéo qui était diffusée précédemment.

Ce choix est dû au fait que dans la majorité des situations d'une activité (ou communication, conversation de manière générale), nous avons tendance à diriger notre

regard vers la personne qui parle, le *locuteur*<sup>29</sup>. La personne qui parle, dirigera son regard vers la personne qui parlait juste avant lui, le *locuteur précédent*, en réponse à une question de ce dernier par exemple.

Si cette solution s'adapte bien à certaines situations et à certaines activités, elle n'est pas généralisable.

Parfois nous avons plus intérêt à voir ce que fait le locuteur, que le locuteur lui même. Lorsqu'un professeur écrit sur le tableau ou projette des transparents, les élèves regardent le tableau ou les transparents projetés. Une fois qu'ils ont mémorisé ce qui est écrit, l'intérêt de regarder le professeur augmente afin qu'ils suivent les explications données par celui-ci. Donc, la diffusion de la personne qui parle aux autres sites ne doit pas être systématique, mais doit tenir compte du contexte de l'activité coopérative en cours.

De même, la personne qui parle ne dirigera pas forcément son regard vers le locuteur précédent. Une personne peut prendre la parole pour diverses raisons, répondre à une question, appuyer ce qui vient d'être dit, changer le cours d'une discussion, ou donner la parole à une autre personne. Dans toutes ces situations, son regard peut être dirigé vers d'autres personnes que le locuteur précédent.

De plus, ce mode ne résout pas le problème de réception vidéo dans une session de travail multipoint. Dans une session multipoint, les utilisateurs aiment voir le maximum de personnes pour avoir une perception de groupe, et avoir connaissance des événements qui peuvent se produire dans les autres sites.

Le mode sonore donne la solution pour recevoir une seule vidéo, en fonction du fait que l'on soit locuteur ou pas. Certains produits, comme Proshare d'Intel, pour pallier ce problème, construisent leur présentation vidéo en formant une mosaïque de quatre vidéo composée du locuteur et des trois derniers locuteurs. Si cette solution construit une présentation vidéo permettant aux utilisateurs de voir plusieurs sites en même temps, la composition obtenue ne correspond pas forcément aux intérêts des utilisateurs. En effet, s'il paraît évident que l'on soit intéressé par la personne qui parle, l'intérêt est moins évident pour les locuteurs précédents. Ceci dépendra du déroulement de l'activité coopérative.

Dans les systèmes utilisant le mode sonore, la commutation de la vidéo n'est effectuée que lorsque le niveau sonore est maintenu pendant deux ou trois secondes, ceci afin d'éviter que les bruits provoqués par certains objets, comme la fermeture d'une porte ou le déplacement d'une chaise, ne provoque la commutation vidéo. Ce délai introduit pour la détection du

---

<sup>29</sup> Le terme anglais Speaker est couramment utilisé.

locuteur, ajouté au délai de transport sur les réseaux, surtout sur des réseaux sans garantie de délai comme Internet, peuvent provoquer une attente plus ou moins longue des participants avant de pouvoir recevoir la vidéo du nouveau Locuteur.

Enfin, le mode sonore n'est plus adéquat lorsque le nombre de sites augmente dans une session de travail. La fréquence de changement de locuteur va augmenter, entraînant avec elle un changement fréquent de la présentation vidéo, même pour de très courte intervention. Ceci peut devenir source de perturbations pour les utilisateurs.

Dans le cas où un grand nombre de participants sont engagés dans une session de travail, le mode directeur est mieux adapté.

### **4.3.2. Le mode directeur**

Dans ce mode, un participant est désigné pour être *Directeur* (chairperson) de la session de travail. Son rôle est de diriger la réception vidéo pour les autres membres du groupe, en attribuant la parole aux participants, soit à ceux qui la demande, soit à son initiative. La personne qui a le *droit* à la parole, sera vue par les autres personnes.

Le système VisioCentre d'enseignement à distance de l'ENIC [Lafon 94] utilise ce mode, encore appelé mode présidentiel. Le système est constitué d'un site Professeur et de plusieurs sites Elèves.

Pendant le cours, l'image du professeur est envoyée à tous les autres sites élèves. Lorsqu'un élève pose une question, le professeur reçoit une notification dans son interface graphique. Le professeur passe la parole à cet élève, dont la vidéo est alors transmise à tous les autres sites.

Nous voyons ainsi que ce mode ne diffère pas du mode sonore dans le sens où la présentation vidéo est toujours dirigée de la même façon, en se basant sur le locuteur. Ce qui change c'est la façon d'avoir le droit à la parole. Dans le mode sonore, l'attribution se fait par détection d'un niveau sonore, alors que dans le mode directeur, l'attribution se fait par un opérateur humain, le Directeur.

Dans ce mode, ainsi que dans le mode sonore, tous les participants auront la même présentation vidéo (sauf pour le locuteur). Dans les situations de face à face, les regards sont dirigés en fonction de l'intérêt que nous portons pour les personnes en face de nous. Ce mode suppose que tous les participants ont exactement les mêmes intérêts. En réalité, ceci n'est pas toujours le cas.

Pour cela, nous devons utiliser un mode permettant une gestion *personnalisée* de la présentation vidéo, pour chaque participant.

Par la suite, nous allons voir que les intérêts des participants vont dépendre de leurs rôles dans l'activité coopérative et des outils qu'ils utilisent.

#### 4.4. Présentation Vidéo Personnalisée

Les postes de travail individuels comme Mac ou Windows, peuvent être personnalisés par l'utilisateur en fonction de ses désirs ou de ses besoins. L'utilisateur peut agir sur son poste pour modifier par exemple le motif du fond et la disposition des icônes sur le bureau, le pointeur de la souris, ou la police de caractère. Les raisons qui peuvent influencer les choix dans la personnalisation du bureau sont :

- L'âge : adulte ou enfant ;
- Le sexe : homme ou femme ;
- Les problèmes de mobilité, de vision ou d'audition ;
- Les connaissances : expert ou novice ;
- La langue ;
- Et bien sûr les goûts particuliers de chacun.

En passant du poste de travail individuel, aux applications de groupe, nous pouvons faire le constat que les membres du groupe diffèrent de par leur rôle et leurs besoins. Dans [Greenberg 91], la notion de personnalisation est alors étendue au travail coopératif pour aboutir à la notion de **Groupware Personnalisable**<sup>30</sup> qui est définie par :

*Un Groupware dont le comportement peut être adapté pour correspondre aux besoins particuliers des participants du groupe (chaque membre du groupe peut observer un comportement différent), et les besoins particuliers du groupe comme un tout (chaque groupe peut observer un comportement collectif différent).*

Pour le cas particulier de la réception vidéo, la diversité des rôles des participants, ainsi que la différence de leur état à un instant donné, due à l'utilisation d'outils ou à la manipulation d'objets particuliers influencent les intérêts de chaque participant à voir ses interlocuteurs.

---

<sup>30</sup> Personalizable groupware.

Pour illustrer le premier point de la définition de Greenberg citée plus haut, prenant l'exemple d'une activité particulière qui est le cours à distance. Dans cette activité, nous avons les rôles de Professeur et d'Elève. Pendant que le professeur explique le cours, il a le même intérêt pour voir les élèves qui se trouvent dans les sites distants, alors qu'un élève particulier aura plus d'intérêt à voir le professeur que les autres élèves. Cet exemple montre que l'on ne doit pas appliquer la même politique de gestion de la présentation vidéo pour tous les participants à une session d'activité coopérative. L'exemple du cours à distance sera traité plus en détail dans la suite du chapitre.

Le dernier point de la définition de Greenberg, traite des besoins du groupe comme un tout. Supposons que nous ayons deux groupes impliqués dans deux sessions distinctes d'une même activité coopérative. Un groupe est restreint à quelques participants, l'autre est large avec un grand nombre de participants. Pour des raisons d'encombrement d'écran, les membres du groupe large ne peuvent pas recevoir toutes les vidéos de leurs interlocuteurs, alors que l'inverse est envisageable pour les membres du groupe restreint. Une politique différente doit être appliquée dans chacun des deux cas, pour éventuellement le même type de rôle.

Nous étendons alors la notion de groupware personnalisable à la notion de **Présentation Vidéo Personnalisée**, qui permet à chaque participant de bénéficier de la présentation vidéo qui correspond le plus à ses centres d'intérêts.

## **4.5. Analyse de l'activité coopérative**

L'analyse de l'activité est souvent utilisée pour le développement des systèmes informatiques [Hoogstoel 95]. C'est une étape indispensable pour la conception d'interfaces utilisateurs [Johnson 92].

De manière générale, l'analyse décompose le problème à résoudre pendant l'activité, en une hiérarchie de tâches devant être accomplies. Parmi les objectifs de l'analyse, nous devons définir :

- Les objectifs à atteindre ;
- Les objets impliqués ;
- Les procédures utilisées ;
- Le temps requis pour accomplir une tâche ;
- La fréquence des opérations ;

De plus pour les activités coopératives, nous devons définir :

- Les rôles des participants ;
- L'interaction entre les participants ;
- La répartition des tâches entre participants.

Le résultat est une description de l'activité permettant de déterminer les éléments nécessaires à l'accomplissement de ses objectifs.

A l'issue de l'analyse de l'activité, nous devons être en mesure de savoir choisir les objets à inclure dans la présentation, les moyens d'interaction des utilisateurs avec le système, ainsi que les médias (texte, image, son, vidéo) les plus adéquats à utiliser.

Pour le cas particulier de la présentation vidéo, l'analyse doit nous aider à acquérir les connaissances nécessaires pour comprendre comment l'intérêt de voir les autres personnes naît, évolue, puis disparaît chez un participant. Le résultat de l'analyse est un modèle permettant de déduire les intérêts des participants dans l'activité, afin de pouvoir gérer la présentation vidéo (Figure 4-2). Le système fournira alors une présentation vidéo qui correspond aux centres d'intérêts de l'utilisateur, donc à ses attentes.

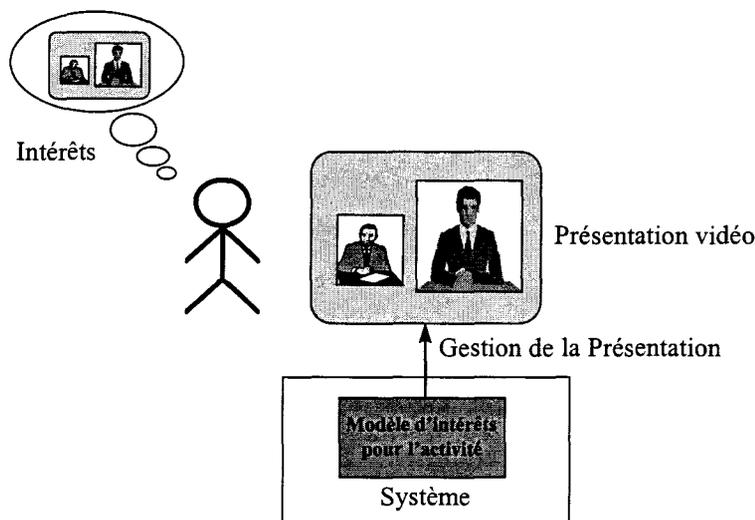


Figure 4-2 : Modèle pour la déduction des intérêts des participants.

Des travaux pour la détermination des intérêts de l'utilisateur sont présentés dans [Yamaashi 95]. Les auteurs estiment le degré d'intérêt DOI<sup>31</sup> qu'a l'utilisateur pour chaque fenêtre vidéo, afin d'optimiser les ressources du système (processeur et réseaux) en les attribuant proportionnellement au DOI. Pour déterminer le DOI de chaque fenêtre, ils appliquent une règle basée sur la distance entre les fenêtres. Les intérêts de l'utilisateur diminuent pour une fenêtre, lorsque la distance entre cette fenêtre et la fenêtre qui a le focus augmente. La fenêtre possédant le focus représente la fenêtre que l'utilisateur regarde. C'est la fenêtre qui l'intéresse le plus. Cette règle s'explique par le fait que lorsque l'utilisateur se focalise sur une région, sa vision est claire au centre de sa vue et floue sur la périphérie. Ceci est dû à la constitution de la rétine de l'œil humain.

Cette méthode ne détermine pas les intérêts de l'utilisateur en vue de l'assister dans sa présentation vidéo. C'est l'utilisateur qui fait ses propres choix sur les vidéos qu'il veut voir en donnant le focus aux fenêtres qui l'intéressent et en effectuant les placements des autres fenêtres sur l'écran. Le système récupère ces données pour déduire les intérêts de l'utilisateur, seulement dans le but d'optimiser les ressources système.

Nous pouvons cependant nous inspirer de cette méthode pour assister l'utilisateur dans le placement des fenêtres vidéo, en appliquant la règle basée sur la distance entre les fenêtres. Nous pouvons mettre par exemple la vidéo qui l'intéresse le plus au centre de l'écran, et les autres fenêtres à une distance inversement proportionnelle à leur DOI. Mais avant, il faudra d'abord déduire les intérêts de l'utilisateur par d'autres méthodes.

Dans [Inoue 95], les auteurs proposent de diriger l'image visuelle d'un système de vidéoconférence en se basant sur l'expérience acquise dans le domaine de l'audiovisuel. Leur but est d'apprendre des leçons sur la façon de présenter des images dans les programmes de télévision et cinéma, qui permettent de garder l'attention des personnes, et de les appliquer pour améliorer les systèmes de vidéoconférence. Pour cela, plusieurs émissions de débats sur télévision ont été analysées.

L'analyse statistique a révélé que le flux des images visuelles dans un programme de débat consiste principalement en un ensemble de prise de vues incluant la personne qui parle, entre lesquelles des prises de vues courtes des autres personnes sont insérées. Des règles de contrôle d'images ont été obtenues à partir de cette analyse puis appliquées à un système de vidéoconférence.

---

<sup>31</sup> DOI : Degree Of Interest.

Cette approche est très intéressante pour toute activité s'approchant des débats télévisés, où l'activité s'appuie principalement sur la conversation, mais elle ne pourrait pas s'appliquer aux activités orientées partage d'objets et de documents.

La deuxième remarque que nous pouvons faire, est que les systèmes de télévision sont unidirectionnels, alors que les systèmes de vidéoconférence sont bidirectionnels. L'analyse des émissions de télévisions permet de bénéficier de l'expérience acquise dans le domaine de l'audiovisuel pour diriger une présentation visuelle pour des observateurs (téléspectateurs), mais ne permet pas de déduire les règles qui permettent de diriger la présentation visuelle des personnes participant aux débats.

Pour cela, *une analyse du contenu* de séquences vidéo de débats est nécessaire pour pouvoir établir des règles générales permettant de diriger la présentation vidéo des participants à un débat. L'analyse du contenu doit s'intéresser à l'étude de la direction des regards des participants.

## **4.6. Les caractéristiques de l'activité pour la gestion vidéo**

L'intérêt représente ce qui est utile à une personne. Dans une réunion, nous avons tendance à regarder les personnes qui nous intéressent le plus à un instant donné, alors qu'une vision périphérique est automatiquement fournie pour les autres personnes à proximité. L'intérêt qui dirige le regard d'une personne pendant une activité dépend principalement des rôles et des états de ses interlocuteurs ainsi que de ses propres rôle et état.

### **4.6.1. Rôles des participants**

*Le rôle est un modèle organisé de conduites relatif à une certaine position de l'individu dans un ensemble interactionnel. Il peut être considéré à différents niveaux de la relation sociale [Rocheblave-Spenlé 69].*

Dans une société, ou dans un groupe, les tâches sont généralement réparties sur chaque membre. Chaque membre aura à assumer des fonctions et des comportements prescrits qui représentent son rôle dans le groupe. Des interactions peuvent exister entre les différents rôles, et une bonne articulation entre eux est nécessaire.

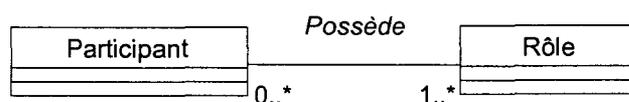
Nous pouvons distinguer le rôle structurel et le rôle fonctionnel. Le rôle structurel correspond à un rôle que peut avoir un participant dans une structure bien établie. Il dépend

généralement de l'activité coopérative. Les rôles de professeur et d'élève dans un cours à distance en sont des exemples. Le rôle fonctionnel par contre, ne dépend pas d'une structure sociale. Il est attribué à un participant pour accomplir une fonction bien précise pendant une session. Le rôle de gestionnaire de conférence ne présuppose pas un rôle dans la hiérarchie d'un groupe social.

Un participant peut avoir plusieurs rôles. Nous pouvons attribuer au professeur le rôle de gestionnaire de conférence pour l'attribution de la parole pendant le cours. Dans les applications d'édition partagée, un participant peut être lecteur et rédacteur en même temps.

Inversement, un même rôle peut être tenu par plusieurs participants. Nous pouvons avoir plusieurs élèves dans un cours.

En utilisant la notation UML, la Figure 4-3 montre la relation entre les notions de participant et de rôle.



*Figure 4-3 : Liaison entre Rôles et Participants*

## **4.6.2. Etats des participants**

Dans une activité coopérative, les participants peuvent utiliser des objets et des outils partagés. Des droits d'accès sont assignés à chaque outil utilisé. Un participant peut avoir le droit de l'utilisation d'un objet en fonction de l'état de ce dernier qui peut être libre ou occupé, et de son rôle. Le rôle d'un participant détermine les droits qu'il a pour l'utilisation des outils. Un observateur dans une table ronde n'a pas le droit à la parole. La gestion des ressources partagées comme l'outil audio, le tableau blanc, ou le télépointeur en fonction des rôles s'en trouve ainsi simplifiée, et du point de vue sémantique, se rapproche plus de la réalité.

### **4.6.2.1. Gestion des outils avec les rôles**

A chaque rôle est assignée une autorisation pour accéder à un outil partagé (Figure 4-4). Cette autorisation se fait avec des priorités et des privilèges.

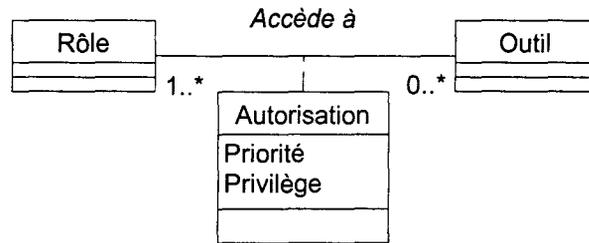


Figure 4-4 : Gestion de ressources partagées avec les rôles.

La priorité est relative aux autres rôles. Un participant peut prendre un outil même lorsqu'il est occupé, à condition que son rôle soit prioritaire à celui du détenteur actuel de l'outil.

Les privilèges définissent les opérations possibles pour agir sur l'outil. Suivant les privilèges assignés à un rôle, ce dernier peut appliquer un certain nombre d'opérations parmi celles qui sont possibles sur l'outil. Dans un outil d'édition, les lecteurs ne peuvent pas écrire sur le document, alors que les rédacteurs peuvent accomplir les deux opérations de lecture et d'écriture.

#### 4.6.2.2. Outils pour la détermination de l'état

L'utilisation d'un outil peut être gérée par un jeton unique, ce qui donne l'exclusivité à un seul participant, ou bien, elle peut être gérée avec plusieurs jetons. Pour le premier cas, nous pouvons donner l'exemple d'un outil audio pour lequel un seul participant a le droit à la parole à un instant donné, alors que pour le second cas, nous pouvons citer un éditeur partagé autorisant plusieurs rédacteurs à la fois.

Un participant peut avoir différents états dans une activité en fonctions des outils qu'il utilise. La prise en main d'un outil par un participant, fait passer ce dernier d'un état de **Non Détenteur de jeton** à un état de **Détenteur de jeton** (Figure 4-5).

L'obtention de l'outil par un participant, peut susciter de l'intérêt pour les autres participants, de telle façon qu'ils voudront par suite le voir dans leur présentation vidéo.

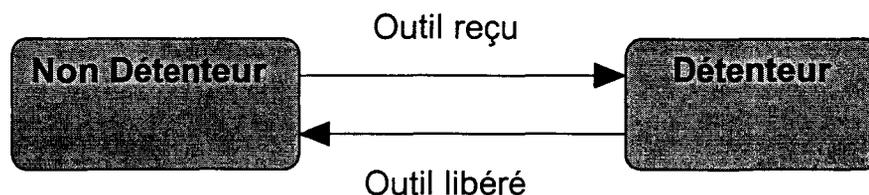


Figure 4-5 : Diagramme d'état d'un participant par rapport à un outil.

Dans certaines activités, le fait de demander un outil peut susciter de l'intérêt vis-à-vis du **Demandeur**. Placer la vidéo de la personne suivante qui prendra la parole dans la présentation vidéo, permet de préparer les participants à un changement de locuteur, évitant ainsi de les brusquer. Pour tenir compte de l'état **Demandeur**, nous modifions le diagramme d'état de la Figure 4-5 pour obtenir celui donné par la Figure 4-6.

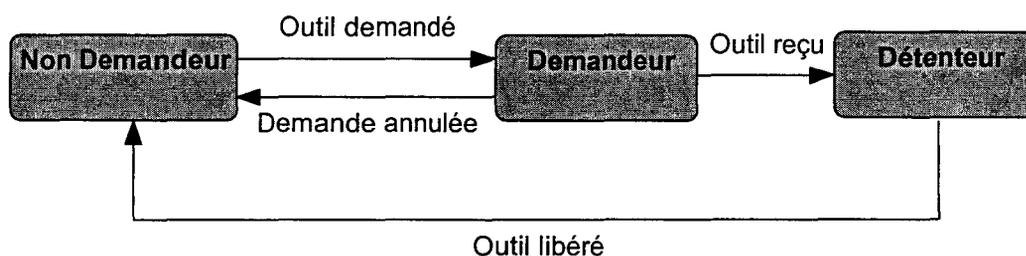


Figure 4-6 : Prise en compte de l'état Demandeur.

De même, suivant le type de l'activité, l'historique de la détention d'un outil peut être important pour en déduire l'intérêt qu'un participant a vis-à-vis des autres membres du groupe. En vidéoconférence, l'état de locuteur précédent est souvent utilisé pour la gestion de la présentation vidéo.

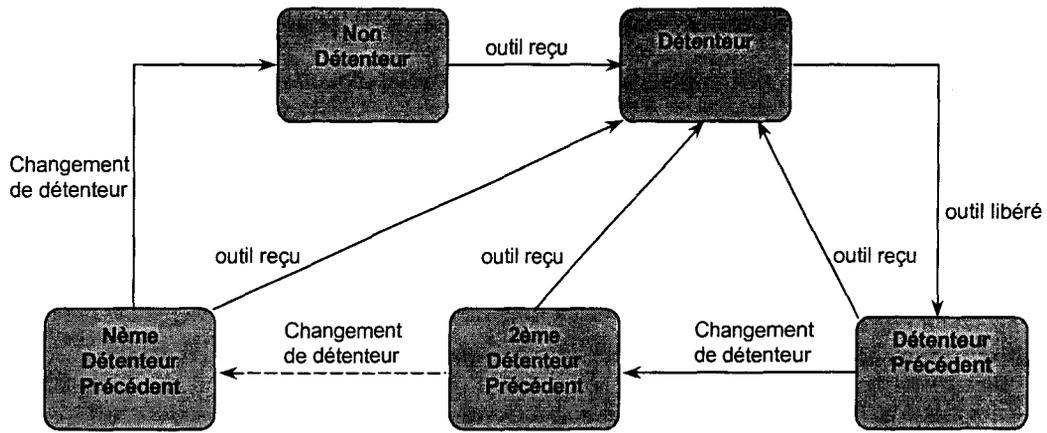


Figure 4-7 : Prise en compte de l'historique de la détention d'un outil.

La Figure 4-7 montre le diagramme d'état pour un participant donné, prenant en compte l'historique de la détention d'un outil. Dans ce diagramme nous supposons que l'importance de l'historique permettant de nous renseigner sur les intérêts du participant pour la réception des autres participants, s'arrête à une profondeur de rang N. Après quoi, l'état du participant passe à celui de *Non Détenteur*.

#### 4.6.2.3. Etats dépendant de plusieurs outils

Jusqu'à maintenant, nous avons défini l'état d'un participant en fonction de l'utilisation d'un seul outil. Cependant pendant une activité coopérative, plusieurs outils peuvent être utilisés au même moment. L'état global du participant sera alors une agrégation de plusieurs sous-états, chacun correspondant à un outil (Figure 4-8).

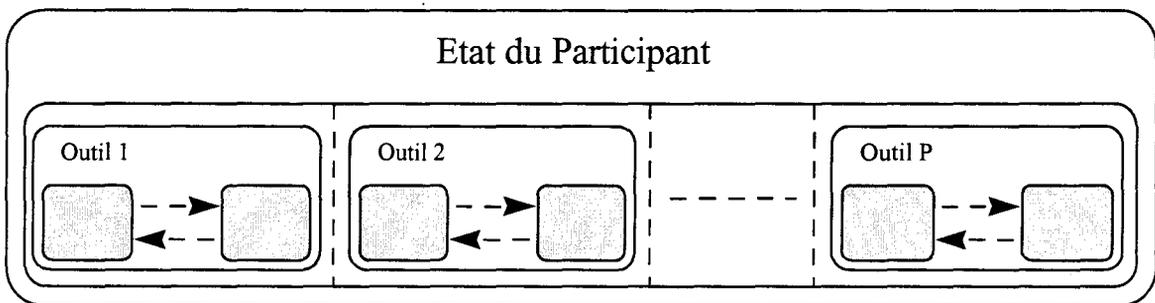


Figure 4-8 : Agrégation d'états en fonction du nombre d'outils.

L'ensemble des états possibles par le participant est le produit cartésien des ensembles d'états, chaque ensemble correspondant à l'ensemble des sous états que peut prendre un participant par rapport à un outil.

L'état d'un participant sera un n-uplet de la forme  $(se1, se2, \dots, seP)$ , où chaque composante représente un sous état.

Prenons un exemple pour illustrer ces notions. Supposons que dans une activité, les participants disposent de deux outils pour coopérer. Un outil audio leur permettant de parler, et un outil d'édition pour faire des annotations.

Par rapport à l'outil audio, un participant est dans l'état **Locuteur** lorsqu'il a la parole, et **Auditeur** dans le cas contraire. Quant à l'outil d'édition, il est **Annotateur** ou **Non Annotateur** suivant qu'il a le jeton pour éditer ou non.

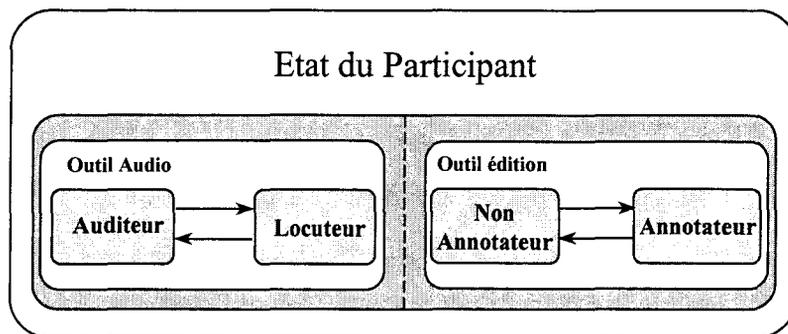


Figure 4-9 : Etat dépendant de deux outils.

Le diagramme d'état composite résultant est donné par la Figure 4-9. L'ensemble des états possibles que peut avoir un participant résultant du produit cartésien entre les états des outils audio et édition est donné par le Tableau 4-1.

|        |                            |
|--------|----------------------------|
| Etat 1 | (Auditeur, Non Annotateur) |
| Etat 2 | (Auditeur, Annotateur)     |
| Etat 3 | (Locuteur, Non Annotateur) |
| Etat 4 | (Locuteur, Annotateur)     |

Tableau 4-1 : Ensemble des états possibles.

Dans certains cas, nous pouvons avoir des contraintes sur l'utilisation des outils pendant l'activité. Dans notre exemple, si nous supposons qu'un participant ne peut pas avoir la parole tant qu'il n'est pas Annotateur, l'état (*Locuteur, Non Annotateur*) ne sera plus possible. L'ensemble des états possibles se réduit alors à ceux du Tableau 4-2.

|        |                            |
|--------|----------------------------|
| Etat 1 | (Auditeur, Non Annotateur) |
| Etat 2 | (Auditeur, Annotateur)     |
| Etat 3 | (Locuteur, Annotateur)     |

Tableau 4-2 : Ensemble des états possibles avec contrainte.

### 4.6.3. Les règles

Comme nous l'avons déjà souligné, le regard d'une personne est dirigé vers ce qui a le plus d'intérêt pour elle à un moment donné. Dans une activité coopérative, les intérêts d'une personne sont liés aux positions qu'ont ses interlocuteurs dans la structure du groupe et des fonctions qu'ils accomplissent, ainsi que de la position et des fonctions de la personne même.

Le but que nous cherchons à atteindre est de concevoir un système permettant d'assister le participant dans la gestion de sa présentation vidéo. Pour cela, le système doit avoir connaissance des intérêts du participant pour composer la présentation vidéo conformément aux souhaits du participant.



Figure 4-10 : Règles pour la déduction des intérêts.

L'analyse de l'activité doit établir des *règles* qui permettent de déduire l'intérêt du participant. Connaissant les paires (*Rôle*, *Etat*) des participants, l'application des règles produit comme résultat les intérêts en terme de priorité entre participants (Figure 4-10). Les intérêts déduits doivent être similaires aux intérêts réels du participant. Ce résultat dépend beaucoup de la qualité de l'analyse de l'activité accomplie.

La syntaxe que nous choisissons pour les règles est la suivante. Etant donné deux rôles  $R\hat{o}le_i$  et  $R\hat{o}le_j$  et deux états  $Etat_x$  et  $Etat_y$ , nous noterons que la paire ( $R\hat{o}le_i$ ,  $Etat_x$ ) est prioritaire sur la paire ( $R\hat{o}le_j$ ,  $Etat_y$ ), par :

$$(R\hat{o}le_i, Etat_x) > (R\hat{o}le_j, Etat_y)$$

Un participant ayant pour rôle et état la paire ( $R\hat{o}le_i$ ,  $Etat_x$ ), sera plus prioritaire pour figurer dans la présentation vidéo qu'un participant ayant la paire ( $R\hat{o}le_j$ ,  $Etat_y$ ).

