

Thèse  
présentée en vue de l'obtention du diplôme de  
Docteur de l'Université du Maine

Spécialité : Acoustique

Par  
Chrysanthie NATHANAIL

**Influence des informations visuelles  
sur la perception auditive.  
Conséquences sur la caractérisation de la qualité  
acoustique des salles.**

Soutenance prévue le 5 octobre 1999 devant le jury composé de :

Durand BEGAULT

Claude DEPOLLIER

Danièle DUBOIS

Catherine LAVANDIER

Jean Dominique POLACK

Olivier WARUSFEL



Thèse  
présentée en vue de l'obtention du diplôme de  
Docteur de l'Université du Maine

Spécialité : Acoustique

Par  
Chrysanthie NATHANAIL

**Influence des informations visuelles  
sur la perception auditive.  
Conséquences sur la caractérisation de la  
qualité acoustique des salles.**



Soutenance prévue le 5 octobre 1999 devant le jury composé de :

Durand BEGAULT

Claude DEPOLLIER

Danièle DUBOIS

Catherine LAVANDIER

Jean Dominique POLACK

Olivier WARUSFEL

# REMERCIEMENTS

Je tiens tout d'abord à remercier Michel Bruneau, la première personne de l'acoustique que j'ai rencontrée en France, pour sa confiance et son encouragement dans ce début, ainsi que pour les démarches administratives qu'il a entreprises à mon égard.

Mes remerciements les plus chaleureux vont à Xavier Meunier qui a travaillé d'interminables heures, souvent le week-end en se privant de sa famille, afin de m'aider à monter le dossier de demande du projet.

Mes remerciements vont également à Catherine Lavandier pour l'enthousiasme qu'elle a montré vis-à-vis du sujet de recherche, ainsi que pour les démarches qu'elle a effectuées afin de rendre possible le déroulement des expériences à l'Université de Cerqu-Pantoise.

Je voudrais remercier Cardyn Drake et Suzanne Winsberg pour m'avoir aidée à m'initier aux analyses statistiques employées dans le cadre ce travail.

Je tiens à remercier tout particulièrement Claude Tailleux, la personne qui m'a confié les techniques de la stéréoscopie. Je le remercie pour sa patience, pour son aide pratique inestimable, ainsi que pour les longues discussions que nous avons eues sur des sujets divers et variés.

Je ne peux oublier ceux qui ont participé à mes expériences et qui souvent se sont gentiment déplacés jusqu'à Cerqu pour subir – je le reconnais – des heures entières de tests. Je les en remercie sincèrement.

Je suis très reconnaissante à Michèle Castellengo qui eut la gentillesse de m'accueillir dans son laboratoire ces deux dernières années, j'apprécie sa discrétion qui fut une force motrice dans mon travail. Le LAM est un lieu d'échange exceptionnel, un toit scientifique chaleureux et humain. Je ne sais pas par lesquels des membres de son équipe commencer et par lesquels finir ; tous m'ont aidée : Corsin, Valérie, Frédérique, Benoît, Bertrand, Myriam, Jean-Yves, Catherine, Claire, Jean, Jean-Marc, Cécile, Nathalie, Gwenaél, Marc-Pierre, nos si gentils stagiaires de cette dernière année...

Je remercie beaucoup mes deux rapporteurs Danièle Dubois et Durand Bequait. La première de s'être plongée dans les eaux exotiques de l'acoustique "objective", et le second dans la langue française manipulée par une grecque. Ils ont été courageux...

Je tiens à remercier vivement Jean-Dominique Polack pour son aide scientifique, pratique et morale. Sa passion pour tout sujet, scientifique et autre, sa curiosité, son esprit caustique, sa rapidité de pensée souvent vertigineuse, son élégance administrative, sa présence tout simplement, m'ont encouragée, motivée, gênée, fait rire, enfin tout ce qui se passe dans une relation saine entre doctorant et directeur de thèse.

Si j'avais pu écrire une page entière pour remercier Olivier Warusfel, je l'aurais fait. Je me contenterai de dire qu'il est la personne qui se trouve derrière tout ce travail depuis la demande du projet jusqu'aux corrections méticuleuses du dernier moment. "Derrière" dans le sens de support, de discrétion, de confiance, de disponibilité, de gentillesse. C'est un homme que j'admire tant scientifiquement qu'humainement.

Ευχαριστω την μητέρα μου και τον πατέρα μου για την αγάπη τους. Είναι αυτοί που μου έδωσαν το ονειρο και την ψυχική δύναμη που χρειάζεται μια τέτοια προσπάθεια.

Je remercie Bertrand, mon mari, pour son énorme soutien pratique et moral, et son amour. Lui dédier ce travail est encore très peu pour exprimer ce que je lui dois...

Quant à Théophilos, mon fils, il n'y est pour rien. Juste pour une grande passion qui dure depuis quinze mois. Mais les passions stimulent...



# SOMMAIRE

<b>RESUME - ABSTRACT</b> .....	<b>11</b>
<b>1 INTRODUCTION</b> .....	<b>13</b>
<b>2 ETAT DE L'ART</b> .....	<b>17</b>
2.1 INTRODUCTION.....	17
2.2 ACOUSTIQUE DES SALLES.....	18
2.2.1 Introduction.....	18
2.2.2 Etudes antérieures.....	18
2.2.3 Les facteurs perceptifs.....	27
2.2.4 Réponse impulsionnelle et critères objectifs.....	29
2.2.4.1 <i>La dimension d'intensité</i> .....	30
2.2.4.2 <i>La dimension temporelle</i> .....	30
2.2.4.3 <i>La dimension spatiale</i> .....	31
2.2.4.4 <i>La dimension fréquentielle</i> .....	33
2.2.5 Corrélations entre facteurs perceptifs et critères objectifs.....	33
2.2.6 Problèmes liés à l'appréciation subjective.....	34
2.2.6.1 <i>Problèmes divers</i> .....	34
2.2.6.2 <i>L'influence de la perception visuelle</i> .....	35
2.2.7 Conclusion.....	38
2.3 PERCEPTION DE LA DISTANCE.....	41
2.3.1 Etudes antérieures.....	41
2.3.2 Indices de distance auditive.....	49
2.3.3 Discussion.....	50
2.3.4 Conclusion.....	53
2.4 PERCEPTION MULTIMODALE.....	54
2.4.1 Notions principales, domaines de recherche.....	54
2.4.2 Interactions sensorielles entre vision et audition.....	56
2.4.2.1 <i>Introduction</i> .....	56
2.4.2.2 <i>Interactions spatiales</i> .....	57
2.4.2.3 <i>Autres influences entre vision et audition</i> .....	61
2.4.3 Méthodes expérimentales et facteurs qui influencent les interactions audiovisuelles.....	62
2.5 DISCUSSION À PARTIR DE L'ÉTUDE BIBLIOGRAPHIQUE - OBJECTIFS.....	66

<b>3</b>	<b>MÉTHODE .....</b>	<b>69</b>
3.1	INTRODUCTION .....	69
3.2	LA SIMULATION SONORE.....	70
3.2.1	La synthèse des scènes sonores.....	70
3.2.2	Le module de traitement de signal du Spatialisateur.....	71
3.2.2.1	<i>Synthèse de la distribution temps / fréquence de l'effet de salle</i> .....	73
3.2.2.2	<i>Synthèse de la distribution spatiale</i> .....	75
3.2.3	Synthèse binaurale et reproduction transaurale.....	78
3.2.3.1	<i>La synthèse binaurale</i> .....	78
3.2.3.2	<i>La reproduction transaurale</i> .....	80
3.2.4	L'interface de contrôle 'perceptif' du <i>Spatialisateur</i> .....	82
3.2.5	Le choix de la méthode de synthèse sonore ( <i>Spatialisateur</i> ).....	85
3.3	LA SIMULATION VISUELLE .....	87
3.3.1	Prises de vues et restitution du relief .....	87
3.3.2	Les vues .....	88
3.4	SALLES ET MATÉRIEL DE SIMULATION. ....	94
3.5	MÉTHODES D'ANALYSE.....	97
3.5.1	Analyse de la Variance .....	97
3.5.1.1	<i>Introduction</i> .....	97
3.5.1.2	<i>Analyse de la variance à un facteur</i> .....	97
3.5.1.3	<i>Analyse de la variance à deux facteurs avec interaction</i> .....	100
3.5.1.4	<i>Analyse de Variance à Mesures Répétées</i> .....	102
3.5.2	L'Analyse en Composantes Principales.....	104
3.5.2.1	<i>Introduction</i> .....	104
3.5.2.2	<i>Tableau et transformation des données - Objectifs</i> .....	104
3.5.2.3	<i>Le nuage des individus</i> .....	105
3.5.2.4	<i>Le nuage des variables, les composantes principales</i> .....	106
3.5.2.5	<i>Relations de dualité</i> .....	107
3.5.2.6	<i>Valeurs propres, qualité de représentation, représentation simultanée</i> .....	108
<b>4</b>	<b>EXPÉRIENCES SUR LA DISTANCE AUDITIVE .....</b>	<b>113</b>
4.1	INTRODUCTION .....	113
4.2	PROCÉDURE EXPÉRIMENTALE GÉNÉRALE.....	115
4.2.1	Réglage subjectif du dispositif transaural. ....	115
4.2.2	Consignes.....	115
4.2.3	Référence .....	116
4.2.4	Sujets .....	116
4.2.5	Stimuli sonores et visuels.....	117
4.2.6	Types de tests, séquences et nombre de présentations .....	117
4.2.7	Ordre de présentation des stimuli sonores et des conditions visuelles.....	118
4.2.8	Réponses, échelle et normalisation .....	118
4.2.9	Durée des tests .....	119

4.3	EXPÉRIENCE PRÉLIMINAIRE .....	120
4.3.1	Objectif.....	120
4.3.2	Méthode.....	120
4.3.3	Résultats .....	121
4.3.4	Discussion.....	124
4.3.5	Conclusion.....	126
4.4	EXPÉRIENCE 1 : ÉTUDE DES STIMULI AUDITIFS .....	128
4.4.1	Objectif.....	128
4.4.2	Méthode.....	128
4.4.3	Résultats .....	130
4.4.4	Discussion.....	132
4.4.5	Conclusion.....	134
4.5	EXPÉRIENCE 2 : ÉVOLUTION DES RÉPONSES.....	135
4.5.1	Objectif.....	135
4.5.2	Méthode.....	135
4.5.3	Résultats .....	136
4.5.4	Discussion.....	140
4.5.5	Conclusion.....	141
4.6	EXPÉRIENCE 3 : INFLUENCE VISUELLE (VUES STATIQUES EN PROJECTION 2-D) .....	143
4.6.1	Objectif.....	143
4.6.2	Résultats .....	143
4.6.3	Conclusion.....	145
4.7	EXPÉRIENCE 4 : INFLUENCE VISUELLE (VUES CHANGEANTES EN PROJECTION 2-D) ...	146
4.7.1	Objectif.....	146
4.7.2	Résultats .....	146
4.7.3	Conclusion.....	146
4.8	EXPÉRIENCE 5 : INFLUENCE VISUELLE (VUES STATIQUES EN PROJECTION 3-D) .....	148
4.8.1	Objectif.....	148
4.8.2	Méthode.....	148
4.8.3	Résultats .....	150
4.8.4	Discussion.....	154
4.9	EXPÉRIENCE 6 : INFLUENCE VISUELLE (VUES CHANGEANTES EN PROJECTION 3-D) ...	157
4.9.1	Objectif.....	157
4.9.2	Méthode.....	157
4.9.3	Résultats .....	158
4.9.4	Discussion.....	164
4.10	DISCUSSION GÉNÉRALE.....	166
4.11	CONCLUSION .....	



<b>5</b>	<b>EXPÉRIENCES SUR UN ENSEMBLE DE FACTEURS PERCEPTIFS .....</b>	<b>173</b>
5.1	INTRODUCTION .....	173
5.2	EXPÉRIENCE PRINCIPALE.....	175
5.2.1	Objectif.....	175
5.2.2	Méthode.....	176
5.2.3	Résultats.....	181
5.2.3.1	<i>La partie unimodale.....</i>	<i>181</i>
5.2.3.2	<i>La partie bimodale – La perception des stimuli sonores.....</i>	<i>185</i>
5.2.3.3	<i>La partie bimodale – L'influence des images.....</i>	<i>192</i>
5.2.4	Discussion.....	202
5.2.4.1	<i>La perception des stimuli sonores.....</i>	<i>202</i>
5.2.4.2	<i>L'influence des images.....</i>	<i>205</i>
5.2.5	Conclusion.....	208
<b>6</b>	<b>DISCUSSION GÉNÉRALE .....</b>	<b>211</b>
6.1	INFLUENCE DES INFORMATIONS VISUELLES SUR L'APPRÉCIATION DE LA QUALITÉ ACOUSTIQUE DES SALLES. ....	212
6.2	EVALUATION SUBJECTIVE MULTIMODALE.....	214
6.3	DISCUSSION SUR LA MÉTHODOLOGIE – DIFFICULTÉS DE LA SIMULATION BIMODALE.....	217
6.4	PERSPECTIVES.....	219
<b>7</b>	<b>CONCLUSION GÉNÉRALE.....</b>	<b>221</b>
<b>8</b>	<b>ANNEXES .....</b>	<b>225</b>
8.1	ANNEXE A : MATÉRIEL DES EXPÉRIENCES.....	225
8.2	ANNEXE B : EXPÉRIENCES SUR LA DISTANCE AUDITIVE .....	226
8.3	ANNEXE C : EXPÉRIENCES SUR UN ENSEMBLE DE FACTEURS PERCEPTIFS .....	237
<b>9</b>	<b>BIBLIOGRAPHIE .....</b>	<b>249</b>

## RESUME

L'objectif de l'étude consiste en premier lieu à évaluer l'influence de la perception visuelle sur la caractérisation de la qualité acoustique des salles ; il vise à une amélioration des corrélations entre impressions auditives et indices objectifs descriptifs de cette qualité. Deux autres objectifs sont, d'une part la mise en place d'une méthode pertinente permettant d'étudier les interactions sensorielles entre vision et audition, d'autre part l'étude de l'influence des informations intermodales sur l'évaluation subjective en général.

Deux approches méthodologiques sont employées. La première, empruntée aux travaux de la psychologie expérimentale, permet d'étudier l'influence de la perception visuelle sur l'attribut perceptif de la distance auditive. La deuxième, inspirée de la méthode habituellement adoptée par les études de la qualité acoustique des salles, explore l'influence visuelle sur un ensemble de sept attributs perceptifs.

Les résultats montrent qu'une même distance auditive est jugée plus éloignée lorsqu'elle est présentée sous des conditions visuelles lointaines. Il s'agit d'une influence faible mais robuste, qui peut correspondre pour un quadruplement de la distance visuelle, à une variation perceptible sur l'indice objectif utilisé pour piloter la sensation de distance auditive dans les expériences. Egalement, une faible influence de la perception visuelle a été révélée sur les attributs sonores de l'enveloppement, de la réverbérance, et de l'impression générale. Du fait que la vision et l'audition utilisent le même cadre de référence spatial, les attributs spatiaux du son semblent plus susceptibles d'être influencés par la perception visuelle.

L'apparition et l'ampleur des interactions dépendent fortement de la qualité de la simulation bimodale, notamment de la compatibilité physique et cognitive des situations simulées. De plus, des paramètres évolutifs et attentionnels relatifs à la procédure expérimentale peuvent s'avérer déterminants.

## ABSTRACT

The objective of this study is to evaluate the influence of visual perception on the assessment of room acoustics quality; it aims at improving the correlation between auditory subjective impressions and objective indices describing this quality. Two other objectives are the development of a method appropriate to study sensory interactions between vision and audition as well as the study of the influence of intermodal information on subjective evaluation in general.

Two experimental approaches are adopted. The first approach, borrowed from the domain of experimental psychology, examines the influence of visual perception on the perceptible attribute of auditory distance. The second approach, inspired methodologies used in studies of room acoustic quality, investigates the visual influence on a set of perceptible attributes.

Results reveal that the same auditory distances are judged to be larger when presented under distant visual conditions. This influence is small but robust. For a multiplication of a factor of four of the visual distance, it corresponds to variations in the objective indices used to drive the impression of auditory distance in our experiments. Similarly, a small influence of visual perception was revealed on the auditory attributes of envelopment, reverberance and general impression. The spatial attributes of sound seem more likely to be influenced by vision because vision and audition use the same frame of reference for spatial judgements.

The actual appearance and the size of interactions depend strongly on the quality of the bimodal simulation, in particular on physical and cognitive compatibility of the simulated situations. Furthermore, evolution and attentional parameters related to the experimental procedure can play a determinant role.



# 1 INTRODUCTION

La qualité acoustique des lieux de spectacle est une notion complexe faisant appel à des réflexions à la fois artistiques et scientifiques. En effet, aujourd'hui encore, la question persiste de savoir si elle correspond au fruit d'un art ou au résultat d'une science bien maîtrisée.

Comme pour la plupart des phénomènes, la réponse n'est jamais simple. Notre point de vue englobe ainsi les deux aspects : il s'agit d'un art que la science essaie – avec un succès notable – d'expliquer. Ceci est par exemple valable pour la peinture : c'est un art, il repose donc essentiellement sur le talent du peintre, mais un peintre peut s'enrichir par l'étude des travaux de ses collègues, et même bénéficier de cours sur la peinture où la théorie des couleurs est enseignée. Ce qui est certain, c'est que le sujet de l'acoustique des salles continue de fasciner artistes et scientifiques, en partie à cause du mystère qu'il semble encore nous réserver.

En tous cas, une règle appartient au domaine de l'acoustique des salles : une salle est considérée avoir une bonne qualité acoustique si elle plaît à ses auditeurs, et non pas uniquement si elle mesurée objectivement comme 'bonne'.

