

## Chapitre 1

# Position du problème: De la spatialisation sonore en contexte de visioconférence

*Montségur est un rocher conique aux versants escarpés qui se dresse au-dessus d'une plaine tourmentée, donnant à première vue l'impression d'une illusion lointaine, d'une chose d'un autre monde. On dirait qu'il est là pour accueillir les armées des anges qui sont peut-être capables, de leur perspective séraphique, de découvrir cet empan de terrain plat où poser une échelle céleste...*

Les Enfants du Graal - P. Berling - Ed. Lattès, 1996



## Sommaire

---

1.1	De la téléconférence à la visioconférence...	27
1.2	Restitution spatialisée de la scène sonore	27
1.3	Mur de Téléprésence	28
1.4	Reproduction 3D d'un champ sonore sur une zone étendue	30
	Références Bibliographiques	31

---

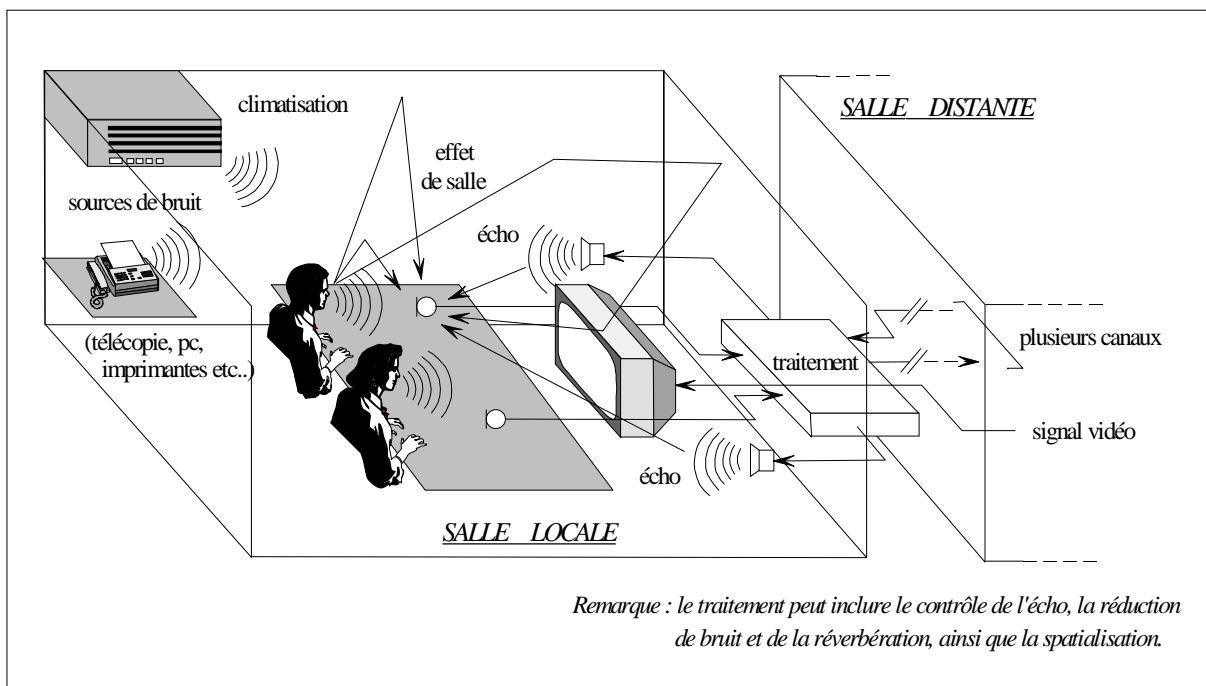


FIG. 1.1 - Principe d'une visioconférence

## 1.1 De la téléconférence à la visioconférence...

La téléconférence est un nouveau service de télécommunication qui se propose d'étendre le principe du dialogue téléphonique entre deux personnes à des groupes d'individus [Mahieux *et al.*, 1998]. Elle consiste à faire communiquer plusieurs groupes de personnes réparties dans des lieux distants. Ce type de service inaugure l'ère des systèmes de *communication de groupe*... Les réunions de travail organisées entre différents sites d'une entreprise constituent une des applications possibles. Dans ces systèmes, la voix de chaque locuteur est captée par un ou plusieurs microphones pour être restituée sur des hauts-parleurs placés dans les locaux distants (cf. Fig. 1.1). Afin d'améliorer la qualité du service, l'image des participants est également restituée au moyen de caméras et d'écrans vidéo. La téléconférence devient alors un moyen de communication audiovisuelle à part entière et on préfère parler de *visioconférence*.

Cependant, la prise et la restitution du son dans ces systèmes de communication de groupe posent des problèmes spécifiques [Miyoshi & Koizumi, 1992], tels que:

- la présence d'un *effet de salle* — à la fois à la prise et la restitution — susceptible d'altérer l'intelligibilité de la parole et à moindre degré d'induire des colorations spectrales,
- le *couplage acoustique* entre les microphones et les haut-parleurs, couplage qui est responsable de l'effet Larsen et des phénomènes d'écho acoustique<sup>1</sup>,