Nous noterons qu'elles ont la même priorité par :

$$(R\hat{o}le_i, Etat_x) = (R\hat{o}le_j, Etat_y)$$

Un participant ayant un rôle donné peut être prioritaire quel que soit son état, sur un participant ayant une paire (Rôle, Etat) donnée. Pour cela nous utiliserons le symbole \* :

$$(R\hat{o}le_i, *) > (R\hat{o}le_j, Etat_y)$$

La même remarque s'applique à un état donné :

$$(*, Etat_x) > (R\hat{o}le_j, Etat_y)$$

## 4.7. Etude de cas : un cours à distance

Pour voir comment appliquer les notions que nous avons vu dans ce chapitre, prenons comme exemple d'activité, un cours à distance.

Dans un cours, deux rôles sont possibles pour les participants, *professeur* et *Elève*. Le rôle de professeur est tenu par un seul participant, alors que plusieurs participants peuvent avoir le rôle d'élève. Le Professeur dispense son cours à l'aide de certains outils, un *outil audio* pour parler, un *tableau blanc* pour faire des annotations, et un *télépointeur* pour la désignation. Les élèves utilisent les même outils pour intervenir pendant le cours pour poser des questions ou répondre à celles posées par le professeur. Ce dernier a la fonction de maître de conférence qui autorise les élèves à utiliser les outils. La Figure 4-11 donne le modèle de classe obtenue.

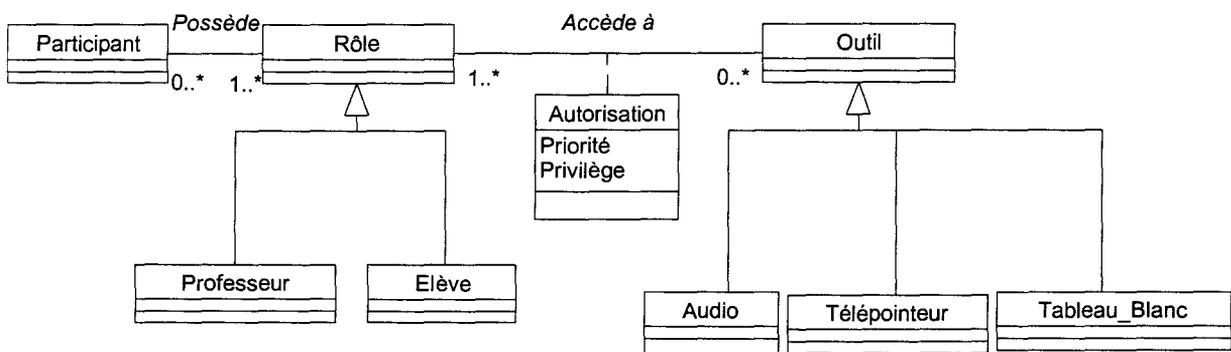


Figure 4-11 : Modèle de classe pour l'activité cours à distance.

### 4.7.1. Les états

Dans un souci de simplification de l'exemple, nous supposons que seul l'outil audio est important pour la détermination des états pouvant influencer les intérêts des participants. Ceci permet, d'une part, d'éviter une explosion combinatoire au niveau des différents états possibles que peut prendre un participant si nous considérons les trois outils en même temps, et d'autre part, de réduire le nombre de règles à établir.

Un seul participant peut utiliser l'outil audio à un instant donné. Le participant qui parle sera dans l'état *Locuteur* et les autres participants seront les *Auditeurs*. En plus de ces deux états, nous utiliserons un troisième état qui est celui de *Locuteur Précédent*. Cet état est souvent utilisé dans les systèmes de vidéoconférence. Lorsque les participants reçoivent la

vidéo du Locuteur, ce dernier reçoit celle du Locuteur Précédent. Nous ne garderons pas de traces dans l'historique de l'utilisation de l'outil audio au-delà de cet état. A un nouveau changement de Locuteur, le Locuteur Précédent devient Auditeur. Le diagramme d'états ainsi obtenu est donné par la Figure 4-12.

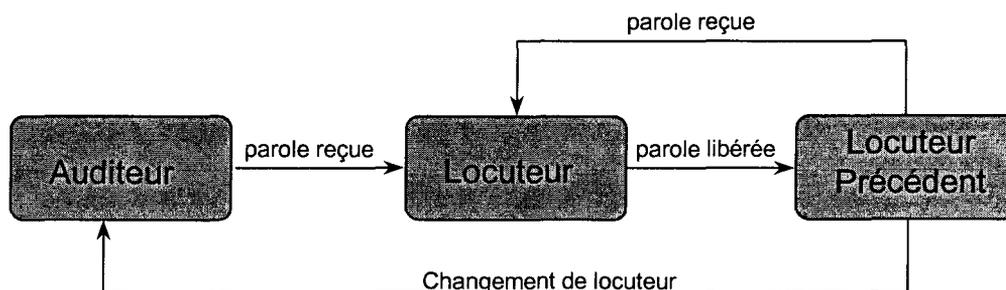


Figure 4-12 : Diagramme d'état d'un participant.

#### 4.7.2. Les règles

Les règles que nous allons énoncer ne sont pas issues d'une analyse de l'activité cours à distance, mais nous les avons seulement établies à des fins pédagogiques pour comprendre cette étude de cas.

Les règles permettent de donner des priorités aux participants en fonction de leurs rôles et états. Cet ordre traduit l'intérêt des participants pour voir leurs correspondants.

Le rôle de Professeur est tenu par un participant unique pendant le cours. Donc tous les participants qu'il verra seront des élèves. Ses intérêts ne se feront pas en distinguant les rôles des participants mais plutôt leur état. Nous admettrons que pour le Professeur, la vue de l'Elève Locuteur est prioritaire sur la vue de l'élève locuteur précédent. Et que ce dernier est prioritaire sur l'élève Auditeur. La règle de priorité qui en résulte est la suivante :

***(Elève, Locuteur) > (Elève, Locuteur Précédent) > (Elève, Auditeur)***

Comme le rôle Elève n'est pas déterminant, nous pouvons le remplacer par le signe \* :

***(\* , Locuteur) > (\* , Locuteur Précédent) > (\* , Auditeur)***

Par contre, pour établir les règles de gestion vidéo pour l'élève, il faut tenir compte de la paire (*Rôle, Etat*). En effet, un élève peut recevoir dans sa présentation vidéo des participants pouvant avoir les deux rôles Professeur et Elève.

Nous distinguons deux cas. Si le professeur est le locuteur, nous appliquons la règle suivante :

***(Professeur, Locuteur) > (Elève, Locuteur Précédent) > (Elève, Auditeur)***

Sinon, le Professeur est moins prioritaire par rapport à l'Elève Locuteur, mais il est prioritaire sur les autres Elèves, quelque soit son état :

***(Elève, Locuteur) > (Professeur, \*) > (Elève, Locuteur Précédent) > (Elève, Auditeur)***

En remarquant que le participant Locuteur est toujours prioritaire quelque soit son rôle, nous pouvons regrouper les deux règles de gestion vidéo de l'élève en une seule règle par :

***(\* , Locuteur) > (Professeur, \*) > (Elève, Locuteur Précédent) > (Elève, Auditeur)***

Rappelons que dans notre exemple, il ne peut y avoir qu'un seul Locuteur à un instant donné. Si un élève est Locuteur, dans la règle ci dessus, le Professeur ne peut être que dans l'un des deux états restant, Locuteur Précédent ou Auditeur.

Dans le cas où plusieurs Locuteurs peuvent exister en même temps, il pourrait y avoir ambiguïté lors de l'application de la règle ci dessus.

### 4.7.3. Scénario

Pour voir comment appliquer les règles établies pour déterminer les intérêts des participants, prenons le cas d'un cours à distance avec un site professeur et trois sites élèves, Petri, Jean Philippe (JP) et Raouf. Une partie du cours suit le scénario suivant :

**Etape 1** : Le professeur explique une partie du cours. C'est le Locuteur ;

**Etape 2** : L'élève Petri pose une question. Petri devient Locuteur et le Professeur devient Locuteur Précédent ;

**Etape 3** : Le Professeur répond. Il est Locuteur et Petri est le Locuteur Précédent ;

**Etape 4 :** Le Professeur poursuit son cours. Il est toujours le Locuteur, mais l'état de Locuteur Précédent de Petri n'est plus pertinent pour le déroulement du cours. Il devient alors Auditeur.

Pendant les quatre étapes du scénario, les élèves Jean Philippe et Raouf ne sont pas intervenus. Ils sont alors restés dans le même état Auditeur.

Le Tableau 4-3 rassemble les intérêts des participants à chaque étape du scénario. Ces intérêts sont déduits après application des règles établies dans le paragraphe précédent.

Chaque ligne représente les intérêts en terme de priorité que possède un participant à chaque étape du scénario. Dans la première étape, les trois élèves sont dans le même état Auditeur. Le Professeur a les mêmes intérêts pour les voir. Pour chaque Elève, le Professeur est prioritaire sur les deux autres Elèves, qui ont la même priorité. Le reste du tableau se lie de la même manière.

| PARTICIPANT   | ETAPE 1              | ETAPE 2              | ETAPE 3              | ETAPE 4              |
|---------------|----------------------|----------------------|----------------------|----------------------|
| Prof          | Petri = JP = Raouf   | Petri > JP = Raouf   | Petri > JP = Raouf   | Petri = JP = Raouf   |
| Petri         | Prof > JP = Raouf    |
| Jean Philippe | Prof > Petri = Raouf | Petri > Prof > Raouf | Prof > Petri > Raouf | Prof > Petri = Raouf |
| Raouf         | Prof > Petri = JP    | Petri > Prof > JP    | Prof > Petri > JP    | Prof > Petri = JP    |

*Tableau 4-3 : Evolution des intérêts des participants pendant le déroulement du scénario.*

La première remarque que nous pouvons faire à partir de ce tableau, est que les participants ont bien des intérêts différents. Une présentation vidéo personnalisée est donc nécessaire pour chaque participant. L'application des règles nous permet de déduire les intérêts des participants pour que le système puisse composer la présentation vidéo attendue.

Nous pouvons aussi voir que les intérêts de Petri n'ont pas changé pendant tout le déroulement du scénario. Comme les élèves Jean Philippe et Raouf n'ont pas pris la parole, le Professeur reste toujours prioritaire pour Petri, même lorsqu'il passe en Locuteur Précédent.

## 4.8. Plus d'intelligence dans les règles

Les règles que nous avons utilisées établissent des priorités entre les participants. Dans certains cas nous pouvons avoir besoin d'autres types de règles complémentaires à ces règles de priorité. Dans l'étude de cas vue dans le paragraphe précédent, le passage entre les étapes 3 et 4 du scénario, ne se manifeste pas par un changement d'états des participants, mais seulement par un changement de contexte. Après avoir répondu à la question de l'élève Petri, le professeur poursuit son cours. C'est toujours le professeur qui est le Locuteur. L'état de Locuteur Précédent de Petri dans l'étape 4 n'est pas pertinent pour l'activité. Mais avec les seules règles de priorité établies, pour le système, Petri restera Locuteur Précédent tant qu'il n'y a pas un changement de Locuteur. Il est difficile pour le système de pouvoir détecter ce changement de contexte.

Une solution serait d'établir d'autres règles qui permettent de récupérer des événements de l'environnement tels que des changements dans l'espace de travail du Professeur permettant de conclure que le cours se poursuit, ou bien d'attendre la fin d'une temporisation lorsque les états des participants ne changent pas. A partir de ces événements, le système décide de faire passer l'élève Petri de l'état Locuteur Précédent à celui d'Auditeur.

Nous pouvons aussi nous inspirer du modèle de la télévision pour établir d'autres règles. Afin d'éviter une lassitude des participants, le système peut décider de changer la présentation vidéo de l'utilisateur après une certaine durée assez longue pendant laquelle il n'y a eu aucun changement, de la même façon que le réalisateur de télévision effectue des changements de l'image transmise aux téléspectateurs.

Le but des règles établies est *d'assister* le participant, et non pas lui *imposer* une présentation figée. Le participant doit avoir la possibilité de modifier la présentation. Comme dans [Yamaashi 95], le participant peut donner le focus à une fenêtre vidéo pour indiquer au système son intérêt. Les règles doivent alors tenir compte des choix du participant pour les faire passer en priorité.

## 4.9. Conclusion

Dans ce chapitre, nous avons souligné la nécessité d'assister le participant à une activité coopérative, dans la composition de sa présentation vidéo. Les deux raisons que nous avons évoquées sont, son incapacité de gérer lui même sa présentation puisqu'il a une visibilité réduite, et pour qu'il se focalise plus sur sa tâche au sein du groupe. Pour cela, nous avons vu

que le système doit avoir connaissance des intérêts du participant pour être en mesure de gérer la présentation vidéo à son profit. Une analyse de l'activité est alors nécessaire pour acquérir ces connaissances. L'analyse doit établir des règles permettant de personnaliser la présentation vidéo. Nous avons vu que ces règles utilisent la paire (Rôle, Etat) pour déduire les intérêts des participants. Ces règles permettent d'obtenir une liste des participants triée par ordre de priorité.

Dans ce chapitre, nous avons traité de la manière dont le système acquiert la connaissance des intérêts du participant pour gérer la présentation vidéo, mais nous n'avons pas abordé le problème de la disponibilité des ressources. La vidéo consomme beaucoup de temps de calcul et de bande passante, or ces ressources sont limitées. Les ressources seront allouées aux vidéo en fonction de leur ordre de priorité. Nous obtenons alors différents niveaux de qualité vidéo. Pour connaître la quantité de ressource à allouer à une vidéo, l'ordre de priorité doit être quantifié en *degré d'intérêt*.

Dans certains cas, nous serons amenés à limiter le nombre de vidéo à recevoir si les ressources disponibles ne suffisent pas.

Le chapitre suivant traitera de ces problèmes et du passage des intérêts et degrés d'intérêts à la qualité de service vidéo.

---

## **5. Solutions pour la Qualité de Service Vidéo**

---

## 5.1. Introduction

Pendant longtemps, l'obstacle majeur qui freinait le déploiement de la plupart des applications multimédia, était la quantité de données nécessaire pour coder la vidéo. Pour avoir une idée, une seconde d'une séquence vidéo Pal, où :

- La fréquence de rafraîchissement est de 25 images par seconde ;
- La taille de chaque image est de 768 colonnes par 576 lignes ;
- un pixel est codé sur 3 octets, chacun correspondant à une couleur primaire (rouge, vert, bleu),

nécessite 33 millions d'octets, soit, environ 32 Mo. Les supports de stockage et les bandes passantes réseaux actuels supportent encore difficilement un tel débit.

Pour cela, plusieurs travaux de recherche ont été faits dans le domaine de la compression d'image, et ont abouti à des normes de compression permettant de réduire considérablement la quantité de données nécessaire pour coder les images. Parmi ces normes, nous pouvons citer JPEG [Wallace 91] pour la compression d'image fixe, MPEG [Le Gall 91] et H.261 [Liou 91] pour la compression vidéo.

Ces techniques de compression sont parfois très complexes, et demandent une grande puissance de calcul. Pour cela, on a souvent recours à une diminution de *la qualité de la vidéo* durant la compression. Par exemple, une vidéo basse fréquence (10 images/seconde) demande moins de calcul qu'une vidéo haute fréquence (25 images/seconde).

De plus, même avec ces techniques de compression, le débit réseau nécessaire pour le transport, reste souvent assez élevé. Dans ce cas, des compressions plus « agressives » sont effectuées afin de baisser le débit, mais au détriment de *la qualité vidéo*.

Les ressources étant limitées, on agit souvent sur *la qualité vidéo* pour optimiser l'utilisation des ressources disponibles.

Dans le chapitre précédent, nous avons vu que chaque participant possède des intérêts différents pour la réception vidéo de ses interlocuteurs. La ressource « écran » est affectée proportionnellement au degré d'intérêt des sites distants.

De la même manière, nous allons affecter les ressources disponibles (réseau et puissance de calcul) en fonction des degrés d'intérêts. Plus un site a de l'intérêt, plus il dispose de ressources, et par conséquent, meilleure est sa qualité vidéo.

Dans le contexte des activités coopératives, le système est distribué, et par conséquent met en communication plusieurs sites. La qualité vidéo reçue par un site va donc dépendre de ses ressources locales ainsi que de celles des sites distants (émetteurs).

Comme les systèmes des utilisateurs sont hétérogènes, ils n'ont pas les mêmes ressources. L'établissement des communications s'effectue alors seulement après une négociation de la *qualité de service : QoS*<sup>32</sup>.

Dans ce chapitre, nous allons aborder la notion de la qualité de service dans les systèmes distribués hétérogènes, puis l'étendre pour définir la qualité de service vidéo.

Nous traiterons la qualité de service vidéo du point de vue du récepteur ainsi que de celui de l'émetteur. Côté récepteur, nous verrons comment passer des intérêts et des degrés d'intérêts des participants à la qualité de service vidéo. Du côté émetteur, nous verrons comment résoudre le problème du choix de la qualité vidéo à émettre, lorsque des requêtes QoS différentes sont émises par les récepteurs.

## 5.2. La qualité de Service : QoS

La notion de qualité de service [Campbell 96], désignée généralement par QoS, a pris naissance dans les systèmes de communication pour décrire les caractéristiques de la transmission des données au niveau des couches basses. Mais avec l'apparition des applications multimédia distribuées, l'utilisation du terme QoS s'est généralisée à tous les niveaux du système [Nahrstedt 95a, Nahrstedt 95b, Coulson 94, Ferrari 94].

En utilisant un protocole comme TCP, la seule garantie que peut offrir un réseau est une transmission de données fiable. Ceci est suffisant pour les applications traditionnelles tels que FTP et Telnet, mais n'est pas adéquat pour les applications ayant des contraintes temporelles. Par exemple, les applications multimédia distribuées ont besoin de communiquer en temps réel et sont très sensibles à la qualité de service qu'elles reçoivent du réseau. Pour ce type d'applications, la notion de QoS doit être étendue des couches de communication jusqu'à la couche application.

---

<sup>32</sup> QoS : Quality of Service.

### 5.2.1. Définitions

Il n'existe pas de définition commune ou formelle pour la QoS. Nous en avons retenu deux. Celle de l'ISO [ISO 95], qui définit la QoS de manière générale par :

*Un ensemble de qualités liées au comportement collectif d'un ou de plusieurs objets*<sup>33</sup>.

Et celle donnée dans [Vogel 95] par :

*La Qualité de Service représente l'ensemble des caractéristiques quantitatives et qualitatives d'un système distribué multimédia, nécessaires pour accomplir la fonctionnalité requise d'une application.*

Et qui est spécifique aux systèmes distribués multimédia.

La définition de Vogel précise que la qualité de service peut être exprimée aussi bien par des caractéristiques qualitatives que quantitatives. Les caractéristiques qualitatives spécifient des services attendus, comme un son stéréo, alors que les caractéristiques quantitatives peuvent être évaluées en termes de mesures concrètes comme la taille d'une image.

Ces définitions sont données du point de vue du système. Comme les systèmes sont conçus pour être exploités par des utilisateurs, nous complétons ces définitions en donnant la QoS du point de vue utilisateur par :

*La Qualité de Service est la qualité **perçue** par un utilisateur pour le service rendu par le système.*

### 5.2.2. La spécification de QoS

La qualité de service est exprimée avec des *paramètres de QoS*. Un paramètre de QoS décrit un attribut particulier d'un système multimédia. Il peut prendre des valeurs dans un intervalle continu et ordonné ou parmi un ensemble de valeurs.

Lorsqu'une QoS est spécifiée, des valeurs sont attribuées à chaque paramètre de QoS reflétant les performances du système multimédia.

Dans [Vogel 95], les paramètres QoS sont classés en cinq catégories : performance, format, synchronisation, coût et utilisateur, auxquels nous pouvons rajouter le niveau de service.

---

<sup>33</sup> A set of qualities related to the collective behaviour of one or more objects.

*Les paramètres orientés performance* permettent de caractériser les performances de flux requises d'un utilisateur. La possibilité de garantir la bande passante, le délai de bout en bout, la gigue, le taux de perte est importante dans une communication multimédia. Pour qu'un système puisse garantir la disponibilité des ressources, il doit avoir une connaissance des caractéristiques du trafic généré.

*Les paramètres orientés niveau de service* spécifie le degré de l'engagement de bout en bout des ressources requises : déterministe, prédictive, ou "best effort"<sup>34</sup>. Alors que la spécification de performance de flux permet à l'utilisateur d'exprimer les performances requises de manière quantitative, le niveau de service permet à ces besoins d'être affinés de manière qualitative. Le niveau de service exprime un degré de certitude avec lequel une QoS requise sera honorée.

*Les paramètres orientés synchronisation* permettent de caractériser le degré de synchronisation entre des flux liés. Par exemple, la synchronisation entre le son et la vidéo n'est pas obligatoirement très forte, mais ne doit pas descendre en dessous d'un certain seuil, pour éviter de perturber la compréhension de l'utilisateur.

*Les paramètres orientés coût du service* spécifient le prix à payer par l'utilisateur pour le niveau de service choisi. S'il n'y a pas de notion de coût de service dans une spécification de QoS, il n'y aura pas de raison pour laquelle l'utilisateur choisisse un niveau autre chose que le niveau de service maximum possible.

*Les paramètres orientés format* permettent de spécifier le format d'un média comme la taille d'une image vidéo, ou la méthode de compression.

*Les paramètres orientés utilisateur* sont purement qualitatifs. Ils peuvent désigner la qualité subjective que peut avoir un utilisateur par rapport à un son ou à une image.

Le Tableau 5-1 donne des exemples de paramètres QoS pour l'audio et la vidéo. Les valeurs d'un paramètre de QoS sont typées et ne peuvent prendre que les valeurs autorisées dans un intervalle. Si cet ensemble définit une relation d'ordre entre ses valeurs, des valeurs peuvent être comparées. Nous pouvons choisir par exemple la valeur la plus *forte* parmi un ensemble de valeurs. Le terme *forte* ne veut pas dire plus grande. Dans le cas de la fréquence de rafraîchissement d'images dans une vidéo, la valeur la plus forte est la valeur la plus grande, par contre dans le cas d'un délai de transport, la valeur la plus forte correspond au délai le plus court.

---

<sup>34</sup> Désigne les services sans aucune garantie de performance comme sur les réseaux IP classiques.

Des relations peuvent exister entre les valeurs de différents paramètres d'un même média. Dans le cas des images vidéo, les paramètres de hauteur et de largeur doivent satisfaire un rapport d'aspect vidéo. Ce dernier peut constituer un troisième paramètre. Deux parmi ces trois paramètres peuvent ainsi être utilisés pour déterminer le troisième suivant les besoins de l'application multimédia.

Nous pouvons aussi avoir des contraintes entre paramètres de différents média, comme le paramètre de synchronisation entre l'audio et la vidéo.

| Paramètre de QoS                    | Intervalle de valeurs                          |
|-------------------------------------|--|
| Taille d'un échantillon audio       | 8 bits, 16 bits                                |
| Fréquence d'échantillonnage audio   | 8 KHz, 11 KHz, 22 KHz, 44.1 KHz                |
| Spatialisation du son               | Mono, Stéréo, 3D                               |
| Format du son                       | G.711, G.722                                   |
| Fréquence d'image vidéo             | 0-30 images/seconde                            |
| Largeur d'image vidéo               | 0-800 pixels                                   |
| Hauteur d'image vidéo               | 0-600 pixels                                   |
| Résolution de couleur d'image vidéo | 8 bits/pixels, 16 bits/pixels, 24 bits/pixels, |
| Rapport d'aspect vidéo              | 4 / 3, 16 / 9                                  |
| Rapport de compression vidéo        | 1 : 1, 2 : 1, 50 : 1                           |
| Synchronisation Audio / Vidéo       | + / - 80 ms                                    |

*Tableau 5-1 : Exemples de paramètres QoS pour l'audio et la vidéo.*

### **5.2.3. La qualité de service à différents niveaux du système**

La spécification de QoS est généralement différente à chaque couche du système. L'utilisateur d'un système de communication multimédia commence à spécifier la qualité de service, en choisissant la qualité du média, vidéo ou sonore par exemple, qu'il souhaite recevoir via l'interface utilisateur de son application multimédia. Les flots de données correspondant à cette qualité vont traverser les différentes couches du système de l'émetteur, partant du périphérique source, jusqu'à la pile de protocole de l'émetteur. Ils seront ensuite acheminés à travers le réseau. Et enfin, vont remonter la pile de protocole du récepteur, jusqu'au périphérique de restitution du média. Chaque module traversé doit fournir une QoS

configurable, basée sur une spécification de QoS. La qualité de service doit être spécifiée et maintenue au niveau de chaque couche du système pour garantir une qualité de service de bout en bout. La Figure 5-1 donne le modèle en couche largement répandu pour la QoS.

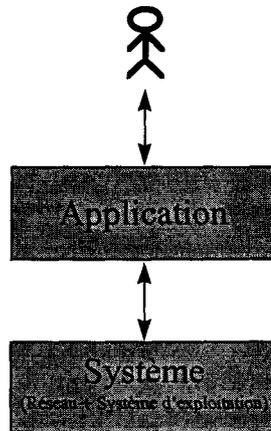


Figure 5-1 : Modèle en couche pour la QoS.

### 5.2.3.1. La QoS au niveau utilisateur

Dans un système multimédia distribué, la spécification de QoS doit être orientée plus utilisateur que système. La spécification QoS au niveau utilisateur est de nature déclarative. L'utilisateur spécifie quels sont ses besoins plutôt que comment ces besoins seront accomplis par les mécanismes de QoS.

La perception humaine de la qualité est subjective. Elle n'est pas quantifiée comme pour les couches du système. Une personne dira qu'une vidéo est de bonne, moyenne ou basse qualité. Deux personnes différentes peuvent avoir des appréciations différentes au sujet d'une même vidéo. Un utilisateur perçoit et choisit une qualité vidéo ou sonore à travers l'interface de l'application multimédia. L'interface graphique de l'application doit présenter la qualité en termes simples et usuels comme une qualité téléphonique ou CD pour l'audio, et une qualité TVHD pour la vidéo. Des considérations de bas niveau comme la synchronisation de flux ou l'ordonnancement des threads dans le système d'exploitation doivent être cachées à l'utilisateur.

Mesurer la perception de la qualité d'une vidéo pour une personne pour une application particulière peut se faire en faisant des tests pour rechercher les paramètres de la vidéo (taille,

fréquences, nombre de couleurs) pour lesquels cette personne trouve que la vidéo a une bonne, moyenne, ou basse qualité.

#### **5.2.3.2. La QoS au niveau application**

Les paramètres de QoS au niveau application décrivent les caractéristiques requises des médias utilisés. Les médias typiques utilisés sont la vidéo et le son. Les paramètres sont relatifs aux images vidéo et aux échantillons sonores. Pour la vidéo, les paramètres typiques sont la taille, le nombre de couleurs, et le nombre d'images par seconde. C'est au niveau application que les contraintes de synchronisation entre les différents médias sont spécifiées.

#### **5.2.3.3. La QoS au niveau système d'exploitation**

Les applications multimédia ont des contraintes temps réel que les systèmes d'exploitation doivent satisfaire. Elles diffèrent des applications temps réel traditionnelles par le fait que le non respect de ces contraintes affecte la qualité de service offerte à l'utilisateur mais n'a pas de conséquences catastrophiques.

Dans une spécification de QoS au niveau système d'exploitation, nous pouvons spécifier que toute tâche doit terminer son exécution avant sa date butoir, comme dans le système YARTOS, un système d'exploitation adapté aux applications de téléconférence [Jeffay 91].

Nous pouvons aussi donner différentes priorités pour l'ordonnancement des threads de traitements. Comme l'oreille humaine est plus sensible à la perte d'échantillons sonores que l'œil humain à la perte d'images, nous pouvons spécifier au système d'exploitation de donner une priorité plus élevée aux threads traitant le son qu'à ceux traitant la vidéo.

#### **5.2.3.4. La QoS au niveau réseau**

Les paramètres de QoS utilisés au niveau réseau décrivent la qualité requise pour le transport des données. Ces données sont caractérisées indépendamment de leur contenu. Le délai de bout en bout, la gigue, la taille des paquets et le taux de perte font partie des paramètres les plus utilisés.

### **5.3. La Qualité Vidéo**

La qualité d'une vidéo est le résultat cumulé de diverses opérations, allant du dispositif d'acquisition (caméra) jusqu'au dispositif de visualisation (écran), en passant par une chaîne de traitement (compression, stockage, restitution, transport).

Cette qualité dépend d'une variété de facteurs comme la taille, la profondeur de couleur, la compression, et la fréquence de rafraîchissement (Figure 5-2).

- **La taille** : spécifie la largeur et la hauteur de l'image. Un rapport fixe entre ces valeurs est généralement utilisé (4/3, 16/9) ;
- **La profondeur de couleur** : spécifie le nombre de couleurs dans une vidéo. Elle est exprimée en nombre de bits par pixel ;
- **La résolution fréquentielle** : agit sur le nombre de coefficient DCT. C'est ce paramètre qui donne l'effet mosaïque pour une image.
- **La résolution temporelle** : à la différence des images fixes, la vidéo a un autre paramètre qui représente la fréquence de rafraîchissement des images dans une vidéo. Ce paramètre est exprimé en nombre d'images par seconde. Comme les images successives constituant une vidéo sont dépendantes de par leur contenu, elles créent une impression de mouvement grâce au phénomène de persistance dans la perception visuelle chez l'homme. A basse fréquence, l'œil perçoit plus une succession d'images qu'un mouvement. Les standards de télévision européen et américain possèdent respectivement 25 et 30 images / seconde.

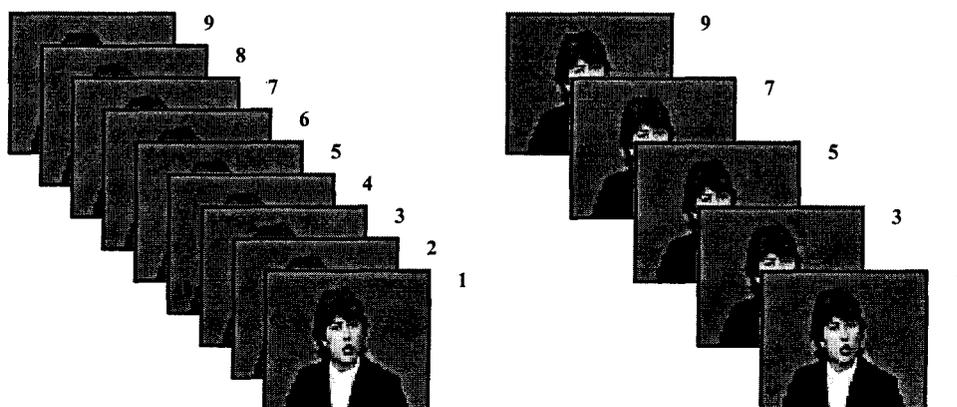
La vidéo utilise une quantité de données conséquente, ce qui nécessite un espace de stockage et une bande passante pour la transmission assez grands. Suivant les besoins, nous agissons sur les facteurs de la vidéo, pour atteindre un compromis entre la qualité de la vidéo et la quantité de ressources disponibles. En baissant la taille ou la profondeur de couleur d'une vidéo, nous baissons aussi sa qualité.



a - Différentes résolutions spatiales.



b - Différentes résolutions fréquentielles.



c - Différentes résolutions temporelles.