Dans notre travail, nous nous intéressons bien évidemment à l'évaluation scientifique de cette qualité. De ce point de vue, deux approches complémentaires sont généralement adoptées :

D'une part, *l'approche perceptive* où la qualité acoustique est évaluée directement par les auditeurs et où l'on tente de comprendre sur quelles impressions ces derniers se basent pour juger cette qualité. Diverses études faisant appel à des analyses multidimensionnelles prouvent qu'il existe un ensemble limité de dimensions perceptives qui suffisent à décrire la variabilité des sensations des auditeurs. Les facteurs perceptifs les plus souvent proposées pour décrire ces dimensions sont : l'intensité sonore, la réverbérance, la précision, l'équilibre tonal et les facteurs liés à des effets spatiaux. Parmi ces derniers, se distinguent la distance et la direction de la source, l'enveloppement, la largeur apparente de la source et l'impression d'espace.

D'autre part, *l'approche objective* qui consiste à décrire de façon physique la modification que subit le signal sonore lors de sa propagation dans la salle. Elle vise plus particulièrement à traduire les facteurs issus de l'approche perceptive en grandeurs physiques mesurables. Les études qui se sont penchées sur ce problème de corrélations entre les domaines perceptif et objectif, ont ainsi permis de définir des indices objectifs corrélés aux facteurs perceptifs, dont les plus courants sont : la force sonore G, la durée de réverbération TR, la précision 80ms C80, l'efficacité latérale LE et le coefficient de corrélation interaural IACC.

L'importance de ces corrélations est évidente si l'on considère que la finalité d'une salle d'écoute est de plaire à ses auditeurs – la qualité acoustique est donc décrite de manière subjective – tandis que les

interventions sur l'acoustique d'une salle ne peuvent être que de nature physique, cette qualité *doit* donc être décrite de manière objective.

L'évaluation et le comportement de ces corrélations a fait l'objet d'une multitude de travaux qui ont mis en lumière leurs limites. Ainsi, deux salles auditivement très différentes peuvent présenter un jeu d'indices objectifs similaires, ce qui prouve qu'ils ne sont pas à même de rendre compte exhaustivement l'appréciation des auditeurs. Une amélioration des corrélations existantes est nécessaire, et cela semble d'autant plus vrai pour les facteurs perceptifs associés aux sensations liées aux effets spatiaux du son.

Le sujet de la présente recherche tire son origine d'une observation rapporté par plusieurs chercheurs en acoustique des salles : l'appréciation auditive de la qualité acoustique d'une salle de la part de ses auditeurs-spectateurs, s'explique mieux si l'on prend en compte certaines informations provenant de la perception visuelle. On remarque en effet que certains facteurs perceptifs, liés surtout à la dimension spatiale du son, comme la distance, l'enveloppement ou l'impression d'espace, évoquent la perception visuelle. Il est donc intéressant de tenter une amélioration des corrélations de ces indices en considérant une influence éventuelle des informations visuelles présentes dans les salles d'écoute, comme la distance des auditeurs par rapport à la scène (ou à la source sonore), aux murs latéraux, au plafond ou aussi le volume de la salle, l'éclairage, les couleurs, les formes, etc.

Malgré des références sporadiques relatives à une telle influence visuelle sur l'appréciation de l'acoustique des salles, très peu de travaux du domaine se confrontent au sujet. Dans la plupart des cas, les chercheurs le mentionnent comme l'un des problèmes de l'évaluation subjective, et tentent a posteriori d'expliquer des corrélations peu élevées ou des résultats peu attendus.

En outre, une pléiade de références bibliographiques provenant du domaine de la psychologie expérimentale et de la psychologie cognitive – mais également assez récemment, de plusieurs domaines de l'audiovisuel – met en évidence l'interdépendance des perceptions visuelle et auditive. Il s'agit en effet de deux processus qui interagissent et se renforcent mutuellement dans une relation complexe. L'appréciation d'un auditeur-spectateur résulte ainsi d'un traitement perceptif faisant intervenir ces deux modalités sensorielles, elle ne doit donc pas être étudiée sur la seule base des données acoustiques.

Par ailleurs, les travaux mentionnés ci-dessus concernent dans leur grande majorité des interactions spatiales entre vision et audition : ces modalités se réfèrent au même espace physique et interagissent fortement dans cette dimension. Si l'on considère que la perception visuelle est traditionnellement supposée l'emporter sur la perception auditive en terme de précision spatiale, on comprend que l'hypothèse d'une influence visuelle sur l'appréciation des attributs spatiaux du son, semble d'autant plus intéressante.

Cette discussion nous amène à présenter l'objectif de ce travail qui est multiple :

Il s'agit d'une part, d'étudier l'influence des informations visuelles sur l'appréciation de la qualité acoustique des salles afin de permettre une meilleure connaissance des souhaits des auditeurs-spectateurs et une amélioration de la pertinence de la caractérisation objective de cette qualité. Faire intervenir des informations issues de la perception visuelle (données architecturales et scénographiques) dans la définition des indices objectifs, permettrait d'affiner les corrélations avec les facteurs perceptifs, et faciliterait le travail et la collaboration des architectes et acousticiens .

D'autre part, notre travail vise à une meilleure compréhension du processus de l'évaluation subjective, évaluation qui implique dans bien des cas des informations issues d'autres modalités

sensorielles que la modalité directement concernée. Dans ce cadre, c'est l'influence des informations visuelles spécialement introduites dans le contexte expérimental qui est étudiée. L'influence d'autres facteurs de nature cognitive ne doit pas être négligée, surtout dans les cas d'expositions multimodales, cas particulièrement intéressants.

Enfin, étant donné le caractère novateur du sujet, l'un des objectifs du travail est d'établir une procédure expérimentale appropriée pour évaluer des relations entre les perceptions auditive et visuelle dans des environnements sonores et visuels complexes réels, tels que les salles de spectacle, mais aussi dans des environnements artificiels, propres aux systèmes de communication tels que le cinéma, la télévision, les ordinateurs "multimédia", etc.

Dans ce même but, les deux disciplines directement concernées par cette étude étant celles de l'acoustique des salles et de la psychologie expérimentale, ce sont les deux approches méthodologiques correspondantes qui s'offrent le mieux pour cette analyse ; le résultat attendu étant évidemment d'obtenir des informations complémentaires, mais aussi de les comparer.

Pour la première approche méthodologique, empruntée à la psychologie expérimentale, un seul attribut perceptif à la fois doit être étudié. C'est l'attribut de la distance auditive de la source sonore qui a été choisi, pour des raisons qui deviennent plus claires à la suite de l'étude bibliographique. Nous nous contentons de mentionner ici que ce choix a été principalement dicté par le fait qu'il s'agit d'un attribut acoustique spatial, qu'il existe par ailleurs des indications concernant l'influence de la distance visuelle sur différents facteurs perceptifs, et qu'enfin, cet attribut offre des possibilités pratiques pour la simulation visuelle.

Afin d'étudier l'influence visuelle sur la perception auditive et de pouvoir attribuer des effets éventuellement trouvés à la perception visuelle, il faut avoir la possibilité de diffuser le même son sous différentes conditions visuelles. Or, les stimulations visuelles et auditives ne sont jamais décorrélées dans les vraies salles de spectacle, ce qui exclut la possibilité d'effectuer des expériences in situ. En ce qui concerne la conception d'un plan expérimental adéquat, une contrainte majeure est donc que les environnements visuels et sonores doivent être décorrélés, c'est à dire contrôlés indépendamment. Le recours à une technique de simulation est donc nécessaire, tant pour l'environnement sonore que pour l'environnement visuel des expériences : la création des scènes sonores est assurée par un processeur d'acoustique virtuelle (le *Spatialisateur* de l'Ircam et Espaces Nouveaux) ; la simulation visuelle est créée à l'aide de la stéréoscopie conventionnelle (prises de vues et projection des images 3-D).

Les résultats permettent, d'une part de tirer des conclusions intéressantes par rapport aux conditions de simulation et à la procédure expérimentale que devraient respecter les recherches visant à étudier les interactions sensorielles, d'autre part, de mettre en évidence une influence faible mais robuste de la perception visuelle sur certains attributs perceptifs, notamment sur ceux liés à la dimension spatiale du son, comme la distance auditive et l'enveloppement sonore.

Enfin, nous présentons la répartition des chapitres suivants de ce travail :

Le deuxième chapitre révisé l'état de l'art dans les domaines de l'acoustique des salles, de la perception de la distance auditive et des interactions sensorielles entre vision et audition. Il présente à la fin une discussion basée sur l'étude bibliographique et détaille les objectifs de ce travail.

Le troisième chapitre présente la méthode expérimentale, en particulier l'outil de simulation sonore, les raisons de son choix, ainsi que la technique de simulation visuelle. Il présente également les salles

de tests, le dispositif expérimental et les méthodes d'analyse employées pour le dépouillement des résultats de ce travail.

Le quatrième chapitre présente les expériences qui traitent de l'influence visuelle sur la perception de la distance auditive. Sept expériences furent organisées en tout (une expérience préliminaire et six expériences principales), elles sont présentées et leurs résultats sont détaillés. Elles permettent d'étudier la méthode expérimentale, d'observer l'influence visuelle sur la distance auditive et de valider les effets trouvés. Des discussions intermédiaires sont présentées afin de mettre en évidence les conclusions les plus marquantes.

Le cinquième chapitre présente les expériences qui traitent de l'influence visuelle sur un ensemble de facteurs perceptifs, descriptifs de la qualité acoustique des salles : la *force sonore*, la *précision*, la *distance auditive*, la *réverbérance*, la *vivacité*, l'*enveloppement* et l'*impression générale*. L'expérience principale est constituée d'une partie uniquement auditive et d'une partie bimodale (simulation visuelle et auditive) ; les résultats de chaque partie sont présentés et discutés séparément. La discussion finale regroupe les principales conclusions.

Enfin, les derniers chapitres concernent la discussion générale portant sur l'ensemble des résultats de nos travaux et sur les perspectives qui s'ouvrent dans les différents domaines.

## 2 ETAT DE L'ART

### 2.1 INTRODUCTION

L'étude de l'influence des interactions sensorielles entre vision et audition sur la qualité acoustique des salles implique directement la discipline de l'acoustique des salles et celle des interactions audiovisuelles, cette dernière étant principalement étudiée dans le cadre de la psychologie expérimentale. La première et la troisième partie de ce chapitre ont donc pour but de réviser l'état de l'art dans ces deux domaines scientifiques en mettant l'accent sur les aspects relatifs au présent travail, notamment les problèmes liés à l'évaluation subjective et à la perception multisensorielle.

Etant donné qu'une grande partie de l'étude porte uniquement sur l'influence des informations visuelles sur la perception de la distance auditive, la deuxième partie de ce chapitre sera consacrée aux travaux spécifiques traitant de la perception de ce facteur de distance. Les raisons de cette attention particulière deviendront plus claires au cours de ce même chapitre. Nous nous contenterons de rappeler ici que :

Premièrement, l'un des buts initiaux de ce travail étant l'étude approfondie de la procédure expérimentale, l'utilisation d'un seul attribut perceptif s'imposait.

Deuxièmement, le choix de la distance auditive est basé sur l'intérêt qu'elle présente :

- du point de vue de la qualité acoustique des espaces, sur le fait qu'étant un attribut spatial du son, elle semblerait davantage susceptible de pouvoir être influencée par des informations visuelles,
- vis-à-vis de contraintes pratiques liées principalement à la simulation du champ visuel.

Après la description de l'état de l'art dans les trois domaines mentionnés précédemment : la qualité acoustique des salles, la perception de la distance auditive et les interactions sensorielles entre vision et audition, une discussion générale réunira les conclusions principales et justifiera les choix des objectifs.

Troisièmement, certaines salles présentent une situation paradoxale : pour exemple, tout au fond du dernier balcon, le son y paraît plus intense qu'à des places plus rapprochées de la scène, alors même que le niveau sonore mesuré y est plus faible<sup>1</sup>.

---

<sup>1</sup> J-D. Polack, communication personnelle.

## 2.2 ACOUSTIQUE DES SALLES

### 2.2.1 Introduction

L'objectif de ce chapitre est de présenter les étapes importantes de l'évolution de l'acoustique des salles, de décrire l'état des connaissances actuelles et d'introduire le contexte dans lequel s'inscrit le présent travail.

Dans ce parcours, nous porterons plus particulièrement notre intérêt sur :

- la notion de *multidimensionnalité* de la qualité acoustique des salles,
- la caractérisation subjective de la qualité acoustique à l'aide de facteurs perceptifs,
- les corrélations entre facteurs perceptifs et critères objectifs,
- les difficultés liées à l'appréciation subjective en général, et l'influence des facteurs issus de la perception visuelle en particulier.

### 2.2.2 Etudes antérieures

Depuis à peu près un siècle et jusqu'à nos jours, un grand nombre d'études se sont intéressées à l'appréciation perceptive de la qualité acoustique des salles. Ces études adoptent :

- soit une approche générale, essayant de définir les phénomènes perceptifs principaux qui structurent cette qualité ainsi que d'associer ces phénomènes à des grandeurs mesurables,
- soit une approche que l'on pourrait qualifier de complémentaire, concernant l'étude séparée des différents aspects perceptifs et leurs liaisons avec des grandeurs physiques.

Un certain nombre des études principales de ce domaine sont présentées ci-après.

C'est grâce à SABINE qu'une expertise empirique sur la construction des salles d'écoute prend l'ampleur d'une science. Dans son fameux article « Réverbération » premièrement publié en 1900 [Sabine, 1992], il constate la complexité du problème de l'acoustique des salles et le qualifie de *three-fold*, c'est à dire comme un problème *triptyque*. Il adopte ainsi une première approche multidimensionnelle en définissant trois conditions nécessaires et suffisantes pour une bonne écoute, concernant les aspects suivants :

1. la sonie (*loudness*),
2. la distorsion de sons complexes : effets d'interférence et de résonance,
3. la confusion : réverbération, échos et bruits extérieurs.

Sabine constate également que la solution du problème de l'acoustique des salles devrait passer par une caractérisation autant quantitative que qualitative. Pour les deux premiers facteurs de son modèle, il ne fournit pas d'indices mesurables, mais pour le troisième, la réverbération, il propose la première solution mathématique qui prévoit l'une des propriétés acoustiques des salles. Il s'agit du *temps de réverbération TR*, qui est une mesure de la *réverbérance*, la sensation auditive liée à l'effet de la salle (physiquement, elle correspond à la durée de décroissance de l'énergie acoustique dans une salle). Sabine l'estime à partir de la connaissance : des capacités d'absorption des matériaux, de leurs surfaces respectives et des dimensions de la salle. Il établit ainsi au début du siècle, une première corrélation et

définit le premier indice objectif, le *temps de réverbération TR*, utilisé encore largement aujourd'hui pour sa facilité de mesure et d'estimation<sup>1</sup>.

BERANEK [1962] publie un livre où il tente une approche générale de la notion de qualité acoustique. Il se base sur des entretiens avec des professionnels de la musique (chefs d'orchestre et critiques musicaux), mais aussi sur sa propre expérience d'écoute afin d'établir un questionnaire et de comparer entre elles 54 salles renommées de son époque. Le questionnaire est composé de 18 termes qu'il distingue en termes indépendants et termes dépendants, comme suit, les termes de Beranek en anglais sont traduits approximativement en français entre parenthèses :

Attributs indépendants positifs :

- intimacy (intimité)
- liveness (vivacité)
- warmth (chaleur)
- loudness of the direct sound  
(puissance sonore du son direct )
- loudness of the reverberant sound  
(puissance sonore du son réverbéré)
- diffusion (diffusion)
- balance (balance)
- uniformity (homogénéité)
- ensemble (ensemble)

Attributs indépendants négatifs :

- echo (écho)
- noise (bruit de fond)
- distortion (distorsion)
- non-uniformity (inhomogénéité)

Attributs dépendants :

- definition, clarity (clarté)
- brilliance (brillance)
- response, attack (attaque)
- texture (texture)
- dynamic range (dynamique)

Basée sur l'évaluation subjective de la qualité des salles, obtenue à l'aide des professionnels de la musique, l'étude de Beranek tente une classification objective. Il associe à ses termes descriptifs 'indépendants' des grandeurs physiques pour lesquelles il définit des plages de valeurs optimales. Les seules grandeurs de nature acoustique qu'il utilise sont les différentes mesures du temps de réverbération. Les autres indices sont issus de la description géométrique et architecturale de la salle. Ainsi, le retard des premières réflexions par rapport au son direct – mesuré géométriquement – décrit

---

<sup>1</sup> Sabine lui-même a beaucoup expérimenté sur la notion de TR optimum : il demandait à des musiciens d'ajouter ou de retirer des coussins, dont il avait au préalable mesuré l'absorption, jusqu'à ce que la réverbération les satisfasse [Sabine, 1993].

la sensation de l'intimité, la puissance sonore du son direct est estimée en faisant intervenir la distance entre l'auditeur et le chef d'orchestre, le degré de diffusion est décrit par les irrégularités architecturales du plafond et des murs, etc.

Beranek établit ainsi à partir de mesures acoustiques ou à partir de longueurs mesurées sur les plans d'une salle, une classification objective de 54 salles (*rating scale*) basée sur une note de préférence globale. Cette note est issue de la somme des notes obtenues sur chaque terme descriptif ; les notes optimales étant évidemment réservées pour les plages de valeurs optimales. Une pondération est attribuée à chaque terme d'après son importance ; dans cette pondération, le terme ayant le plus de poids (40%) est l'intervalle de temps entre l'arrivée du son direct et celle de la première réflexion (Initial Time Delay Gap, ITDG). Cet intervalle décrit, d'après Beranek, l'impression de l'intimité (ou de présence) du son, la sensation de se sentir proche des musiciens, d'être impliqué dans l'œuvre musicale. En ce qui concerne le poids des autres termes, chez Beranek, la vivacité (*liveness*) et la chaleur (*warmth*) comptent pour 15% de la note globale chacune, la puissance sonore du son direct contribue à 10%, la puissance sonore du son réverbéré à 6%, la balance et l'homogénéité ensemble à 6%, etc.