- les perturbations des *bruits* d'environnement (ventilation, appareillage électronique, ordinateurs...).

Différentes solutions existent. Privilégier une prise de son en *champ proche* ou utiliser des microphones *directifs* par exemple permet de s'affranchir partiellement des effets de salle et d'accroître le rapport signal à bruit. Spécifiquement développées pour ce type d'applications, les *antennes acoustiques* offrent dans le même esprit une prise de son à la fois directive et contrôlable par des moyens électroniques.

Enfin, il est possible de faire intervenir en aval de la prise de son des *algorithmes de traitement du signal* qui réalisent des opérations de déréverbération, d'annulation d'écho ou de réduction de bruit [Mahieux *et al.*, 1998]...

## 1.2 Restitution spatialisée de la scène sonore

Un autre axe de recherche concerne la *spatialisation de la scène sonore* [Miyoshi & Koizumi, 1992] [Mahieux *et al.*, 1998]. Il s'agit de restituer la position et les éventuels mouvements des différents interlocuteurs. Cette idée intervient d'abord dans un souci de *cohérence avec la localisation visuelle* sur les écrans vidéo. En outre, des tests menés pour évaluer la qualité des systèmes de visioconférence ont montré qu'une restitution sonore monophonique dégrade sensiblement la perception de la qualité globale du système, même si la restitution de l'image est parfaite [Chateau, 1999].

La spatialisation sonore permet aussi d'*améliorer l'intelligibilité de la parole*, car l'oreille humaine est capable de distinguer des sources émettant simultanément dès lors qu'elles occupent des positions distinctes dans l'espace. Ainsi une source de bruit est perceptivement moins gênante, lorsque le système auditif peut la discriminer spatialement de la source utile. Ce sont les propriétés du *masquage binaural* que l'on retrouve dans l'effet "Cocktail Party" [Brian, 1997] [Plomp, 1977]. De plus, avec la restitution spatialisée de la scène sonore, plusieurs conversations peuvent coexister en parallèle sans interférer entre elles. Plus généralement, la spatialisation des voix des locuteurs contribue à *soulager l'effort d'attention* de leurs interlocuteurs pour le suivi de la conversation. La spatialisation de la scène sonore en visioconférence se traduit donc par une amélioration du confort global des utilisateurs.