*Figure 5-2 : Différentes qualité vidéo.*

Pour établir une qualité vidéo, des paramètres de format doivent être choisis au niveau application. Ces choix doivent correspondre aux qualités perçues par les utilisateurs. La méthode communément utilisée [Alfano 97], consiste à présenter aux utilisateurs des vidéos avec différents paramètres de format. Les utilisateurs doivent alors attribuer une note de un à

cinq, correspondant à des appréciations de qualité mauvaise à excellente (Tableau 5-2). Cette méthode permet ainsi de faire une *projection* de la QoS utilisateur sur la QoS application.

| Qualité    | Note | Dégradation de la vidéo  |
|------------|------|--------------------------|
| Excellente | 5    | Imperceptible            |
| Bonne      | 4    | Perceptible, non gênante |
| Moyenne    | 3    | Un peu gênante           |
| Pauvre     | 2    | Gênante                  |
| Mauvaise   | 1    | Très gênante             |

Tableau 5-2 : Différents niveaux de qualité vidéo.

L'hétérogénéité des systèmes utilisateur fait que l'interprétation des niveaux de qualité est différente suivant chaque utilisateur. Plus précisément, elle est relative aux ressources de chacun. Le niveau de qualité le plus élevé chez l'un peut correspondre au niveau le plus bas chez un autre.

## 5.4. La QoS du point de vue récepteur

Nous avons vu que chaque participant décide de diriger son regard vers les personnes qui l'intéressent le plus. Par analogie, dans notre étude, nous avons pris le récepteur comme étant l'initiateur de l'établissement de la spécification de la QoS vidéo qu'il désire recevoir.

Pour chaque interlocuteur, il doit alors recevoir une qualité vidéo proportionnelle au degré d'intérêt qu'il a pour lui.

| Intérêt | Qualité       |
|---------|---------------|
| Fort    | Bonne         |
| Moyen   | Intermédiaire |
| Faible  | Basse         |

Tableau 5-3 : Relation entre Intérêt et Qualité.

A chaque degré d'intérêt, va correspondre une qualité vidéo. Ces degrés peuvent être spécifiés avec une granularité fine. Dans notre cas nous considérons seulement trois degrés, fort, moyen et faible. A ces trois degrés d'intérêt vont correspondre trois niveaux de qualité vidéo, bonne, intermédiaire et basse (Tableau 5-3). En fonction du degré d'intérêt déterminé par les règles établies pour l'activité, le système va déterminer la qualité vidéo correspondante.

Pour établir une communication entre l'émetteur et le récepteur correspondant à la qualité vidéo choisie, des *mécanismes de QoS* doivent être utilisés.

### **5.4.1. Les mécanismes pour la QoS**

Les paramètres de QoS permettent de spécifier une qualité de service. Mais pour établir la QoS spécifiée, des mécanismes doivent être fournis par l'architecture du système. Nous pouvons classer les mécanismes utilisés pour garantir une QoS en mécanismes pour la phase d'établissement et mécanismes pour la phase opérationnelle. Ces mécanismes peuvent se trouver à différents niveau du système.

#### **5.4.1.1. Les mécanismes pour la phase d'établissement**

Ces mécanismes sont utilisés seulement en phase d'établissement de la QoS. Un autre appel à ces mécanismes ne se fera que lorsqu'il y a un changement au niveau des spécifications de QoS.

*Négociation de QoS* : une fois qu'une QoS est spécifiée par une partie, elle doit le négocier avec les autres parties [Rothermel 97]. Cette négociation peut être verticale ou horizontale. Une négociation verticale se fait entre les différentes couches d'un système d'un même site, alors qu'une négociation horizontale se fait entre couches de même niveau de systèmes de sites différents.

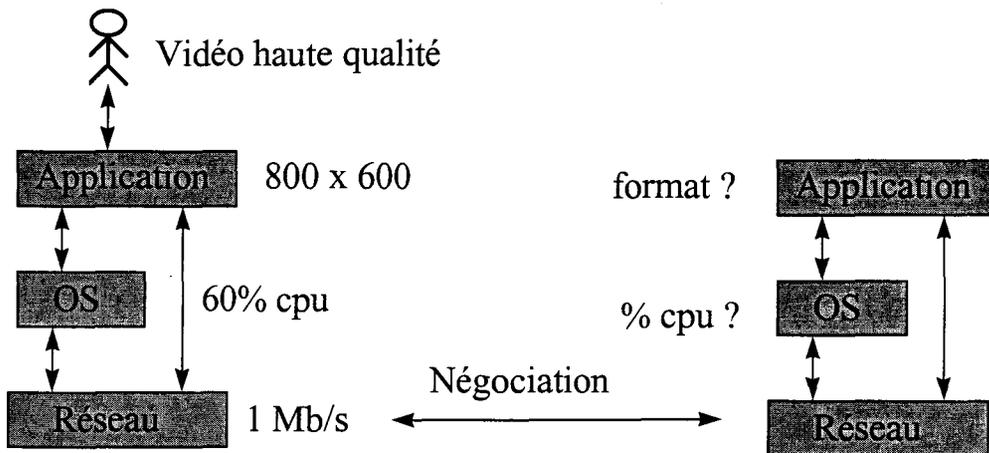


Figure 5-3 : Négociation de QoS.

La Figure 5-3 montre que les négociations horizontales et verticales sont nécessaires à tous les niveaux. Une négociation entre les applications doit d'abord se faire pour le format vidéo. Si l'on effectue une négociation horizontale seulement au niveau des couches basses réseaux, nous ne pouvons en déduire le format vidéo que l'émetteur doit envoyer, puisqu'une spécification de QoS réseau peut correspondre à plusieurs spécifications de QoS application.

Pour une communication point à point, il est relativement aisé de faire une négociation. Comme l'établissement d'une QoS dépend des ressources disponibles, la négociation se fera en fonction des ressources de la partie qui a le moins de ressources.

La négociation de QoS est beaucoup plus complexe dans une situation multipoint. Dans un groupe, les différents sites peuvent avoir des ressources hétérogènes. Négocier la QoS en fonction des ressources du site qui a le moins de ressources peut pénaliser plus d'un site, qui auront des ressources non exploitées.

**Projection des paramètres QoS entre différentes couches** : un système de communication multimédia possède plusieurs composants organisés en couche. Au niveau de chaque couche, des paramètres QoS sont déterminés. Les valeurs attribuées aux paramètres QoS d'une couche, correspondent aux paramètres QoS « demandés par » la couche supérieure, et aux paramètres QoS « demandés à » la couche inférieure.

La projection de QoS permet d'effectuer les fonctions de conversion entre représentations de QoS à différents niveaux du système. Par exemple, un certain niveau de qualité subjective,

vu par un participant, va se traduire par un certain temps processeur au niveau système d'exploitation, et une certaine bande passante.

Pour satisfaire une QoS, un certain nombre de ressources doivent être utilisées. Des fonctions de déterminations des ressources nécessaires doivent être utilisées. Ces fonctions déterminent seulement la quantité de ressources nécessaires à la QoS demandée. Elles ne font pas d'allocation de ressource, ni de renégociation de QoS dans le cas où les ressources disponibles ne peuvent pas satisfaire la QoS demandée.

**Contrôle d'admission** : le contrôle d'admission est responsable de la comparaison des ressources nécessaires résultant de la QoS désirée, avec les ressources disponibles du système. La décision pour que la requête soit acceptée dépend généralement de la politique de gestion des ressources et de leur disponibilité. Une fois que le contrôle d'admission a été complété avec succès sur les ressources d'un module particulier, les ressources locales sont immédiatement réservées, et seront engagées plus tard lorsque le contrôle d'admission de bout en bout (accumulation des tests saut par saut) est accepté.

**Réservation de ressources** : en se basant sur la projection QoS et le contrôle d'admission à chaque ressource locale des modules traversés (CPU, mémoire, périphériques d'entrée/sortie, commutateur, routeurs,...), des ressources sont allouées de bout en bout. La réservation de ressources est faite une fois la négociation de la QoS acceptée.

#### **5.4.1.2. Les mécanismes pour la phase d'opération**

Ces mécanismes servent à assurer le maintien de la QoS contractée dans la phase d'établissement. Les mécanismes les plus utilisés sont :

**Surveillance** : permet à chaque niveau du système de traquer la QoS en cours satisfaite par la couche inférieure. Ce mécanisme fait généralement partie de la boucle de rétroaction du mécanisme de maintenance de QoS pour maintenir la qualité de service accomplie par les modules surveillés.

**Maintenance** : le mécanisme de maintenance compare la qualité de service surveillée avec celle contractée. S'il y a une différence, ce mécanisme y remédie en faisant un ajustement des ressources des modules pour maintenir la QoS.

**Filtrage** : un filtre permet de transformer les données pendant leur progression dans le réseau, dans le but d'altérer quelques propriétés concernant la QoS. Les mécanismes de

filtrage permettent de résoudre le problème de l'hétérogénéité de récepteurs dans le cas d'un flux multicast. Chaque récepteur recevra les flux en fonction de ses capacités.

Le filtrage peut se faire de trois façons :

- Par suppression de données du flux en cours. Un exemple est un filtre qui enlève l'information de couleur d'un flux vidéo, laissant seulement l'information de luminance ;
- Par translation (transcodage). Le filtre altère les propriétés du flux en faisant une conversion des données du flux. C'est le cas par exemple lorsqu'un flux entre dans le filtre au format JPEG, et en sort au format H.261.
- Filtrage implicite. C'est le cas lorsque le flux de données est transmis dans des connexions séparées. La QoS peut ainsi être altérée par simple établissement ou rupture de certaines connexions qui forment le flux de données. Des méthodes d'encodage vidéo en couches existent, comme MPEG 2, où chaque couche sera envoyée sur un canal.

**Contrôle de flux** : le contrôle de flux est largement utilisé dans les applications réseaux. L'émetteur doit souvent ajuster son débit de sortie en fonction du retour qu'il reçoit, soit du récepteur si ce dernier n'arrive pas à suivre le même rythme que l'émission, soit du réseau s'il y a congestion sur certains liens. Le protocole RTP [Schulzrinne 96] implémente le contrôle de flux entre applications. Le récepteur utilise des primitives du protocole RTCP pour indiquer à l'émetteur l'état de la réception. L'émetteur adapte son flux en conséquence. L'application de vidéoconférence sur Mbone « vic » implémente le protocole RTP au niveau application.

**Ordonnement de flux** : en fonction de l'importance que représente un flux, il peut être ordonné en priorité par rapport à d'autres flux. Ceci peut se faire au niveau local, où un temps processeur plus grand est attribué à un flux, ou au niveau du réseau, dans lequel un flux peut être acheminé en priorité.

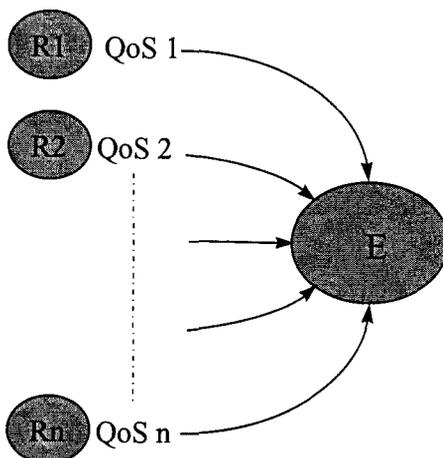
**Synchronisation de flux** : nous pouvons avoir une synchronisation intra flux ou inter flux. La synchronisation intra flux permet de respecter la fréquence de production d'un flux au moment de sa restitution, ainsi que l'ordre de ses échantillons. Par contre, la synchronisation inter flux resynchronise des flux distincts. Le cas le plus commun est la synchronisation entre image et son.

## 5.5. Qos du point de vue émetteur

Dans une session de communication, chaque nœud est composé de deux parties, le *récepteur* et l'*émetteur*. Nous avons vu dans le paragraphe précédent, que c'est la partie

récepteur qui initie l'établissement de la communication en envoyant des requêtes de QoS vidéo aux autres participants avec les paramètres vidéo correspondant à ses intérêts. La partie émetteur, se charge d'envoyer la vidéo aux autres participants avec les paramètres de qualité demandés.

Pour une communication point à point, l'émetteur envoie la vidéo avec les paramètres demandés par le récepteur. Par contre, dans notre cas, la communication s'effectue en mode multipoint pour faire coopérer plusieurs participants. L'émetteur (E) peut ainsi recevoir plusieurs demandes de QoS différentes, provenant chacune d'un récepteur (R<sub>i</sub>) (Figure 5-4).



*Figure 5-4 : Requêtes QoS différentes*

Deux raisons peuvent amener les participants à spécifier des requêtes de QoS différentes :

- La différence entre les intérêts des participants pour voir le même participant ;
- La différence entre les ressources des participants.

### **5.5.1. Différence entre les intérêts des participants**

Chaque récepteur spécifie sa requête de demande de vidéo en fonction de ses intérêts. Si nous reprenons l'exemple du cours à distance vu dans le chapitre précédent (§ 4.7), les participants n'ont pas les mêmes intérêts à chaque étape du déroulement du cours.

En spécifiant trois niveaux de qualité vidéo, un niveau haut, un niveau intermédiaire, et un niveau bas, les présentations vidéo attendues par les participants correspondant au Tableau 4-3 des intérêts (§ 4.7.3 ), sont données par la Figure 5-5.

Dans les étapes deux et trois, l'Elève Petri et le Professeur P sont vus différemment. Alors que le Professeur et les Elèves sont intéressés par voir l'Elève qui pose la question (Petri), ce dernier est plus intéressé par voir la personne à qui il pose la question, c'est à dire le Professeur.

Dans la deuxième étape, le Professeur reçoit deux requêtes de QoS différentes, alors que dans la troisième étape, c'est l'Elève Petri qui reçoit deux requêtes de QoS différentes.

Dans ces deux cas, quelles sont les requêtes prioritaires ?

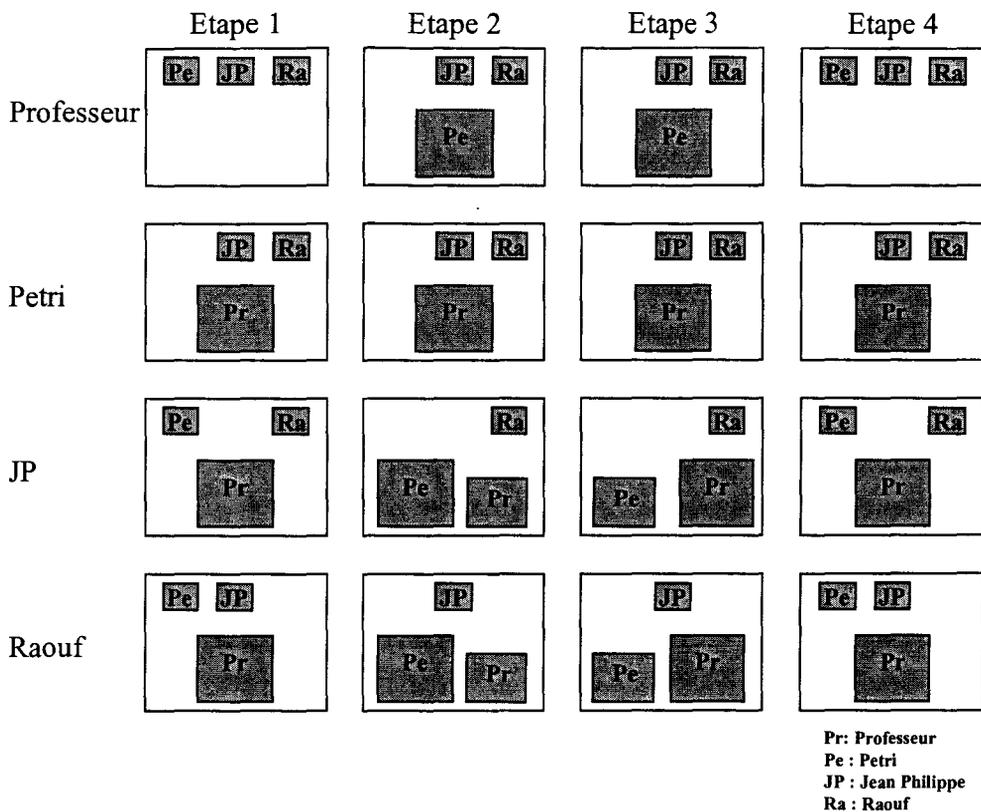


Figure 5-5 : Présentation vidéo attendues par les participants.

### 5.5.2. Différence entre les ressources des participants

Si les systèmes de travail coopératif enlèvent la barrière de distance entre personnes éloignées géographiquement, l'hétérogénéité des systèmes crée une distance entre les ressources des sites des participants.

En fonction de ses intérêts, le récepteur établit des priorités entre les participants qu'il veut voir. Mais ce sont les ressources qui permettent de quantifier ses priorités en requêtes de QoS

vidéo à envoyer à l'émetteur. Les paramètres de qualité vidéo qui formeront la requête seront choisis non seulement pour être conformes aux intérêts du participant, mais aussi pour une utilisation optimale des ressources disponibles.

Comme nous nous plaçons dans un milieu où les systèmes utilisateurs sont hétérogènes, deux sites ayant le même degré d'intérêt pour recevoir un participant, émettront des requêtes différentes, s'ils ont des ressources différentes.

Inversement, deux requêtes similaires ne traduisent pas le même degré d'intérêt lorsqu'elles proviennent de deux sites ayant des ressources différentes.

### **5.5.3. Détermination des vecteurs QoS optimaux**

Le Tableau 5-4 regroupe un ensemble de standards de formats vidéo utilisés dans les systèmes de vidéoconférence. A l'origine, ces standards sont issus des mondes de la télévision, de la vidéoconférence, et des ordinateurs. La représentation graphique de cet ensemble de formats montre la dispersion des valeurs de leurs paramètres (Figure 5-6).

|                 | Largeur | Hauteur |
|-----------------|---------|---------|
| SIF             | 352     | 288     |
| CCIR 601        | 720     | 576     |
| High-1440 level | 1440    | 1152    |
| High level      | 1920    | 1152    |
| Pal/Secam       | 768     | 576     |
| sub-QCIF        | 128     | 96      |
| QCIF            | 180     | 144     |
| CIF             | 360     | 288     |
| 4CIF            | 786     | 704     |
| 16CIF           | 1408    | 1152    |
| indeo           | 320     | 240     |

*Tableau 5-4 : Formats vidéo standards.*

Il en résulte que de manière général, l'émetteur peut recevoir plusieurs requêtes QoS vidéo différentes. Ceci pose le problème de savoir quelle est la QoS vidéo à choisir. Du point de vue ressources, l'émetteur ne peut pas satisfaire chaque demande. Cela nécessiterai un codage vidéo spécifique pour chaque QoS demandée.

La première solution qui puisse être envisagée, est de choisir la QoS la plus faible. C'est la solution la plus simple, et qui est, traditionnellement, implicitement adoptée, puisque à

l'établissement de la communication, tous les sites participants se calent sur les capacités du site le moins puissant. Chaque récepteur recevra un flux vidéo dont la qualité correspondra à celle demandée par le site ayant demandé la QoS la plus faible. Cette politique peut pénaliser les sites qui ont demandé des QoS plus élevées. Leurs ressources seront sous exploitées ou bien leurs intérêts ne seront pas respectés.

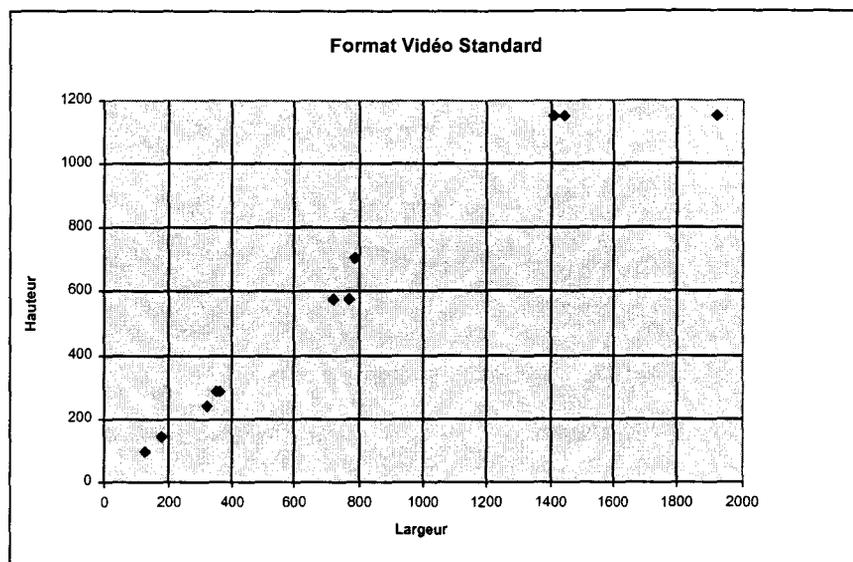


Figure 5-6 : Représentation graphique montrant la dispersion des formats vidéo standards.

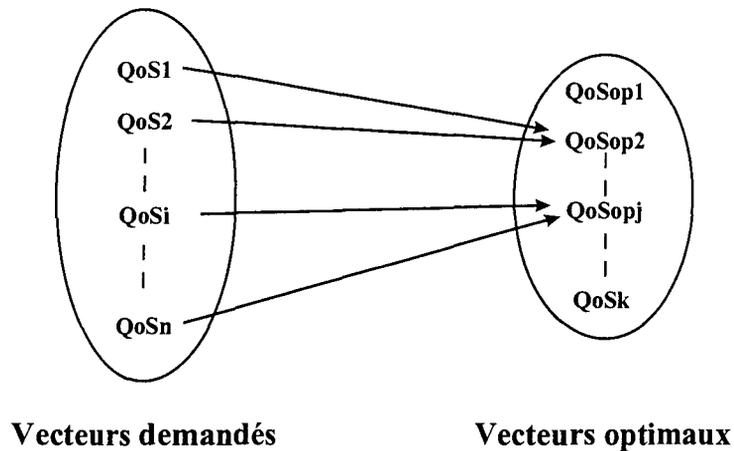
En utilisant les techniques de codage en couche réalisées au niveau de l'émetteur et les passerelles de transcodage se trouvant sur le réseau, nous pouvons augmenter le nombre de QoS vidéo à émettre.

Ainsi, suivant la disponibilité de ressources, d'autres solutions peuvent être envisagées. En supposant que le nombre de QoS possibles vaux  $K$ , nous pouvons choisir les  $K$  vecteurs de telle sorte qu'ils soient uniformément répartis entre les deux vecteurs extrêmes demandés. Le récepteur choisira le vecteur le plus proche de sa demande. Cette solution permet de répartir les ressources disponibles sur le spectre totale de QoS demandées.

La solution précédente peut ne pas être adéquate si les demandes ne sont pas réparties de manière uniforme entre les deux demandes extrêmes. Une amélioration de la solution serait de choisir les  $K$  vecteurs QoS à émettre en fonction des QoS les plus demandées, pour satisfaire le plus de demandes possible.

Notre but est donc d'arriver au schéma de la Figure 5-7. A chaque vecteur  $QoS_i$  demandé par un récepteur, nous devons faire correspondre un vecteur optimal  $QoS_{opj}$ , parmi les vecteurs déterminés pour une satisfaction maximale des récepteurs. Le récepteur  $i$  ayant demandé la  $QoS_i$  recevra une vidéo de  $QoS_{opj}$ . Cette correspondance se fait telle que la distance entre les deux vecteurs soit minimale.

Les demandes que l'on fait correspondre au même vecteur QoS optimal sont donc proches.



*Figure 5-7 : Correspondance entre Vecteurs QoS Demandés et Vecteurs QoS optimaux.*

Pour déterminer les vecteurs QoS optimaux, nous devons alors grouper les demandes en  $K$  classes, puis pour chaque classe, déterminer le vecteur QoS optimal.

Pour déterminer les classes, posant le problème comme suit :

Etant donné un nombre  $N$  de vecteurs de  $P$  paramètres, correspondants aux  $N$  requêtes des participants, chaque requête possédant  $P$  paramètres pour la qualité vidéo, comment obtenir une répartition de ces  $N$  objets en  $K$  classes, où  $K$  serait le nombre de QoS vidéo que l'on peut fournir grâce au codage en couche et aux passerelles.

Une fois ces classes obtenues, il faut trouver les  $K$  vecteurs de QoS optimaux.

Pour gérer la complexité dans les systèmes utilisateurs et le réseau, nous pouvons considérer que les flux de données correspondent à un ensemble de services et de classes de trafic sur le réseau. Ses flux peuvent être caractérisés, et toutes les demandes de QoS, viendraient dans l'une de ces classes.

Cette approche peut être restrictive pour de nouvelles applications qui peuvent apparaître, ayant des flux de données spécifiques, ainsi que de nouveaux systèmes ayant des caractéristiques qui ne correspondent pas aux classes déjà préétablies. Les classes de trafic peuvent changer constamment lorsque de nouvelles applications sont développées.

Pour cela nous allons traiter ce problème de manière générale, c'est-à-dire que les différentes QoS utilisateurs peuvent être très variées.

Pour répartir les requêtes QoS en classes, nous utilisons un algorithme de classification automatique particulier (voir Annexe III). Nous créons ainsi une partition de K classes des N requêtes. Les vecteurs optimaux de chaque classe peuvent être calculés simplement en prenant la moyenne, le maximum, le minimum, ou tout autre critère. Dans notre cas, nous prendrons la valeur inférieure pour prendre en compte les sites qui n'ont pas assez de ressources. Les sites qui ont demandé des QoS plus élevées recevront une QoS légèrement inférieure, et ne seront pas beaucoup pénalisés.

La solution adoptée jusqu'ici, prend en compte seulement le point de vue statistique, sans prendre en compte l'aspect activité coopérative. L'analyse statistique des données cache tout l'aspect du travail coopératif. Cette solution doit être complétée pour prendre en considération le bon déroulement de l'activité coopérative.

#### **5.5.4. Solution du point de vue activité coopérative**

La solution du choix des QoS vidéo à émettre doit favoriser le bon déroulement de l'activité coopérative. Avec la solution statistique « pure », nous pouvons arriver à deux résultats. Dans un cas favorable, les données des requêtes QoS seront telles que la classification va aboutir à des classes bien distinctes (Figure 5-8-a). Ces classes seront optimales au sens statistique du terme.

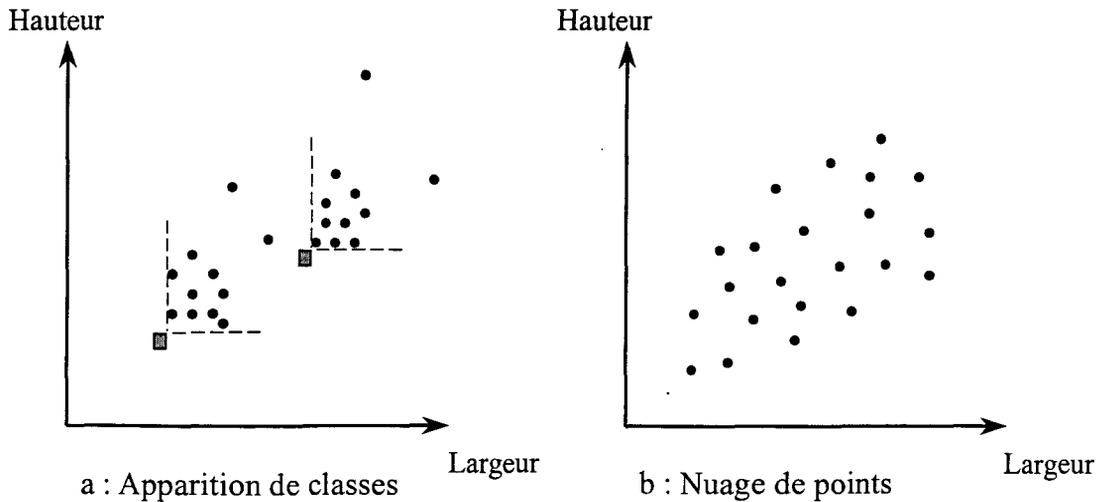


Figure 5-8 : Répartition des requêtes QoS dans un espace à deux dimensions. Exemple de requêtes pour la dimension de la vidéo.

Prenons une activité coopérative dans laquelle plusieurs participants sont engagés. A la manière des émissions de débats télévisés, les participants sont regroupés en deux catégories. La première comprend un nombre restreint de participants qui animent activement le débat, alors que la seconde catégorie est composée de la majorité des participants, qui sont principalement des observateurs, mais qui peuvent toutefois intervenir pour poser des questions. Dans ce cas, les observateurs sont en surnombre, et la classification sera principalement effectuée avec les données des requêtes émises par les observateurs. Le résultat ne sera pas optimal au sens de l'activité coopérative, puisque les participants qui participent le plus au débat risquent de ne pas avoir leur requête satisfaite.

Le choix des QoS doit tenir compte des rôles des participants dans l'activité coopérative.

Dans le cas défavorable, le résultat de la classification ne donne aucune classe apparente (Figure 5-8-b). Les requêtes des participants peuvent former un nuage de points, qui ne donne pas une classification optimale même au sens statistique.

Cette solution initiale, traite toutes les requêtes de manière égale. Pour arriver à des résultats meilleurs, nous devons introduire des paramètres pour tenir compte de l'activité coopérative. Ces paramètres sont le degré d'importance et le degré d'intérêt des participants.

### 5.5.4.1. Degré d'importance

Dans une activité coopérative, les participants ont différents rôles, et utilisent différents outils pour exécuter leur tâche. En fonction de son rôle et des outils qu'il utilise, un participant devient plus ou moins important par rapport aux autres participants pour la session de travail.

Dans le choix de la QoS à envoyer, l'émetteur doit tenir compte de l'importance du demandeur en attribuant à sa requête une priorité.