Le choix des termes descriptifs de Beranek ainsi que la distinction entre termes indépendants et termes dépendants reste discutable. En effet, il considère comme indépendants les termes pour lesquels il arrive à trouver des corrélations avec des grandeurs physiques, les autres termes étant considérés dépendants. Il existe aujourd'hui des techniques d'analyse multidimensionnelle qui permettent d'extraire des dimensions indépendantes d'une structure perceptive sur laquelle se base l'appréciation de la qualité acoustique, mais n'étant pas au point à l'époque, ni l'indépendance ni la dépendance des termes de Beranek n'est vraiment assurée.

Par ailleurs, il est possible que les musiciens interrogés aient utilisé des termes différents pour décrire une même impression, ou encore qu'ils aient utilisé un même terme pour décrire deux impressions différentes ; c'est un problème connu de *sémantique*<sup>1</sup>, c'est à dire lié à l'étude du langage considéré du point de vue du sens. Certains termes ont ainsi très probablement 'some meaning in common' comme le citent Hawkes et Douglas [1971]. Il s'agit en effet d'un problème courant de l'appréciation verbale nécessitant normalement des analyses linguistiques spécifiques.

Et puis, les musiciens interrogés se basent sur leur mémoire pour répondre, méthode qui peut évidemment introduire des erreurs dues justement aux limites de la mémoire auditive.

Dans son classement des salles, Beranek fait l'hypothèse que les attributs subjectifs descriptifs de la qualité acoustique des salles peuvent être additionnés de manière linéaire pour donner une note globale d'appréciation. Premièrement, la complexité de la qualité acoustique d'une salle ne peut pas être représentée par une seule note, donc sur un seul axe, deux salles ayant obtenu la même note pouvant très bien avoir un caractère acoustique différent, et par conséquent, elles peuvent *sonner* différemment.

Deuxièmement, l'hypothèse avancée par Beranek suggère qu'un défaut dans une salle peut être compensé par la qualité d'autres attributs perceptifs, par exemple, l'existence d'un écho peut être compensée par un niveau sonore ou un temps de réverbération optimal. Or, on sait que la qualité d'une salle est dans le cas général déterminée par l'absence de défaut [Knudson, 1931], [Vian, 1983], [Kahle, 1995] et que ces défauts ne peuvent que partiellement être compensés par des valeurs optimales sur les autres facteurs. Il existe donc des exigences *qualitatives* (manque de défauts) et ce n'est que d'après leur satisfaction qu'une approche *quantitative* devient vraiment intéressante.

---

<sup>1</sup> Du mot grec *sémantikos* : «qui signifie», «important».

Malgré ce classement « unidimensionnel » des salles, et un certain manque de rigueur quant au choix des attributs perceptifs et sur la distinction entre attributs indépendants et attributs dépendants, les travaux de Beranek restent très importants. Le nombre élevé des termes descriptifs utilisés a permis de constater la complexité de la caractérisation de la qualité acoustique des salles et de préparer le terrain pour introduire la notion de multidimensionnalité. La description des principaux effets subjectifs à l'aide des mesures physiques a introduit l'idée que cette qualité peut être évaluée de manière objective et a incité à la recherche de correspondances entre les domaines subjectif et objectif.

Enfin, Beranek introduit la notion de retard des premières réflexions par rapport au son direct. Il ne distingue pas les réflexions provenant des murs latéraux de celles provenant du plafond (si les réflexions provenant du plafond viennent avant les réflexions latérales elles sont aujourd'hui considérées comme nuisibles [Marshall, 1967]), mais il était le premier à signaler les effets subjectifs des premières réflexions et contribue à l'élaboration d'un premier groupe d'indices qui, avec le temps, a connu une importance croissante.

BERANEK dans un article publié récemment [Beranek, 1992], réduit et réorganise les attributs subjectifs présentés dans son livre. Il en extrait sept principaux qu'il considère nécessaires pour une bonne acoustique. Il ne base pas son choix sur une analyse particulière, mais plutôt sur sa longue expérience, en constatant simplement que 'la plupart des tests d'écoute sont en accord'. Ces sept attributs sont présentés en anglais :

1. *Reverberance*
2. *Loudness*
3. *Spaciousness*
4. *Clarity*
5. *Intimacy*
6. *Warmth*
7. *Hearing on stage*

Il fournit également une liste des caractéristiques acoustiques objectives nécessaires pour une bonne acoustique, reproduite ici en anglais :

1. *Acceptable reverberation time*
2. *Adequate loudness*
3. *A short initial-time-delay gap (ITDG) between the direct sound and the first reflection reaching the listener*
4. *A number of early lateral reflections that immediately follow the first reflection – these contribute to the feeling of “spaciousness”*
5. *A diffuse sound field created by an adequate number of reflections from all angles plus irregularities and ornamentation to eliminate acoustical “glare”*
6. *The ratio of energy in the first 80 ms to that in the next 2 s*
7. *Warmth of the sound by properly shaping the reverberation-time curve at low frequencies*

HAWKES et DOUGLAS [1971] constatent d'abord que la qualité acoustique des salles ne se juge pas sur le seul axe de “bonne / mauvaise” – la dimension appelée hédonique – comme Beranek l'avait suggéré, mais sur un ensemble de dimensions perceptives qui reflètent les différents composants de cette qualité. Ils présentent ainsi pour la première fois, une approche rigoureuse pour analyser les réponses subjectives obtenues dans les salles d'écoute et établissent la notion de multidimensionnalité.

Ils réalisent des expériences en conditions réelles de concert, la plupart dans le Royal Festival Hall à Londres. Leur questionnaire est composé de 16 paires de termes descriptifs opposés (échelles différentielles sémantiques) issus de la liste de Beranek.

Une analyse factorielle a permis de définir 5 dimensions orthogonales dans le Royal Festival Hall :

*Resonance (Reverberance)*

*Definition*

*Proximity*

*Balance and Blend*

*Brilliance*

Les auteurs ont trouvé que le nombre des facteurs indépendants révélés variait en fonction de la salle, du type de musique jouée, du degré de variation dans le contenu musical et dans l'interprétation.

Les résultats subjectifs ont pu être corrélés avec des différences objectives observées entre les quatre salles étudiées<sup>1</sup> : la réverbérance (*Reverberance*) a été corrélée au temps de réverbération *TR*, la brillance (*Brilliance*) à un long temps de réverbération dans les hautes fréquences, et l'intimité (*Intimacy*) à plusieurs paramètres : la proximité de l'orchestre, un court ITDG (initial time delay gap) et un rapport élevé entre la hauteur et la largeur de la salle (salle haute et étroite).

SCHROEDER, GOTTLÖB et SIEBRASSE ont publié une étude comparative d'une vingtaine de salles de concert [Schroeder et al. cités par Cremer et Muller, 1982], [Schroeder et al., 1974]. Une campagne de mesures et une série de tests subjectifs en laboratoire ont été menées parallèlement. Les sujets devaient donner des jugements de préférence.

L'enregistrement anéchoïque d'un extrait musical de Mozart était diffusé sur deux haut-parleurs placés sur la scène des salles à tester. Des enregistrements par tête artificielle ont été effectués en différents emplacements dans chaque salle. Les tests ont eu lieu en laboratoire en conditions anéchoïques où les signaux enregistrés ont été diffusés sur deux haut-parleurs à l'aide de la technique de 'cross-talk cancellation' (technique transaurale). Cette technique permet d'affecter à chaque oreille le signal sonore qui lui correspond, c'est à dire à l'oreille droite le signal enregistré par l'oreille droite de la tête artificielle, et de même pour l'oreille gauche. Les champs sonores étaient présentés par paires, et la puissance sonore subjective était rendue constante (égalisée) parce que, d'une part il a été mis en évidence qu'elle dominait les jugements de préférence, et d'autre part, le niveau sonore arbitraire des enregistrements, et par conséquent le niveau arbitraire de référence dans les expériences, ne permettait pas une restitution sonore fidèle (en ce qui concerne le niveau sonore).

Une analyse factorielle a permis d'extraire 4 facteurs significatifs, mais uniquement le premier d'entre eux était stable. Tous les sujets ont montré une corrélation positive avec ce facteur qui correspondait alors à un consensus entre les sujets. Les autres facteurs ont été attribués par les auteurs aux goûts personnels. Seules les corrélations de ce premier facteur avec les grandeurs objectives ont été prises en considération et sont présentées ci-dessous :

- La corrélation la plus élevée a été observée pour le rapport entre l'énergie précoce (< 50 ms) et l'énergie totale (rapport 'Definition' D50 d'après Thiele) mais elle était négative, ce qui indique qu'un son moins clair et donc plus réverbérant était préféré.

---

<sup>1</sup> Les dimensions issues de l'analyse globale sur les 4 salles étudiées sont : 'resonance', 'evenness', 'intimacy', 'definition', 'enjoyment' et 'brilliance'.

- La corrélation positive avec le *temps de réverbération* n'est pas surprenante, il semblerait d'ailleurs que cette grandeur pourrait être utilisée comme une mesure inverse du rapport entre l'énergie précoce avant 50 ms et l'énergie tardive, comme Barron l'a suggéré [Barron, 1978].
- La troisième mesure objective en corrélation également négative avec le facteur de consensus était celle du *coefficient de corrélation binaurale* ("*InterAural Cross Correlation*" ou IACC), un critère qui décrit la similarité des signaux sonores au niveau de chaque oreille de l'auditeur. Cette similitude est entre autre liée à la largeur de la salle : plus la salle est large plus le critère IACC augmente. Dans cette expérience, des valeurs basses de l'IACC étaient préférées, c'est à dire des impressions acoustiques propres à des salles plus étroites (entraînant des réflexions latérales plus fortes).

Ces résultats sembleraient être influencés par la méthode d'excitation des salles. La directivité acoustique des haut-parleurs était bien plus forte que la directivité d'un orchestre. Ainsi, un signal sonore diffusé sur haut-parleurs crée un champ sonore bien moins réverbérant que le champ sonore créé par un véritable orchestre. Dans cette étude, ce sont très probablement des champs sonores peu réverbérants qui ont incité les auditeurs à souhaiter un son moins précis – et donc plus réverbérant – et une plus grande taille subjective de la salle. Si les résultats de cette étude semblent influencés par la méthode expérimentale, ils sont néanmoins importants dans la mesure où ils montrent que dans le cas des lieux d'écoute peu réverbérants, le temps de réverbération devient déterminant.

Dans cette étude, la puissance sonore subjective des champs sonores a été égalisée parce qu'il a été observé qu'elle avait tendance à influencer les réponses. Ce fait a certes démontré l'importance de ce facteur sur la préférence, mais il a empêché l'étude de la structure réelle de l'espace perceptif, privé ainsi de l'une de ses dimensions fondamentales.

Par ailleurs, comme mentionné auparavant, il existe des salles qui ont des qualités acoustiques différentes, mais qui sont pourtant aussi bonnes les unes que les autres. Des tests subjectifs portant uniquement sur la préférence permettent de déterminer les facteurs liés à cette préférence – donc à un axe du type 'bon / mauvais' – mais ils empêchent l'apparition de facteurs qui peuvent ne pas être liés à la préférence, et qui contribueraient à une description plus complète de la qualité acoustique.

WILKENS et ses collaborateurs du groupe de Berlin ont mené une série de travaux importants présentés dans le livre de Cremer et Müller [Cremer et Müller, 1982]. Au cours d'une tournée de l'Orchestre Philharmonique de Berlin, des mesures acoustiques et des enregistrements par tête artificielle ont été réalisés – dans les salles non occupées – en des emplacements semblables pour chacune des six salles étudiées. Les salles ont également été excitées à l'aide d'une paire d'enceintes directives afin de comparer les résultats des deux méthodes d'excitation.

La reproduction du son pour les tests subjectifs a été réalisée en laboratoire sur casque audio. Le questionnaire proposé comprenait 19 paires de termes sémantiques opposés, reproduites sur la **figure 2.1** ci-après ; les sujets utilisaient des échelles différentielles de 1 à 6 pour donner leur jugement.

Un premier résultat de cette étude montre que les deux méthodes d'excitation – excitation par un véritable orchestre ou par une paire d'enceintes – présentent des différences perceptives significatives. Notamment, l'excitation par la paire d'enceintes semble créer des champs sonores qui sont perçus moins réverbérants, avec une taille subjective de la salle réduite (thèse de Wilkens cité dans [Barron, 1988]). Ceci dépend évidemment de la directivité des enceintes et des microphones d'enregistrement; il semblerait en effet que ce soit le diagramme de directivité des enceintes qui "fixe" le niveau de réverbération. Par ailleurs, ce résultat semble expliquer aussi les résultats issus des expériences de Schroeder, Gottlob et Siebrasse cités précédemment. La demande de davantage de réverbération et d'effets spatiaux de la part des sujets dans ces expériences semblerait être liée à la méthode d'excitation des salles.

	1	2	3	4	5	6
1. <i>klein</i> (small)	-----					<i>groß</i> (large)
2. <i>angenehm</i> (pleasant)	-----					<i>unangenehm</i> (unpleasant)
3. <i>undeutlich</i> (unclear)	-----					<i>deutlich</i> (clear)
4. <i>weich</i> (soft)	-----					<i>hart</i> (hard)
5. <i>brilliant</i> (brilliant)	-----					<i>matt</i> (dull)
6. <i>rund</i> (rounded)	-----					<i>spitz</i> (pointed)
7. <i>kräftig</i> (vigorous)	-----					<i>gedämpft</i> (muted)
8. <i>gefällt</i> (appealing)	-----					<i>gefällt nicht</i> (unappealing)
9. <i>stumpf</i> (blunt)	-----					<i>scharf</i> (sharp)
10. <i>diffus</i> (diffuse)	-----					<i>konzentriert</i> (concentrated)
11. <i>aufdringlich</i> (overbearing)	-----					<i>zurückhaltend</i> (reticent)
12. <i>hell</i> (light)	-----					<i>dunkel</i> (dark)
13. <i>verschwommen</i> (muddy)	-----					<i>klar</i> (clear)
14. <i>trocken</i> (dry)	-----					<i>hallig</i> (reverberant)
15. <i>schwach</i> (weak)	-----					<i>stark</i> (strong)
16. <i>höhenbetont</i> (emphasised tremble)	-----					<i>nicht höhenbetont</i> (tremble not emphasised)
17. <i>tieffenbetont</i> (emphasised bass)	-----					<i>nicht tieffenbetont</i> (bass not emphasised)
18. <i>schön</i> (beautiful)	-----					<i>häßlich</i> (ugly)
19. <i>leise</i> (soft)	-----					<i>laut</i> (loud)

**Figure 2.1:** Echelles différentielles sémantiques proposées par le questionnaire de Wilkens (traduction de l'allemand en anglais par Wilkens)

Les réponses ont été soumises à une analyse multidimensionnelle qui a révélé trois facteurs, expliquant presque 90% de la variance des réponses. La relation des termes descriptifs avec ces facteurs a permis l'interprétation suivante, donnée par ordre d'importance décroissante :

- *ampleur ou puissance du son*
- *précision d'écoute*
- *timbre*

Le travail de cette équipe présente la particularité que les salles ont été excitées à l'aide d'un véritable orchestre, permettant ainsi de respecter le réalisme des champs sonores. De plus, les mesures acoustiques et les enregistrements utilisés pour les tests subjectifs étaient effectués dans de mêmes conditions d'occupation des salles (la seule différence était la présence de l'orchestre lors des enregistrements<sup>1</sup>). Une plus grande confiance peut alors être accordée quant aux corrélations établies.

<sup>1</sup> Le réalisme des champs sonores est évidemment assuré dans les études optant pour des tests en conditions réelles de concert, mais les mesures acoustiques et les tests subjectifs dans ces études sont généralement effectués dans différentes conditions d'occupation des salles (mesures objectives en salles vides et tests subjectifs en salles occupées), ce qui n'est pas le cas ici.

Ces corrélations sont proposées par Lehmann et Wilkens [Cremer et Müller, 1982] et sont : la *force sonore*  $G$  pour l'ampleur du son, le *temps central*  $t_s$  pour la précision et la *penne du temps de réverbération EDT* en fonction de la fréquence (*Early Decay Time*) pour le timbre.

L'étude sur les préférences subjectives a révélé deux groupes de sujets : un groupe qui semblait avoir une préférence pour des sons amples, donc forts et réverbérants, et un groupe qui a montré une nette préférence pour des sons précis. La structure de l'espace perceptif de ces deux groupes, étudiée séparément, était composée des trois mêmes facteurs mentionnés auparavant, ce qui prouve que malgré les préférences particulières des individus, ils procèdent à la caractérisation de la qualité acoustique de la même manière, c'est à dire en se basant sur les mêmes dimensions perceptives.

Les facteurs perceptifs révélés par Wilkens et son équipe représentent trois dimensions principales de la perception : l'ampleur du son relève principalement de la dimension de l'intensité, la précision de la dimension temporelle, et le timbre de la dimension fréquentielle. La dimension spatiale, elle, n'apparaît pas de manière explicite, ni dans la liste des termes, ni dans les dimensions issues des tests<sup>1</sup>.

Une explication serait basée sur le fait que la diffusion binaurale du son au travers du casque présente au moins deux inconvénients potentiels, liés justement à la reproduction spatiale : la localisation du son "à l'intérieur de la tête" et les inversions "avant - arrière" [Blauert, 1996]. C'est peut-être à cause de cette représentation spatiale limitée que les aspects spatiaux du son sont absents de la liste, et évidemment des résultats.

Les travaux du groupe de Berlin confirment les hypothèses avancées par des études antérieures : que la perception de la qualité acoustique des salles est structurée à partir de facteurs perceptifs communs ; par ailleurs, l'observation de sous-groupes de sujets a permis de vérifier la stabilité du concept de facteur perceptif.