Dans le domaine de la visioconférence, la spatialisation sonore est donc devenue un sujet à l'ordre du jour. D'ores et déjà, le système *Varèse*, système de visioconférence de groupe qui a vu le jour au C.N.E.T.,

1. Lorsqu'un locuteur parle, sa voix est diffusée dans la salle distante. Elle est alors captée par les microphones situés dans cette salle, ce qui implique qu'elle est partiellement retransmise dans le local initial. Le locuteur perçoit donc sa propre voix réémise par les hauts-parleurs, moyennant un temps de retard lié aux délais de transmission et de traitement et qui peut varier entre 200 ms et 1 s... Ce phénomène constitue l'écho acoustique. Il est perceptivement très gênant dès que le retard est supérieure à 50 ms. Par ailleurs, le système de prise et restitution du son définit dans ces conditions un système bouclé. Lorsque le gain de la boucle est supérieur ou égal à 1, le système part en boucle et émet un sifflement à sa fréquence de résonance. Ce phénomène est connu sous le nom d'effet Larsen.

propose une restitution sonore spatialisée qui, bien qu'elle soit basée sur une approche très simplifiée, a clairement démontré l'intérêt de la spatialisation sonore pour la visioconférence [Mahieux *et al.*, 1998]. Dans le système Varèse, l'effet de spatialisation est obtenu en plaçant derrière l'écran vidéo un nombre suffisant de haut-parleurs de façon à associer à chaque locuteur un haut-parleur. Tant que l'écart angulaire entre la localisation auditive et visuelle des personnes reste inférieur à  $11^\circ$ , les images sonores et visuelles sont subjectivement confondues [Thurlow & Jack, 1973] [Chateau, 1999]. Cette condition est scrupuleusement respectée dans les studios Varèse. Les images des locuteurs distants sont projetées sur des écrans dont la base mesure entre 1,5 et 2 mètres. Les participants étant situés à plusieurs mètres (de l'ordre de 2 à 5 mètres) de l'écran, il suffit de disposer trois haut-parleurs derrière chaque écran pour obtenir un écart angulaire maximal de  $15^\circ$  en azimut entre l'image sonore et visuelle des personnes distantes pour l'ensemble des positions d'écoute. Opérationnel depuis plusieurs années, le système Varèse a fait la preuve, dans son utilisation au quotidien au sein du C.N.E.T., ainsi que pour les présidents du consortium Global One, que la spatialisation sonore constitue un des enjeux majeurs des futurs systèmes de visioconférence [Mahieux *et al.*, 1998].

### 1.3 Mur de Téléprésence

L'introduction de la spatialisation sonore dans les nouveaux services de télécommunications procède aussi du souci d'accroître le *réalisme* de la restitution des scènes distantes. Le concept de "*mur de téléprésence*" reprend cette idée à un niveau beaucoup plus ambitieux: il définit une interface de communication audiovisuelle si transparente qu'elle donne aux différents participants l'illusion qu'ils partagent le même espace sonore et visuel, alors qu'ils sont situés en des lieux géographiquement bien distincts (cf. Fig. 1.2) [Miyoshi & Koizumi, 1992]... On cherche ainsi à se rapprocher des conditions "naturelles" d'une conversation en vue d'une interactivité optimale. On touche au domaine de la *réalité virtuelle*.

À l'échelle d'une communication téléphonique, l'intérêt du mur de téléprésence peut paraître discutable, mais il faut se rappeler qu'une des principales utilisations de la visioconférence concerne l'organisation de réunions à distance qui durent parfois plusieurs heures. Dans ces conditions, les utilisateurs deviennent très sensibles à tous les défauts du système et un désagrément mineur peut être perçu comme très gênant à l'issue d'une séance d'une journée. Parce que le service de visioconférence est essentiellement destiné à être utilisé sur la durée, la téléprésence constitue donc un enjeu primordial. À plus long terme, on envisage d'appliquer le principe du mur de téléprésence à un concept de *cafétéria virtuelle* afin de constituer des espaces d'échanges informels entre différents sites d'une entreprise en bénéficiant de la convivialité d'un lieu de détente. Il faut en effet avoir présent à l'esprit que c'est bien souvent dans le cadre de ces lieux d'échanges informels que les projets d'une entreprise prennent naissance ou progressent...