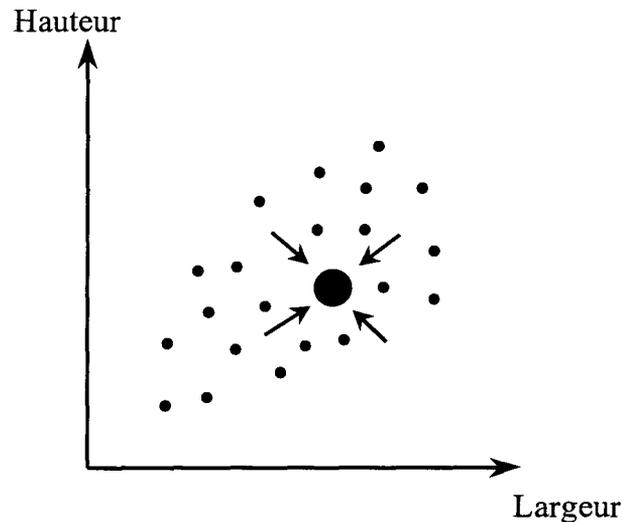


Figure 5-9 : Degré d'importance.

Pour cela, à chaque requête sera associée une pondération traduisant le degré d'importance du participant demandeur. De cette façon, les requêtes ne sont plus traitées de manière égale, et la recherche des vecteurs QoS optimaux sera drainée par les requêtes dont les pondérations sont plus grandes (Figure 5-9).

Reprenons notre exemple du cours à distance. A l'étape trois, le professeur répond à une question qui a été posée par l'élève Petri dans la seconde étape. La partie émetteur de Petri reçoit trois requêtes, l'une provenant du Professeur et les deux autres des élèves Jean Philippe et Raouf. La requête du Professeur correspond à une vidéo de haute qualité, alors que les deux autres requêtes des élèves Jean Philippe et Raouf correspondent à une qualité intermédiaire. En adoptant la solution statistique pure, nous favoriserons les requêtes des élèves Jean Philippe et Raouf. Comme le professeur doit répondre à la question posée, sa position est plus

importante à cet instant pour l'activité. Pour cela, nous introduisons la notion d'importance du participant pour donner au professeur une priorité plus élevée par rapport aux autres élèves.

#### **5.5.4.2. Degré d'intérêt**

L'établissement de la QoS par le récepteur se fait en fonction de son degré d'intérêt vis à vis de l'émetteur. Ce dernier peut donc exploiter cette information pour améliorer le choix du QoS optimal à émettre. Le degré d'intérêt va servir pour pondérer la requête du participant correspondant. De cette façon, le choix de la QoS optimale va favoriser les requêtes émises par les participants qui ont plus intérêt à voir l'émetteur.

La perception de l'environnement d'un participant possédant un degré d'intérêt bas, ne sera pas beaucoup influencée si sa requête n'est pas totalement satisfaite.

Par contre, les participants pour lesquels il a plus d'intérêt auront des requêtes dont les pondérations seront plus grandes, et par conséquent auront plus de chance d'être satisfaites.

#### **5.5.4.3. Détermination des facteurs de pondération**

Le problème de détermination des pondérations est toujours un exercice difficile. Dans notre cas, l'analyse de l'activité considérée doit nous permettre de déterminer les priorités entre les participants pendant la session de travail. Les valeurs des pondérations peuvent être faites expérimentalement en prenant des valeurs différentes et ne garder que les valeurs les plus satisfaisantes.

#### **5.5.5. Remarque**

Dans la résolution du problème de requêtes multiple de QoS, nous avons volontairement opté pour une solution générale où nous donnons comme hypothèse que les requêtes QoS sont très variées, et qu'elles ne font pas partie d'un groupe de classes préétablies.

Donc, si les requêtes faites correspondent à quelques formats standards, le choix de la QoS optimale sera plus simple.

Aussi, la complexité du problème est réduite en sachant, que les vecteurs QoS ne vont pas occuper tout l'espace de données possible, puisque certaines contraintes sur les paramètres existent. Si nous prenons comme exemple la taille de la vidéo, les valeurs prises par la largeur et la hauteur varient entre une valeur maximale et une valeur minimale. De plus, le rapport entre ces deux paramètres est généralement compris dans un intervalle (Figure 5-10).

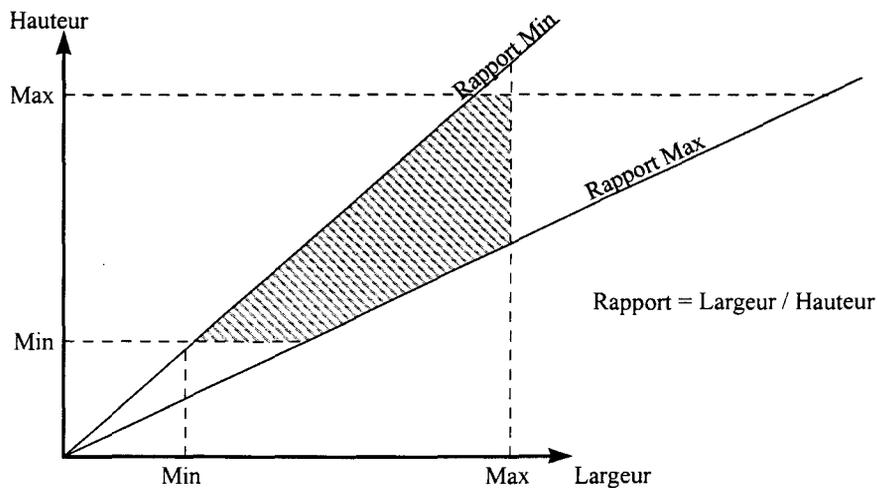


Figure 5-10 : Intervalles de variation de la taille de la vidéo.

Les valeurs minimales, maximales, et le rapport de ces paramètres sont déterminées par des règles d'ergonomie et par les contraintes technologiques. Des règles ergonomiques peuvent fixer les valeurs minimales de la longueur et de la largeur en dessous desquelles la vidéo n'est pas exploitable par les utilisateurs, alors que les valeurs maximales peuvent être dues à des limites technologiques comme la taille de l'écran de visualisation.

## 5.6. Conclusion

Dans ce chapitre, nous avons d'abord défini la qualité de service, puis nous avons montré comment faire le lien entre les degrés d'intérêt des participants et les qualités vidéo que peut offrir le système.

Dans notre étude, nous avons trouvé que pour l'établissement de liaisons vidéo satisfaisant une certaine qualité de service, l'activité coopérative doit être prise en compte, autant du point de vue de l'émetteur que celui du récepteur. Le récepteur utilise les intérêts qu'il a pour spécifier la QoS vidéo à négocier avec les sites distants. Pour le récepteur, nous avons vu qu'il était confronté au problème de la réception de plusieurs requêtes QoS Vidéo. Pour cela, il utilise le degré d'intérêt qu'ont les récepteurs pour le voir, ainsi que l'importance des participants pour le bon déroulement de l'activité, afin de choisir la QoS à envoyer.

Comme chaque activité possède ses propres règles pour la détermination des intérêts des utilisateurs, nous verrons dans le chapitre suivant, l'architecture choisie pour s'adapter à différents types d'activités coopérative.

---

## **6. Conception et mise en œuvre d'une Architecture adaptable à l'activité coopérative**

---

## **6.1. Introduction**

Dans les deux chapitres précédents, nous avons fait la proposition d'un système permettant de gérer la présentation vidéo du participant dans une activité coopérative. D'une part, cette gestion correspond aux intérêts du participant, et d'autre part, elle est contrainte par la qualité de service possible avec les ressources disponibles.

Chaque activité coopérative possède ses spécificités, comme les rôles et les règles de gestion de la présentation vidéo. Il en résulte alors que la gestion de la présentation vidéo des participants est différente dans chaque type d'activité.

La conception d'une architecture fermée ne tenant compte que d'un seul type d'activité coopérative, conduit à des systèmes dédiés. Un utilisateur désirant participer à différentes activités coopératives devra installer une application spécifique pour chacune d'elle.

Pendant la conception de l'architecture du système, nous devons alors prendre en considération cette diversité des activités coopératives, en séparant les aspects de l'activité coopérative du reste de l'application. L'application permettant à l'utilisateur de coopérer, doit s'adapter à différents types d'activités. Lorsqu'un participant se joint à une activité, son système récupère les spécificités de l'activité qui viennent se rajouter à l'architecture globale de l'application.

Les technologies des objets répartis [Orfali 96] comme CORBA<sup>35</sup> [OMG 95], DCOM<sup>36</sup> [Eddon 98, Sessions 97, Brockschmidt 95] et JavaBeans [Englander 97], permettent d'avoir des composants connectables dynamiquement.

Dans ce chapitre, nous commencerons par donner des notions sur les Frameworks, puis nous montrerons comment les utiliser dans notre architecture afin de s'adapter à différentes activités coopératives. Des détails sur les choix techniques d'implémentation sont expliqués, et les expériences réalisées avec notre prototype de gestion vidéo sont décrites et commentées.

---

<sup>35</sup> CORBA : Common Object Request Broker Architecture.

<sup>36</sup> DCOM : Distributed Component Object Model.

## 6.2. Les Frameworks

Si la programmation orientée objet est venue pour combler les lacunes de la programmation structurée en extensibilité, maintenabilité et réutilisabilité, en permettant aux développeurs de travailler à un plus haut niveau d'abstraction avec des objets et des classes, la programmation orientée Framework permet de maximiser les bénéfices de la programmation orientée objet [Taligent 93].

Une définition des Frameworks est donnée par Johnson :

*Un Framework est un ensemble de classes qui forme un modèle abstrait pour donner solution à une famille de problèmes similaires [Johnson 88].*

Les Frameworks orienté objet ont été promus dans le but d'aider les développeurs à produire des logiciels plus rapidement, et à se concentrer plus sur la valeur ajoutée que sur les parties répétitives. Un Framework est décrit comme un modèle réutilisable exprimé par un ensemble de classes abstraites et de la manière dont leurs instances collaborent [Roberts 96]. Il capture l'expertise nécessaire pour résoudre les problèmes dans un domaine particulier. Il cache ainsi la partie de la conception qui est commune à toutes les applications dans ce domaine, et laisse les parties qui doivent être spécifiques.

Une importante caractéristique des Frameworks, est que le code ajouté par l'utilisateur pour façonner le Framework, sera appelé par ce dernier, au lieu que se soit le code écrit par l'utilisateur qui fait appel aux services du Framework. Ce mécanisme est souvent appelé « le principe de Hollywood », « ne nous appelez pas, on vous appelle ». Ceci dispense le programmeur de la tâche du contrôle du flux d'exécution.

Nous pouvons distinguer deux types de Frameworks, les boîtes blanches et les boîtes noires. Un Framework est appelé boîte blanche, s'il est basé sur l'héritage. Le développeur d'application personnalise la boîte blanche en dérivant de nouvelles sous classes et en surchargeant les fonctions membres. Le terme « boîte blanche » fait référence à la visibilité. Due à l'héritage, la structure interne de la classe parent est souvent visible à ses sous classes.

L'héritage est défini de manière statique à la compilation. Nous ne pouvons pas modifier l'implémentation héritée des classes parentes à l'exécution. Aussi, l'implémentation des sous classes est liée à celles des classes parentes. Cette dépendance limite la flexibilité et la réutilisabilité.

Un Framework est appelé boîte noire, s'il est basé sur la composition d'objet. Les nouvelles fonctionnalités sont obtenues par assemblage ou composition d'objets. Un développeur d'application utilise alors cet assemblage de composants pour spécifier le comportement de l'application. La composition d'objets est connue pour être plus flexible que l'héritage. La raison principale est que la composition d'objets peut être changée dynamiquement, alors que les structures d'héritage sont figées au moment de la compilation. La composition d'objet nécessite que les objets composés doivent avoir des interfaces bien définies, pour que l'utilisateur n'ait à connaître que l'interface externe.

Les détails des objets ne sont pas visibles, et n'importe quel objet peut être remplacé par un autre objet de type compatible à l'exécution.

### **6.3. Architecture ouverte**

Dans le but de fournir une architecture permettant de s'adapter dynamiquement au type d'activité coopérative, nous adoptons le principe des Frameworks boîtes noires qui permettent d'accéder dynamiquement aux services d'un composant à l'exécution.

Pour cela, nous séparons les composants de notre architecture, en deux parties. Une partie commune aux différentes activités, et une partie spécifique à l'activité.

La partie commune représente le Framework de l'application. Elle regroupe les composants pour :

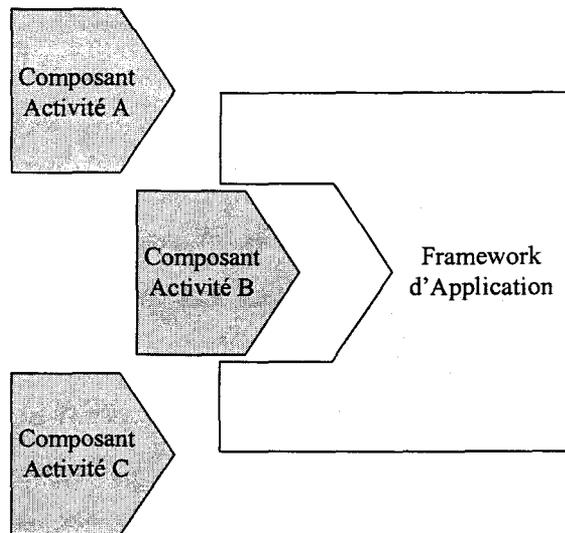
- La gestion des périphériques d'acquisition et de restitution vidéo ;
- La gestion de session : création de session, entrée et sortie d'une session, fermeture de session ;
- La partie communication : transmission et réception vidéo, et données de contrôle, notification ;
- les mécanismes de QoS ;

Dans la partie adaptable, nous mettrons toutes les spécificités de l'activité coopérative, à savoir :

- les rôles spécifiques à l'activité ;
- les autorisations qu'à chaque rôle pour l'utilisation des outils dans l'activité : sa priorité, ses privilèges ;
- les règles de détermination des priorités traduisant les intérêts du récepteur ;

- les règles de détermination des priorité sur les requêtes de QoS faites à l'émetteur ;

Pour permettre à notre architecture de s'adapter dynamiquement à différents types d'activités coopératives, le composant activité est un composant connectable, encapsulant toutes les spécificités de l'activité coopérative (Figure 6-1).



*Figure 6-1 : Composant Activité connectable.*

## **6.4. Interface entre composant Activité et le Framework d'application**

Un Framework et un composant interagissent par l'intermédiaire d'une interface. Cette interface permet au Framework, consommateur de services, de bénéficier des services offerts par le fournisseur de services, le composant.

L'un des obstacles majeur pour l'utilisation des Frameworks est l'apprentissage. Avant de pouvoir personnaliser un Framework, un utilisateur doit bien comprendre le principe d'utilisation du Framework, pour pouvoir ensuite ajouter son code personnel. Pour cela, il est recommandé que l'interface soit simple, afin de faciliter la réutilisation des Frameworks.

Nous avons défini une interface entre le composant activité et le reste de l'architecture. Le composant activité définit les rôles, les règles traduisant les intérêts des récepteurs et les règles pour déterminer les participants dont les requêtes QoS sont prioritaires.

Une interface simple et réduite doit exister entre le composant activité et l'architecture d'application, pour faciliter la réalisation de nouveaux composants activité (Figure 6-2).

Les services que peut demander le Framework au composant sont :

#### ***Requête de la liste des rôles***

Lorsqu'un participant veut entrer dans une activité, il doit choisir son rôle dans l'activité au sein du groupe. Pour avoir une connaissance des rôles possibles, l'application doit obtenir la liste des rôles possibles en envoyant une requête pour le composant activité.

#### ***Requête des autorisations sur les outils***

Permet au système de demander si un participant ayant un rôle donné possède les droits d'utilisation d'un outil, et quels sont ses privilèges et ses priorités.

#### ***Requête de la priorité des participants***

Pour composer sa présentation vidéo, le Framework doit connaître les intérêts que possède le participant. Il doit fournir au composant activité la liste des participants avec leur rôle et leur état. Le composant activité applique les règles pour déterminer les intérêts du récepteur et retourne la liste des participants triée par ordre de priorité.

Le Framework lance cette requête lorsqu'il reçoit les événements de :

- changement d'état d'un participant ;
- entrée ou sortie d'un participant dans la session.

#### ***Requête de la priorité des participants pour l'émetteur***

Pour déterminer la qualité avec laquelle il doit émettre sa vidéo, l'émetteur doit prendre en considération les requêtes de QoS vidéo qui lui ont été envoyées par les autres participants. Pour choisir la QoS optimale du point de vue activité coopérative, le Framework demande au composant le degré d'importance de la requête pour l'activité.

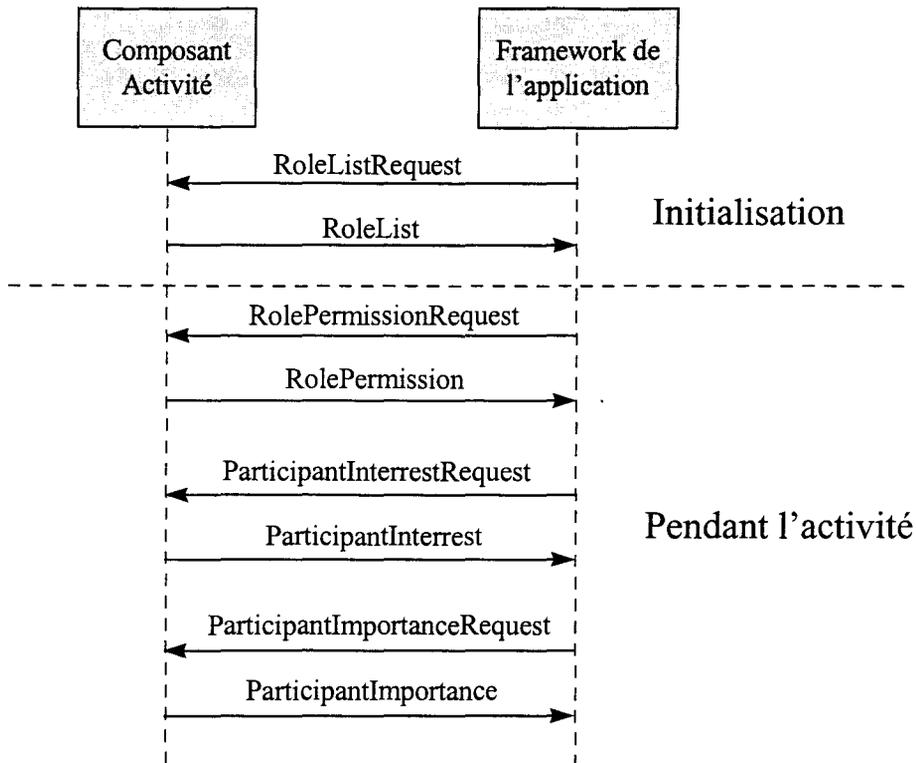


Figure 6-2 : Interface entre le composant activité et le Framework d'application.

## 6.5. Les composants du Framework

Dans le Framework, nous distinguons deux parties duales, l'émetteur et le récepteur. La partie récepteur permet de choisir pour le participant les qualités vidéo provenant des sites qui l'intéressent, alors que la partie émetteur envoie la vidéo du site local aux sites intéressés après avoir déterminé les vecteurs QoS optimaux.

### 6.5.1. Le récepteur

La Figure 6-3 montre les différents composants permettant de déterminer la requête QoS du côté du récepteur.

Un *gestionnaire de groupe* maintient la liste des participants avec les rôles correspondants et les outils associés. Il maintient également le profil de chaque participant. Le profil représente les capacités qu'a le participant en terme de QoS vidéo. Le profil d'un participant nous évite de lui demander une QoS plus élevée que ses capacités. Le profil est mis à jour périodiquement ou à chaque fois qu'il y a un changement, tout en évitant de surcharger le réseau.

Le **composant activité** lance l'exécution des règles pour la détermination des intérêts du participant lorsqu'il y a un changement au niveau du groupe (changement de la personne qui parle ou de l'attribution de certains outils).

Le **gestionnaire de ressources** gère les ressources locales. Il renseigne sur la disponibilité des ressources. Il peut faire de la réservation de ressource pour les applications.

Après avoir choisi les QoS optimales, le **négociateur de liaison** doit établir les liaisons permettant de recevoir les vidéos avec les qualités correspondantes. Pour garantir, la QoS vidéo, le **négociateur de liaison** doit effectuer une projection de QoS vidéo en QoS réseau, en terme de bande passante et de délai. Ceci pourra être utilisé pour faire de la réservation de ressources réseau, à l'aide de protocoles comme RSVP [Zhang 93].

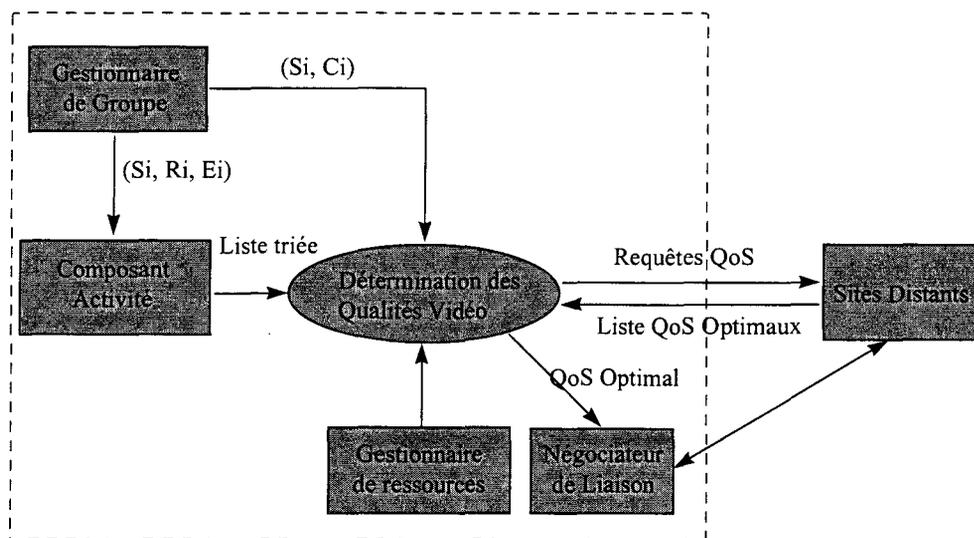


Figure 6-3 : Architecture côté récepteur.

Pour déterminer la requête de QoS vidéo, une liste des participants triée par ordre de priorité est fournie par le composant d'activité. En fonction des ressources locales et des capacités des émetteurs, une sélection de sites à recevoir est faite, puis les vecteurs de QoS sont établis et envoyés aux émetteurs correspondants. Une liste de QoS possibles est fournie par chaque émetteur. Le récepteur choisit alors la QoS s'approchant le plus de sa requête.

Finalement le négociateur de liaison établit les liaisons nécessaires.

### 6.5.2. L'émetteur

L'émetteur reçoit différentes requêtes de QoS vidéo. Il doit déterminer les vecteurs QoS optimaux. Pour cela, il doit faire appel au gestionnaire de ressources pour connaître ses capacités, et au composant activité pour déterminer les requêtes QoS prioritaires en fonction de l'importance des participants pour la coopération (ParticipantImportanceRequest).

La liste des vecteurs QoS optimaux sera alors transmise aux sites demandeurs, qui choisiront alors la QoS la plus proche de leur demande, en cherchant la distance minimale entre la QoS demandée et la liste des QoS optimales.

Les paramètres de QoS optimaux seront transmis au *Codeur Vidéo* pour effectuer la compression vidéo, ainsi qu'aux passerelles de transcodage. Une réservation de ces passerelles devra être faite au préalable.

Finalement le gestionnaire de liaison prend la main pour établir les liaisons en se basant sur les mécanismes de QoS nécessaires de niveau réseau entre émetteur et récepteur (Figure 6-4).

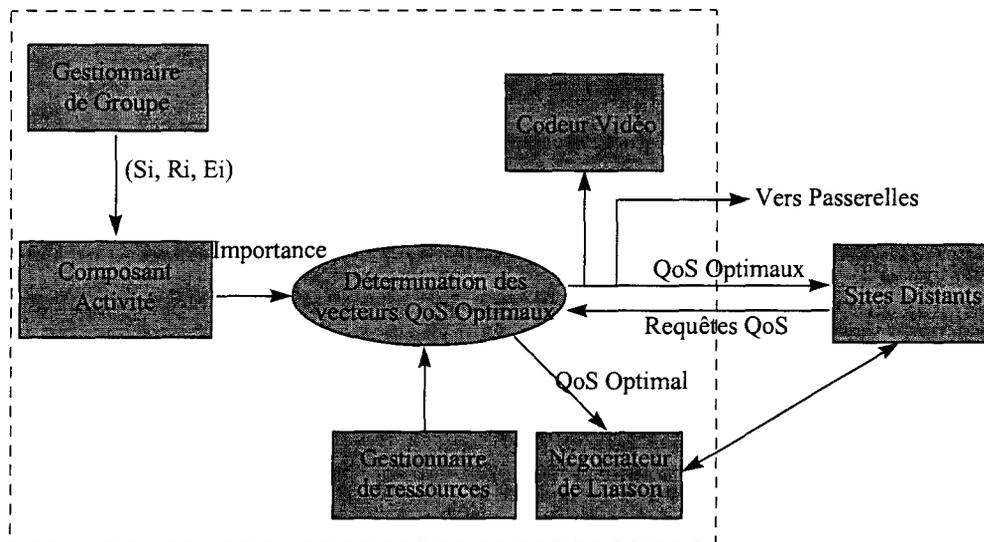


Figure 6-4 : Architecture côté émetteur.

## 6.6. Déploiement du système

Notre système est décomposé en deux applications distinctes, le *Serveur de session* et l'application *Participant* (Figure 6-5).

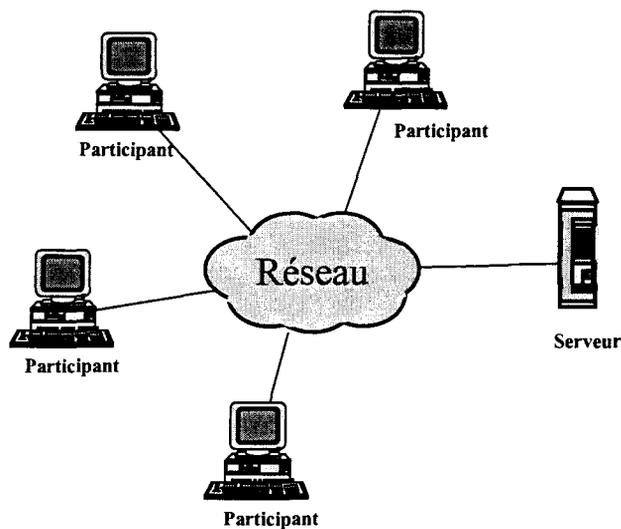


Figure 6-5 : déploiement du système.

Le serveur de session est un nœud central permettant de référencer les sessions d'activités en cours, ou programmées. Il se charge des notifications d'événements dans une session, comme l'entrée ou la sortie d'un participant pendant la session.

Un participant désirant rejoindre une session particulière, se connecte au serveur de session, afin de pouvoir récupérer les informations nécessaires. Le participant choisit la session qui l'intéresse, puis rentre au sein du groupe. Il récupère alors le composant activité permettant de l'assister dans la présentation vidéo, ainsi que les informations sur les autres participants, comme leur identité, leur profil de ressource, et les adresses multicast de diffusion. Le serveur l'enregistre dans la session, et fait une notification aux autres membres du groupe. Ces derniers récupèrent l'objet stockant les données relatives au nouveau participant, puis recalculent les priorités concernant la présentation vidéo. Les effets de l'entrée du nouveau participant dans la session sur leur présentation vidéo, dépendent de sa contribution à l'activité.

De même, lorsqu'un participant quitte la session, le serveur fait une notification aux autres participants.

Le serveur tient à jour les états des participants renseignant sur l'attribution des outils à un moment donné. Ceci permet à un participant de récupérer le contexte de la session au moment où il adhère au groupe, et de composer la présentation vidéo en conséquence.

## 6.7. La transmission vidéo

Dans nos travaux, nous avons utilisé la pile TCP/IP pour la transmission des données. Au niveau transport, cette pile offre deux protocoles, TCP et UDP. Le protocole TCP est un protocole pour la transmission fiable des données, puisqu'il intègre des mécanismes de contrôle d'erreur. Le contrôle sur erreur peut être nécessaire pour des données comme le transfert de fichier. La transmission des données se fait point à point en mode connecté. Pour des communications multipoints, nous devons combiner des communications point à point pour former un domaine multipoint. Les données à envoyer seront dupliquées sur chaque liaison point à point. Sur RNIS, la duplication des données pour les participants est généralement réalisé par un MCU.

Le contrôle d'erreur et la duplication des données introduisent un délai supplémentaire dans la transmission de données, ce qui n'est pas souhaitable pour le transport des médias continus, et particulièrement l'audio et la vidéo. De plus, la duplication de données introduit un gaspillage de la bande passante. Dans les larges groupes, la duplication de données devient impraticable.

A la différence de TCP, UDP est un protocole de transport sans connexion. Il ne gère pas le contrôle d'erreur sur les pertes de paquets. De ce fait, élimine le délai introduit par TCP pour le contrôle d'erreur. Cependant, nous devons ajouter la gestion des erreurs au niveau application si cela s'avère nécessaire.

Le protocole UDP est alors plus léger que le protocole TCP du fait qu'il n'y a pas de transmission de données relatives à la gestion du flux. Ceci permet un gain en bande passante pour les données utile.

Les contraintes de temps réel de la vidéo (et des médias continus de manière générale), ne permettent pas la ré-émission des paquets perdus. Le protocole UDP est alors plus adéquat pour le transport des médias continus. De plus, comme le flux vidéo est rafraîchi de manière continue, la perte de certains paquets est tolérée.

Une autre raison majeure de choisir UDP comme protocole de transport de la vidéo, est la possibilité de transmettre les données en mode multicast. La diffusion de la vidéo au sein de la conférence, se faisant en mode multicast, permet :

- D'éviter un nœud central pour dupliquer les paquets vidéo aux participants.

- De gagner en bande passante, puisque les paquets vidéo seront dupliqués le plus près possible du destinataire.

### 6.7.1. Emission vidéo

Chaque émetteur transmet sa vidéo sur une ou plusieurs adresses multicast, suivant qu'il utilise ou pas du codage en couche.

Comme l'en-tête contient un champ sur deux octets spécifiant la taille du paquet (en-tête comprise) la taille d'un paquet peut alors atteindre 65 ko. Cependant, comme UDP est sans contrôle d'erreur, et comme le paquet sera fragmenté dans les couches basses, il est préférable d'envoyer des paquets plus petit, correspondant à la taille des MTU<sup>37</sup>. Cette taille varie en fonction du réseau. Elle est de 1500 octets pour Ethernet, mais est généralement fixée à 576 octets pour Internet. Si un fragment est perdu, tout le paquet UDP est perdu. Donc, plus le paquet est grand, plus la probabilité de sa perte augmente.