BARRON a publié en 1988 les résultats d'une étude menée sur onze salles de concerts britanniques [Barron, 1988]. Son but était de comparer ses résultats – obtenus en conditions réelles de concert – avec les résultats des travaux antérieurs, et plus particulièrement avec ceux du groupe de Berlin obtenus par des expériences en laboratoire.

Des auditeurs experts étaient priés de répondre à un questionnaire réduit, composé de neuf échelles différentielles, présenté sur la **figure 2.2**.

Les termes du questionnaire ont été choisis afin que les différentes échelles soient jugées indépendamment et que *l'impression générale* représente la seule variable dépendante.

L'analyse des réponses a, dans la plupart des cas, vérifié ces indépendances, sauf dans trois cas d'intercorrélations : entre réverbérance et enveloppement, entre enveloppement et intimité et entre intimité et puissance sonore. Le diagramme Venn de ces corrélations est reproduit à la **figure 2.3**.

Les corrélations observées mènent à la conclusion que l'enveloppement, introduit dans le questionnaire pour représenter les effets spatiaux, est interprété différemment par les sujets. Il semble exister un groupe de sujets qui l'associe à la réverbérance – et qui juge probablement les composants spatiaux du son réverbéré – et un autre groupe qui l'associe à l'intimité ; ce dernier juge un aspect du son commun à l'enveloppement et à l'intimité qui, d'après l'auteur, semblerait être la puissance sonore.

---

<sup>1</sup> Elle pourrait pourtant se manifester de manière implicite par les termes du questionnaire *small-large*, ainsi que par la dimension liée à l'ampleur du son.

CLARITY	Muddy	Clear
REVERBERANCE	Dead	Live
ENVELOPMENT	Expansive	Constricted
INTIMACY	Remote	Intimate
LOUDNESS	Loud	Quiet
BALANCE:		
	Weak	Loud
Treble re. mid-frequencies	----- -----	
Bass re. mid-frequencies	----- -----	
Singers/Soloists re. orchestra	----- -----	
BACKGROUND NOISE:		
	Inaudible	Acceptable
	Tolerable	Intolerable
OVERALL IMPRESSION:		
	Very poor	Poor
	Mediocre	Reasonable
	Good	Very good
	Excellent	
	N. B.: scale extremes refer to experience limits	

Figure 2.2 : Questionnaire proposé par Barron lors de son étude portant sur onze salles de concerts britanniques (reproduit de [Barron, 1988]).

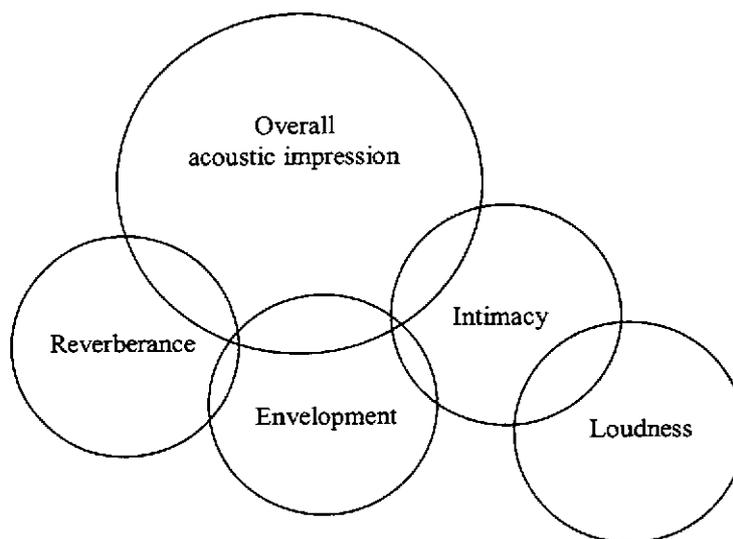


Figure 2.3 : Diagramme Venn de corrélations avec coefficients de corrélation supérieurs à 0.40 (reproduit de [Barron, 1988]).

Les résultats sur l'impression générale ont révélé trois corrélations significatives : avec l'intimité, la réverbérance et l'enveloppement (figure 2.3). L'enveloppement est très apprécié par l'ensemble des auditeurs, mais étant donné sa double interprétation, Barron distingue deux groupes d'auditeurs : ceux qui montrent une préférence pour la *réverbérance* et ceux qui préfèrent une bonne *intimité* du son.

L'étude des préférences dans les travaux de Berlin a aussi révélé deux groupes de sujets : ceux qui préféreraient une bonne *puissance sonore* (ampleur du son) et ceux qui préféreraient une bonne *précision*. Aucune ambiguïté quant à l'interprétation des facteurs perceptifs n'a été observée, mais dans le questionnaire de l'étude de Berlin, il n'y avait pas de facteur décrivant la dimension spatiale, notamment l'enveloppement. Une autre différence apparaît entre les deux études : les travaux britanniques ne révèlent pas la *précision*; très probablement parce que les salles de concerts britanniques présentent en général une bonne *précision* – elles varient donc peu sur cette dimension – ou aussi parce que la méthode de reproduction du son dans l'étude de Berlin favorisait la perception de cet attribut.

Finalement, dans son livre publié en 1993, Barron cite comme les plus importantes les impressions subjectives suivantes : *Précision, Réverbérance, Enveloppement, Intimité* et *Puissance sonore*. Il précise que cette liste n'est ni complète, ni définitive, mais qu'elle offre une base de référence utile pour des investigations ultérieures [Barron, 1993].

En ce qui concerne les corrélations avec des critères objectifs, les plus pertinents semblent être l'EDT (*Early Decay Time*) pour la réverbérance, et le niveau sonore pour la puissance sonore (critères observés également dans l'étude de Berlin). Les taux de corrélation entre la *précision* et le temps de réverbération (TR) ou l'EDT sont très faibles. L'enveloppement est corrélé avec le niveau sonore et, dans une moindre mesure, avec la fraction d'énergie précoce latérale  $L_F$  (Early Lateral Energy Fraction) définie par Barron dans [Marshall et Barron, 1981]. L'intimité est corrélée avec le niveau sonore mais aussi avec la distance entre la source et l'auditeur.

### 2.2.3 Les facteurs perceptifs

Dans la section précédente, nous avons pu constater qu'un nombre limité de facteurs perceptifs est suffisant pour modéliser la perception de la qualité acoustique des salles de concerts. Ces facteurs sont considérés comme des repères dans l'espace multidimensionnel de la perception et sont communs à tous les sujets. Par contre, l'importance attribuée à chacun d'entre eux ainsi que les plages de valeurs préférées peuvent varier d'un sujet à l'autre.

Les différents travaux traitant de l'appréciation subjective de la qualité acoustique des salles semblent souvent proposer des jeux de facteurs différents. Rappelons que les facteurs révélés par les travaux du groupe de Berlin [Cremer et Muller, 1982] sont *l'ampleur du son*, la *précision* et *l'équilibre spectral*. La dimension spatiale semble être absente<sup>1</sup>. Les travaux de Barron [Barron, 1988] révèlent encore la *puissance sonore*, en revanche, ils tiennent compte de la *réverbérance*, de *l'intimité* et de *l'enveloppement* – donc de la dimension spatiale – mais ils ne révèlent pas la *précision* ni la dimension spectrale. En fait, nous avons pu constater que le type de champ sonore testé (salle, musique, interprétation), les termes du questionnaire, la méthode expérimentale (surtout dans le cas de reproduction des champs sonores en laboratoire) et l'analyse employée, sont à l'origine des différences observées et peuvent être responsables de la disparition d'un facteur important ou de l'apparition d'un facteur secondaire.

Malgré ces différences, les études concordent de manière générale pour suggérer qu'un nombre limité de 5 à 6 facteurs serait suffisant pour expliquer la variabilité des impressions.

---

<sup>1</sup> Elle pourrait en partie être présente dans la notion de l'ampleur du son.

Ces facteurs sont :

1. La *puissance sonore* décrivant l'impression de l'intensité subjective globale.

2. L'*impression d'espace* qui décrit l'impression liée à la position et à l'étendue des sources sonores dans l'espace (réflexions et réverbération comprises). Elle décrit l'impression liée à la taille subjective de la salle, l'impression d'être dans un espace à trois dimensions<sup>1</sup>. Il est généralement admis qu'il existe au moins deux effets distincts [Barron et Marshall, 1977], [Griesinger, 1997] :

- *L'enveloppement* qui décrit le degré d'immersion des auditeurs dans le son, l'impression d'être entouré par le son. Une certaine ambiguïté existe par rapport à ces définitions : pour plusieurs auteurs *l'impression d'espace* et *l'enveloppement* décrivent une sensation identique ou très similaire<sup>2</sup> [Barron, 1971], [Barron, 1988].
- *la largeur apparente de la source* (Apparent Source Width, ASW) qui est plus liée à la source sonore, elle décrit l'élargissement subjectif de la source provoqué par la salle [Keet, 1968], [Griesinger, 1977].

3. L'*intimité* (aussi *présence*, *distance* ou *proximité*) qui décrit le degré d'identification avec l'œuvre, l'impression d'y être impliqué ou détaché, ou encore l'impression d'être proche ou loin de la source sonore. Elle semble apparaître uniquement lors d'expériences réalisées en conditions réelles de concert.

4. La *réverbérance* ou *impression de réverbération*, impression liée à la présence et à l'évolution du champ réverbéré, c'est à dire du son diffus dans une salle, non localisable, qui atteint les auditeurs après le son direct et les premières réflexions (discrètes et diffuses).

5. La *précision d'écoute* (aussi *définition*, *transparence* ou *netteté*), liée à la faculté de percevoir distinctement les différents composants des sons musicaux ou autres.

6. L'*équilibre tonal* ou *équilibre spectrale* qui décrit la sensation liée à l'équilibre entre basses, moyennes et hautes fréquences et à l'homogénéité du discours musical en général. Le *timbre* est une caractéristique similaire mais bien plus complexe englobant à la fois des notions spectrales et temporelles. Enfin, la *vivacité* et la *brillance* sont liées à la prééminence des hautes fréquences.

Un consensus sur l'importance relative des « *facteurs d'appréciation d'une bonne acoustique* » semble exister<sup>3</sup>, [Cremer et Müller, 1982], [Polack, 1982]. La hiérarchie généralement admise de ces facteurs est présentée ci-dessous, par ordre d'importance décroissante :

- Amplitude sonore
- Impression d'espace
- Précision
- Timbre ou Equilibre tonal
- Réverbérance

---

<sup>1</sup> Nous notons ici la description réputée de l'impression spatiale donnée par le Directeur du Concertgebouw d'Amsterdam 'the sensation of spatial impression corresponds to the difference between feeling *inside* the music and looking *at* it, as through a window' [Marshall, 1967].

<sup>2</sup> En effet, il semblerait que l'impression d'espace aux basses fréquences correspondrait plus à l'enveloppement tandis que l'impression d'espace aux hautes fréquences correspondrait plus à la largeur apparente des sources [Barron et Marshall, 1977].

<sup>3</sup> Marshall, communication personnelle.

Cette hiérarchie peut évidemment être influencée par la salle, le type de musique, l'interprétation, etc., et donc, la "véritable" hiérarchisation – celle qui correspond à une vraie salle – ne peut être établie que dans des situations réelles de concert [Kahle, 1995].

Enfin, et si l'on considère les quatre dimensions fondamentales : intensité, temps, espace et fréquence de tout phénomène physique – et donc de toute perception d'un phénomène physique – on remarque que la *puissance sonore* relève de la dimension d'intensité, la *réverbérance* et la *précision* principalement de la dimension temporelle, l'*impression d'espace* de la dimension spatiale et l'*équilibre tonal* de la dimension fréquentielle. L'*intimité* semble être surtout liée à la dimension d'intensité et à la dimension spatiale. Cette structure ne vient pas en contradiction avec la structure de l'espace perceptif décrit précédemment, elle en est tout simplement parallèle et nous sera utile pour l'organisation du chapitre suivant.

## 2.2.4 Réponse impulsionnelle et critères objectifs

L'appréciation subjective de la qualité acoustique d'une salle représente l'approche la plus pertinente pour sa caractérisation puisque le but de chaque salle est justement – et tout simplement – de satisfaire ses auditeurs. Cette appréciation ne pouvant pourtant pas se faire avant l'étape finale de la construction ou de la réhabilitation acoustique d'une salle, le passage par l'étape intermédiaire d'une description physique de la qualité acoustique est nécessaire. Les mêmes auteurs qui se sont intéressés à la définition des facteurs perceptifs, descriptifs de la qualité acoustique, ont mené parallèlement des études afin de définir des grandeurs physiques (mesurables ou estimables) qui puissent traduire de manière objective ces principaux effets subjectifs.

Ces grandeurs physiques, appelées improprement critères objectifs – ce sont en fait des indices – peuvent être issues dans leur grande majorité<sup>1</sup> de la réponse impulsionnelle de la salle, c'est à dire de la réponse de la salle à une impulsion (Dirac). Cette mesure décrit complètement la transformation apportée au signal sonore par la salle entre un point d'émission et un point de réception. Ainsi, dans l'hypothèse qu'une salle soit un système linéaire invariant dans le temps, si la réponse impulsionnelle est connue, il est possible de prédire la réponse de la salle à n'importe quelle autre excitation. Ceci est valable pour de mêmes emplacements de la source et du récepteur et de mêmes directivités respectives. On suggère alors que les valeurs d'un jeu d'indices objectifs est valable pour des positions et orientations bien déterminées de la source et du récepteur dans une même salle, car les valeurs de plusieurs critères pourraient varier de manière significative dans le cas contraire.

Il existe un très grand nombre d'indices objectifs proposés par les différents auteurs ; beaucoup de ces indices sont cependant communs à tous, à de faibles variations près. Ceux dont les corrélations avec les effets subjectifs semblent les mieux établies selon la littérature, sont présentés ci-après (cf. § 2.1.5). Ils sont ici distingués en quatre groupes d'après la dimension physique qu'ils décrivent : indices décrivant l'intensité, l'espace, le temps et la fréquence.

---

<sup>1</sup> Certains indices décrivant la perception spatiale nécessitent la mesure de la réponse impulsionnelle binaurale, enregistrée à l'aide d'une tête artificielle ou avec un microphone bidirectionnel.

### 2.2.4.1 La dimension d'intensité

Pendant longtemps, seules les variations relatives du niveau sonore à l'intérieur d'une salle ont été considérées, cependant l'accent a récemment été mis sur l'importance du niveau absolu.

Le premier indice est la *force sonore*  $G$ , définie par Lehmann [Cremer et Müller, 1982][Beranek, 1992]. Il s'agit du niveau sonore à un emplacement donné de la salle, normalisé par la puissance de la source (elle s'exprime par dix fois le logarithme du rapport de l'énergie émise par une source sonore impulsionnelle et omnidirectionnelle mesurée à un emplacement donné, sur l'énergie totale de cette même source mesurée dans une chambre anéchoïque et corrigée pour une distance standard de 10 m).

$$G = 10 \cdot \log \left( \frac{\int_0^x p^2(t) dt}{\int_0^x p_A^2(t) dt} \right) \quad \text{dB}$$

### 2.2.4.2 La dimension temporelle

Deux groupes d'indices sont principalement liés à la dimension temporelle : les indices décrivant la *durée de réverbération* et les indices décrivant la *précision*.

#### 2.2.4.2.1 Indices décrivant la durée de réverbération

La définition du *temps de réverbération*  $TR$  est donnée par Sabine [Sabine, 1932]. Il représente l'intervalle de temps que met l'énergie sonore globale dans une salle pour décroître d'un facteur de  $10^6$  (60 dB) après l'arrêt de toute émission sonore. Sabine a permis d'estimer le  $TR$  en fonction du volume  $V$  de la salle, de sa surface  $S$  et de l'absorption moyenne par mètre carré  $\alpha$ , la relation est la suivante :

$$TR = 0,16 \cdot V / (S \cdot \alpha)$$

La courbe de décroissance étant rarement régulière, plusieurs pentes peuvent y être mesurées et plusieurs mesures de la durée de réverbération existent. La plus importante d'entre elles est – avec le  $TR$  classique – la *durée de réverbération initiale* (*Early Decay Time, EDT*) qui décrit le début de la décroissance et correspond au temps nécessaire pour une décroissance de 0 à -10 dB (multipliée par un facteur 6) [Jordan, 1980], [Beranek, 1992]. Son usage répandu est justifié par le fait que le début de la décroissance n'est jamais statistique et diffère donc d'une place à l'autre.

#### 2.2.4.2.2 Indices décrivant la précision

Plusieurs indices ont été proposés pour décrire la précision. Il s'agit :

- soit du rapport énergétique entre l'énergie précoce – arrivant dans un certain intervalle de temps après le son direct qui est également inclus – et l'énergie tardive arrivant après cet intervalle ;
- soit du rapport énergétique entre l'énergie précoce et l'énergie totale.