En pratique, la réalisation d'un mur de téléprésence suppose de concevoir un système de visualisation et de sonorisation capables de restituer en 3 dimensions les scènes distantes. Il importe que les positions et les mouvements des différents interlocuteurs soient restitués dans les locaux distants. De plus, les personnes participant à une visioconférence doivent pouvoir se déplacer librement dans la salle en conservant une perception 3D correcte — sur le plan sonore et visuel — de leurs interlocuteurs. Un rendu réaliste passe en outre par une reproduction de l'image des individus à l'échelle 1:1, c'est à dire en grandeur nature.

Le cahier des charges d'un mur de téléprésence s'attache donc à deux points:

- ▷ *restituer une scène sonore et visuelle en 3 dimensions*, ce qui implique que les informations de position et de mouvement des individus, à la fois en direction et en distance, soient convenablement saisies et reproduites,
- ▷ *offrir une zone de restitution étendue* pour permettre les déplacements des utilisateurs.

En outre, le système de restitution doit être *totalelement transparent du point de vue des utilisateurs* et en particulier il ne doit leur imposer ni contraintes, ni le port d'un équipement spécifique, tel qu'un casque ou des lunettes.

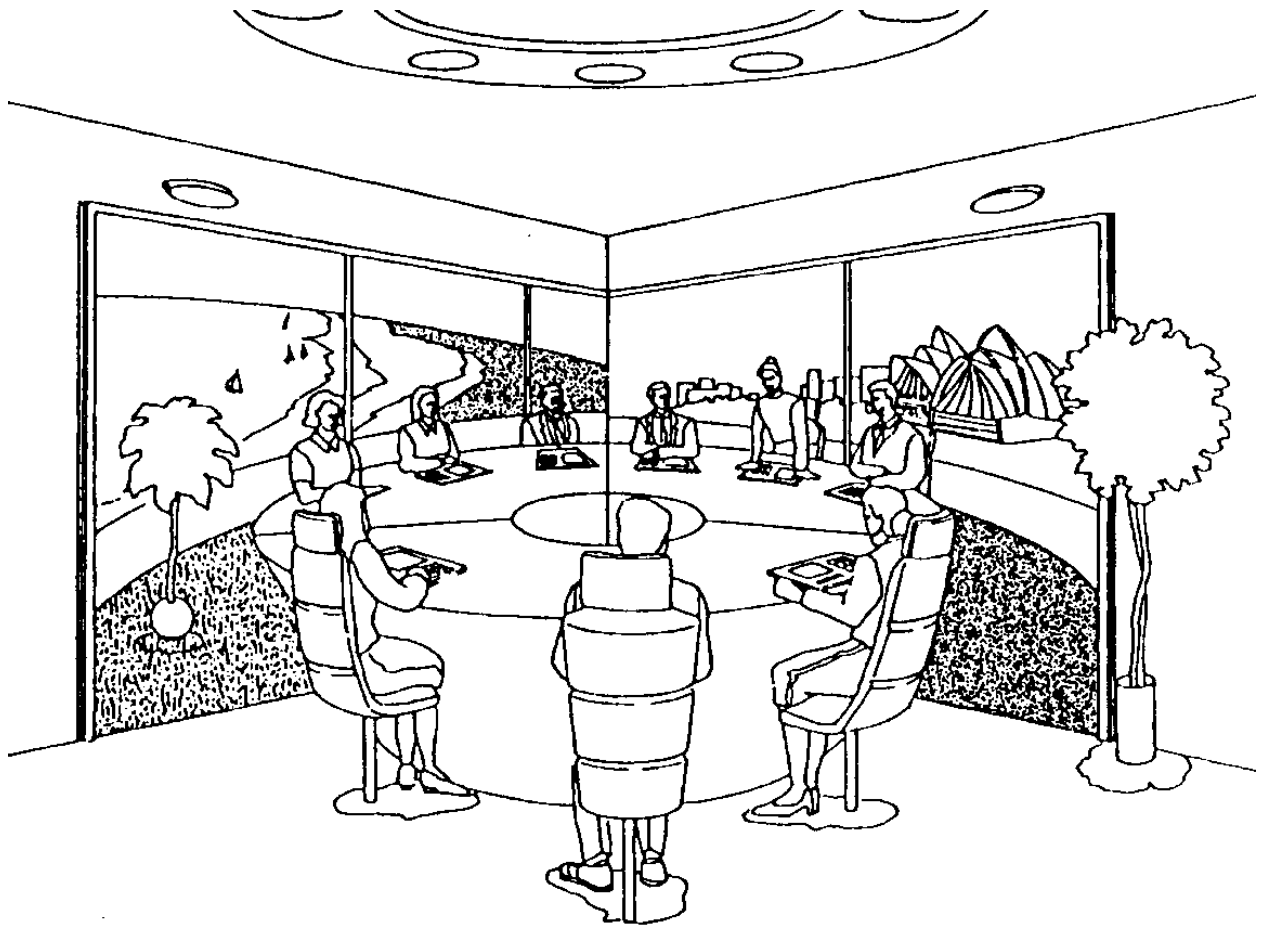


FIG. 1.2 - Concept du mur de téléprésence: Les participants ont l'illusion de partager le même espace que leurs interlocuteurs distants (localisés à Singapour et Sydney), d'après [Miyoshi & Koizumi, 1992].

## 1.4 Reproduction 3D d'un champ sonore sur une zone étendue

Dans la perspective de la téléprésence, nous nous concentrerons dans le cadre de la présente étude sur le problème de la reproduction 3D de la *scène sonore*, en laissant de côté tous les aspects concernant la