Pour atteindre un maximum d'efficacité, il est alors nécessaire que la taille du paquet UDP avoisine la taille des MTU. De cette façon, nous diminuons le coût de calcul généré par les opérations de fragmentation et de réassemblage.

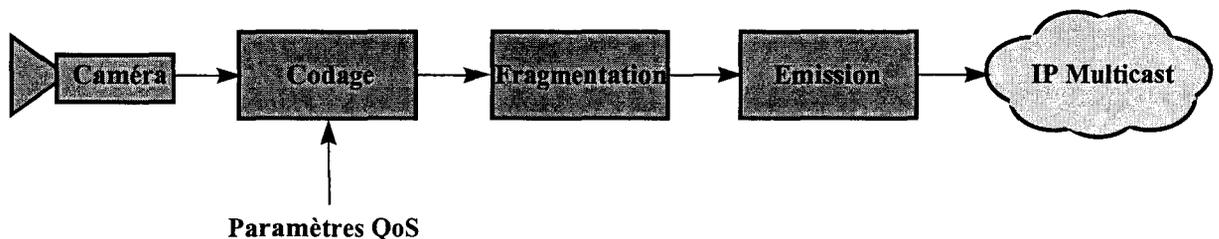


Figure 6-6 : Chaîne d'émission vidéo.

La Figure 6-6 représente la chaîne d'émission de la vidéo. Après l'acquisition, la vidéo est codée en tenant compte des paramètres QoS Vidéo déterminés. Avant l'émission de la vidéo, une opération de fragmentation est effectuée. Cette opération consiste à diviser l'image en plusieurs paquets, permettant ainsi leur transport avec UDP. Le découpage en paquet permet, s'il y a perte de paquets, de ne pas perdre toute une image par exemple.

---

<sup>37</sup> MTU : Maximal Transmission Unit.

## 6.7.2. Réception vidéo

La chaîne de réception de la Figure 6-7 est effectuée pour chaque site distant figurant dans la présentation vidéo. Dans notre implémentation, nous avons affecté un thread à chaque site distant pour le traitement de la chaîne de réception. Des priorités sur l'ordonnancement des threads peuvent être établies en fonction des priorités déterminées pour la réception des participants correspondants.



Figure 6-7 : Chaîne de réception vidéo.

Les paquets reçus sont rassemblés dans l'ordre de leur numéro de séquence enregistré dans l'en-tête du paquet. L'image compressée ainsi obtenue passe dans le décodeur adéquat. Plusieurs décodeurs peuvent être présent sur le système local. Enfin, l'image décompressée est affichée dans la présentation du participant.

## 6.8. Choix du langage d'implémentation

Pour le développement de nos travaux, deux langages avaient retenu notre attention, le C++ un langage ayant donné ses preuves, et Java un langage plein de promesses.

### 6.8.1. Le langage retenu : le C++

Pour le développement de nos travaux, le langage C++ [Stroustrup 86] a été retenu pour deux raisons principales. D'une part, parce qu'il s'interface naturellement avec le langage C [Kernighan 78], et d'autre part, c'est un langage orienté objet.

Le langage C++ est un sur-ensemble du langage C, gardant ainsi une compatibilité avec ce dernier, et bénéficiant de ses points forts en tant que langage performant et « proche du matériel ». Nous avons besoin de ces avantages, pour la manipulation des données vidéo au niveau de « l'octet » pour le traitement et le transport. D'ailleurs, le langage C a été baptisé « langage de programmation système » car il sert beaucoup à écrire des compilateurs et des

systèmes d'exploitation, dont le système UNIX qui fut à l'origine du langage. Il est souvent choisi pour le développement de programmes nécessitant des accès directs au matériel, puisqu'il permet la manipulation d'objets matériels comme les adresses mémoire.

La facilité du langage C à accéder au matériel, a fait que la plupart des constructeurs de périphériques multimédia, comme les cartes d'acquisition sonore ou vidéo, fournissent des kits de développement sous forme de bibliothèques de fonctions en C. Ceci permet ainsi l'exploitation de ces périphériques par des programmeurs d'applications spécifiques. Or, nous avons besoin de ces bibliothèques pour le pilotage des cartes d'acquisition vidéo, comme la spécification du format vidéo en fonction de la QoS désirée par exemple.

De plus, le C++ est un langage orienté objet. Il ajoute au langage C le concept de classe, donnant aux programmeurs les avantages de la programmation objet comme l'abstraction de données et la réutilisation du code grâce au mécanisme d'héritage.

Les langages orientés objet nous permettent de réfléchir à une solution en utilisant des concepts dans le domaine du problème. Dans notre cas, nous avons conçu notre système en utilisant des classes du domaine des activités coopératives comme le Participant, le Rôle et l'Outil.

### **6.8.2. Le langage Java, l'autre candidat**

Au début de nos travaux, le langage Java commençait à émerger et faisait l'objet de plusieurs études et d'articles dans la communauté informatique [Gosling 95, Flanagan 97, Eckel 98]. Nous ne pouvions pas rester indifférent à ce langage. C'est un langage orienté objet livré avec une librairie de classes. Ceci permet de diminuer la charge de travail des développeurs d'applications et de standardiser leurs développements.

Une des caractéristiques du langage qui nous intéressait est son orientation vers le réseau en fournissant un ensemble de classes (java.net) pour la communication entre applications. De plus, il s'intègre naturellement dans les navigateurs Web grâce aux applets Java. L'application coopérative peut ainsi être téléchargée sous forme d'applet à partir d'un serveur au moment où le participant se joint à la session d'activité coopérative [Shirmohammadi 98]. Le bénéfice est que l'utilisateur n'a pas à se soucier de l'installation de l'application, ni à se préoccuper de sa mise à jour puisqu'il est certain de posséder la dernière version en la téléchargeant du serveur référençant la session.

L'autre avantage de Java qui nous intéressait, est son indépendance vis-à-vis de la plate-forme matérielle. Un programme Java est compilé pour obtenir du « p-code ». Ce dernier est exécuté sur une machine virtuelle qui l'interprète et le transforme en code de la machine cible. Une implémentation de la machine virtuelle spécifique à chaque plate-forme matérielle permet alors à un programme Java d'être exécuté sur différentes plates-formes (Figure 6-8).

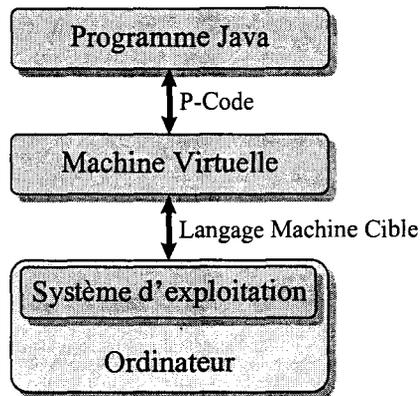


Figure 6-8 : Principe de la machine virtuelle Java.

Cependant, malgré ses nombreux avantages, le langage Java souffrait de certains problèmes de « jeunesse » qui ont fait qu'il n'a pas été retenu pour le développement de nos travaux.

Pour le transport des flux vidéo, nous avons besoin d'un langage pouvant répondre aux exigences temps réel des médias continus. Le langage Java présentait des déficiences au niveau des performances qui sont dus à l'interprétation du p-code par la machine virtuelle. Comme le C++ est un langage compilé, ses performances sont supérieures à celle de Java.

L'autre handicap de Java, est son incapacité à accéder au matériel. Dans notre cas, cela veut dire que nous ne pouvions pas piloter une carte d'acquisition vidéo avec le langage Java. La raison de cette limite est que la machine virtuelle n'implémentait pas des classes permettant l'accès aux périphériques multimédia.

La solution pour contourner ce problème, était d'utiliser les méthodes natives NMI<sup>38</sup> qui permettent à un programme Java de faire appel à des fonctions écrites avec un langage comme

<sup>38</sup> NMI : Native Method Interface. Cette interface de programmation a évolué aujourd'hui vers JNI pour Java Native Interface.

le C par exemple. Pour nos besoins, le pilotage de la carte vidéo s'effectue avec le programme développé en C par exemple, mais qui est dépendant de la plate-forme. Ceci fait perdre l'avantage de l'indépendance de Java vis-à-vis de la plate-forme.

Aujourd'hui, le langage Java a beaucoup évolué. Des améliorations ont été portées sur la machine virtuelle pour accroître ses performances. De plus, l'utilisation des techniques de compilation à la volée JIT<sup>39</sup> pour transformer les programmes Java en code directement exécutable par la machine, améliore très sensiblement les performances. Avec la compilation JIT, les performances de Java s'approchent de celle du C++ [Mangione 98].

Une autre amélioration du langage consiste dans le rajout d'un nouvel ensemble de classes : les Java Media Framework (JMF). Ces classes permettent à un programme Java de capturer et de jouer des flux multimédia, sans passer par un autre langage par le biais de l'interface de programmation JNI. Actuellement, la version JMF 1.0 ne permet pas la capture audio et vidéo. Ces fonctionnalités seront incluses dans la version JMF 2.0 qui sera disponible dans le deuxième semestre 1999.

Les améliorations apportés au langage Java tant sur le point des performances que sur les fonctionnalités pour la manipulation des données multimédia, nous permettent aujourd'hui d'envisager de porter nos développement sur Java.

---

<sup>39</sup> JIT : Just In Time.

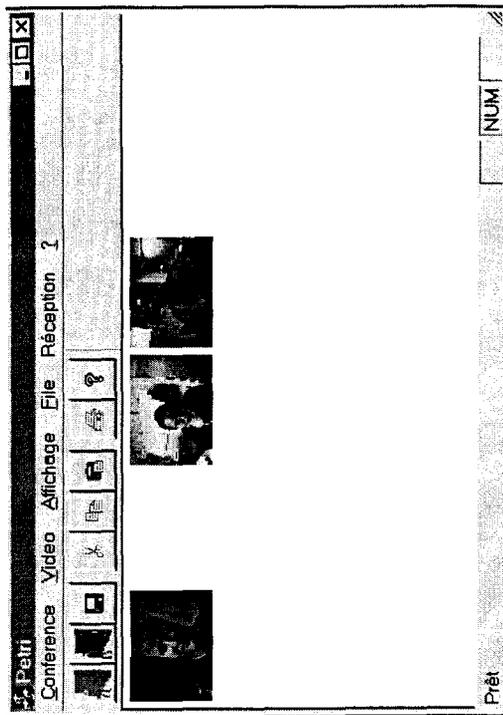
## **6.9. Exemple de Présentation vidéo**

Pour donner un exemple de présentation vidéo, nous reprenons le scénario du cours à distance vu dans le chapitre 3. Après une courte phase d'initialisation, Stan (le Professeur) commence à dispenser son cours. Après un certain temps, il est interrompu par Petri qui veut poser une question. Stan lui répond, ensuite poursuit le cours.

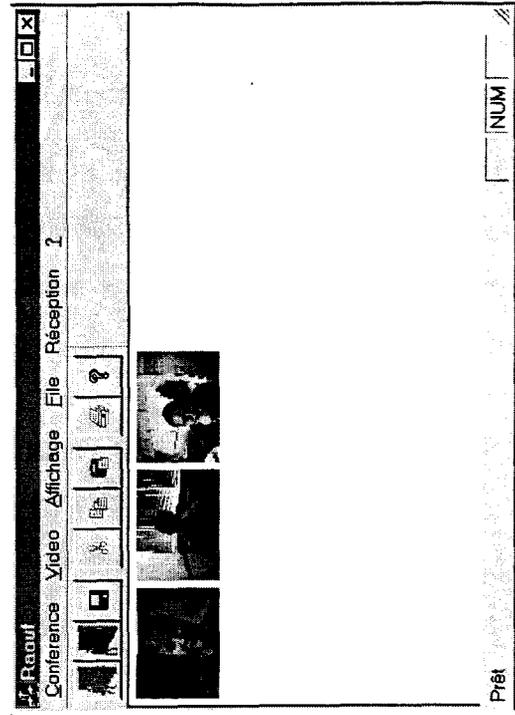
Dans la phase d'initialisation, les participants n'ont pas d'intérêt particulier pour leur interlocuteurs. Les vidéos sont reçues avec la même qualité vidéo. Ensuite, suivant le déroulement du cours, des qualités vidéo sont attribuées en fonction des intérêts des participants. Pour cette exemple, nous avons agit sur la résolution spatiale pour obtenir trois qualités vidéo.

Dans la présentation vidéo, nous avons inclus seulement les vidéos des sites distants. La vidéo du participant local ne figure pas dans la présentation vidéo. Nous pouvons éventuellement la rajouter, notamment pour lui permettre de bien se positionner pour rester dans le champ capté par la caméra. La décision de faire figurer la vidéo locale dans la présentation vidéo pourra dépendre de l'encombrement de l'écran par exemple.

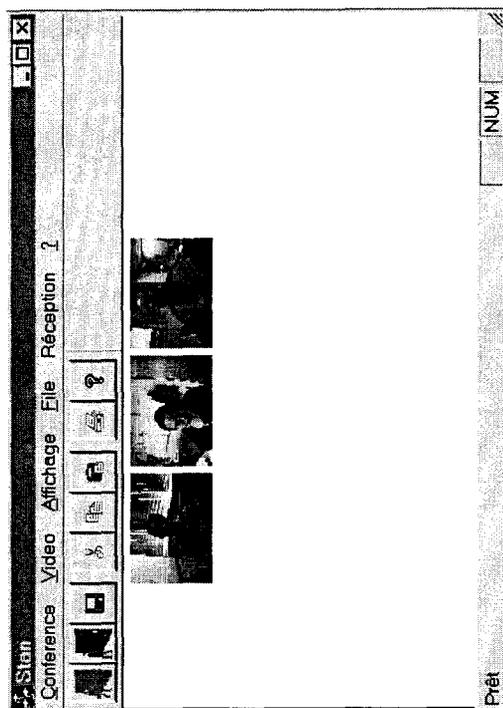
Dans les pages suivantes, les présentations vidéo de chaque participant sont représentées à chaque phase du scénario.



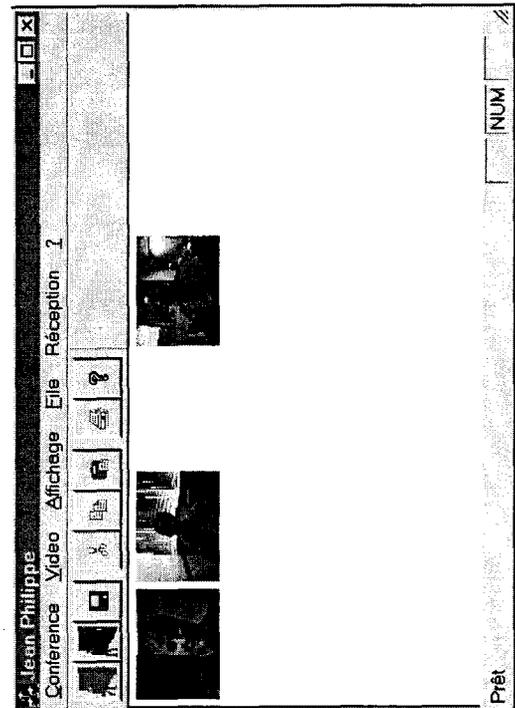
Présentation Vidéo de Petri



Présentation Vidéo de Raouf



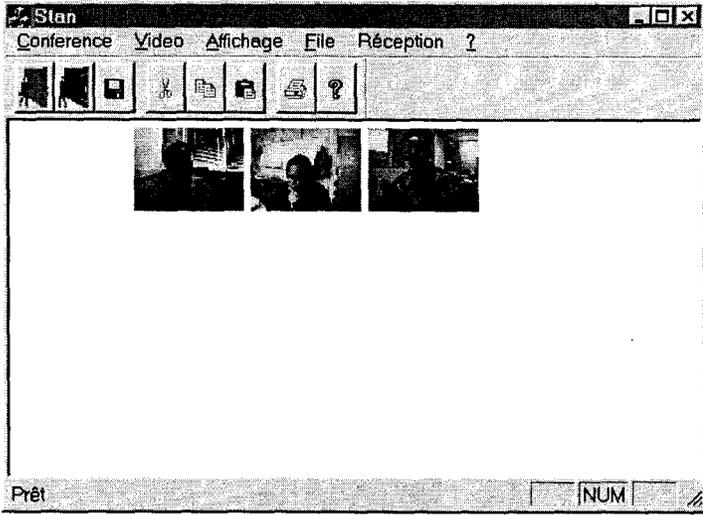
Présentation Vidéo de Stan



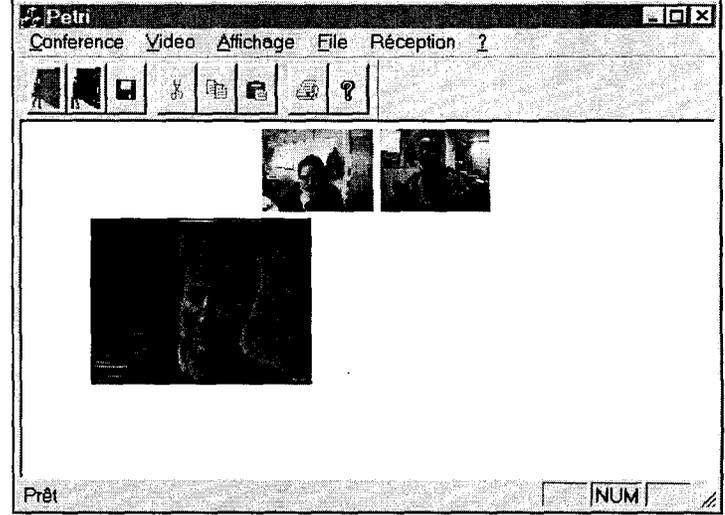
Présentation Vidéo de Jean Philippe

**Phase d'initialisation** : le cours n'a pas encore commencé. Les participants n'ont pas d'intérêts particuliers à regarder leurs correspondants dans les sites distants.

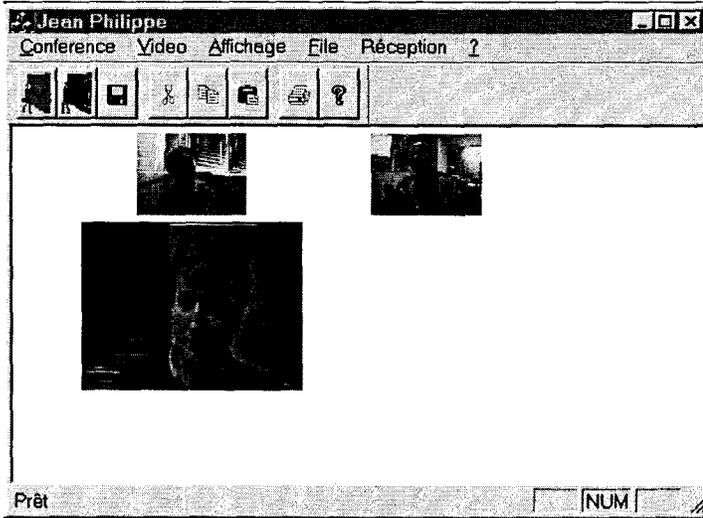
Phase 1 : Stan (le Professeur) explique le cours.



Présentation Vidéo de Stan



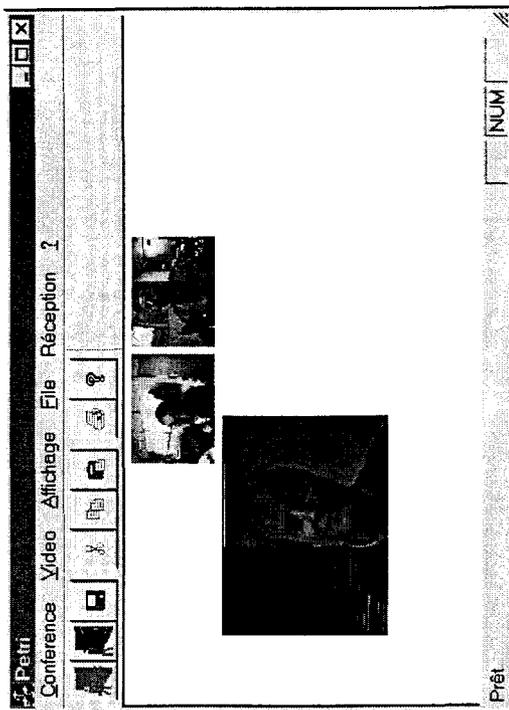
Présentation Vidéo de Petri



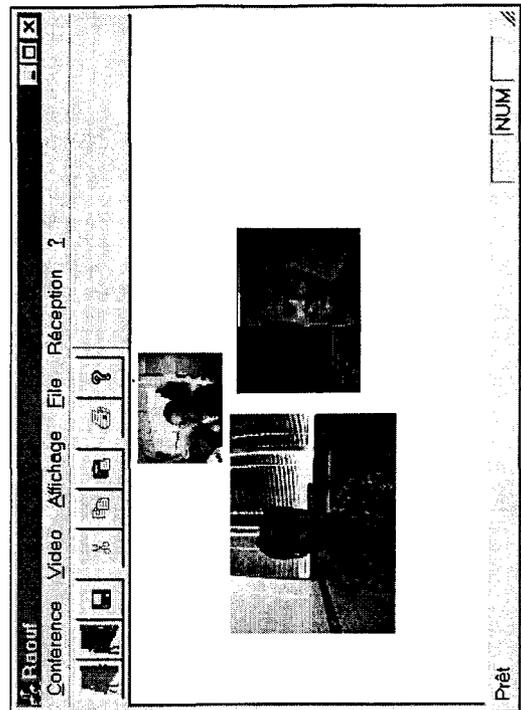
Présentation Vidéo de Jean Philippe



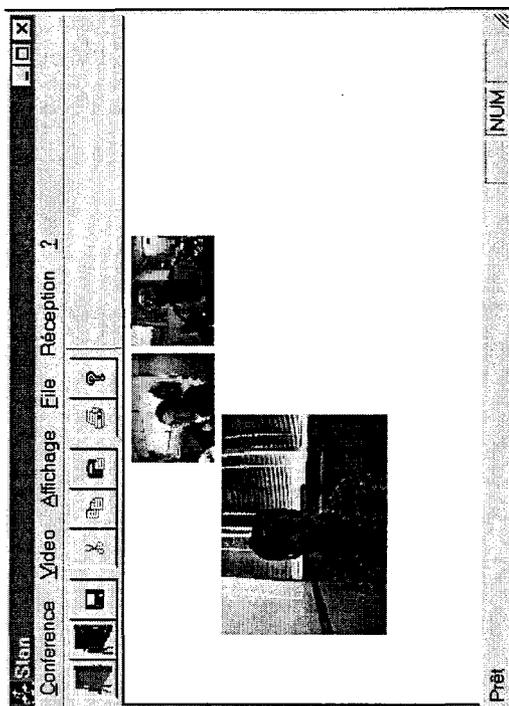
Présentation Vidéo de Raouf



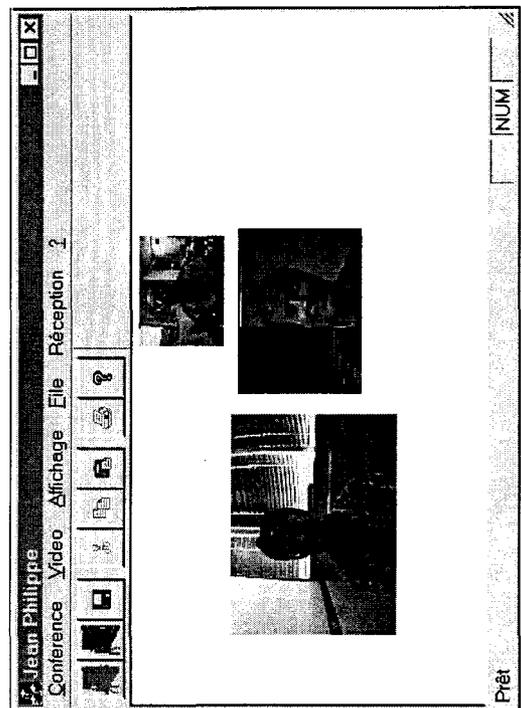
Présentation Vidéo de Petri



Présentation Vidéo de Raouf



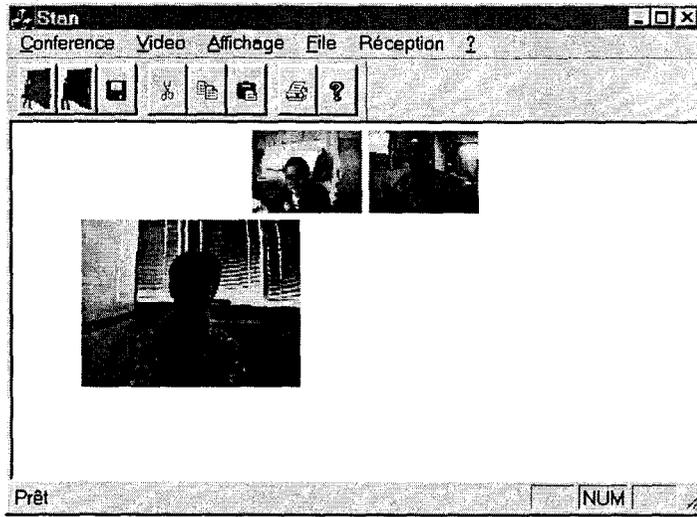
Présentation Vidéo de Stan



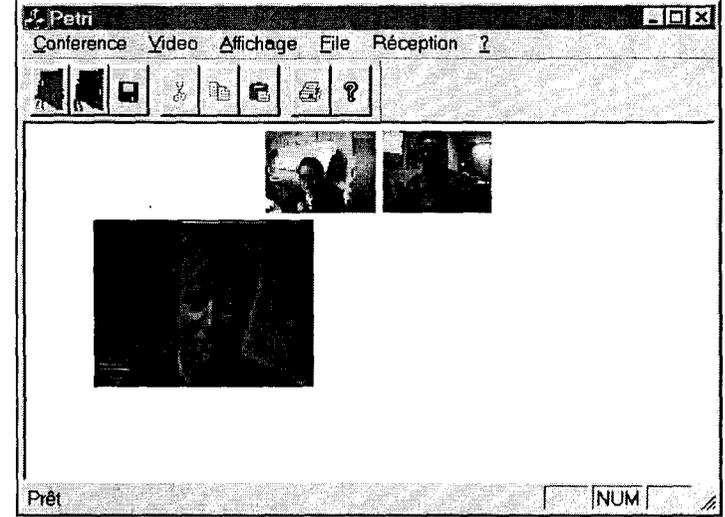
Présentation Vidéo de Jean Philippe

**Phase 2** : Petri pose une question. Une meilleure qualité vidéo lui est alors attribuée dans les présentations vidéo des autres.

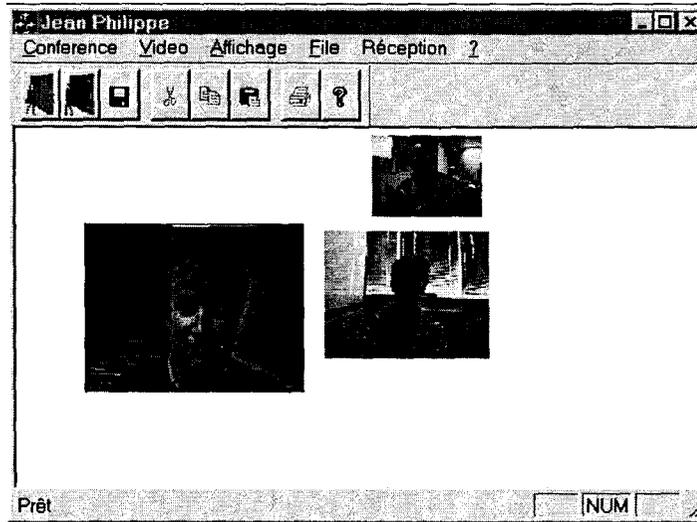
Phase 3 : Stan répond à Petri.