Le principe de base est toujours le même : considérer l'énergie sonore initiale comme utile et l'énergie sonore tardive comme nuisible ou secondaire. Les indices les plus importants parmi eux sont la *Précision C80*, la *Définition D50* et le *temps central*  $t_s$  ou  $t_c$  :

La *Précision C80* ('Clarity' en anglais), proposée par Alim, Reichardt et Schmidt, est le rapport entre l'énergie sonore arrivant dans les 80 premières millisecondes et l'énergie arrivant après cet intervalle de temps [Kuttruff, 1979], [Cremer et Müller, 1982].

$$C = 10 \log_{10} \left( \frac{\int_0^{80 \text{ ms}} p^2(t) dt}{\int_{80 \text{ ms}}^{\infty} p^2(t) dt} \right) \quad \text{dB}$$

La *Définition D50* proposée par Thiele est le rapport entre l'énergie sonore arrivant dans les 50 premières millisecondes et l'énergie totale [Kuttruff, 1979].

$$D = \frac{\int_0^{50 \text{ ms}} p^2(t) dt}{\int_0^{\infty} p^2(t) dt} \cdot 100 \quad \%$$

Le *temps central*  $t_s$  traduit la position du centre de gravité de l'énergie de la réponse impulsionnelle. Il a été proposé comme un indice de remplacement, statistiquement mieux défini que les rapports énergétiques qui font intervenir des intervalles de temps fixes (50 ou 80 ms) et présentent un certain caractère arbitraire [Cremer et Müller, 1982], [Beranek, 1992].

$$t_s = \left( \frac{\int_0^{\infty} t \cdot p^2(t) dt}{\int_0^{\infty} p^2(t) dt} \right)$$

### 2.2.4.3 La dimension spatiale

Il existe deux groupes d'indices utilisés pour décrire la dimension spatiale :

Le premier groupe concerne des indices décrivant la proportion d'énergie sonore large bande arrivant latéralement par rapport à celle arrivant de toutes les directions. Deux de ces indices sont les plus répandus :

- l'*efficacité latérale* (Lateral Efficiency, LE) introduite par Jordan [Jordan, 1980]. Elle décrit le rapport entre l'énergie latérale arrivant entre 25 et 80 ms et l'énergie totale, provenant donc de toutes les directions, dans l'intervalle de 0 à 80 ms. L'énergie latérale est mesurée à l'aide d'un

microphone omnidirectionnel et un microphone bidirectionnel dont l'axe de sensibilité maximale est perpendiculaire à l'axe longitudinal de la salle.

$$LE = \frac{\int_{80 \text{ ms}}^{80 \text{ ms}} p_8^2(t) dt}{\int_0^{80 \text{ ms}} p^2(t) dt}$$

où  $p_8$  est la pression acoustique mesurée avec le microphone bidirectionnel (directivité en forme de 8).

- la fraction d'énergie latérale  $L_f$ , indice qui se calcule, introduit par Barron [Barron et Marshall, 1981], [Barron, 1993].

$$L_f = \frac{\int_0^{80 \text{ ms}} p^2(t) \cos \theta dt}{\int_0^{80 \text{ ms}} p^2(t) dt}$$

Le deuxième groupe d'indices est basé sur le degré de corrélation des signaux reçus par les deux oreilles des auditeurs ; ils sont généralement mesurés en large bande. Le plus répandu est la *cohérence interaurale* (Interaural Cross Correlation, IACC) introduite entre autres par Damaske et eto [Damaske et Ando, 1972], [Beranek, 1992] :

$$IACC(\tau) = \frac{\int_{t_1}^{t_2} p_L(t) p_R(t + \tau) dt}{\sqrt{\int_{t_1}^{t_2} p_L^2(t) dt \cdot \int_{t_1}^{t_2} p_R^2(t) dt}}$$

$$IACC = \max(|IACC(\tau)|), \quad -1 \text{ ms} \leq \tau \leq 1 \text{ ms}$$

IACC décrit le degré de ressemblance des signaux sonores au niveau des deux oreilles des auditeurs. Si ces signaux sont identiques, IACC est égale à 1 ; s'ils sont complètement différents IACC est égale à 0. Cette mesure nécessite des réponses impulsionnelles binaurales mesurées à l'aide d'une tête artificielle.

IACC peut être mesurée sur différentes plages temporelles. Ainsi,  $IACC_E$  correspond à une mesure de cet indice entre 5 et 80 ms, tandis que le  $IACC_L$  correspond à une mesure entre 80 ms et 3 s [Beranek, 1992].

#### 2.2.4.4 La dimension fréquentielle

Les indices décrivant la distribution fréquentielle de l'énergie sonore sont les moins bien définis [Jordan, 1980]. Divers indices ont été proposés pour caractériser la réponse d'une salle aux basses fréquences, ou aux fréquences aiguës, comme le rapport des durées de réverbération des fréquences moyennes sur celles des basses fréquences (*Bass ratio* de Beranek) [Beranek, 1996], ou encore, comme le rapport des énergies des différentes bandes d'octave [Barron, 1988]. Souloudre et Bradley ont récemment proposé – et testé en laboratoire – deux critères exprimant la réponse aux basses fréquences et aux aiguës [Souloudre et Bradley, 1995]. Aucun indice ne semble pourtant être généralement adopté. La pratique consiste simplement à observer la variation de l'énergie sonore, du TR ou de l'EDT, dans les différentes bandes d'octave.

#### 2.2.5 Corrélations entre facteurs perceptifs et critères objectifs

Les corrélations entre les facteurs perceptifs et les indices objectifs sont souvent loin d'être parfaites. A l'exception de la puissance sonore et de l'impression de réverbérance – pour lesquelles les indices objectifs correspondants sont pertinents – les autres impressions, surtout celles liées à la dimension spatiale et à la dimension fréquentielle, sont caractérisées par des corrélations relativement basses, souvent complexes et instables.

Malgré les difficultés, dues principalement à la nature de la caractérisation subjective, il existe un certain nombre d'indices consacrés par la littérature pour leurs bonnes corrélations avec les impressions subjectives. Les corrélations pertinentes sont révélées pour :

- La *puissance sonore* corrélée avec la *force sonore*  $G$  (travaux de Lehmann cités dans [Cremer et Müller, 1982], [Barron, 1988]).
- L'*impression d'espace* qui peut être décrite à l'aide de deux indices spatiaux : la cohérence interaurale large bande IACC et l'efficacité latérale LE. L'  $IACC_E$  (IACC mesurée entre 5 et 80 ms) semble plus adaptée, mais la corrélation élevée entre ces deux indices et la relative simplicité à mesurer l'efficacité latérale, fait que son usage est plus répandu. L'impression d'espace est également corrélée au niveau sonore global [Keet, 1968], [Barron, 1971], [Barron, 1993].

Enfin, il semblerait que ce sont principalement les basses fréquences qui contribuent à l'*enveloppement*, les hautes fréquences étant plutôt responsables de la sensation d'un *élargissement apparent de la source* [Barron et Marshall, 1981], [Blauert et al., 1984], [Griesinger, 1997].

- La *précision* qui est décrite par des indices comme C80 ou le temps central  $t_s$ , mais les corrélations ne sont pas toujours élevées. Elle est généralement inversement corrélée à différentes mesures de la durée de réverbération.
- L'*impression de réverbérance* qui est traduite par le temps de réverbération TR classique, mais des corrélations plus fortes ont été observées avec la durée de réverbération initiale EDT [Atal et al., 1965], [Jordan, 1980]. De plus, C80 peut aussi être une mesure (négative) de l'impression de réverbération.
- L'*intimité* qui semble principalement liée au niveau sonore [Barron, 1988]. Elle semble également liée à des aspects spatiaux du son : elle est souvent liée à l'*enveloppement* et à la *proximité* de la

source et des corrélations significatives avec la distance entre source et récepteur ont été trouvées [Hawkes et Douglas, 1971], [Barron, 1988]. De plus, un aspect spectral du son semble être présent : un son riche en hautes fréquences est un son perceptivement proche puisque les hautes fréquences sont rapidement absorbées par l'air, elles sont ainsi plus atténuées pour les signaux lointains [Coleman, 1968]. Enfin, Beranek propose un indice temporel : le retard temporel initial (Initial Time Delay gap, ITDG), l'intervalle de temps entre l'arrivée du son direct et celle de la première réflexion, grandeur qu'il a introduite en 1962. Il constate que le ITDG doit être utilisé en combinaison avec un rapport énergétique (énergie précoce / énergie tardive,  $C_t$ , où  $t = 50$  ms, 80 ms etc.) et avec l' $IACC_E$  mesurée entre 5 et 80 ms [Beranek, 1992].

## 2.2.6 Problèmes liés à l'appréciation subjective

### 2.2.6.1 Problèmes divers

Les difficultés rencontrées lors de l'établissement des corrélations entre indices acoustiques et impressions subjectives sont principalement dues à la nature complexe de l'appréciation subjective. Nous citons ici quelques phrases de Marshall<sup>1</sup> en anglais :

« The concert experience... is subjective, personal, and different for each person and from performance to performance. It is subject to mood, acoustics, presence of other audience, comfort, air conditions and view and above all it is intergraded. » Il entend par "intergraded" que les sensations provoquées lors cette expérience ne peuvent pas être dissociées, elles forment un événement perceptif solide et condensé dans le temps.

Il continue : « For each listener the concert experience depends not on an assemblage of discrete acoustical factors but on involvement in an Acoustical Gestalt<sup>2</sup> which like all Gestalten transcends the sum of its identifiable parts. This notion is easily widened to encompass the whole of the designed environment in which the concert takes place, in that the acoustical design is inseparable from the architecture. Indeed the fact of the enclosure, the building itself plays a central role in providing the Gestalt. »

L'appréciation de la qualité acoustique d'une salle est premièrement un processus fortement subjectif : il concerne des appréciations formulées par des personnes ou des groupes de personnes ayant chacune des modalités d'écoute qui leur sont propres. Aussi, les préférences individuelles sur la salle, mais aussi sur des qualités acoustiques autres que la qualité de la salle, comme le style de la musique, l'œuvre musicale, la qualité de l'orchestre ou le style de l'interprétation, peuvent influencer ce jugement.

Tout jugement subjectif étant éminemment relatif, l'appréciation d'un événement sonore dépend des sensations auditives qui l'ont précédé [Stevens, 1975], [Canevet, 1994/1995]. La mémoire auditive – à

---

<sup>1</sup> Marshall, communication personnelle.

<sup>2</sup> Le mot allemand 'Gestalt' correspond à *contour*, *silhouette*, dans ce cas il semble mieux traduit en français par *forme* ('La psychologie de la forme', [Guillaume, 1937]).

court ou à long terme<sup>1</sup> – sert donc de référence aux individus, donne lieu à des comparaisons souvent inconscientes et influence forcément l'appréciation de chaque nouvelle sensation.

De plus, toutes les expériences auditives précédentes sont, suivant leur impact, inscrites dans une sorte de base de référence et sont finalement – avec bien d'autres paramètres auditifs et non-auditifs – intégrées dans une structure cognitive qui forme l'expérience d'écoute.

La mémoire auditive et l'expérience d'écoute, mais aussi la réputation d'une salle ou d'un orchestre, peuvent ainsi créer des attentes par rapport à un événement et influencer l'appréciation de la qualité sonore.

D'autres paramètres non-auditifs pouvant influencer l'appréciation subjective peuvent être la disposition psychologique, des aspects culturels, l'influence du contexte, la sensation de confort, etc.

Et puis, percevoir et juger un son est une chose, communiquer ce jugement en est une autre. La méthode même de recueil des réponses peut influencer les résultats. Dans le cas des interviews ou des questionnaires – la méthode la plus répandue pour les études de la qualité acoustique des salles – le problème de la *sémantique* se pose, c'est-à-dire qu'un même terme descriptif peut être interprété différemment par les différents sujets. Ou au contraire, les sujets peuvent parler de la même sensation auditive en employant deux (ou plusieurs) mots différents. L'origine de ce phénomène est évidemment liée à des modes propres d'expression verbale, des habitudes ou des représentations linguistiques etc., mais aussi à l'expérience d'écoute et à la connaissance de l'œuvre musical, et donc à la finesse avec laquelle l'auditeur est capable d'analyser une scène sonore. En effet, plus on est « habitué » à une expérience – ou autrement dit, plus le niveau d'expertise est élevé – plus les nuances que l'on peut y discerner sont nombreuses<sup>2</sup>.

Préférences interindividuelles, mémoire et expérience, attentes, disposition psychologique, facteurs cognitifs et culturels, confort, sont tous des facteurs qui peuvent influencer l'appréciation auditive. Il manque alors les informations provenant des autres modalités sensorielles, et notamment celles provenant de la perception visuelle. Elles peuvent évidemment influencer l'appréciation auditive de la qualité acoustique des salles, d'autant plus que ces dernières sont également conçues afin de plaire visuellement à leurs spectateurs. Ces considérations sont séparément discutées ci-après.

### 2.2.6.2 L'influence de la perception visuelle

La manière dont les sons sont appréhendés dans un espace ne peut pas être indépendante des informations non-sonores, physiques, perçues par des modalités sensorielles autres que la modalité auditive. En effet, il est généralement admis que des informations issues de la perception visuelle dans des salles de concerts – ou faudrait-il dire de spectacle ? – comme des aspects architecturaux, la position des auditeurs par rapport à la scène ou par rapport aux parois, les couleurs, l'éclairage etc., peuvent influencer l'appréciation de la qualité acoustique.

Il faut noter ici que l'appréciation subjective de l'environnement visuel n'est pas non plus un processus perceptif déterministe sans particularités propres, et qu'il comporte des différences

---

<sup>1</sup> Il s'agit par exemple de mémoire à court terme lorsque des tests perceptifs en laboratoire ont lieu ou lorsque de différents emplacements dans une même salle sont testés, et de mémoire à long terme lorsque l'acoustique de deux salles différentes est jugée in situ, ce qui nécessite souvent des intervalles de temps importants.

<sup>2</sup> Approche théorique du développement perceptif sur 'l'affinement croissant de la différenciation perceptive grâce à des apprentissages naturels et artificiels' adopté par E.Gibson [Casalis, 1991].

interindividuelles. Cependant, l'impression visuelle peut de manière analogue au processus de l'appréciation auditive, être globalement décrite à l'aide de grandeurs visuelles mesurables (comme par exemple la distance de la scène ou des murs latéraux, le volume de la salle, les couleurs, etc.).

Il n'existe pas à notre connaissance de preuves rigoureuses ni de recherches spécifiques dans le domaine. L'influence visuelle est en général traitée comme un biais : un facteur qui *aurait dû* être contrôlé ou qui *n'a pas pu* être contrôlé. Elle donne lieu à des références courtes et sporadiques qui tentent *a posteriori* d'expliquer des corrélations pas toujours très élevées, ou le plus souvent, qui la mentionnent simplement comme encore une limite de l'appréciation subjective auditive.

Un auteur cite : « Recueillir des jugements par interview présente plusieurs inconvénients : la sémantique, la mémoire acoustique à long terme, les influences acoustiques autres que la qualité acoustique de la salle et l'influence d'autres paramètres non-acoustiques ». Il mentionne dans cette dernière catégorie "les aspects architecturaux", informations obtenues visuellement [Kahle, 1995].

Dans un autre texte : « Peuvent entrer dans l'appréciation [de la qualité acoustique des salles] des paramètres fortement déterminants n'ayant pas de relation directe avec la qualité acoustique d'un lieu, comme des dispositions psychologiques, l'aspect visuel, la sensation de confort ou ce qu'on a entendu dire sur la salle » [Vian, 1996].

Griesinger cite : « ... the author believes that in room acoustics the visual impression often dominates the aural impression... » [Griesinger, 1997].

Une équipe de chercheurs français chargée d'étudier l'emplacement optimal d'un nouvel orgue dans la Primatiale St Jean de Lyon, s'est vue confrontée au problème suivant : lors des tests perceptifs destinés à choisir la meilleure des trois places qui étaient étudiées, l'emplacement de l'orgue existant a été trouvé plutôt satisfaisant alors que ce sont justement des plaintes sur son acoustique qui avaient incité l'étude ! Or, lors des tests perceptifs réalisés à l'aide d'une série de haut-parleurs émettant les signaux musicaux de l'orgue, il s'agissait du seul emplacement accompagné de la vue de l'orgue existant (la série de haut-parleurs était placée au pied de l'orgue). C'est à la cohérence entre stimulation auditive et perception visuelle de l'orgue, que les chercheurs ont attribué ces résultats inattendus [Rocquigny, 1993].

Une autre étude française et deux travaux britanniques semblent se préoccuper un peu plus du problème et, dans les deux cas, publient des corrélations d'un indice visuel – la distance entre source et récepteur – avec des facteurs perceptifs auditifs :

POLACK et al. [1992] ont réalisé des évaluations subjectives pour 7 places du Théâtre Municipal du Mans à l'aide d'un questionnaire (adaptation en français du questionnaire de Barron (Figure 2.2) et, dans leur étude de corrélations avec les indices objectifs, ils constatent : « Detailed analysis of the answers reveals classification of positions according to *Overall Impression* to mirror classification according to sound level, especially when low frequency sound level is considered. However, both distance to the source and visual impression can also explain this classification, a not so surprising result since only at positions where listeners overlook the stage does the direct sound experience no masking ».

Polack, dans une discussion à propos des mêmes résultats cite « I have further comments which is on measuring envelopments, and the most amazing things which come out of our evaluation is that in small halls, ...,the best [envelopment] evaluation is in the middle of the hall... but according to all simulations which have been done, the best envelopments should be on the side. So my question is not what we are judging for envelopment, it is not related to our visual impression from the hall ?, in the middle you are enveloped by the hall » [SAS, 1994].

HOWKES et DOUGLAS [1971] (cf. 2.1.2) dans leur étude des dimensions perceptives basée sur les appréciations d'auditeurs dans 4 salles britanniques, trouvent que *Intimacy* et *Proximity* se confondent et tendent à exprimer souvent des dimensions perceptives semblables (corrélation entre échelles *distant / close* et *public / intimate* > 0.4). La dimension *enjoyment* (disliked very much / enjoyed very much) n'a pas pu être corrélée à des mesures acoustiques, la raison avancée par les auteurs était que ce facteur semblerait influencé par des préférences personnelles sur le type de la musique et sur l'interprétation, mais aussi *inconsciemment* par des aspects non-acoustiques, comme l'éclairage, le chauffage et le confort des fauteuils.

Ils constatent pour la totalité de leurs résultats : « The variation in mean factor scores indicates how the subjective response to acoustics differs with position in the hall. In many cases the scores seem to be a function of distance from the orchestra. Of course other independent factors must also influence the four sets of means obtained (Definition, Reverberance, Balance and Blend, Proximity). Otherwise the four factors would not be uncorrelated ».

Ils trouvent finalement une corrélation significative pour la totalité de leurs résultats (toutes salles confondues), entre *Intimacy* et la distance des auditeurs par rapport à l'orchestre.