chaîne vidéo du système de visioconférence. En d'autres termes, le problème que nous avons traité à travers ce travail de thèse concerne la *reproduction 3D d'un champ sonore sur une zone étendue* appliquée au contexte de visioconférence. Il existe plusieurs méthodes pour reproduire un champ spatialisé. Nous allons les décrire à présent, en mettant en évidence à la fois leurs performances, en terme de restitution sonore 3D et d'étendue de la zone d'écoute, et leur pertinence du point de vue du contexte applicatif de visioconférence.



## Références Bibliographiques

- BRIAN C. J. MOORE. (1997). *An Introduction to the Psychology of Hearing*. Academic Press, San Diego.
- CHATEAU N. (Mars 1999). Contribution of Perceived Audiovisual Spatial Fusion to Subjective Audiovisual Quality. *In: Collected Papers from the Joint Meeting "Berlin 99" (137<sup>th</sup> Meeting of the Acoustical Society of America / 2<sup>nd</sup> Convention of the European Acoustics Association)*.
- MAHIEUX Y., PETIT J.-P., EMERIT M., LE TOURNEUR G. & THOMAS J.-P. (1998). Le traitement du son pour la téléconférence. *L'écho des RECHERCHES*, 65-76.
- MIYOSHI M. & KOIZUMI N. (1992). NNT's Research on Acoustics for Future Telecommunication Services. *Applied Acoustics*, **36**, pp. 307-326.
- PLOMP R. (1977). Acoustical Aspects of Cocktail Parties. *Acustica*, **38**, pp. 186-191.
- THURLOW W.-R. & JACK C.E. (1973). Certain Determinants of the "Ventriloquism Effect". *Perceptual and Motor Skills*, **36**, pp. 1171-1184.



## Table des Illustrations

1.1	Principe d'une visioconférence . . . . .	26
1.2	Concept du mur de téléprésence: Les participants ont l'illusion de partager le même espace que leurs interlocuteurs distants (localisés à Singapour et Sydney), d'après [Miyoshi & Koizumi, 1992]. . . . .	29

---