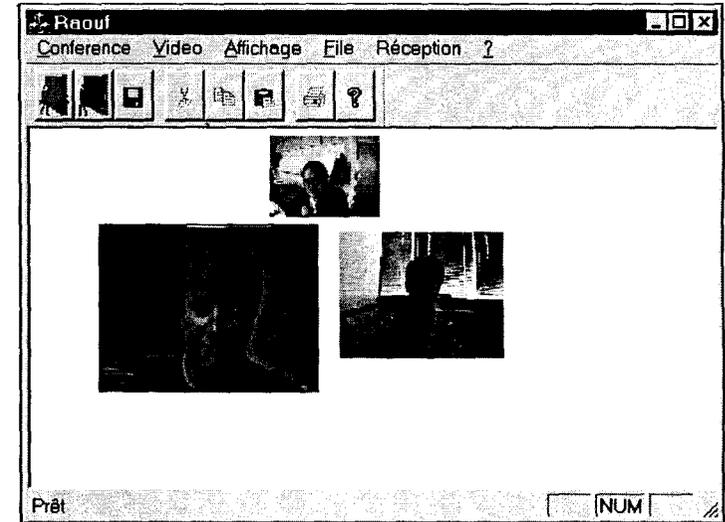
Présentation Vidéo de Stan



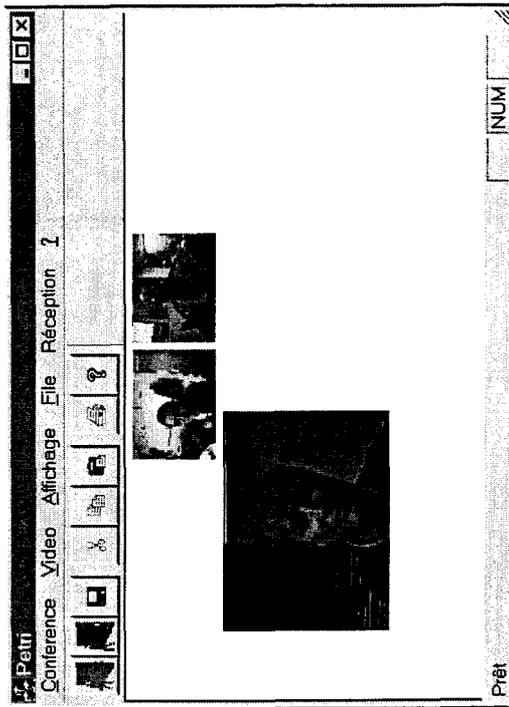
Présentation Vidéo de Petri



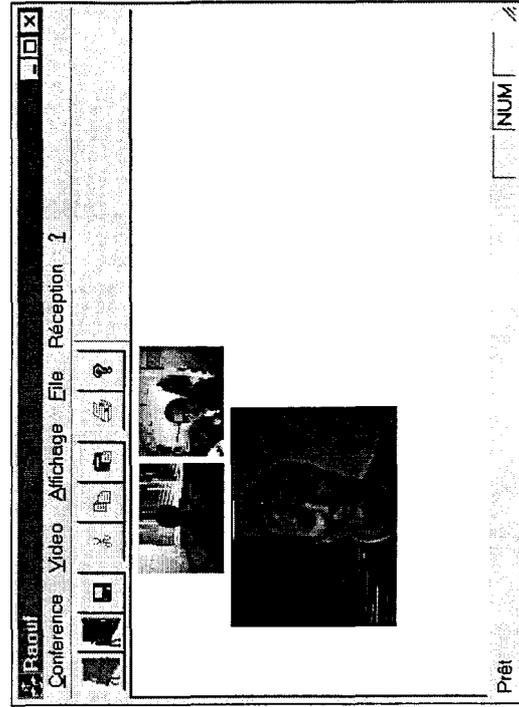
Présentation Vidéo de Jean Philippe



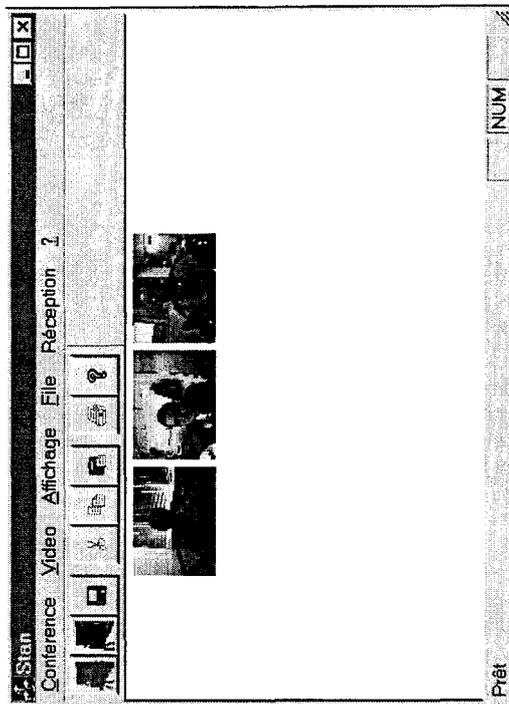
Présentation Vidéo de Raouf



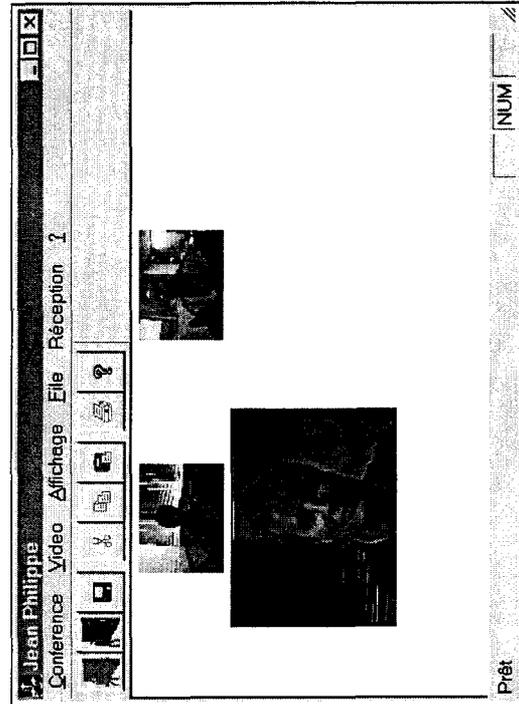
Présentation Vidéo de Petri



Présentation Vidéo de Raouf



Présentation Vidéo de Stan



Présentation Vidéo de Jean Philippe

**Phase 4** : Stan poursuit l'explication du cours. L'état de Petri redevient Auditeur.

## 6.10. Validation fonctionnelle

Pour valider les développements effectués, nous avons expérimenté nos travaux en réalisant une maquette de test. Les objectifs que nous nous sommes fixés étant la gestion de la QoS vidéo et l'adaptation à différentes activités coopératives.

### 6.10.1. Gestion de la QoS Vidéo

La gestion de la QoS vidéo se situe à deux niveaux :

- Au niveau du récepteur, lui permettant une gestion vidéo en fonction de ses intérêts vis-à-vis de ses correspondants ;
- Au niveau de l'émetteur, lui permettant une gestion vidéo en fonction des différentes requêtes QoS vidéo qu'il peut recevoir.

Pour offrir des qualité vidéo différentes pour les utilisateurs, notre maquette doit gérer le codage en couche. Pour l'acquisition de la vidéo, nous avons utilisé des cartes miro DC 20<sup>40</sup>. Ces cartes permettent de faire de la compression matérielle au standard M-JPEG (voir Annexe I). Dans ce format, chaque image d'une séquence vidéo est encodée indépendamment des autres au format JPEG. On parle alors de codage intra-image. Pour l'obtention de différentes qualité vidéo, nous avons opté pour un codage en couches temporel (Figure 6-9).

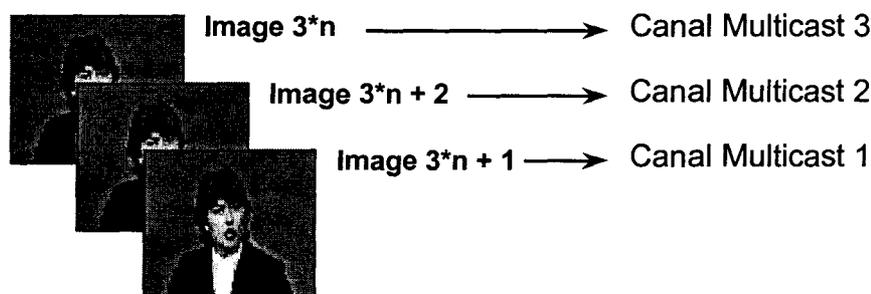


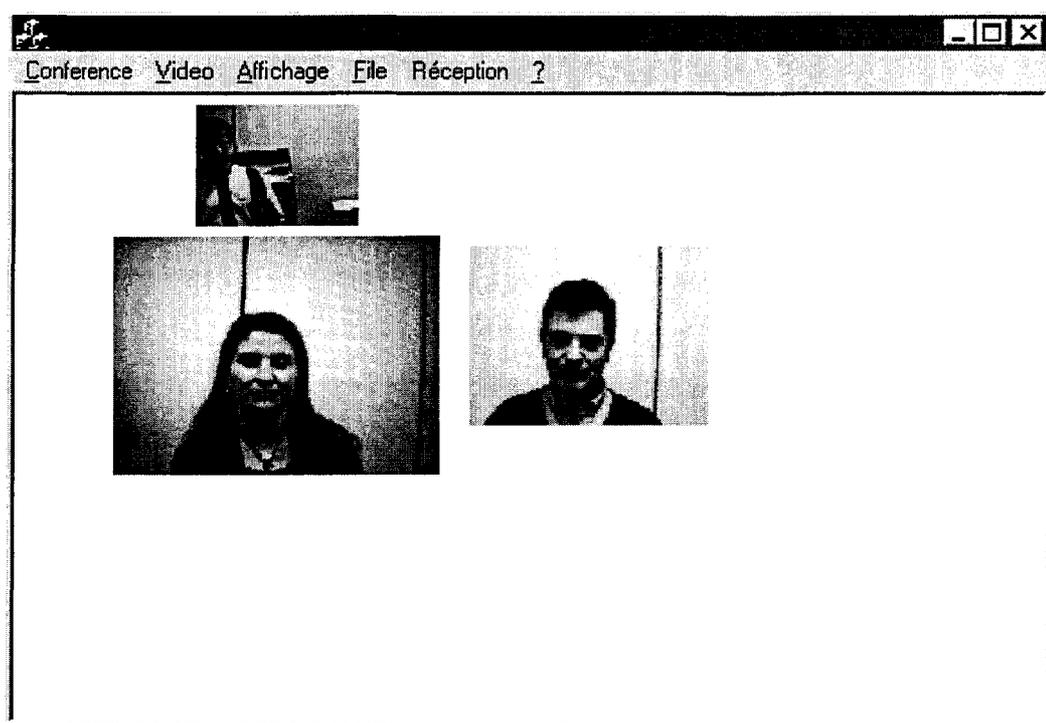
Figure 6-9 : Codage en couche temporel.

Dans nos tests, pour obtenir du codage en couches temporel, nous avons séparé le flux vidéo émis en trois couches. Trois images successives sont émises sur trois canaux multicast

<sup>40</sup> <http://www.miro.com>

différents. Ainsi, le récepteur pourra obtenir trois niveaux de qualité vidéo en s'abonnant respectivement à un, deux ou trois canaux multicast.

En plus du codage en couche temporel, nous agissons à la réception sur les tailles des vidéos. L'allocation de l'espace écran pour l'affichage d'une vidéo s'effectue en fonction de son degré d'intérêt pour le récepteur. Nous avons alors établi trois tailles vidéo allant de la plus petite à la plus grande. Les tailles petite, moyenne et grande sont associées respectivement aux vidéo reçues sur un, deux et trois canaux multicast. La Figure 6-10 montre un exemple de réception vidéo après la combinaison des actions spatiale et temporelle sur les vidéos.



*Figure 6-10 : Combinaison Spatio-temporelle pour la détermination d'une QoS Vidéo.*

### **6.10.2. Adaptation à différentes activités coopératives**

Nous avons testé notre prototype pour deux types d'activité coopérative, pour des conversations de groupes et pour des cours à distance.

### 6.10.2.1. Conversation de groupe

Pour l'activité de conversation de groupe, nous avons couplé notre prototype de gestion vidéo avec un pont audio analogique permettant de transmettre les voix des participants. Ces derniers ont le même rôle, celui de *Conférencier*.

Etant donné que tous les participants ont le même rôle, celui de Conférencier, les règles que nous avons établies dépendent seulement des outils utilisés dans la session de coopération. Comme l'activité est une conversation de groupe, les règles pour la gestion de la présentation vidéo des participants dépendent exclusivement de la possession de la parole.

La règle ainsi établie est la suivante :

*(\* , Locuteur) > (\* , Locuteur Précédent) > (\* , Auditeur)*

Nous donnons donc comme prioritaire le participant qui parle à un instant donné, avec une profondeur d'historique égale à un, correspondant au *Locuteur Précédent*.

### 6.10.2.2. Cours à distance

Dans cette activité, nous avons couplé notre prototype vidéo avec TuttelCast, un outil de diffusion de page HTML sur Internet (Figure 6-11).

TuttelCast est une application développée à l'ENIC pour répondre à des besoins d'enseignement à distance [Colin 98]. Comme pour un cours basé sur des transparents, le professeur affiche des pages HTML successives en les commentant. La page commentée à un instant donné est automatiquement diffusée à tous les sites Elèves.

De la même manière que pour l'activité de conversation de groupe, un pont audio analogique est utilisé pour permettre au professeur et aux élèves de parler et de s'entendre.

Pour les règles de gestion vidéo, nous avons utilisé les mêmes règles que celles énoncées dans le paragraphe 4.7.2, à savoir :

La règle pour le professeur est :

*(\* , Locuteur) > (\* , Locuteur Précédent) > (\* , Auditeur)*

Et la règle pour les élèves est :

(\*, Locuteur) > (Professeur, \*) > (Elève, Locuteur Précédent) > (Elève, Auditeur)

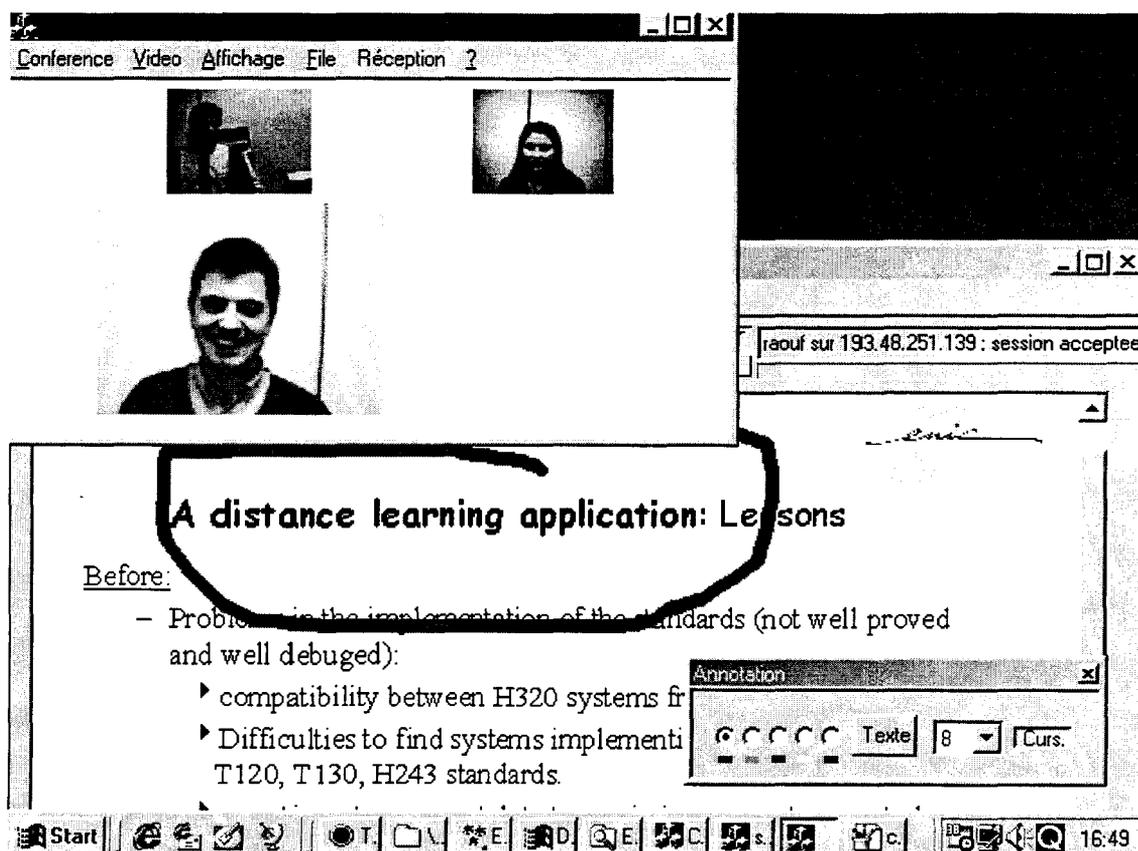


Figure 6-11 : Couplage avec TutelCast.

### 6.10.3. Résultats

Les expériences effectuées avec notre prototype vidéo ont été réalisées en réseau local au sein de notre laboratoire de l'ENIC. Les utilisateurs du prototype ont apprécié la gestion vidéo offrant différentes qualités en fonction des rôles et des états des participants. En particulier, cette gestion vidéo avec différentes qualités permet d'attirer l'attention des participants dès qu'un changement d'état d'un participant se produit. Par exemple, le passage de la parole d'un participant à un autre est immédiatement remarqué, alors que dans les systèmes traditionnels dans lesquels les vidéos reçues ont la même qualité, les participants doivent détecter eux mêmes les changements de locuteur. Ceci peut introduire un délai entre le moment du changement et celui de sa prise en compte par les participants, pouvant ainsi être source de perturbation pour le bon déroulement de la session de travail. Ce délai est principalement fonction de la qualité vidéo reçue. Il croît lorsque la qualité vidéo diminue.

Les personnes habituées à utiliser TuttelCast sans vidéo, trouvent que le couplage avec notre prototype vidéo permet de donner une conscience de groupe en créant « une salle de cours virtuelle ». En effet, notre système vidéo permet à un élève de voir le professeur, ainsi que les autres élèves. Ce point a été évoqué lors d'une enquête réalisée pour l'évaluation du système VisioCentre de l'ENIC [Hoogendoorn 98]. Les élèves regrettaient de ne pas voir les élèves des autres sites. Ils n'avaient pas conscience de leur présence, sauf au moment des questions où la vidéo du professeur est remplacée par celle de l'élève qui pose la question.

Avec notre système de gestion vidéo, les participants sont en permanence conscients de la présence des membres du groupe.

## **6.11. Conclusion**

Le système que nous avons réalisé est composé de deux applications. Le serveur de session et le participant. Le serveur permet de référencer les sessions d'activités coopératives en cours, et gère les notifications des événements pour les participants. L'application participant est installée sur le site participant. Elle est scindée en deux parties séparant les aspects communs d'une activité, des aspects spécifiques. Pour pouvoir adapter le système à différents type d'activités coopératives, nous avons adopté le principe des Frameworks, permettant à un participant de récupérer dynamiquement un composant activité encapsulant les aspects spécifiques de l'activité.

Au niveau implémentation des services vidéo, nous avons utilisé la pile de protocoles réseau la plus répandue actuellement grâce au réseau Internet, la pile TCP/IP. Nous avons choisi le protocole de transport UDP au détriment du protocole TCP, d'une part, parce qu'il est plus léger du fait qu'il n'effectue pas de contrôles d'erreurs sur les pertes de paquets, et d'autre part, parce qu'il permet la transmission de données en mode multicast, réalisant ainsi un gain considérable de bande passante réseau.

Les expériences réalisées avec notre prototype vidéo, ont montré l'apport de nos propositions pour la gestion de la présentation vidéo dans les sessions d'activités coopératives. Ces expériences ont validé la nécessité d'attribuer des QoS vidéo différentes pour les vidéos reçues. En effet, notre prototype permet d'attirer les regards des utilisateurs dès qu'un changement se réalise pendant la session de travail. Par exemple, à un changement

de Locuteur, l'attention des participants est captée, en attribuant une meilleure qualité vidéo pour le Locuteur.

Dans nos travaux, nous nous sommes focalisé sur la partie présentation vidéo. Le composant activité que nous avons présenté est une délégation pour la partie vidéo. Le composant activité doit être complété pour prendre en considération les aspects de travail coopératif, comme la gestion de persistance des objets partagés.

---

## **7. Conclusion et perspectives**

---

**D**ans les travaux présentés dans cette thèse, nous avons abordé le problème de la gestion de la présentation vidéo des participants dans les activités coopératives.

Plus précisément, nous avons proposé un système permettant d'assister le participant dans la composition de sa présentation. Les limites des ressources (écran, processeur, réseau) ne permettent pas à l'utilisateur de recevoir toutes les vidéos des sites distants, surtout lorsqu'un grand nombre de participants sont impliqués dans l'activité.

Pour cela nous avons proposé un modèle permettant une *gestion personnalisée de la présentation vidéo*. Ce modèle détermine les intérêts des participants pour voir leurs interlocuteurs afin de composer la présentation vidéo. Il utilise des règles qui déterminent les intérêts à partir des rôles et les états des participants. Les intérêts sont présentés sous forme de priorités entre participants.

Une fois les intérêts déterminés, le système doit établir les liaisons vidéo correspondantes. La qualité vidéo de chaque site sera proportionnelle à la priorité qui lui est accordée.

Nous avons été confrontés au problème de réception de requêtes QoS différentes au niveau de l'émetteur. Le fait que l'on se trouve dans un milieu hétérogène, et que les participants peuvent avoir des intérêts différents, font qu'un site pris du point de vue émetteur va recevoir des requêtes de QoS vidéo différentes. Pour résoudre ce problème, nous avons pris en compte les aspects de l'activité coopérative en pondérant les requêtes avec le degré d'intérêt et le degré d'importance du participant demandeur. L'originalité de notre proposition, est de faire *remonter les décisions du choix de la QoS vidéo du niveau système au niveau utilisateur*.

Dans nos travaux, nous avons également abordé le problème de la diversité des activités coopératives. Comme chaque activité coopérative possède ses propres caractéristiques en termes de rôles et d'outils utilisés, la gestion de la présentation vidéo sera différente pour chaque activité. Afin d'éviter qu'un utilisateur n'installe autant d'applications sur son système que d'activités coopératives dans lesquelles il peut participer, nous offrons *une architecture*

---

*ouverte, adaptable au type d'activité coopérative.* Notre architecture utilise le principe des Frameworks et des composants connectables. Ce dernier va encapsuler toutes les spécificités de l'activité, alors que le Framework comprend les parties communes à toutes les activités.

En termes de perspectives à nos travaux, nous prévoyons un certain nombre d'améliorations à rajouter à notre prototype de gestion vidéo, ainsi que de son intégration à d'autres systèmes utilisés à l'ENIC et à TRIGONE.

### ***Validation d'usage***

Nous avons expérimenté notre prototype de gestion vidéo au sein de notre laboratoire pour valider les fonctionnalités offertes par le système. Ses expériences peuvent être considérées comme étant effectuées dans des conditions « idéales », vu que nous avons utilisé notre réseau local pour le transfert de la vidéo entre les participants.

Nous prévoyons donc une exploitation réelle de notre prototype vidéo, avec un groupe d'étudiants inscrits dans le cadre de TutTelVisio®, la formation à distance dispensée par l'ENIC.

Dans l'expérience, les élèves suivront des cours à distance, au moyen de l'outil TuttelCast pour la diffusion de transparents, d'un pont audio pour la parole, et de notre prototype de gestion vidéo pour l'image. Les élèves auront principalement des accès modem.

Nous sommes en train de préparer des questionnaires qui seront distribués aux Professeurs et aux Elèves pour :

- Donner leur avis sur l'apport de la vidéo par rapport à l'audio seul ;
- Exprimer leur ressenti à la différence de la qualité vidéo.

Les résultats de cette expérience permettront une validation d'usage de notre maquette de gestion de la QoS vidéo.

### ***Analyse de l'activité***

Dans notre proposition, nous n'avons pas abordé le problème de la détermination des pondérations qui correspondent aux degrés de priorité ou aux degrés d'importance des participants. Ces valeurs sont spécifiques à chaque activité. Elles peuvent être déterminées après une analyse de l'activité considérée. Des travaux sur l'analyse des activités coopératives sont en cours au laboratoire Trigone. Les résultats de ces travaux vont nous aider à mieux évaluer notre proposition.

### ***Site multi-utilisateur***

Dans notre proposition, nous avons implicitement fait correspondre un participant à un site. Nous prévoyons d'affiner notre proposition pour qu'elle prenne en considération les sites multi-utilisateurs. Assister la présentation vidéo pour les systèmes salle de conférence, est une tâche encore plus difficile. Un tel système peut accueillir plusieurs participants. Ces participants peuvent avoir des rôles différents, et donc des intérêts différents en ce qui concerne les vidéos à recevoir.

Comme ces participants se partagent les mêmes dispositifs de visualisation, le problème est de choisir en fonction de quels participants dans la salle sera produite la présentation vidéo.

Le même problème peut se poser pour choisir quelle vidéo sera émise dans un site salle de conférence, puisque les caméras sont partagées.

Le problème sera moins difficile à résoudre si la salle de conférence comprend plusieurs écrans de visualisation et plusieurs caméras.

Une solution possible consiste à établir une hiérarchie entre les (Rôle, Etat) des participants dans un même site, et c'est le rôle qui se trouve au sommet de la hiérarchie (à un moment donné) qui imposera ses choix.

### ***Enrichir Scénic***

Scénic est un agent responsable de la gestion spatio-temporelle des interfaces graphiques dans les systèmes de travail coopératif [Carlier 98]. Basé sur la métaphore du théâtre, il permet la composition de la présentation des utilisateurs. Cependant, Scénic ne gère pas la notion de QoS vidéo. Il agit localement au site utilisateur pour assembler les éléments constituant l'interface graphique.

Nous envisageons d'enrichir Scénic, en lui apportant notre connaissance de la Qualité de Service. L'intégration de notre modèle de gestion de QoS vidéo à Scénic, permettra de faire apparaître la notion de QoS dans la présentation réalisée par Scénic, tout en laissant transparent les aspects de négociation de la QoS avec le réseau et les sites distants.

### ***Evolution du VisioCentre***

Actuellement, le VisioCentre s'appuie sur une architecture utilisant le Réseau RNIS pour le transport des données. Pendant un cours, les sites Elèves et Professeur sont reliés en mode multipoint à travers un pont Multipoint. Pour joindre un cours, chaque site doit effectuer une connexion avec le pont. Cette architecture n'est pas flexible, puisque RNIS ne permet pas à

---

des sites hétérogènes d'établir des liaisons entre eux, sans obliger tous les sites à s'aligner sur le site ayant l'accès réseau le plus faible.

Des études ont débuté pour faire évoluer le système VisioCentre, afin qu'il puisse prendre en compte l'hétérogénéité des utilisateurs, aussi bien des sites amphithéâtres que des sites individuels. Pour cela, la couche communication du VisioCentre doit être modifiée pour utiliser la pile de protocole TCP/IP. Dans ce cadre, nous préconisons l'utilisation de notre architecture de gestion de la QoS vidéo, qui permettra à chaque participant, Professeur ou Elève, de bénéficier d'une présentation vidéo compatible avec ses intérêts et ses ressources.



---

# 8. Annexe I : la Compression

## Vidéo

---

Dans le domaine de la compression d'image, nous pouvons distinguer la compression d'image fixe et la compression vidéo. Les méthodes de compression exploitent les redondances existantes dans les images. Les régions voisines d'une image étant généralement fortement corrélées, les algorithmes de compression utilisent cette redondance spatiale pour réduire la quantité d'information pour coder l'image.

En plus de la redondance spatiale, les algorithmes de compression vidéo exploitent la redondance temporelle existante dans une séquence vidéo. Des techniques, comme la détection et la compensation de mouvement permettent de coder seulement les différences entre les images successives d'une séquence vidéo. Ce type de codage est dit inter-images (interframe). Il permet de réduire considérablement la taille du flux vidéo. Cependant, ces techniques de compression sont parfois très complexes, et demandent une grande puissance de calcul.

Avec la prolifération des algorithmes de compression, des organismes comme l'IUT et l'ISO ont établi certains standards de compression, pour limiter les risques d'incompatibilité. Chaque standard de compression est spécifique à un certain type d'application. Certaines applications ont besoin d'un codage temps réel au détriment du facteur de compression comme la vidéoconférence, alors que d'autres ont besoin d'un maximum de compression alors que l'encodage n'est pas fait en temps réel comme la VOD (video on demand) ou les applications sur CD-ROM.

Dans cet annexe, nous décrivons différentes techniques de compression, ainsi que les standards qui les utilisent.

## **8.1. Les techniques de compression**

Les techniques de compression peuvent être classés en deux classes, les techniques sans perte et les techniques avec perte.

Les techniques de compression sans perte (compactage) permettent de réduire le volume d'information d'un objet sans aucune perte. Après décompression, ces techniques permettent de retrouver exactement le même objet original avant compression. Elles sont généralement utilisées pour les applications où aucune perte n'est tolérée comme pour la compression des fichiers exécutables ou textes. Elles présentent comme avantages une simplicité d'implémentation et ne demandent pas une grande puissance de calcul. Par contre, elles ne permettent pas d'atteindre des facteurs de compression très élevés. Les codages RLE et Huffman sont des techniques de compression sans pertes.

Pour les techniques de compression avec perte, une partie de l'information est perdue au cours du codage. Ces techniques sont beaucoup utilisées dans la compression d'images fixes et de séquences vidéo, où l'on exploite les limites du système visuel humain pour distinguer certaines couleurs. Parmi ces techniques avec pertes, il y a la quantification de coefficients DCT et la compensation de mouvement qui permettent de réduire considérablement la quantité d'information pour stocker des images fixes ou des séquences vidéo, sans pour cela que l'œil humain s'en aperçoive. Cependant, ces techniques sont très gourmandes en puissance de calcul, et bien souvent, leur mise en œuvre est faite avec des circuits spécialisés, tels que ceux développés par Intel ou C-Cube.

### **8.1.1. Le codage RLE**

Parmi les techniques de base de compression les plus utilisées, on trouve le Codage RLE ("Run Length Encoding"). C'est le plus simple des algorithmes de compression. Son principe consiste à détecter une donnée ayant un nombre d'apparitions consécutives qui dépasse un seuil fixé

Ce codage est intéressant pour des données comportant peu de valeurs différentes, et de longues séquences, mais est peu intéressant pour des données textuelles ou de type image

photographiques ayant de nombreuses valeurs et de courtes séquences. Les formats d'image comme PCX et TIFF utilisent le codage RLE.

### 8.1.2. Le codage de Huffman

Le codage de Huffman est quant à lui une méthode de compression statistique de données qui permet de réduire la longueur moyenne du codage d'un alphabet.

Le principe de l'algorithme de Huffman consiste à recoder les octets rencontrés dans un ensemble de données sources comme un fichier ou une image, avec des valeurs de longueurs binaires variables. Huffman propose de recoder les données qui ont une occurrence très faible sur une longueur binaire supérieure à la moyenne, et recoder les données très fréquentes sur une longueur binaire très courte.

Ce codage est très efficace pour des données textuelles ou pour des images monochromes.

### 8.1.3. La transformée DCT

Une transformée est une opération permettant de transformer un ensemble de données d'un domaine source vers un autre domaine destination. En compression, la motivation d'utiliser des transformées est de pouvoir représenter l'information de manière plus compacte dans le domaine destination. Ceci permettra ensuite d'éliminer les coefficients qui ne stockent pas beaucoup d'informations.

Il existe plusieurs types de transformées mathématiques. La DCT (Discrete Cosine Transform), transformée en cosinus discrète, a été retenue dans la plupart des standards de compression d'images, parce qu'elle offre le meilleur compromis entre la concentration de l'information d'une image dans un nombre réduit de coefficients DCT, et le coût de calcul par rapport aux autres transformées.

En appliquant la DCT à un bloc de 8 x 8 pixels, nous obtenons une matrice Y de 8 x 8 coefficients DCT, chacun donné par :

$$y_{kl} = \frac{c(k)c(l)}{4} \sum_{i=0}^7 \sum_{j=0}^7 x_{ij} \cos\left(\frac{(2i+1)k\pi}{16}\right) \cos\left(\frac{(2j+1)l\pi}{16}\right)$$

où  $k, l = 0, 1, \dots, 7$

et  $c(0) = \frac{1}{\sqrt{2}}$ ,  $c(k) = 1$  si  $k \neq 0$ .