BARRON [1988] dans son étude concernant plusieurs salles britanniques trouve également une corrélation élevée entre *l'intimité* et la distance entre la source et le récepteur ( $r = -0.59$ ) (notons que la corrélation de *l'intimité* avec les différentes mesures de niveau sonore est à peine plus élevée :  $r = 0.61$ ). Barron ne tente pas d'explication, se posant la question sur la nature de la sensation à laquelle les auditeurs répondent : « This begs the question, though, whether listeners are in fact responding to an acoustic sensation or not when they judge *intimacy* ».

De plus, Barron trouve une corrélation, étonnamment *positive*, entre le niveau sonore subjectif et la distance des auditeurs avec la source. Cela indique que les mêmes niveaux sonores sont jugés plus forts aux emplacements plus lointains. L'explication donnée par l'auteur est que les auditeurs créent des attentes sur le niveau sonore, en fonction de leur emplacement et de leurs connaissances sur le comportement acoustique des salles de concert. Ils jugent ainsi le niveau sonore d'après ces attentes en effectuant une sorte de 'compensation' de la distance qui les sépare de la source. Cette compensation est estimée par Barron à 0.09 dB/m, atténuation typique due à la distance dans les salles de concert.

Dans un travail plus récent Barron fait la première tentative d'incorporer une information provenant de la perception visuelle dans un indice ; il produit ainsi une équation de régression pour la puissance sonore [Barron, 1996]:

$$\text{Loudness} = 2.96 \cdot (\text{Tsl} + 0.076 \cdot \text{Dist}) + 35$$

Il obtient une corrélation élevée quand la puissance sonore est liée non seulement au niveau sonore global (Total sound level, Tsl), mais aussi à la distance entre source et récepteur ( $r = 0.82$  contre  $r = 0.77$ ).

La puissance sonore peut donc bénéficier de corrélations plus élevées en y incorporant des informations de provenance visuelle. Si même la puissance sonore, attribut perceptif si robuste et bien corrélé à des mesures physiques peut être influencée par des facteurs qui sont à la fois visuels et cognitifs (attentes), il est évident que la question se pose pour les autres facteurs perceptifs, dans leur majorité beaucoup moins bien représentés par des grandeurs physiques acoustiques.

Enfin, Barron remarque que ce type d'information ne serait pas révélé dans le cas d'expériences en laboratoire sans simulation visuelle.

Ajoutons ici que les travaux ayant opté pour une simulation en laboratoire s'affranchissent partiellement du problème puisque les conditions visuelles restent les mêmes. Elles sont pourtant rarement décrites : si par exemple le plan de la salle est souvent donné, en revanche la hauteur, les couleurs, la lumière, les distances des sujets par rapport aux parois etc. sont négligés (surtout si la restitution du son est effectuée sur casque audio ou si les expériences ont eu lieu dans des conditions anéchoïques). Si l'on accepte que les conditions visuelles particulières de chaque expérience peuvent influencer les résultats, cela rend les comparaisons entre ces différentes expériences assez délicates.

Et puis, nous pourrions nous demander si ce n'est pas en partie à cause des conditions visuelles relativement restreintes des salles d'expériences (anéchoïques ou non) que les effets spatiaux du son y semblent avoir le plus de mal à se manifester (travaux de Schroeder et de Wilkens par exemple (cf. § 2.1.2))

Par ailleurs, on remarque que les termes mêmes qui sont utilisés pour décrire certains facteurs perceptifs – liés surtout à la dimension spatiale du son – évoquent la perception visuelle. Ainsi l'impression d'espace, la proximité, la largeur de source et même l'intimité, représentent des sensations auditives mais ont aussi des équivalents visuels directs ; ce qui n'est pas le cas pour la puissance ou la précision du son par exemple. En effet, dans le chapitre concernant les interactions entre audition et vision – modalités d'ailleurs qualifiées de spatiales – nous pourrions constater qu'elles utilisent le même cadre de référence pour les jugements spatiaux, c'est-à-dire l'espace qui nous entoure (cf. § 2.3).

En effet, dans le cas général où les auditeurs sont en même temps spectateurs, la perception simultanée de cet équivalent visuel est inévitable. Des représentations mentales<sup>1</sup> entre attributs équivalents des perceptions auditive et visuelle peuvent donc influencer le jugement des auditeurs ; ils répondent ainsi à une sensation plus générale – une sensation bimodale – bien que pensant répondre uniquement à la sensation auditive ou unimodale.

On cite ici Vian [1983] qui en proposant une liste des "impressions subjectives ou qualitatives attribuées à une salle" définit l'intimité en nous renvoyant directement son équivalent visuel :

« Intimité ou Présence : l'écoute revêt un caractère intime lorsque l'on a l'impression d'être proche des instrumentistes, par analogie avec l'intimité visuelle offerte par les petites salles ».

Et puis, il n'y a pas seulement les informations que l'auditeur-spectateur aperçoit visuellement de son emplacement, mais aussi les informations visuelles qu'il a pu obtenir sur la salle lors de son entrée ou pendant l'entracte par exemple. Ces informations, en fonction de son expérience des salles de concert, lui créent des attentes par rapport à la qualité sonore du lieu ; ces attentes pourraient évidemment influencer son appréciation.

## **2.2.7 Conclusion**

Un grand progrès a incontestablement été réalisé dans les dernières décennies sur le domaine de la qualité acoustique des salles d'écoute. Il apparaît aujourd'hui évident que cette qualité ne se définit pas sur une seule dimension perceptive – du type 'bonne / mauvaise' par exemple – mais sur un ensemble limité d'attributs perceptifs traduisant les sensations auditives principales et – en règle général

---

<sup>1</sup> C'est à dire la manière dont l'information est stockée ou 'représentée' dans la mémoire.

indépendantes – provoquées chez les auditeurs présents dans une salle. Ce jeu de facteurs perceptifs a pu être défini et hiérarchisé.

Parallèlement aux travaux qui se sont penchés sur la définition des facteurs perceptifs, de nombreuses recherches ont été réalisées afin de trouver des grandeurs physiques capables de décrire au mieux ces sensations auditives. Une pléiade d'indices objectifs a ainsi été proposée offrant des corrélations qui en général, peuvent être considérées satisfaisantes, mais qui dans plusieurs cas peuvent encore bénéficier d'améliorations.

Ainsi, si les corrélations obtenues pour les sensations liées à l'intensité et à la dimension temporelle sont suffisamment robustes, comme par exemple celles pour la puissance sonore et pour la réverbérance, celles obtenues pour les sensations liées à la dimension d'espace et à la dimension fréquentielle semblent moins bien établies. Or, la hiérarchisation des facteurs perceptifs liés à une bonne qualité acoustique nous permet de constater que la demande d'une sensation sonore riche en effets spatiaux est très importante ; ces effets spatiaux apparaissent parmi les mieux placés dans les préférences des auditeurs, juste après l'exigence primordiale d'un niveau sonore satisfaisant.

Il n'est donc pas étonnant qu'après une longue période où les recherches sur la qualité acoustique se sont principalement orientées vers les aspects temporels des champs sonores créés dans les salles, la recherche de ces dernières décennies se penche principalement, et avec un intérêt croissant, sur les aspects spatiaux de cette qualité. Si l'impression d'espace en tant qu'enveloppement et largeur de source apparente est un peu plus étudiée, la perception de la distance dans les salles de concerts semble relativement négligée<sup>1</sup>. Ceci malgré la forte corrélation – voire même chez certains auteurs, l'identification – qu'elle présente avec l'intimité, un attribut auditif d'importance primordiale.

Les corrélations faibles trouvent leur origine essentiellement dans la particularité de l'appréciation subjective. Ainsi, nous avons vu que des paramètres acoustiques autres que ceux liés à la qualité acoustique de la salle peuvent l'influencer, tels que le style de la musique, l'œuvre musicale, la qualité de l'orchestre ou encore le style de l'interprétation. De plus, les préférences individuelles, la disposition psychologique, la sémantique (problèmes liés à la signification des termes ou des événements), la mémoire, l'expérience, les attentes créées par rapport à un événement, d'autres facteurs cognitifs et culturels, le confort, etc., sont des facteurs qui peuvent également influencer l'appréciation auditive. Enfin, des informations provenant d'autres modalités sensorielles et – partant du principe que *par définition* les auditeurs sont également spectateurs – ce sont notamment les informations visuelles qui peuvent être à l'origine de jugements d'appréciation auditive biaisés.

Parmi tous ces paramètres influençant l'appréciation subjective de la qualité acoustique, seules les informations visuelles<sup>1</sup> sont de nature physique et peuvent en conséquence être décrites à l'aide d'indices facilement mesurables et reproductibles. Ces informations visuelles (les aspects architecturaux, la position des auditeurs par rapport à la scène ou par rapport aux parois de la salle, les couleurs, l'éclairage, etc.) permettent, de manière analogue à l'appréciation auditive, de décrire l'appréciation subjective visuelle des auditeurs-spectateurs.

En effet, un certain nombre de travaux qui n'ont pas étudié le sujet spécifiquement, ont pourtant observé des influences de nature visuelle sur leurs résultats. Les facteurs visuels qui ont changé durant

---

<sup>1</sup> Ceci est probablement dû au fait que les études qui visent à homogénéiser l'écoute dans une salle, ont principalement pour objectif de maintenir une *présence* aussi grande que possible pour tous les spectateurs (ce qui correspond à une distance auditive courte).

<sup>2</sup> La sensation de confort pourrait aussi être décrite en partie à l'aide de grandeurs physiques, mais c'est une notion bien plus complexe englobant des informations provenant de toutes les modalités sensorielles : informations visuelles, thermiques, olfactives, tactiles, vibratoires, etc.

ces expériences étaient la distance entre l'auditeur et la scène et l'emplacement de l'auditeur par rapport aux parois de la salle. Les facteurs subjectifs auditifs qui semblent être influencés dans ces études étaient la proximité de la source sonore (ou éloignement, distance, présence, etc.), l'enveloppement, l'intimité, mais aussi l'impression générale et assez étonnamment, dans l'une des études, la puissance sonore.

Les impressions liées à la dimension spatiale du son semblent donc plus susceptibles d'être influencées par la modalité visuelle. Si à cette observation, on associe des conclusions empruntées au domaine des interactions sensorielles, où les deux modalités de l'audition et de la vision sont qualifiées de *spatiales* et où il est généralement admis qu'elles interagissent fortement, surtout dans cette dimension (spatiale), on comprend l'importance d'une étude de l'influence de la perception visuelle sur les attributs spatiaux des champs sonores. Des descripteurs physiques visuels pourraient donc être incorporés à l'élaboration d'indices objectifs dans le but d'affiner les corrélations existantes.

## 2.3 PERCEPTION DE LA DISTANCE

### 2.3.1 Etudes antérieures

La localisation des sources sonores se base sur une représentation sphérique, à savoir, elle se caractérise par les trois coordonnées suivantes : la distance, l'azimut et le site (souvent improprement appelé élévation par anglicisme). Nous nous intéresserons plus particulièrement dans cette partie à la perception de la distance.

Quelques uns des travaux effectués antérieurement dans le domaine de la perception auditive de la distance sont passés en revue ci-après. Ils présentent à la fois les procédures expérimentales employées et les résultats correspondants, les deux éléments étant liés entre eux.

Les recherches initiales dans le domaine ont porté sur la perception de la distance sous des conditions anéchoïques :

COLEMAN [1962] a étudié la localisation en distance d'un son inconnu (bruit blanc) en tant que fonction du nombre de ses présentations répétées. L'expérience se déroulait à l'extérieur sur un lac gelé couvert de neige. Quatorze haut-parleurs étaient utilisés dans le plan médian de face, la distance qui les séparait était approximativement de 0,65 m (2 ft), le plus proche était à 1,65 m (5 ft) des sujets et le plus éloigné à 10 m. Seuls les 10 haut-parleurs du milieu étaient utilisés. Le sujet et les haut-parleurs étaient surélevés afin de limiter les effets de sol masquants, mais de manière à permettre un trajet direct entre chaque émetteur et la tête du sujet. Le signal sonore était supposé être inconnu pour les 20 sujets : 1 seconde de bruit large bande avec une pression acoustique de 65 dB, probablement linéaire, mesurée à 33 cm (1ft) du haut-parleur. Chaque sujet recevait en tout 100 stimuli (10 stimuli par haut-parleur) ; la séquence des stimuli était ordonnée de façon à ce que, pour chacun des sujets, le même haut-parleur apparaisse en positions n°1, n°11 et n°100. Les sujets étaient divisés en quatre groupes correspondants à un haut-parleur particulier (3, 5, 7 ou 9 m). La tâche des sujets consistait à identifier le haut-parleur émetteur pour les 100 stimuli.

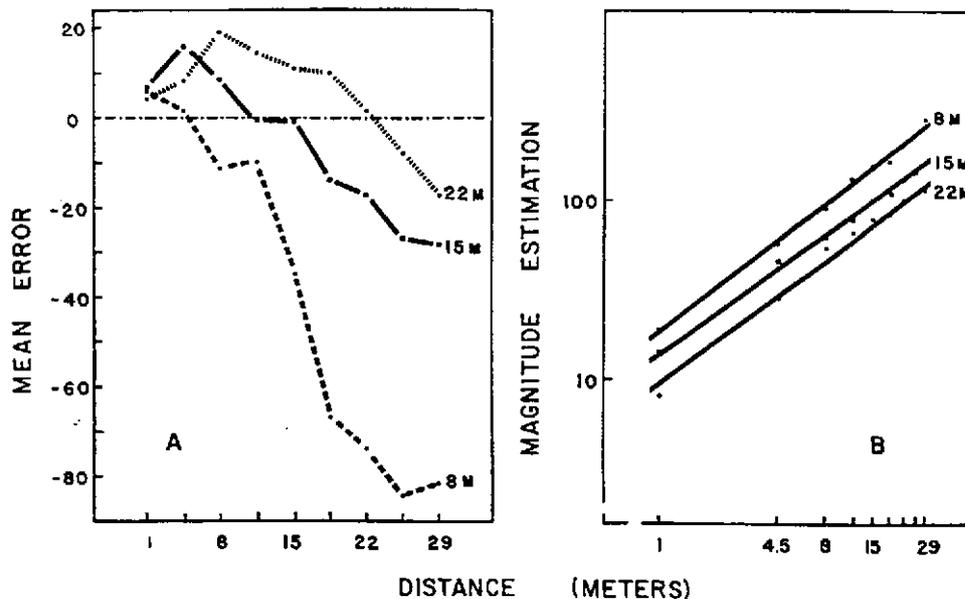
Les résultats ont montré que les jugements de distance à la première présentation des stimuli n'étaient pas liés à la distance physique. La distance la plus proche de 3 m était légèrement surestimée tandis que les trois autres étaient sérieusement sous-estimées. Ainsi, le premier son était localisé entre 2,2 et 3,6 m, quel que soit le haut-parleur activé. Néanmoins, pour les stimuli n° 11 et n°100, l'Analyse de Variance a révélé une différence significative entre les réponses concernant les 4 haut-parleurs. Cela montre que, au moins à partir du 11<sup>ème</sup> stimulus (2<sup>ème</sup> présentation du même haut-parleur), les erreurs de localisation deviennent beaucoup plus faibles. Dans tous les cas, les distances perçues pour les 4 haut-parleurs restaient inférieures aux distances physiques. Coleman a remarqué ainsi qu'un certain entraînement était nécessaire de la part des sujets afin qu'ils puissent utiliser l'intensité d'un signal sonore inconnu comme indice pour juger sa distance.

COCHRAN, TROOP et SIMSON [1968] ont étudié la capacité à estimer la distance sous des conditions d'écoute relativement naturelles. La procédure utilisée était la méthode de l'estimation de grandeur (*magnitude estimation method*) de différentes distances par rapport à une distance de référence.

La première expérience, effectuée à l'extérieur sur herbe, était menée grâce à un orateur réel qui parlait à cinq distances différentes (de 1 à 29 m) en cinq azimuts de l'avant vers l'arrière avec des intervalles de 45°. La distance de référence, de 15 m, était présentée tous les cinq stimuli. Chacun des vingt sujets jouait 75 stimuli. Les sujets avaient les yeux bandés, la moitié d'entre eux pouvait bouger la tête,

l'autre moitié n'y était pas autorisée. Les résultats – exploités à l'aide de l'analyse de variance – ont surtout montré une tendance à sous-évaluer les distances, d'une manière généralement proportionnelle à la distance jugée, mais toutefois avec une différence significative entre les réponses de chaque individu. Sur un graphe en coordonnées logarithmiques, les réponses apparaissent comme une fonction linéaire de la distance de la source ; ce qui signifie que la perception de la distance suit une loi de puissance. Il n'y avait pas de différence de réponse selon les azimuts ou selon les mouvements de la tête.

Dans une deuxième expérience, également en extérieur, différentes distances de référence étaient utilisées, mais un seul azimut. Les stimuli dans cette expérience – comme pour la troisième – étaient enregistrés et reproduits sur haut-parleurs, mais comportaient les mêmes phrases que dans le premier cas. Les résultats ont montré une différence significative entre les réponses selon les trois distances de référence employées. La tendance à sous-estimer les distances lointaines diminuait avec une distance de référence croissante. Les graphes logarithmiques réalisés pour les trois distances de référence étaient également linéaires ; les pentes étaient indépendantes de la distance de référence (figure 2.4).



**Figure 2.4 :** A. Erreur moyenne pour l'estimation des sujets utilisant l'une des trois distances de référence. B. Tracé log-log des estimations des grandeurs moyennes normalisées par rapport aux trois distances de référence (reproduit de [Cochran et al 1968]).

La troisième expérience fut menée à l'intérieur de la chapelle de Albion College (USA) qui possède vraisemblablement une durée de réverbération importante. Les yeux de douze sujets ont été bandés avant leur entrée dans le lieu. Six distances de référence étaient utilisées. La tendance à sous-estimer les distances lointaines pour une distance de référence courte, était presque doublée par rapport à l'expérience 2. Il y avait également une tendance à surestimer les courtes distances lorsque des distances de référence éloignées étaient employées. Finalement, les trois distances les plus lointaines – parmi les huit utilisées en tout – étaient très difficilement discernées entre elles, elles étaient pratiquement jugées identiques.

Dans un certain nombre d'expériences, GARDNER [1969] a étudié la capacité à évaluer la distance de signaux parlés dans un espace anéchoïque. La première expérience fut réalisée avec cinq haut-parleurs visibles par les sujets, face à eux, entre 3,3 m et 9,9 m de distance, au niveau des yeux. Le signal, qui était de la parole enregistrée, était diffusé uniquement par les haut-parleurs placés à 3,3 m et à 9,9 m.

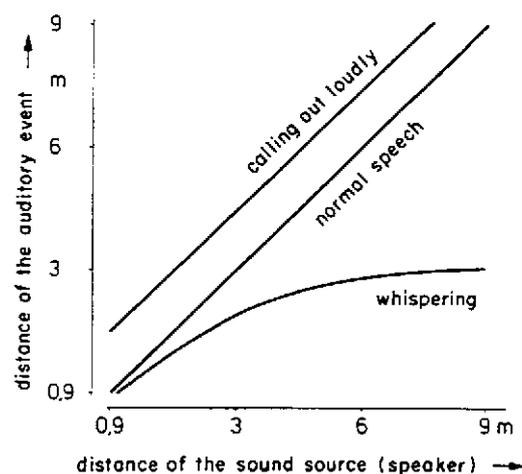
Un seul haut-parleur était activé par session (la moitié des sujets n'écoutait que le haut-parleur proche, l'autre moitié que le haut-parleur lointain). Le niveau sonore mesuré à la place des auditeurs était le même pour les deux haut-parleurs. Il variait de manière aléatoire entre 20 dB(B) et 65 dB(B). La tâche des cinq sujets consistait à indiquer lequel des 5 haut-parleurs visibles était la source sonore.

Les résultats montrent que la distance perçue ne dépend que du niveau sonore et non pas de la position du haut-parleur. Une expérience très similaire réalisée avec le haut-parleur le plus proche à 1,5 m a donné les mêmes résultats.

Il est à noter que *l'effet de proximité* (proximity-image effect) observé par le même auteur dans un travail informel précédent [Gardner, 1968], semble apparaître également dans ces expériences : les signaux les plus forts étaient localisés à l'endroit le plus rationnel, c'est-à-dire sur le haut-parleur le plus proche. Ce phénomène, principalement lié à des informations issues de la perception visuelle, est manifeste pour les niveaux sonores élevés mais peut apparaître également pour les niveaux sonores bas ; ceux-ci sont alors attribués aux haut-parleurs lointains.

L'utilisation de la parole réelle a conduit à des résultats quelque peu différents. Pour cette expérience, les dix sujets avaient les yeux bandés et un orateur se déplaçait dans la salle, il parlait en quatre positions comprises entre 1 m et 9,9 m, face aux sujets. Pendant les déplacements de l'orateur, un bruit masquant était diffusé. Quatre niveaux de parole étaient employés : murmure, voix basse, niveau de conversation et cri. Pour chaque type de parole, on tentait de garder un même niveau sonore pour "la source", quelle que soit la distance. En conséquence, l'intensité était un indice de la distance dans cette expérience.

Les réponses ont donné une forte corrélation entre la position réelle de l'orateur et sa position apparente. Les deux voix normales (voix basse, conversation) étaient localisées de la même façon, avec simplement une tendance à sous-estimer le signal le plus distant. Le signal crié était localisé plus loin que la parole normale, un effet qui peut être causé par une sorte d'association du *besoin du cri* avec des distances plus éloignées. L'effet inverse a été observé pour le cas du murmure pour lequel la distance était très sous-estimée, sauf pour la position la plus proche. L'explication donnée par l'auteur se base sur le contenu élevé en hautes fréquences du murmure (cf. figure 2.5).



**Figure 2.5 :** Distance apparente de l'événement sonore pour trois types de parole : murmure (whispering), conversation normale (normal speech) et cri (calling out loudly), en fonction de la distance du haut-parleur (reproduit de Blauert [1996]).

En utilisant la procédure qui vient d'être décrite, mais avec quatre haut-parleurs servant de sources, réglés pour un niveau sonore d'écoute constant, aucune corrélation entre distance réelle et distance apparente ne fut mise en évidence.

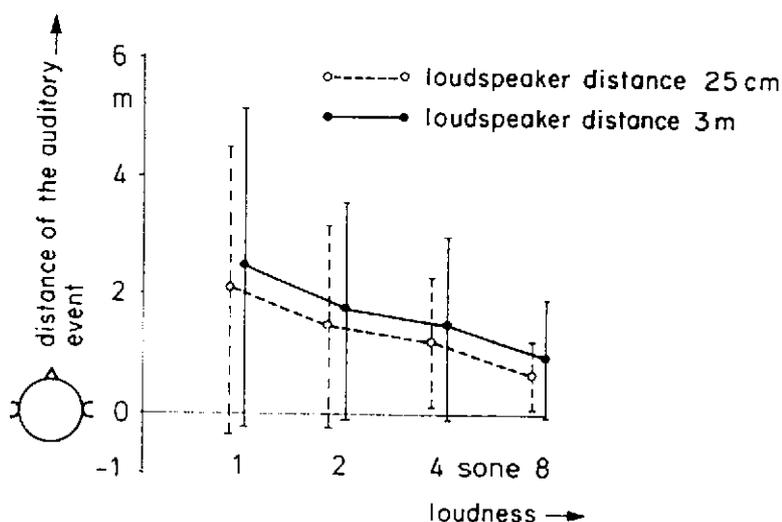
Le travail de LAWS résumé dans Nielsen [1991] et Blauert [1996], concerne l'évaluation de la distance apparente de la source sonore et non pas celle de la distance de la source sonore. Egalement, il a étudié la localisation du son dans la tête (*inside-the-head locatedness IHL*).

Trois expériences étaient menées dans une chambre anéchoïque sans lumière :

- avec un haut-parleur, une fois à 0,25 m et une fois à 3 m devant le sujet ;
- avec un casque audio diffusant : soit avec le signal d'origine non égalisé pour la présence du casque, soit avec le signal d'origine égalisé (ce qui correspond à une simulation en champ libre du haut-parleur à 3 m).

Pour s'assurer que les 27 sujets étaient capables de juger la distance, des tests ont été effectués en vue d'une estimation de la distance visuelle. Ceux-ci étaient concluants puisque de très petits écarts étaient relevés sur les distances : à 0.25 , 0.50, 1, 2 et 4 m du front des sujets.

Lors des tests d'écoute, un bruit blanc à quatre niveaux de sonie différents était utilisé, à 1, 2, 4 et 8 sones (ce qui correspond à 40, 50, 60 et 70 phones), le niveau était ajusté pour chaque sujet par comparaison avec un son pur à 1 kHz. Les résultats pour les haut-parleurs placés à 0.25 m et à 3 m indiquèrent la même tendance, à savoir une perception des sons plus proche avec l'accroissement de la sonie. Bien que statistiquement non significative, il y avait une petite différence de résultats sur les deux haut-parleurs : les sons émis par le plus proche étaient perçus en moyenne 30 cm plus près que lorsqu'ils étaient émis par le plus éloigné (figure 2.6).



**Figure 2.6** : Distance apparente d'un événement sonore en fonction de la sonie de bruit blanc pour deux distances de haut-parleurs. Valeurs moyennes et écart-types pour 27 personnes (résultats de Laws, reproduits de Blauert [1996]).

Pour résumer, les différences entre les distances apparentes des sources sonores ont été trouvées bien inférieures aux écarts réels entre ces sources (haut-parleurs), tel que le décrit le schéma ci-après (figure 2.7).

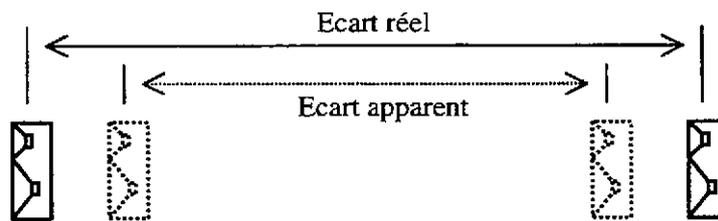


Figure 2.7 : Ecart réel et écart apparent des sources sonores (d'après les résultats de Laws présentés dans Nielsen [1991] et Blauert [1996]).

A l'écoute au casque audio du son sans filtrage, les réponses sur la distance étaient soit très courtes – devant ou derrière la tête – soit situées dans la tête. L'événement sonore était donc localisé dans la tête et était indépendant du niveau de sonie. A l'écoute au casque du signal égalisé (en amplitude et en phase), l'image de la source sonore était ramenée devant et les distances étaient comprises entre un tiers et la moitié des distances données pour le haut-parleur à 3m. Ceci semble pouvoir être attribué à l'utilisation d'un filtre d'égalisation basé sur une moyenne des courbes des 12 auditeurs (et donc pas sur un filtrage individuel).

Un point commun à tous les tests, avec haut-parleurs et avec casque audio, est que les réponses variaient énormément en fonction des sujets.

Des travaux plus récents étudient la perception de la distance en conditions réverbérantes :

MERSHON et KING [1975] avant de décrire leurs expériences font quelques observations intéressantes vis-à-vis de la perception de la distance :

Premièrement, ils constatent qu'en ce qui concerne la tâche des auditeurs, il est important de déterminer si ceux-ci doivent estimer la *distance* de la source sonore ou bien la *distance apparente* de la source sonore ou de l'événement sonore. Le premier choix – qui dans un autre contexte a été caractérisé de « cognitif » [Mershon et Gogel, 1975] – suggère implicitement que d'autres indices que la distance auditive apparente doivent être pris en compte. Cette approche peut être adaptée pour l'évaluation précise de la performance de localisation mais elle est inadaptée pour l'étude de la perception elle-même. L'appréciation *apparente* de la distance est la seule information fiable en ce qui concerne cette dernière.

Deuxièmement, et par rapport aux conditions visuelles des expériences, ils remarquent que la présentation des haut-parleurs aux sujets peut influencer les distances *perçues*, ils citent d'ailleurs comme exemple de ce biais visuel *l'effet de proximité* [Gardner, 1968].

Enfin, ils distinguent les indices utilisés pour la perception de la distance en facteurs *égocentriques* et facteurs *exocentriques*<sup>1</sup> : égocentriques sont les indices absolus, et exocentriques sont les indices

<sup>1</sup> Mershon et King utilisent ici des notions empruntées à la psychologie expérimentale. Il est généralement admis dans ce domaine que la localisation spatiale dépend du système de référence utilisé. On distingue habituellement

relatifs. Un indice absolu donne une information sur les propriétés du son lors de sa première présentation ; un indice relatif ne donne que des informations sur les variations (voir aussi [Nielsen, 1993], [Brungart, 1993]).

Afin de mettre en évidence la différence entre indices absolus et relatifs, Mershon et King ont réalisé deux expériences où deux indices ont été testés simultanément : l'intensité et le rapport d'énergie entre le champ direct et le champ réverbéré. Dans la première expérience, mise en place en conditions réverbérantes, les 80 sujets avaient les yeux bandés, ils ont écouté deux bruits large bande de niveaux acoustiques de 50 et 60 dB(A) au niveau des oreilles (**tableau 2.1**). Deux haut-parleurs étaient placés face à eux, à 2,74 m et à 5,49 m. Chaque sujet n'a entendu qu'un seul haut-parleur. L'expérience était organisée afin que la moitié des sujets entende le bruit le plus fort en premier et vice versa. La tâche de chacun des sujets consistait à juger la distance apparente qui le séparait de la source sonore (en pieds), et de la reporter par écrit (les yeux toujours bandés !).

Les résultats de la première présentation des sons dans cette première expérience ont montré qu'il n'y avait statistiquement pas de différence significative sur la distance apparente entre les quatre conditions (voir schéma ci-dessous), ni entre bas et fort niveau sonore, ni selon la position du haut-parleur.

<i>1<sup>ère</sup> présentation des sons (expérience 1)</i>		
	<b>50 dB(A)</b>	<b>60 dB(A)</b>
<b>2,74 m</b>	20 sujets	20 sujets
<b>5,49 m</b>	20 sujets	20 sujets

**Tableau 2.1** : Première expérience de Mershon et King [1975]

La seconde expérience fut réalisée en chambre anéchoïque, les haut-parleurs étant installés de la même façon que dans la première expérience. Là aussi, la moitié des sujets a entendu l'un des deux haut-parleurs, l'autre moitié a entendu le second (**tableau 2.2**). Trois niveaux sonores étaient utilisés en tout, mais seulement deux niveaux étaient présentés à chaque sujet (65 et 55 ou 55 et 45 dB(A)). Une présentation systématique du premier signal sonore était réitérée suite à 4 présentations du second signal sonore (six stimuli en tout). L'ordre de présentation était inversé pour l'autre moitié des sujets. Cela a permis d'étudier s'il y avait un effet de détérioration lié à l'indice de l'intensité. La tâche des sujets était identique à celle de la première expérience.

<i>1<sup>ère</sup> présentation des sons (expérience 2)</i>			
	<b>65 dB(A)</b>	<b>55 dB(A)</b>	<b>45 dB(A)</b>
<b>2,74 m</b>	20 sujets	20 sujets	
<b>5,49 m</b>		20 sujets	20 sujets

**Tableau 2.2** : Deuxième expérience de Mershon et King [1975]

deux types de référence : une référence centrée sur le sujet, basée ainsi sur la position du corps dans l'espace (localisation égocentrique), et une référence centrée sur l'environnement (localisation exocentrique) [Richelle et al., 1994], [Casalis, 1991].

Comme pour la première présentation à la première expérience, la première présentation dans cette seconde expérience n'a pas révélé de différences perceptives significatives entre les quatre conditions, et ce, malgré la différence de 20 dB(A) entre deux conditions (65 dB(A) à 2,74 m et 45 dB(A) à 5,49 m).

Il est par conséquent peu probable que l'intensité puisse servir d'indice absolu pour la distance. De plus, les résultats de la sixième présentation des stimuli ont renforcé la conclusion précédente en démontrant que l'intensité est un indice absolu inefficace, pas seulement lors d'une présentation initiale mais aussi après une longue durée sans changement.

Des résultats contraires ont été trouvés pour le rapport d'énergie entre le champ direct et le champ réverbéré : les jugements de la distance étaient globalement plus éloignés en conditions réverbérantes qu'en conditions anéchoïques. Mershon et King ont donc avancé que ce rapport énergétique pourrait être interprété comme un indice absolu de la distance. En ce qui concerne la distance réelle des sources, il n'y a qu'en conditions réverbérantes que les distances perçues étaient proches de la distance physique.

Enfin, Mershon et King font remarquer la possibilité d'existence d'une « tendance de distance spécifique » (*specific distance tendency SDT*) de l'écoute, qui peut être présente sous conditions d'informations réduites des indices absolus, et engendrerait une estimation des stimuli à une distance relativement courte (environ 2 m). Le phénomène analogue existant en vision, les auteurs se demandent si ce n'est pas plutôt une tendance de la perception en général (*'perceptual tendency'*), indépendante de la modalité spatiale impliquée (visuelle ou auditive). Ce dernier aspect est lié à l'observation générale que les distances proches sont surestimées tandis que les distances lointaines sont sous-estimées, à condition qu'une variation systématique sur un indice absolu efficace soit réalisée.

MERSHON et BOWERS [1979] ont poursuivi les recherches sur le rapport d'énergie entre le champ direct et le champ réverbéré en tant qu'indice absolu de la distance. Le principe des expériences était de diffuser un seul son par sujet (pour un azimut donné) dans une pièce, afin de voir si divers degrés de réverbération provoqués par différentes distances de haut-parleurs, pourraient servir d'indice absolu pour la distance. Une phase d'entraînement aux conditions d'écoute dans la salle de tests précédait les expériences. Tous les tests étaient réalisés dans la même salle de classe dont le temps de réverbération était d'environ 0,7 s.

Lors de la première expérience, deux présentations d'un bruit large bande de 5 secondes étaient introduites, l'une d'en face (0°), l'autre de côté (90°). Cinq haut-parleurs étaient placés à cinq distances différentes 0,55, 1, 2, 4 et 8 m, mais chaque sujet n'écoutait qu'un seul haut-parleur par direction (0° ou 90°). La moitié des sujets a d'abord reçu le son à 0° et ensuite le son à 90°; l'ordre a été inversé pour la moitié restante des sujets. Le niveau sonore était constant de 60 dB(A) au niveau des oreilles quel que fut le haut-parleur en marche. Les 200 sujets avaient les yeux bandés, leur tâche consistait à inscrire la distance apparente de la source sonore. L'analyse de variance effectuée sur les résultats a montré une différence significative des réponses pour les diverses positions de haut-parleurs. Les distances les plus courtes étaient surestimées et les plus éloignées étaient sous-estimées. Il n'y avait pas de différence significative entre l'exposition à 0° ou à 90°, ni entre le premier et le deuxième son.

Les deuxième et troisième expériences étaient réalisées dans la même salle que la première, avec 3 positions frontales des haut-parleurs (1, 3 et 6 m), et un stimulus correspondant à une onde carrée de 200 Hz (de 62 dB(A) environ) d'une durée de 5 secondes. Pour la deuxième expérience, la moitié des 120 sujets a reçu des consignes dans la salle même des tests, tandis que l'autre moitié ne connaissait rien sur les conditions d'écoute de la salle. La tâche des sujets consistait à donner oralement leur estimation de la distance apparente. La troisième expérience était similaire à la deuxième, mais cette fois les 120 nouveaux sujets devaient reporter la distance apparente par écrit. Les résultats de ces

expériences n'ont pas montré de différence statistiquement significative entre les quatre procédures (première expérience, deuxième avec et sans préparation, troisième expérience). Cela signifie que, dans ce cadre, il n'y a pas de différence sur la perception de la distance si l'environnement sonore est connu ou non des sujets avant l'exposition aux stimuli.