Pour retrouver le bloc de pixels original, nous devons appliquer la transformation DCT inverse sur les 64 coefficients DCT. La transformation DCT inverse est donnée par :

$$x_{ij} = \sum_{k=0}^7 \sum_{l=0}^7 y_{kl} \frac{c(k)c(l)}{4} \cos\left(\frac{(2i+1)k\pi}{16}\right) \cos\left(\frac{(2j+1)l\pi}{16}\right)$$

La transformée DCT n'est pas une technique de compression par elle-même. Après l'application de la DCT à un bloc de pixels, nous obtenons autant de coefficient DCT que de pixels de départ. C'est son association avec la quantification qui permet de réduire le volume de données.

#### 8.1.4. La quantification

Le processus d'élimination des coefficients DCT correspondants aux fréquences qui ne stockent pas une information significative est appelé quantification. La quantification est la principale source de perte d'information dans les schémas de codage basés sur la DCT.

La quantification est utilisée pour réduire le nombre de bits nécessaire pour stocker un coefficient DCT. Pour cela, une matrice de quantification est utilisée. Chaque valeur de cette matrice est un facteur de quantification qui sert à diviser le coefficient DCT correspondant. Le résultat est alors arrondi à l'entier le plus proche avant d'être stocké.

Pour le décodage, la valeur quantifiée stockée est multipliée par le facteur de quantification, pour retrouver le coefficient DCT de départ.

La détermination des éléments de la matrice de quantification est basée sur des expériences sur la perception humaine. Il s'agit de faire une série de tests pour savoir à partir de quelles valeurs la dégradation de l'image devient perceptible. Les tests ont révélé que l'œil humain reste insensible à l'élimination des hautes fréquences.

Des tables de quantification types sont utilisées dans les standards de compression comme JPEG ou H.261. Ces tables de quantification permettent de faire une compression de l'image sans que l'œil humain s'en aperçoive, mais le facteur de compression n'est pas élevé. Pour augmenter le facteur de compression, il faut multiplier avec un facteur les éléments de la table de quantification, mais au prix d'une perte de qualité qui sera perceptible par l'œil humain.

### 8.1.5. Compensation et estimation de mouvement

La compensation de mouvement est le processus permettant de compenser les déplacements d'objets d'une image à une autre. Son principe consiste à prévoir l'image courante en fonction de l'image précédente, puis coder la différence entre les deux images, permettant ainsi d'augmenter considérablement le taux de compression.

La sortie d'un bloc de compensation de mouvement peut s'exprimer par :

$$e(x, y, t) = I(x, y, t) + I(x - u, y - v, t - 1)$$

où,  $e(x, y, t)$  représente la prédiction résiduelle,  $I(x, y, t)$  la valeur du pixel à la position  $(x, y)$  de l'image à l'instant  $t$ , et  $I(x - u, y - v, t - 1)$  la valeur du pixel correspondant à la position  $(x - u, y - v)$  de l'image à l'instant  $t - 1$ .

Pour cela, la compensation de mouvement est précédée par l'estimation de mouvement, le processus qui recherche les pixels correspondant entre les images successives. L'objectif de l'estimation de mouvement est de déterminer un bloc dans l'image de référence qui ressemble le plus à un bloc donné de l'image courante. La sortie du bloc d'estimation de mouvement, est le vecteur de mouvement  $(u, v)$  qui définit le mouvement relative d'un bloc d'une image à une autre.

La qualité de l'estimation de mouvement dépend de la taille du bloc et de la région de recherche du bloc correspondant. En augmentant la taille de la région, nous augmentons les chances de trouver le bloc ressemblant, mais nous augmentant aussi le temps de calcul. En augmentant la taille du bloc, nous diminuons le temps de recherche, mais nous diminuons aussi les chances pour trouver le bloc qui ressemble le plus.

Pour les standards de compression, des choix ont été fait pour faire un compromis entre la qualité de l'estimation de mouvement et le temps de calcul.

La compensation de mouvement est effectuée au niveau du codeur et du décodeur. L'estimation de mouvement quant à elle se fait seulement au niveau du codeur.

## 8.2. Les standards de compression

Plusieurs standards de compression ont été établis. La plupart de ces standards utilise la transformation DCT. Nous allons présenter dans la suite trois standard des plus utilisés

actuellement. Pour les images fixes, nous présenterons JPEG [Wallace 91], et pour la vidéo, H.261 [Liou 91] et MPEG [Le Gall 91].

### **8.2.1. La compression JPEG**

JPEG (Joint Photographic Experts Group), est un format de compression d'image fixe. La Figure 8-1 montre le diagramme en bloc du codeur JPEG pour une seule composante couleur de l'image. Le processus est répété pour chaque composante de l'image.

L'image est d'abord divisée en bloc de 8 x 8 pixels. Chaque bloc subi une transformation vers le domaine fréquentiel. Pour cela, on applique à chaque bloc une transformation du type DCT. On obtient ainsi les coefficients de la DCT. Jusqu'ici, l'image n'a subie aucune compression.

Le résultat de cette transformation est un bloc de 64 coefficients, dans lequel les coefficients correspondants aux hautes fréquences sont très petit ou s'approche de zéro.

L'étape suivante est la quantification. C'est la principale source de compression. Une matrice de quantification 8 x 8 est utilisée pour chaque composante de l'image. Chaque élément de la matrice de quantification est une valeur entre 1 et 255 et définit le pas de quantification pour le coefficient de DCT correspondant. Une quantification agressive permet d'accroître la compression. Toutefois, elle peut aussi introduire des artefacts indésirable et des distorsions sur l'image décompressée. A la sortie de la quantification, plusieurs valeurs de la matrice obtenue sont nulles. Ces valeurs nulles correspondent généralement aux fréquences élevées, qui ne sont pas perçues par l'oeil.

Ce bloc de coefficients est stocké selon une séquence en zigzag de telle sorte que les coefficients à basse fréquence sont stockés en premier et les coefficients à hautes fréquences en derniers. Ceci afin de faire un codage d'entropie efficace.

La dernière étape du codeur JPEG est le codage d'entropie. Les coefficients continus (DC) sont traités différemment des coefficients AC. Puisque les valeurs des coefficients DC de blocs adjacents sont fortement corrélées, ce n'est pas la valeur elle même qui est utilisée pour le codage d'entropie, mais plutôt la différence avec le coefficient DC du bloc précédent.

Le codage d'entropie est sans perte. Il se fait en deux étapes, la première est un codeur prédictif pour le coefficient continu (DC, élément [0, 0] de la matrice) et un codage RLE pour les coefficients AC. La seconde étape encode les données suivant leur caractéristiques statistiques. Un codage de Huffman ou bien un codage arithmétique est alors utilisé. Le codage Huffman est le plus utilisé.

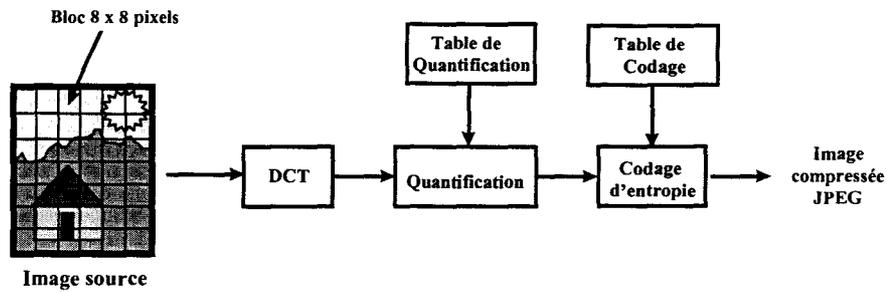


Figure 8-1 : Diagramme bloc de l'encodeur JPEG.

La Figure 8-2 montre le diagramme bloc d'un décodeur JPEG. Un décodage d'entropie est suivi d'une quantification inverse, puis d'une DCT inverse pour reconstituer l'image.

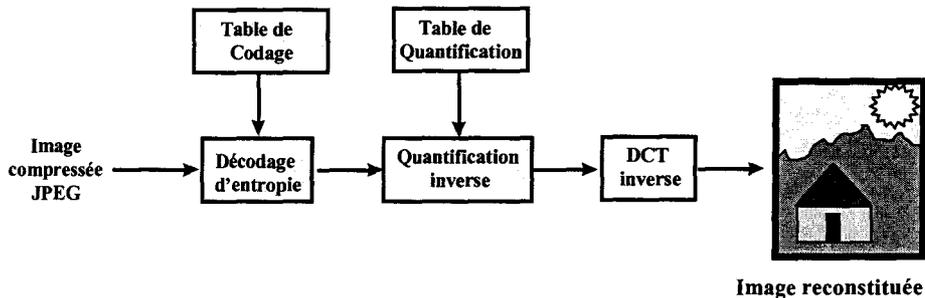


Figure 8-2 : Diagramme bloc du décodeur JPEG.

Bien que le standard JPEG a été établi pour la compression des images fixes, il est beaucoup utilisé pour la compression de séquence vidéo. On parle alors de M-JPEG (Motion JPEG), qui consiste en une succession d'images JPEG.

M-JPEG encode chaque image d'une séquence vidéo comme une image JPEG standard indépendante. Ceci a comme avantage un accès rapide et aléatoire à une image dans une séquence vidéo. Comme chaque image est indépendante, des algorithmes de compensation de mouvement ne sont pas utilisés. Le temps de calcul est ainsi réduit. En revanche, le facteur de compression n'est pas élevé puisque la redondance temporelle n'est pas éliminée.

### 8.2.2. La compression H.261

Le CCITT a développée la norme de compression vidéo H.261 pour les applications de vidéoconférence sur le réseau RNIS. Cette norme fait partie de la famille H.320. Elle permet d'encoder de la vidéo à des débits de  $p \times 64$  kbps, la valeur de  $p$  dépendant du nombre de canaux  $B$  utilisés qui peut atteindre 30 pour le réseau RNIS. Le débit maximal qui peut être atteint est alors de 2 Mbps.

H.261 supporte deux formats vidéo :

- Le format CIF (Common Intermediate Format) qui a une résolution de 360 x 288 pixels pour la luminance (Y) du flux vidéo et une résolution de 180 x 144 pixels pour les deux composantes de chrominance ( $C_R$  et  $C_B$ ) de la vidéo.
- Le format QCIF (Quarter-CIF) qui a une taille égale au quart de celle du format CIF, 180 x 144 pixels pour la luminance et 90 x 72 pixels pour la chrominance.

En fonction du nombre de canaux utilisés et du format choisi, H.261 peut atteindre 30 images par seconde.

H.261 possède deux modes de fonctionnement, le mode intra-frame dans lequel chaque image est codée indépendamment des autres images du flux comme dans une image JPEG, et le mode inter-frame où un algorithme de compensation de mouvement est utilisé.

La Figure 8-3 montre le schéma synoptique général d'un codeur H.261. Une image vidéo est d'abord translatée dans le format CIF puis stockée dans une mémoire. Suivant le mode choisi, la transformation DCT est appliquée à un bloc de 8 x 8 pixels pour une intra-image, ou bien elle est appliquée à la différence entre l'image courante et l'image précédente dans le cas du mode inter-image. Ensuite, une quantification et un codage d'entropie sont effectués.

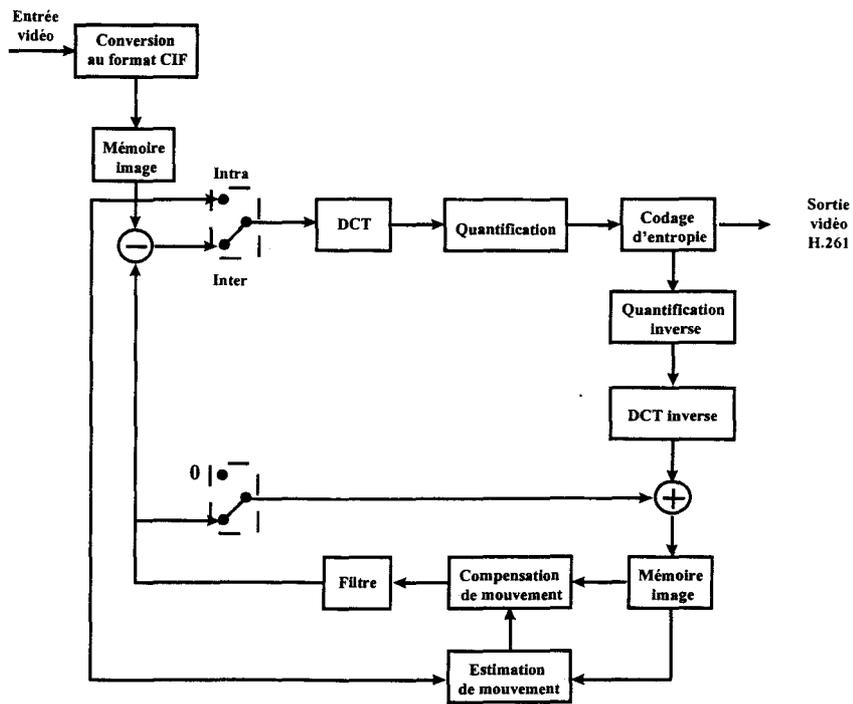


Figure 8-3 : Schéma synoptique d'un codeur H.261.

Le flux de données compressées est organisé hiérarchiquement en quatre couches : image, groupe de blocs, macro-bloc, et bloc (Figure 8-4).

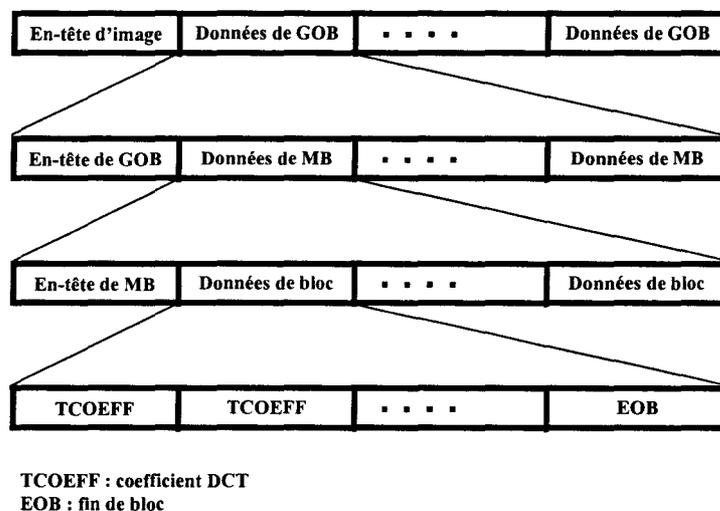


Figure 8-4 : Structure hiérarchique d'une image H.261.

L'image est la couche de haut niveau. Chaque image est divisée en groupe de blocs (GOB). Les images CIF et QCIF contiennent respectivement douze et trois GOBs. Chaque GOB est divisé en 33 macro-blocs. Un macro-bloc consiste en six bloc de 8 x 8 pixels, dont quatre pour la luminance Y, et deux pour la chrominance  $C_B$  et  $C_R$ .

L'UIT a développé une nouvelle norme, H.263. C'est principalement une amélioration de H.261 pour son utilisation sur le réseau téléphonique à travers des modems de type 28.8 kbps. De nouveaux formats ont été ajoutés, le sub-QCIF (128 x 96), le 4CIF (704 x 786), et le 16CIF (1408 x 1152).

### 8.2.3. La compression MPEG

Basés sur la norme H.261, les membres du groupe de travail MPEG (Motion Picture Expert Group) ont abouti à la norme MPEG-1. Comme c'est un standard de compression pour les images animées, son principe de compression consiste à coder les différences entre les images successives. Sa principale différence avec H.261 est que MPEG utilise une prédiction bidirectionnelle qui consiste à rechercher le bloc similaire dans l'image précédente et l'image suivante dans le flux vidéo.

Dans MPEG, trois types d'image sont définis, pour offrir un compromis entre un codage efficace et un accès aléatoire aux images dans une séquence compressé (Figure 8-5).

#### **L'image I (I-frame)**

L'intra-image (l'intra-picture) est codée indépendamment des autres images. Elle est transmise dans sa totalité selon une compression similaire à JPEG. Son taux de compression n'est pas très élevé, mais elle permet un accès aléatoire dans une séquence vidéo.

#### **L'image P (P-frame)**

L'image prédite (predicted) est codée à partir des images précédentes I ou P en utilisant les techniques de compensation de mouvement. Elle stocke la différence entre l'image courante et les images précédentes I ou P.

#### **L'image B (B-frame)**

L'image prédite bidirectionnelle (interpolée) utilise les images I et P précédentes et suivantes pour prédire l'image courante. Les images B possède le plus grand facteur de compression. Pour chaque bloc de l'image courante, on recherche les deux blocs qui lui ressemble le plus dans les images précédentes et suivantes. Une moyenne est calculée à partir

de ces deux blocs, puis la différence avec le bloc courant est utilisée pour la compression. Les images B ne sont pas utilisées pour la prédiction d'autres images B ou P.

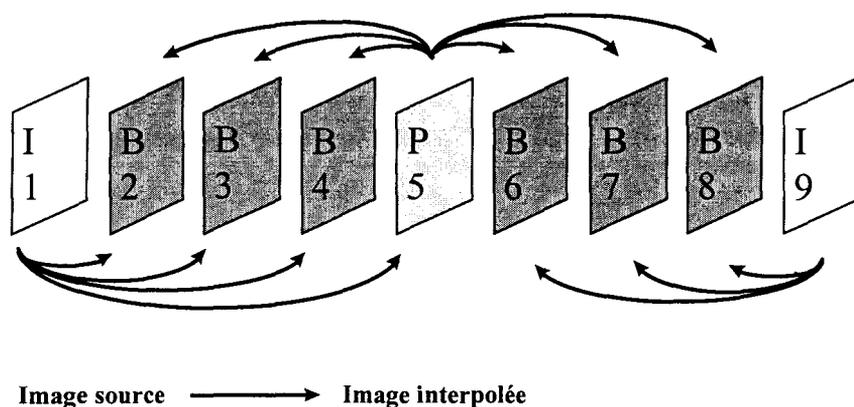


Figure 8-5 : Exemple d'une séquence MPEG montrant la dépendance entre les images I, P et B.

La norme MPEG se décline en plusieurs versions répondant chacune à des usages spécifiques.

### MPEG-1

MPEG-1 est développée pour exploiter des images animées à partir de supports de stockage tels que le CD ROM. Elle permet un débit de 1.5 Mbps. Elle est optimisée pour le format SIF (Standard Interchange Format) qui a une résolution de 352 x 240 pixels à 30 images par seconde pour NTSC, et une résolution de 352 x 288 à 25 images par seconde pour le PAL et le SECAM.

### MPEG-2

MPEG-2 est la seconde phase de travail du groupe MPEG. Avec MPEG-2, le but est de définir un standard générique qui peut être appliqué à une large gamme d'applications pour des débits pouvant atteindre 10 Mbps. Un autre groupe a été désigné pour la définition de MPEG-3 qui pouvait servir de point de départ à une future norme de télévision numérique haute définition (TVHD), mais très vite ses travaux ont été fusionnés avec MPEG-2.

MPEG-2 propose plusieurs schémas de codage. Un schéma peut être défini par un niveau et un profil. Le profil donne des détails sur la manière de coder la vidéo, comme l'utilisation

du codage hiérarchique, alors que le niveau spécifie les caractéristiques, comme la dimension de la vidéo. Quatre niveaux ont été définis :

- Niveau bas (Low level) : format SIF (352 x 288) à 30 images par seconde ;
- Niveau principal (Main level) : format CCIR 601 (720 x 576) à 30 images par seconde ;
- Niveau haut-1440 (High-1440 level) : 1440 x 1152 à 60 images par seconde ;
- Niveau haut (High level) : 1920 x 1152 à 60 images par seconde ;

#### **MPEG-4**

MPEG-4 qui est en cours de normalisation, n'est pas juste un standard de compression. Prévus au départ pour la transmission et à la réception d'images animées à très bas débit, passant par exemple par une ligne téléphonique classique, elle a vu son champ d'action s'élargir pour prendre en considération plusieurs types applications. Elle diffère des autres méthodes de compression par le fait qu'elle soit basée sur une approche objet. L'image n'est pas découpée en bloc, mais plutôt en objets réels constituant l'image. Ceci permet d'avoir une grande flexibilité au niveau du décodage. L'utilisateur pourra choisir les objets à décoder pour composer sa vidéo local. Il pourra choisir de ne pas recevoir un fond inutile. Pour cela, MPEG-4 incorporera un langage de description appelé MSDL (MPEG-4 System & Description Language) pour déterminer le contenu d'un flux vidéo.

---

# 9. Annexe II : Codage en couches

---

Le codage en couches permet de disposer de plusieurs qualités vidéo à partir d'un flux initial unique. Il permet ainsi à chaque utilisateurs d'exploiter ce flux en fonction des capacités de son système.

Nous présentons dans cet annexe, deux techniques de codage en couches, le codage progressif et le codage hiérarchique.

## 9.1. Le codage progressif

Le codage progressive permet à un utilisateur de transmettre une image en plusieurs couches. La première couche est une approximation de l'image, mais chaque couche suivante améliore l'image.

Le codage progressive utilise les mêmes techniques utilisées dans le codage utilisé en JPEG de base, appelé aussi codage séquentiel. Cependant, il introduit un buffer entre les étapes de quantification et de codage d'entropie pour stocker la totalité des coefficients de l'image. Le codage d'entropie est alors effectué autant de fois que de couches nécessaires sur le contenu du buffer.

Deux méthodes peuvent être choisies pour sélectionner les informations par couche.

### 9.1.1. Par sélection spectral

Une sélection des coefficients quantifiés est faite pour être transmis. Par exemple, comme il y a 64 coefficients par bloc, dans un premier temps, seulement le coefficient DC et les trois premiers coefficients DC (correspondant à l'ordre en zigzag) sont transmis, puis, dans la

deuxième couche, les 16 coefficients AC suivant, et enfin, dans la dernière couche, les 44 coefficients AC restant sont transmis.

### **9.1.2. Par approximation successive**

Une sélection de bit pour chaque coefficient est transmise pour chaque couche au lieu de transmettre le coefficient quantifié en entier. Par exemple, dans la première couche, les trois bits les plus significatifs sont transmis, puis dans la deuxième et troisième couche, le restant des bits est transmis.

Les deux méthodes peuvent être mixées, ce qui permet à l'utilisateur de choisir la « progression » de manière très flexible.

## **9.2. Le codage hiérarchique**

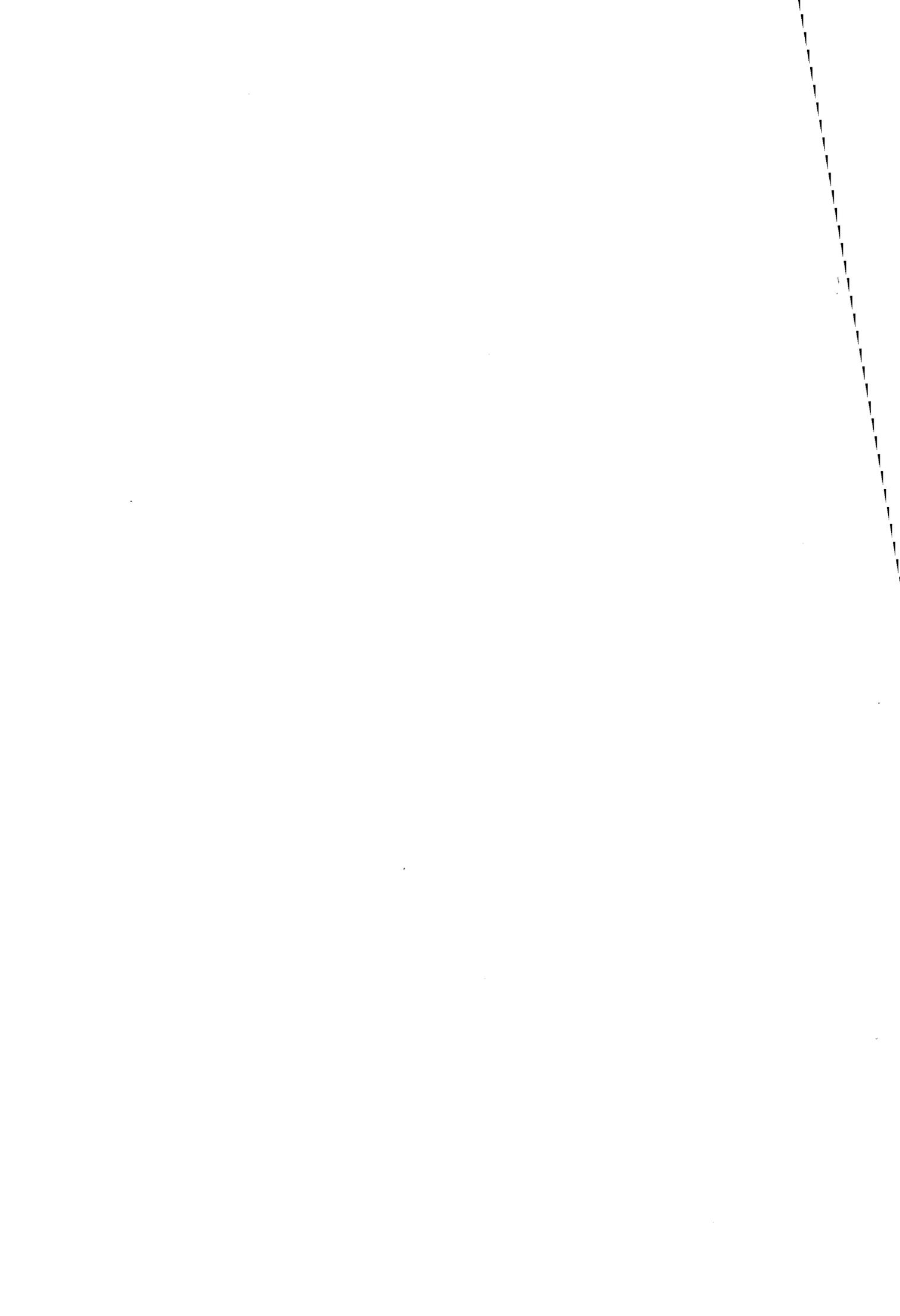
Le codage hiérarchique permet à l'utilisateur de décoder une version à basse résolution d'une image, sans pour cela décoder totalement l'image originale, puis la réduire.

Ce codage peut être utilisé en combinaison avec le mode séquentiel ou progressif.

Un certain nombre d'étapes est effectué dans un encodeur hiérarchique. Premièrement, l'image est réduite par un facteur de deux horizontalement, verticalement, ou sur les deux dimensions, autant de fois que le nombre de couches désiré. Le résultat est la résolution minimale de l'image qui peut être décodée. Ensuite, le codeur compresse cette image réduite en utilisant le codage séquentiel ou progressif. L'image compressée obtenue est alors transmise. Après, le codeur décompresse l'image compressée, multiplie par un facteur de deux horizontalement, verticalement, ou deux sur les deux dimensions, en utilisant un filtre d'interpolation qui est aussi utilisé dans le décodeur. Le résultat est alors comparé avec l'image originale de la même taille, ou avec l'image originale elle-même, si on est déjà remonté à la taille de départ. La différence de cette comparaison est encodée en utilisant la même méthode de compression mentionnée précédemment. Dans cet encodage, une table de quantification différente peut être utilisée, parce que la différence de deux images a d'autres caractéristiques (statistiques) que l'image elle-même. Ce processus de remontée et de calcul de différence est effectué jusqu'à ce que l'on atteigne la résolution d'origine de l'image.

Le codage hiérarchique nécessite autant de buffer que de résolutions d'image pour stocker les différentes images réduites, et un buffer au niveau du décodeur. De plus, si le décodeur a besoin de l'image avec la plus haute résolution, un plus grand calcul doit être fait au niveau du codeur et du décodeur que celui fait pour le codage séquentiel.





---

# 10. Annexe III : Classification

---

Une classification est un regroupement d'objets semblables. Chaque objet étant caractérisé par  $p$  variables, une classification se fait selon un critère rationnel. Ainsi, selon le critère choisi, un objet peut appartenir à plusieurs classes.

La classification automatique est une branche de l'analyse de données, dont les techniques sont destinées à produire des groupements d'objets ou d'individus décrits par un certain nombre de variables ou de caractères.

Le recours aux techniques de classification automatique est sous-tendu par quelques idées générales concernant le champ d'observation. On suppose que certains regroupement doivent exister, ou au contraire, comme dans notre cas, on exige que certains regroupements soient effectués.

Il existe plusieurs familles d'algorithmes de classification :

- Les algorithmes ascendants (ou encore agglomératifs) qui procèdent à la construction des classes par agglomération successive des objets deux à deux, et qui fournissent une hiérarchie de partitions des objets ;
- Les algorithmes descendants (ou encore divisifs) qui procèdent par dichotomies successives de l'ensemble des objets, et qui peuvent encore fournir une hiérarchie de partitions ;
- Les algorithmes conduisant directement à des partitions, comme les méthodes d'agrégation autour de centres mobiles.

Comme dans notre problème, il s'agit de répartir les vecteurs de qualité en un nombre de classes correspondant au nombre de qualité vidéo que l'on peut obtenir en utilisant le codage hiérarchique et le transcodage, se sont les algorithmes de la dernière catégorie qui nous intéresse.

On exposera dans la suite, les techniques d'agrégation autour de centres mobiles, apparentées à la méthode des « Nuées dynamique ».

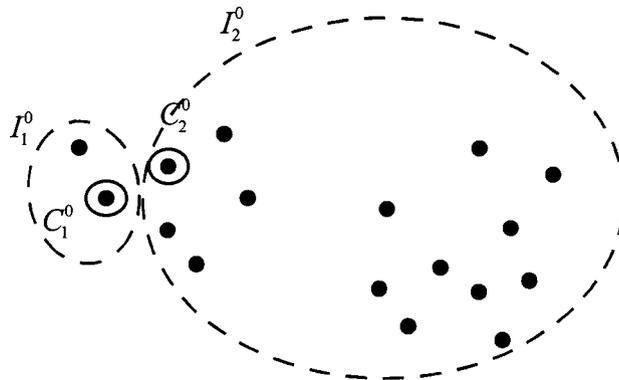
### 10.1. Agrégation autour de centres mobiles

L'algorithme que nous présentons peut être imputé principalement à [Forgy 65].

Soit à partitionner un ensemble  $I$  de  $n$  individus caractérisés par  $p$  paramètres ou variables. On suppose que l'espace  $R^p$  supportant les  $n$  points individus est muni d'une distance euclidienne appropriée notée  $d$ . on désire constituer au maximum  $q$  classes. Les étapes de l'algorithme sont :

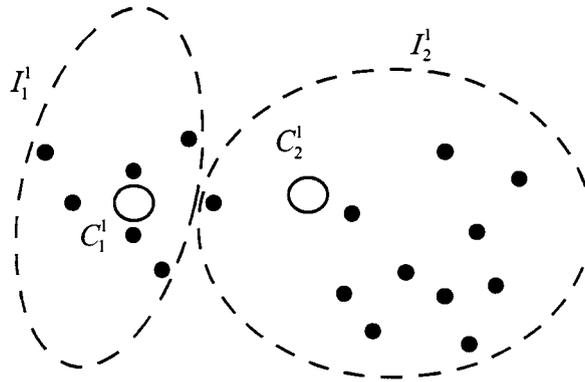
Etape 0 :

On détermine  $q$  centres provisoires de classes (par exemple par tirage pseudo-aléatoire). Ces centres  $\{C_1^0, C_2^0, \dots, C_k^0, \dots, C_q^0\}$  induisent une partition de  $I$  en  $q$  classes :  $\{I_1^0, I_2^0, \dots, I_k^0, \dots, I_q^0\}$ . Ainsi, l'individu  $i$  appartient à  $I_k^0$  si le point  $i$  est plus proche de  $C_k^0$  que de tous les autres centres.

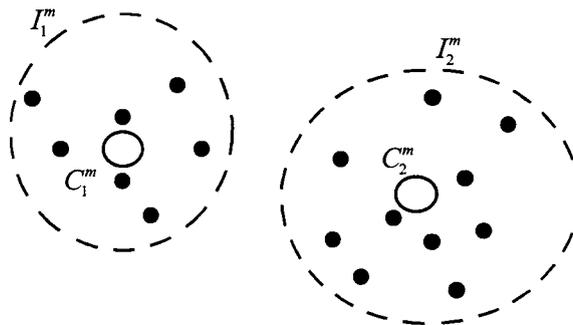


Etape 1 :

On détermine  $q$  nouveaux centres de classes :  $\{C_1^1, C_2^1, \dots, C_k^1, \dots, C_q^1\}$  en prenant les centres de gravité des classes  $\{I_1^0, I_2^0, \dots, I_k^0, \dots, I_q^0\}$ . Ces nouveaux centres induisent une nouvelle partition construite selon la même règle, et formée des classes notées :  $\{I_1^1, I_2^1, \dots, I_k^1, \dots, I_q^1\}$ .

Etape m :

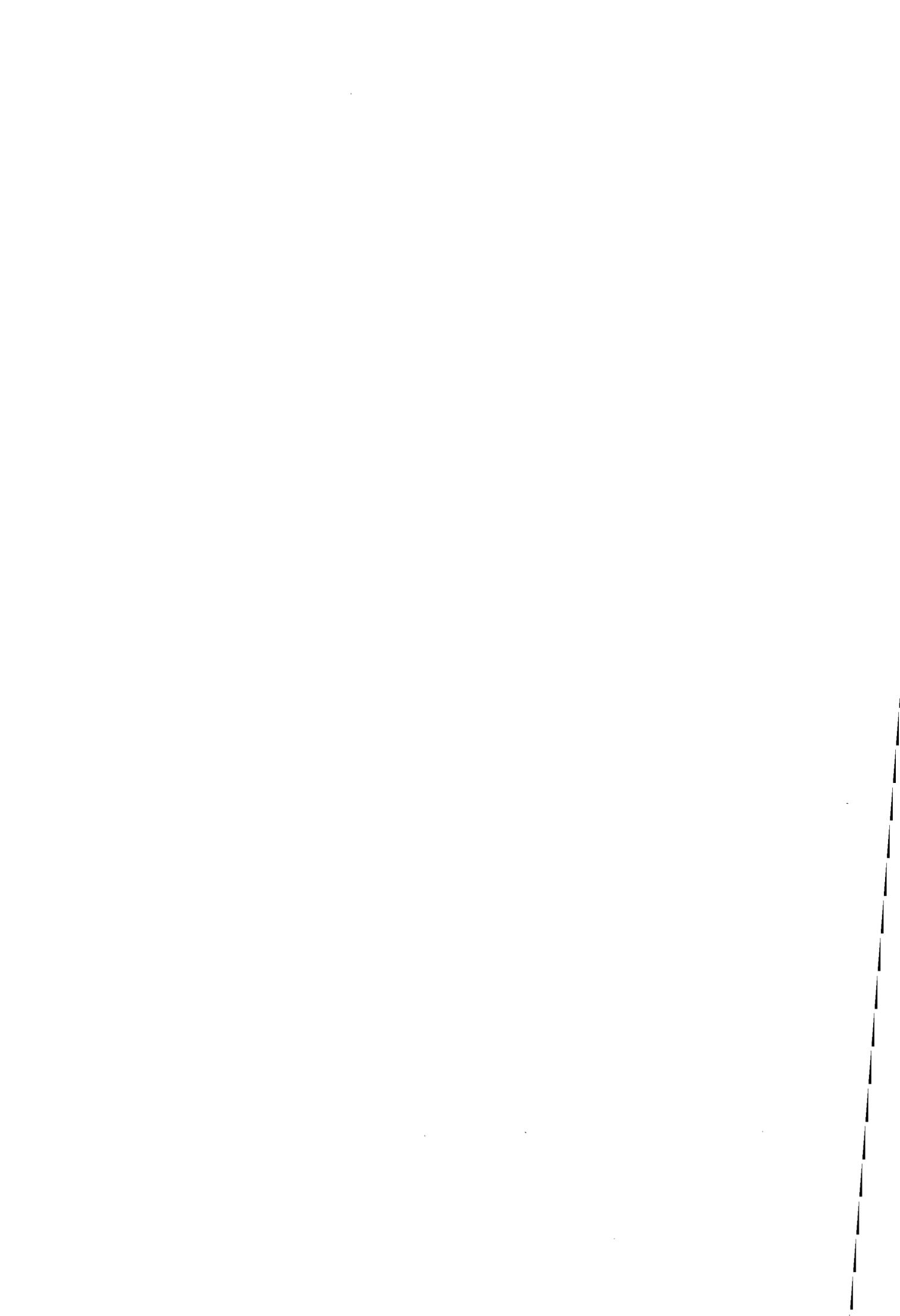
On détermine  $q$  nouveaux centres de classes :  $\{ C_1^m, C_2^m, \dots, C_k^m, \dots, C_q^m \}$  en prenant les centres de gravité des classes  $I_1^{m-1}, I_2^{m-1}, \dots, I_k^{m-1}, \dots, I_q^{m-1}$ . Ces nouveaux centres induisent une nouvelle partition construite selon la même règle, et formée des classes notées :  $\{ I_1^m, I_2^m, \dots, I_k^m, \dots, I_q^m \}$ .



L'algorithme s'arrête soit lorsque deux itérations successives conduisent à la même partition, soit lorsqu'un critère convenablement choisi (par exemple la variance intra-classe) cesse de décroître de façon sensible, soit encore parce qu'un nombre maximal d'itérations a été fixé a priori.

Il existe d'autres variantes de cet algorithme comme la technique des nuées dynamiques [Diday 72] ou la méthode K-means (K-moyennes) [MacQueen 67].

Il est nécessaire de préciser qu'une classification automatique obtenue sur un ensemble n'est jamais la seule classification de cet ensemble, mais une classification parmi beaucoup d'autres, établie à partir de variables et de méthodes choisies intentionnellement. Dans tous les cas, la partition obtenue dépendra du choix initial des centres à l'étape 0.



# Bibliographie

[Alfano 97]

M. Alfano, *User requirements and resource control for cooperative multimedia applications*. In Proceeding of ECMAST '97, Milan, Italy, May 21-23, 1997.

[Altenhofen 93]

M. Altenhofen, *The BERKOM Multimedia Collaboration Service*. Proceedings of ACM Multimedia 93, 1993.

[Amir 95]

E. Amir, S. McCanne, H. Zhang, *An Application Level Video Gateway*. ACM Multimedia, San Francisco, CA, November 1995.

[Amir 96]

Elan Amir, Steven McCanne and Martin Vetterli, *A Layered DCT Coder for Internet Video*. IEEE International Conference on Image Processing, Lausanne, Switzerland, September, 1996.

[ATM Forum 95]

ATM Forum, *User Network Interface (UNI) Specification, Version 3.1*, Prentice-Hall, 1995.

[ATM Forum 96]

ATM Forum, *ATM Traffic Management Specification, Version 4.0*, Prentice-Hall, 1996.

[Bly 93]

S. A. Bly, S. R. Harrison, S. Irwin, *Media Spaces : Bringing People Together in a Video, Audio, and Computing Environment*. Communication of the ACM, Vol.36, No.1, January 1993.

[Boulkroune 96a]

R. Boulkroune and P. Vincent, *A Multipoint Multimedia Conference Management System*. TDP96 La Londe Les Maures, France, June 1996.

[Boulkroune 96b]

R. Boulkroune, P. Vincent, *Une Architecture pour les Applications Coopératives Synchrones*. Calculateurs Parallèles Volume 8 - n° 3 / 1996.

[Boyer 94]

D. G. Boyer and M. E. Lukacs, *The Personal Presence System - A Wide Area Network Resource for the Real Time Composition of Multipoint Multimedia Communications*. Proceedings of ACM Multimedia 94, 1994.

[Brockschmidt 95]

K. Brockschmidt, *Inside OLE 2*. Microsoft Press, second edition, 1995.

[Bruce 96]

V. Bruce, *The Role of the face in communication : implications for videophone design*. *Interacting with Computers*, Volume 8, Number 2, June 1996.

[Bull 78]

P. Bull, *The Interpretation of Posture through an Alternative Method to Rôle Play*. *British Journal of Social and Clinical Psychology*, 17, 1978.

[Buxton 92]

W. Buxton, *Telepresence: Integrating Shared Task and Person Spaces*. Proceedings of Graphics Interface' 92, 1992.

[Campbell 96]

A. Campbell, C. Aurrecoechea, L. Hauw, *A Review of QoS Architectures*. Proceedings of 4th IFIP International Workshop on Quality of Service, IWQS'96 (invited paper), Paris, France, March, 1996.

[Carlier 98]

P. Carlier, *ScenIC : une proposition pour l'assistance à la composition spatio-temporelle des interfaces des environnements de TCAO*. Thèse de Doctorat en Informatique, Université des Sciences et Technologies de Lille, 1998.

[Chaddha 95a]

N. Chaddha, G. A. Wall and B. Schmidt, *An End to End Software Only Scalable Video Delivery System*. Proceedings of the Fifth International Workshop on Network and Operating System Support for Digital Audio and Video, Durham, New Hampshire, April 18-21, 1995.

[Chaddha 95b]

N. Chaddha, *Software only Scalable Video Delivery System for Multimedia Applications over Hetrogeneous Networks*. In Proceedings of the IEEE International Conference on Image Processing, Washington, DC, October, 1995.

[Chapanis 75]

A. Chapanis, *Interactive Human Communication*. *Scientific American*, Vol. 232, March, 1975.

[Clark 92]

W. J. Clark, *Multipoint Multimedia Conferencing*. *IEEE Communications Magazine*, May 1992.

[Colin 98]

J. F. Colin, P. Vincent, *TuttelCast: a Web presentation tool*. Leverage Workshop Cambridge, January, 1998.

[Comer 95]

D. Comer, *Internetworking With TCP/IP, Volume 1 : Principales, Protocols, and Architecture*. Third Edition, Prentice-Hall, Englewood Cliffs, N. J., 1995.

[Coulson 94]

G. Coulson, A. Campbell, P. Robin, G. Blair, M. Papathomas, D. Hutchison, *The Design of a QoS Controlled ATM Based Communications System in Chorus*. IEEE Journal on Selected Areas in Communications, Special Issue on ATM Local Area Networks, 1994.

[Croisy 95]

P. Croisy, *Collecticiel temps réel et apprentissage coopératif: des aspects sociaux et pédagogiques jusqu'au modèle multi-agent de l'interface de groupe*. Thèse de Doctorat en Informatique, Université des Sciences et Technologies de Lille, 1995.

[Decina 86]

M. Decina, CCITT Recommendation on the ISDN : A Review. IEEE Journal on Selected Areas in Communications, May, 1986.

[Deering 91]

S. Deering, *Multicast Routing in a Datagram Internetwork*. PhD thesis, Stanford University, December 1991.

[Delgrossi 93]

L. Delgrossi, C. Halstrick, D. Hehmann, R. G. Herrtwich, O. Krone, J. Sandross, C. Vogt, *Media Scaling for Audiovisual Communication with the Heidelberg Transport System*. Proceedings of the ACM Multimedia 93, Anaheim, California, 1-6 August 1993.

[De Prycker 93]

M. De Prycker, *Asynchronous Transfer Mode – Solution for Broadband ISDN*. Second Edition, Ellis Horwood, London, 1993.

[Diday 72]

E. Diday, *Optimisation en Classification Automatique et Reconnaissance des Formes*. Rev. Fr. Inf. Rech. Opér. 6<sup>ème</sup> année, p.61-95, novembre 1972.

[Diot 97]

C. Diot, W. Dabbous, and J. Crowcroft, *Multipoint Communication : A Survey of Protocols, Functions, and Mechanisms*. IEEE Journal on selected areas in communications, Vol. 15, No. 3, April, 1997.

[Dommel 97]

H. P. Dommel, and J. J. Garcia-Luna-Aceves, *Floor control for multimedia conferencing and collaboration*. Multimedia systems, Volume 5 Number 1, p. 23-38, January 1997.

[Dorcey 95]

T. Dorcey, *Cu-SeeMe Desktop Videoconferencing Software*. ConneXions, Volume 9, No 3, March, 1995.

[Dourish 91]

P. Dourish, *Godard : A Flexible Architecture for A/V Services in a Media Space*. EuroPARC Technical Report EPC-91-134, Rank Xerox EuroPARC, Cambridge, UK, 1991.

[Dourish 92]

P. Dourish, V. Bellotti, *Awareness and Coordination in Shared Workspaces*. Proceedings of CSCW'92, p.107-114, 1992.

[Dourish 93]

P. Dourish, *Culture and Control in a Media Space*. Proceedings of the Third ECSCW, Milan, Italy, 13-17 September 1993.

[Eckel 98]

Bruce Eckel, *Thinking in Java*. Prentice-Hall, 1998.

[Eddon 98]

G. Eddon, H. Eddon, *Inside Distributed Com*. Microsoft Press, March, 1998.

[Ekman 75]

P. Ekman, W. Friesen, *Unmasking the face*. Englewood Cliffs, NJ, Prentice Hall, 1975.

[Ellis 91]

C. A. Ellis, S. J. Gibbs, and G. L. Rein, *Groupware: Some issues and experiences*. Communications of the ACM, vol. 34 no. 1, 1991.

[Englander 97]

R. Englander, M. Loukides, *Developing Java Beans*. O'Reilly & Associates, June, 1997.

[Eriksson 94]

H. Eriksson, *Mbone: The multicast backbone*. Communications of the ACM, August, 1994.

[Feit 96]

S. Feit, *TCP/IP : Architecture, Protocols, and Implementation With IPv6 and IP Security*. Second Edition, McGraw-Hill, October, 1996.

[Ferrari 94]

D. Ferrari, A. Banerjea, and H. Zhang, *Network Support for Multimedia - A Discussion of the Tenet Approach*. Computer Networks and ISDN Systems, Vol. 26, 1994.

[Flanagan 97]

D. Flanagan, *Java in a Nutshell: A Desktop Quick Reference for Java Programmers*. O'Reilly & Associates, Second Edition, May 1997.

[Forgy 65]

E. W. Forgy, *Cluster Analysis of Multivariate Data : Efficiency Versus Interpretability of Classifications*. Biometrics, 21, p. 768-79, 1965.

[Gale 90]

S. Gale, *Human aspects of interactive multimedia communication*. Interacting with computers, 2, p.175-189, 1990.

[Gajewska 94]

H. Gajewska, J. Kistler, M. S. Manasse, and D. D. Redell, *Argo : A System for Distributed Collaboration*. Proceeding of the ACM Multimedia'94, San Francisco, CA, USA, 10-1994.

[Gaver 92a]

W. Gaver, *The affordances of media spaces for collaboration*. In Proceedings of CSCW'92, Toronto, Canada, pp. 17-24, November, 1992.

[Gaver 92b]

W. Gaver, T. Moran, A. MacLean, L. Löfstrand, P. Dourish, K. Carter, and W. Buxton, *Realising a Video Environment : EuroPARC's Rave System*. in Proceedings of the ACM, Conference on Human Factors in Computing Systems CHI'92, Monterey, California, May, 1992.

[Goodwin 81]

C. Goodwin, *Conversational Organization: Interaction between speakers and hearers*. New York: Academic Press, 1981.

[Gosling 95]

J. Gosling, H. McGilton, *The Java Language Environment : A White Paper*. Technical Report, Sun Microsystems, October, 1995.

[Greenberg 91]

S. Greenberg, *Personalizable groupware: Accomodating individual roles and group differences*. Proceedings of the Second European Conference on Computer Supported Cooperative Work (EC-CSCW '91), Amsterdam, The Netherlands, September 25-27, 1991.

[Greenberg 92]

S. Greenberg, M. Roseman, D. Webster, and R. Bohnet, *Human and technical factors of distributed group drawing tools*. Interacting with Computers, 4(1), pp. 364-392, December, 1992.

[Grudin 88]

J. Grudin, *Why CSCW Applications Fail: Problems in the Design and Evaluation of Organizational Interfaces*. Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'88), Portland, Oregon, September 26-28, 1988.

[Gutwin 96]

C. Gutwin, M. Roseman, S. Greenberg, *A Usability Study of Awareness Widgets in a Shared Workspace Groupware System*. Proceedings of the ACM 1996 Conference on CSCW, Boston, p. 258-268, 1996.

[Handley 93]

M. J. Handley, P. T. Kirstein, M. A. Sasse, *Multimedia integrated conferencing for European researchers (MICE) : piloting activities and the conference management and multiplexing centre*. Computer Networks and ISDN Systems 26, p.275-290, 1993.

[Hoffman 96]

D. Hoffman and M. Speer, *Hierarchical video distribution over internet-style networks*. IEEE International Conference on Image Processing, Lausanne, Switzerland, September, 1996.

[Hoogendoorn 98]

P. Hoogendoorn, *Design of a Group Memory System Based on Computer Conferencing in a Distance Training*. University of Twente, The Netherlands, April, 1998.

[Hoogstoel 95]

F. Hoogstoel, *Une approche organisationnelle du travail coopératif assisté par ordinateur. Application au projet Co-Learn*. Thèse de Doctorat en Informatique, Université des Sciences et Technologies de Lille, 1995.

[Huitema 95]

Christian Huitema, *Et Dieu créa l'Internet*. Editions Eyrolles, Paris, Mai, 1995.

[Isaacs 93]

E. A. Isaacs and J. C. Tang, *What video can and can't do for collaboration: a case study*. Proceedings of the ACM Multimedia 93, Anaheim, CA, 1993.

[ISO 95]

*Information Technology - Quality of Service - Framework*. ISO/IEC CD 13236.2, October, 1995.

[Inoue 95]

T. Inoue, K. Okada, Y. Matsushita, *Learning from TV programs : Application of TV presentation to a videoconferencing system*. Proceedings of the UIST 95, Pittsburgh, PA, USA, November 14-17, 1995.

[ITU 90]

ITU-T, *H.320, Narrow-band visual telephone systems and terminal equipment*. ITU-T Recommendation, 1990.

- [ITU 93a]  
ITU-T, *T.123, Protocol Stack for Audiographics and Audiovisual Teleconference Applications*. ITU-T Recommendation, 1993.
- [ITU 93b]  
ITU-T, *T.122, Multipoint Communication Service for Audiographic and Audiovisual Conferencing Service Definition*. ITU-T Recommendation, 1993.
- [ITU 94]  
ITU-T, *T.125, Multipoint Communication Service Protocol Specification*. ITU-T Recommendation, 1994.
- [ITU 95a]  
ITU-T, *T.124, Generic Conference Control*. ITU-T Recommendation, 1995.
- [ITU 95b]  
ITU-T, *T.126, Multipoint Still Image and Annotation Protocol*. ITU-T Recommendation, 1995.
- [ITU 95c]  
ITU-T, *T.127, Multipoint Binary File Transfer Protocol*. ITU-T Recommendation, 1995.
- [ITU 96]  
ITU-T, *T.120, Data Protocols for Multimedia Conferencing*. ITU-T Recommendation, 1996.
- [Jeffay 91]  
K. Jeffay, D. L. Stone, D. Poirier, *YARTOS: Kernel support for efficient, predictable real-time systems*. In *Proceeding of IFAC/IFIP, Workshop on Real-Time Programming*, Atlanta, Georgia, May, 1991.
- [Johnson 88]  
R. E. Johnson, *Designing Reusable Classes*. *The Journal of Object-Oriented Programming*, Vol.1, No.2, 1988.
- [Johnson 92]  
P. Johnson, *Human Computer Interaction : Psychology, Task analysis and Software Engineering*. McGraw-Hill Book Company, London, 1992.
- [Karmouch 93]  
A. Karmouch, *Multimedia distributed cooperative system*. *Computer Communications*, Vol. 16 No 9, September, 1993.
- [Karsenty 93]  
A. Karsenty, *GroupDesign : un collecticiel synchrone pour l'édition partagée de documents*. Thèse de Doctorat en Informatique, Université Paris XI Orsay, 1993.

[Karsenty 94]

A. Karsenty, *Le Collecticiel : de l'Interaction Homme-Machine à la Communication Homme-Machine-Homme*. Technique et sciences informatiques, Volume 13 No1, 1994.

[Kernighan 78]

B. W. Kernighan, and D. M. Ritchie, *The C Programming Language*. Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[Kraut 93]

R. E. Kraut, R. S. Fish, R. W. Root, B. L. Chalfonte (1990), *Informal Communication in Organizations : Form, Function and Technology*. Readings in Groupware and Computer Supported Cooperative Work: Assisting Human-Human Collaboration, Ronald M. Baecker (ed.), Morgan Kaufmann, San Mateo, CA, 1993.

[Lafon 94]

J. L. Lafon, *A French experiment in distance learning by ISDN: Le VisioCentre de Formation*. ISDN applications in education and training, 1994.

[Laubach 98]

M. Laubach, J. Halpern, *Classical IP and ARP over ATM*. RFC 2225, April, 1998.

[Le Gall 91]

D. Le Gall, *MPEG : A Video Compression Standard for Multimedia Applications*. Communications of the ACM, Volume 34, No 4, April, 1991.

[Liou 91]

M. Liou, *Overview of the px64 kbit/s Video Coding Standard*. Communications of the ACM, Volume 34, No 4, April, 1991.

[Macedonia 94]

M. R. Macedonia, D. P. Brutzman, *MBone provides audio and video across the Internet*. IEEE Computer magazine, April, 1994.

[MacQueen 67].

J. MacQueen, *Some methods for Classification and Analysis of Multivariate Observations*. Proceedings of the 5<sup>th</sup> Berkeley Symposium, 1967.

[Mangione 98]

C. Mangione, *Just In Time for Java vs. C++*. NC World Magazine, Volume 2, Issue 1, January, 1998.

<<http://www.ncworldmag.com/ncworld/ncw-01-1998/ncw-01-jperf.html>>

[McCanne 95]

S. McCanne and Van Jacobson, *vic : A Flexible Framework for Packet Video*. Proceedings of the ACM Multimedia, San Francisco, November, 1995.

- [McCanne 96]  
S. McCanne, Scalable compression and transmission of internet multicast video. PhD dissertation, University of California, Berkeley, CA, December, 1996.
- [McDaniel 97]  
S. E. McDaniel, T. Brinck, *Awareness in Collaborative Systems*. ACM CHI'97 Workshop organized by Susan E. McDaniel and Tom Brinck, Atlanta, Georgia, March 22-27, 1997.
- [Metcalfe 76]  
R. M. Metcalfe, D. R. Boggs, *Ethernet : Distributed Packet Switching for Local Computer Networks*. Communications of the ACM, Vol. 19, No 7, July, 1976.
- [Miah 94]  
Tunu Miah, *Multimedia Conferencing Architecture and User Interfaces for Collaborative Environment*. Proceedings of the Eurographics in Graz, Austria, June 6-9 1994.
- [Minzer 89]  
S. Minzer, *Broadband ISDN and Asynchronous Transfer Mode (ATM)*. IEEE Communications Magazine, Vol. 27, No 9, September, 1989.
- [Nahrstedt 95a]  
K. Nahrstedt and J. M. Smith, *The QoS Broker*. IEEE MultiMedia, Vol. 2, No. 1, Spring 1995.
- [Nahrstedt 95b]  
K. Nahrstedt, R. Steinmetz, *Resource Management in Networked Multimedia Systems*. IEEE Computer, May, 1995.
- [Onvural 93]  
R. O. Onvural, *ATM Networks : Performance Issues*. Artech House Publishers, Boston, 1993.
- [Orfali 96]  
R. Orfali, D. Harkey and J. Edwards, *The Essential Distributed Objects Survival Guide*, John Wiley and Sons, Inc., 1996.
- [OMG 95]  
OMG, *The Common Object Request Broker : Architecture and Specification*. Technical Report, Object Management Group, July 1995.
- [Perednia 95]  
D. A. Perednia, A. Allen, *Telemedicine technology and clinical applications*. Journal of the American Medical Association Vol. 273, No. 6, February 1995.

[Pujolle 97]

G. Pujolle, *Les Réseaux, Deuxième édition*. Eyrolles, 1997.

[Rocheblave-Spenlé 69]

A. M. Rocheblave-Spenlé, *La notion de rôle en psychologie sociale*. PUF, 1969.

[Rodden 91]

T. Rodden, *A survey of CSCW systems*. *Interacting with Computers*, vol. 3 no. 3, 1991.

[Root 88]

R. Root, *Design of a Multi-Media Vehicle for a Social Browsing*. Proceedings of ACM, CSCW'88, Portland, Oregon, 1988.

[Rothermel 97]

K. Rothermel, G. Dermler, W. Fiederer, *QoS Negotiation and Reservation for Distributed Multimedia Applications*. Proceedings of the IEEE International Conference on Multimedia Computing and Systems'97, Ottawa, Ontario, Canada, June 3-6, 1997.

[Schooler 96]

E. M. Schooler, *Conferencing and Collaborative Computing*. *Multimedia Systems*, Volume 4 Issue 5, October 1996.

[Schulzrinne 96]

H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, *RTP: A transport protocol for real-time applications*. RFC-1889, Audio-Video Transport Working Group, Internet Engineering Task Force, January 1996.

[Sessions 97]

R. Sessions, *COM and DCOM: Microsoft's Vision for Distributed Objects*, Wiley Computer Publishing, December 1997.

[Shirmohammadi 98]

S. Shirmohammadi, J. C. Oliveira and N. D. Georganas, *Applet-Based Telecollaboration: A Network-centric Approach*. *IEEE Multimedia Magazine*, Volume 5, Number 2, 1998.

[Socolofsky 91]

T. Socolofsky, C. Kale, *A TCP/IP Tutorial*. RFC 1180, Spider Systems Limited, January, 1991.

[Stefik 87a]

M. Stefik, D. G. Bobrow, G. Foster, S. Lanning, and D. Tatar, *WYSIWIS revised: Early experiences with multiuser interfaces*. *ACM Transactions on Office Information Systems*, Vol. 5, No. 2, April, 1987.

[Stefic 87b]

M. Stefik, G. Foster, D. G. Bobrow, K. Kahn, S. Lanning, and L. Suchman, *Beyond the chalkboard: Computer support for collaboration and problem solving in meetings*. Communications of the ACM, Vol. 30, No. 1, January, 1987.

[Stroustrup 86]

B. Stroustrup, *The C++ Programming Language*. Addison Wesley Series in Computer Science, Reading, Massachusetts, 1986.

[Taligent 93]

Taligent Inc., *Leveraging Object-Oriented Frameworks*, a Taligent White Paper, 1993.  
<<http://www.software.ibm.com/ad/taligent/>>

[Tanenbaum 96]

A. Tanenbaum, *Computer Networks, 3rd Edition*. Prentice Hall, 1996.

[Tang 92]

J. C. Tang and E. A. Isaacs, *Why Do Users Like Video ? Studies of Multimedia-Supported Collaboration*. Sun Microsystems Laboratories Inc., Technical Report TR-92-5, 1992.

[Tang 94]

J. C. Tang, E. A. Isaacs, M. Rua, *Supporting distributed groups with a montage of lightweight interactions*. Proceedings of CSCW'94, Chapel Hill, North Carolina, USA, October 22-26, 1994.

[Turletti 94]

T. Turletti, *The INRIA Videoconferencing System (IVS)*. ConneXions - The Interoperability Report, Vol. 8, No 10, October, 1994.

[Vilers 92]

P. Vilers, *Etude d'un système de téléconférence en temps réel : application à un jeu de rôle en gestion de production*. Thèse de doctorat en Informatique, Université des Sciences et Technologies de Lille, 1992.

[Vogel 95]

A. Vogel, B. Kerherve, G. Von Bochmann and J. Gecsei, *Distributed Multimedia and QOS: A Survey*. IEEE MultiMedia, Vol. 2, No. 2, Summer 1995.

[Wallace 91]

G. K. Wallace, *The JPEG Still Picture Compression Standard*. Communications of the ACM, Volume 34, No 4, April, 1991.

[Whittaker 97]

S. Whittaker, B. O'Conaill, *The role of vision in Face to Face and Mediated Communication*. In Video-Mediated Communication, Eds. K. Finn *et al*, Lawrence Erlbaum Associates, New Jersey, p. 23-49, 1997.

[Yamaashi 95]

K. Yamaashi, Y. Kawamata, M. Tani, H. Matsumoto, *User-Centred Video: Transmitting Video Images Based on the User's Interest*. Proceedings of CHI'95, Denver, Colorado, May 7-11, 1995.

[Yousfi 96]

F. Yousfi, *PLACO : Modélisation par Workflow et Conception d'un Système de Planification Coopérative, Application aux Unités de Soins*. Thèse de Doctorat en Informatique, Université des Sciences et Technologies de Lille, 1996.

[Zhang 93]

L. Zhang, S. Deering, D. Estrin, S. Shenker, D. Zappala: *RSVP: A New Resource ReSerVation Protocol*. IEEE Network Magazine, Vol. 7, No. 5. pp. 8-18, September, 1993.

[Zhou 95]

J. Zhou, B. Cousin, *Multi-video-scheme Videoconferencing Applications*. publication interne IRISA No 978, Décembre, 1995.

[Zundel 96]

K. M. Zundel, *Telemedicine: History, Applications, and Impact on Librarianship*. Bulletin of the Medical Library Association, Vol. 84 No 1, 1996.