Les résultats des trois expériences montrent que le rapport d'énergie entre le champ direct et le champ réverbéré peut servir d'indice absolu pour la distance, bien que ce ne soit pas avec une grande précision.

Dans des travaux plus récents, WAGENAARS [1990] a étudié la localisation – en azimut et en distance – dans une salle de séjour normale, et donc en champ plutôt réverbérant. Le dispositif des expériences était composé de 16 haut-parleurs placés dans un plan horizontal autour du sujet, avec des intervalles de 45° et des rayons de 1 m et 2 m. La tâche des sujets consistait à sélectionner l'un des 16 haut-parleurs visibles, qui leur semblait correspondre à la position de la source sonore. Divers signaux étaient utilisés, à bande étroite ou large bande, transitoires et non transitoires. Les résultats ont montré que le bruit, les impulsions et la musique peuvent être très bien localisés. Les sons purs avec au moins une variation impulsionnelle étaient localisés plutôt bien, tandis que les sons purs avec des temps de montée et de descente assez longs (*onsets* et *offsets* de 1s) sont localisés plutôt mal. Il est intéressant de noter que même en conditions réverbérantes, ces signaux à bande fine sans impulsion ne peuvent pas être localisés précisément. Finalement, lorsque des erreurs de localisation étaient relevées, celles-ci concernaient dans leur grande majorité l'appréciation de la distance. L'indice pour les jugements de distance dans ces expériences était principalement l'intensité, mais la réverbération a dû également jouer un rôle.

NIELSEN [1993] a réalisé une série d'expériences afin d'étudier la perception de la distance dans différentes salles: une chambre anéchoïque, une salle d'écoute normalisée (IEC 268-13 ; TR ≈ 0,3 - 0,4 s) et une salle de classe. Des haut-parleurs situés en différents emplacements (en azimut et en distance) dans ces salles diffusaient un signal connu qui était une voix féminine (5 s).

En plus des différentes salles, les indices testés étaient : le rapport énergétique son direct / son réfléchi, la sonie et l'azimut. Afin de faire varier l'indice lié aux réflexions, quatre distances (1, 1.71, 2.92 et 5 m) de haut-parleurs ont été utilisées pour chacune des directions 0° et 45°, un seul haut-parleur a été placé à 90°, et un autre à 180° (en tout, les haut-parleurs dans chaque salle étaient au nombre de 10, sauf pour la salle d'écoute où ils n'étaient que 9). Le niveau sonore était constant à la position d'écoute sauf quand cet indice variait spécifiquement. Trois niveaux de sonie étaient utilisés, mesurés à la place d'écoute : 58, 68 et 78 phones. Les 32 sujets devaient juger la distance apparente de la source. Ils ne pouvaient pas voir les haut-parleurs mais ils disposaient de repères au sol pour les guider dans leurs jugements.

Les résultats ont été dépouillés à l'aide de l'Analyse de Variance fournissant des informations complémentaires sur les interactions entre les différents facteurs expérimentaux (indices mentionnés ci-dessus). Les moyennes des jugements de distance donnés par les sujets ne correspondent généralement pas à la distance réelle des sources et la dispersion des résultats est importante. Ils montrent une nette différence entre les résultats obtenus pour la chambre anéchoïque et les résultats obtenus pour les deux autres salles. Pour ces deux salles réverbérantes, la perception de la distance d'une source semble être presque constante, indépendante du niveau sonore. L'influence des réflexions sur la perception de la distance semble donc claire et se manifeste à l'aide d'un indice principal : le rapport des énergies du son direct et du son réverbéré. Cet indice varie de manière similaire avec la perception de la distance.

L'influence du niveau est importante dans la chambre anéchoïque et dans la salle d'écoute normalisée. Dans la salle de classe, elle est remarquablement inférieure ; ceci peut être dû à la réverbération bien

plus élevée dans cette salle. La distance apparente dans les chambres réverbérantes semble donc presque indépendante du niveau sonore.

Pour les directions des sources autres que 0°, la distance apparente est supérieure (à celle pour 0°). En outre, les résultats de Nielsen reproduisent les résultats d'autres études concernant la plage de variation de la distance apparente – le plus souvent inférieure à celle des distances réelles des sources (les distances proches sont surestimées et les distances lointaines sont sous-estimées) – et la grande dispersion des réponses [Cochran et al., 1968], [Mershon et Bowers, 1979].

A la suite de ses expériences principales, le même auteur a effectué deux expériences 'de contrôle' identiques à ces premières mais où les haut-parleurs étaient visibles par les auditeurs [Nielsen, 1991]. A la première expérience, les sujets devaient comme auparavant, juger la distance apparente de la source, tandis qu'à la deuxième ils devaient identifier le haut-parleur émetteur. Dans les deux cas, la dispersion des résultats était remarquablement inférieure à celle des expériences principales (la dispersion des résultats de l'expérience *d'identification* était encore plus faible). Nous pouvons donc faire l'hypothèse que dans ces expériences, les sujets sont influencés inconsciemment par la vue des haut-parleurs et ils se sont appuyés davantage sur la distance 'réelle' – ou visuelle – des sources pour juger la distance apparente (dans les expériences principales, le seul indice absolu pour juger la distance 'réelle' des sources était le rapport des énergies du son direct et du son réverbéré). Ces résultats vérifient les observations avancées par Mershon et King [1975] (cf. ci-dessus) concernant l'influence des facteurs visuels et l'importance des consignes données aux sujets, à savoir, si ceux-ci doivent juger à l'aide d'une échelle continue ou bien discrète (juger la distance apparente ou identifier la source sonore émettrice).

Enfin, on cite ici une étude qui a spécifiquement testé l'influence visuelle sur la perception de la distance auditive et dans laquelle il a été trouvé que, pour des distances courtes, la distance auditive est influencée par la distance visuelle. Des expériences dans lesquelles les sujets étaient invités à juger si une cible sonore ('auditory' rattle) était à portée de la main ou non, ont été réalisées. Les sujets ont montré une tendance accrue à juger que la cible sonore était à portée de la main quand sa présentation était accompagnée d'une animation visuelle équivalente ('visual' rattle) plus proche de 8 cm que la cible sonore réelle [Brown et al., 1998].

### 2.3.2 Indices de distance auditive

Plusieurs propriétés du signal sonore varient avec la distance à la source et le problème est de définir celles qui sont les plus pertinentes d'un point de vue perceptif. Les quatre indices les plus généralement admis sont présentés ci-après [Coleman, 1963], [Blauert, 1996], [Jouhaneau, 1997], [Polack, 1982], [Nielsen, 1993].

- *La sonie*. Il s'agit d'un indice fin et le plus largement accepté pour la distance. En conditions anéchoïques, la pression acoustique, étroitement liée à la sonie, est inversement proportionnelle à la distance de la source. En conditions réverbérantes, la pression acoustique varie moins qu'en conditions anéchoïques. La valeur de la sonie en tant qu'indice de la distance dépend de la connaissance de la sonie de la source elle-même et ne permet donc pas l'estimation de sa distance absolue [Coleman, 1962], [Mershon et King, 1975].
- *Le spectre fréquentiel*. L'absorption du son par l'air est plus forte aux hautes fréquences. Par conséquent, le spectre d'un son peut donner une indication sur la distance de la source [Coleman, 1963 et 1968] : une baisse du contenu en hautes fréquences d'un son, augmente sa distance auditive apparente. Comme pour la sonie, la valeur de cet indice dépend de la connaissance de la

source et plus particulièrement de son contenu spectral ; le spectre est donc un indice de la distance relative. Il est en effet plus facile d'estimer la distance apparente d'une voix connue que celle d'un bruit non identifié. Dans une salle où se produisent de nombreuses réflexions, le spectre du son réfléchi varie en fonction de l'absorption des murs (dépendant de la fréquence), mais aussi en fonction de la distance de propagation du son dans l'air.

- *Le rapport d'énergie entre le champ direct et le champ réverbéré* du son. Il varie en fonction de la distance sous des conditions réverbérantes. En effet, l'énergie du champ réverbéré reste constante dans la salle tandis que celle du son direct décroît avec la distance. Ce rapport dépend de la pièce, par conséquent la connaissance de cet espace est nécessaire. Souvent, cette connaissance est acquise inconsciemment en pénétrant dans la pièce. Le rapport de réverbération peut donner des informations sur la distance dès la première présentation d'un son et il ne dépend pas forcément de la connaissance du signal émis ; il peut donc être considéré comme un indice absolu [Mershon et King, 1975], [Mershon et Bowers, 1979]. Enfin, comme la réverbération est la plupart du temps dépendante de la fréquence, le spectre de fréquence du son change également avec le degré de réverbération.
- *Les différences binaurales*. Deux sortes de différences existent : des écarts importants de niveau ou de phase entre les signaux parvenant à chaque oreille, ainsi que des différences sur la structure fine des deux signaux (surtout pour des distances proches, jusqu'à 25 cm). Pour une source sonore très proche, l'effet de masque de la tête devient significatif ; également, la distance entre les deux oreilles est relativement grande par rapport à la distance de la source, ce qui génère des différences de niveau plus importantes (dues aux différences des trajets du son vers les deux oreilles) [Coleman, 1963], [Blauert, 1996]. Les différences binaurales sont indépendantes de la source sonore, elles sont donc des indices absolus de la distance. Enfin, notons que les différences binaurales sont plus importantes dans le cas de sources sonores positionnées de manière asymétrique par rapport aux auditeurs (à condition qu'ils ne puissent pas tourner la tête). Dans le cas de présentation des sources le long de l'axe frontal ( $0^\circ$ ) et afin d'étudier les autres indices révélateurs de la distance, les différences binaurales peuvent être négligées [Mershon et King, 1975].

### 2.3.3 Discussion

L'expérience prouve que l'évaluation de la distance d'une source sonore chez l'Homme – « en l'absence d'information visuelle » comme cite un auteur – est relativement imprécise. Il s'agit d'un processus de localisation sonore complexe, bien plus complexe que celui de la localisation en azimut [Blauert, 1996].

Les indices utilisés pour la perception de la distance se distinguent en indices absolus et en indices relatifs. Pourtant, que ce soit en présence d'indices absolus ou d'indices relatifs, notre système perceptif semble *avoir besoin de repères* ou de *références* afin de juger la distance sonore. Ces repères servent à effectuer des comparaisons; il semble donc exister un aspect relatif fort dans tous types de jugements de la distance. Nous justifions cette constatation à la suite :

Pour les *indices absolus* – indices qui peuvent fournir une information sur la distance d'une source lors de sa première présentation – nous constatons que l'information est obtenue grâce à un processus de comparaisons immédiates :

- les *différences binaurales* fournissent un indice basé sur la comparaison des deux signaux présents aux oreilles des auditeurs,

- le *rapport d'énergie entre le champ direct et le champ réverbéré* se base sur la comparaison de ces deux énergies : précoce et tardive.

De plus, les *indices absolus* impliquent une certaine expérience, ils utilisent des connaissances – innées ou établies intuitivement dans le système perceptif – comme une sorte de repère cognitif. Ainsi, nos connaissances précédemment acquises sur la ressemblance des signaux auditifs au niveau des deux oreilles vis-à-vis de la distance de la source correspondante, ainsi que nos connaissances relatives aux rapports énergétiques entre son direct et son réverbérant dans des espaces clos, servent indéniablement comme une référence pour les jugements de la distance.

Pour les *indices relatifs* – les indices qui ne donnent des informations que sur les variations de la distance – le fait que le système perceptif ait besoin de repères semble évident ; ce qui semble plus important est l'identification et les mécanismes d'acquisition de ces repères. On cite ici :

- La *comparaison directe* entre les stimuli successifs qui varient sur un même (ou sur plusieurs) indice physique de distance auditive, comme par exemple l'intensité ou le spectre fréquentiel.
- La *perception visuelle* de la salle de tests ou des sources sonores – ou la *connaissance* de la salle et de la position des sources sonores – qui donne un repère incontestable pour les jugements de la distance auditive. Ces indices peuvent minimiser la dispersion des résultats [Nielsen, 1991] ou donner lieu à des phénomènes comme le phénomène 'd'image de proximité' (cf. 2.2.1.; [Gardner, 1968]).

Les informations visuelles représentent, dans la plupart des cas, un indice involontaire et par conséquent, un effort particulier est apporté par les chercheurs afin d'éliminer leur influence : on observe souvent que les yeux des sujets sont occultés pendant les expériences [Mershon et King, 1975] [Mershon et Bowers, 1979], parfois ils sont occultés *avant* même leur entrée dans la salle de tests [Cochran et al., 1968]. Les expériences ont souvent lieu dans des conditions d'obscurité [Laws, 1972] ou bien les sources sonores et la salle d'expériences sont cachés des sujets à l'aide de rideaux [Nielsen, 1991]. Enfin, d'autres chercheurs placent plusieurs haut-parleurs devant les sujets, mais activent seulement un nombre limité d'entre eux (souvent les haut-parleurs du milieu ou les deux haut-parleurs des extrémités) [Coleman, 1962], [Gardner, 1969].

- La *connaissance des signaux sonores* est un facteur primordial sur la perception de la distance. La voix humaine par exemple – surtout si elle appartient à un orateur réel – est beaucoup plus facilement repérable en distance qu'un bruit large bande. Cette connaissance pourrait être innée, acquise intuitivement ou acquise à la suite d'un entraînement spécifique.
- L'*entraînement ou l'apprentissage* des auditeurs avant ou pendant les tests. Il peut aider les auditeurs à obtenir des informations et à se familiariser avec le signal sonore ou les propriétés acoustiques de la salle. De plus, la présentation successive des signaux variant sur un indice de distance relative (par exemple l'intensité), peut apprendre aux sujets à utiliser cet indice pour juger la distance [Coleman, 1962], [Mershon et Bowers, 1979].

A ce propos, on cite ici quelques phrases de Blauert [1996] : '*Adaptation and learning are observed in studies of directional hearing in the median plane and particularly in studies of distance hearing. This means that localization may change as a function of time*'.

- D'autres *repères de nature cognitive* se référant au contexte des expériences en général et à l'expérience vécue sont présentés ci-après afin de montrer à quel point ils peuvent être nombreux et variés.

Les *consignes* données aux sujets : est-ce que ceux-ci doivent estimer la *distance apparente* de la source sonore, la *distance de la source sonore* ou bien la *source sonore* (tests d'identification par

exemple). Les deux dernières consignes semblent limiter la dispersion des résultats ; les sujets se basent davantage sur des facteurs visuels et cognitifs pour juger la distance de la source.

Cochran et al. [1968] ont travaillé sur la notion de *distance de référence*. Ils ont trouvé des différences significatives selon les trois distances de référence employées : la tendance générale consistant à sous-estimer les distances lointaines diminuait avec une distance de référence croissante (cf. 2.2.1). Ces résultats nous semblent très importants dans la mesure où ils montrent à quel point la perception de la distance est relative et a besoin de repères. Ils indiquent également comment la perception peut être influençable, elle peut « glisser » ou « s'adapter » (cf. le terme 'adaptation' employé par Blauert [1996]).

Par ailleurs, nous pouvons même nous demander si la tendance générale à sous-évaluer les distances, d'une manière globalement proportionnelle à la distance jugée, n'est pas en partie liée à une notion de distance de référence : c'est la position de l'auditeur qui serait alors considérée comme référence.

Un autre problème lié à la perception de la distance est celui de l'*ordre de présentation des stimuli*. Le jugement d'un son n'est pas indépendant du son (ou des sons) qui le précède. Ainsi, pour éviter les effets d'ordre, si par exemple deux stimuli – ou deux conditions expérimentales en général – doivent être présentés aux sujets, la moitié d'entre eux doit recevoir l'un des stimuli d'abord et l'autre moitié l'autre stimulus en premier. En effet, nous remarquons que toute démarche expérimentale rigoureuse se doit d'alterner l'ordre de présentation des stimuli de manière systématique. Evidemment, si le nombre de stimuli ou de conditions est plus élevé que deux, plusieurs ordres aléatoires doivent être créés afin d'être présentés aux sujets [Wagenaars, 1990], [Nielsen, 1993].

Enfin, un autre aspect cognitif semble être révélé par les résultats de Gardner [1969] : la distance du signal crié était localisée plus loin que la parole normale, tandis que la distance du murmure était localisée plus proche. Les auteurs suggèrent qu'il s'agit probablement d'une sorte d'association du cri avec des distances plus longues. On pourrait également supposer une association du murmure avec les distances les plus proches (en plus de son contenu en hautes fréquences (cf. § 2.3.1)).

Les différents indices interagissent : ils collaborent ou rivalisent entre eux. Il semble évident que même un indice sonore pertinent peut être renforcé par la présence d'un autre indice sonore, la présence d'un indice visuel ou encore d'un repère cognitif, à condition bien sûr que ces indices soient concordants entre eux. La localisation en distance se fait alors avec plus de certitude et de précision. Si au contraire, ces indices sont discordants entre eux et discordants avec les repères cognitifs, la perception d'un événement devient plus complexe, la prédominance d'un indice sur un ou plusieurs autres dépend bien sûr des indices eux mêmes, du contexte particulier ainsi que des différences interindividuelles.

Enfin, on cite ici quelques phrases de Jouhaneau se référant à la localisation de direction, mais qui concernent également la localisation en distance [Jouhaneau, 1997] : '*Dans la pratique, il n'est pas rare de remarquer que la précision de la localisation d'une source sonore est grandement améliorée quand le cerveau dispose d'informations complémentaires : vision, connaissance ou reconnaissance du signal, conditionnement, expectative, sixième sens, etc. ... Parmi les paramètres étudiés, il semble que ce soient la vision et le degré de connaissance préalable des signaux utilisés qui constituent les indices les plus pertinents de l'aide à la détection.*'

### 2.3.4 Conclusion

La perception de la distance auditive est un processus complexe, évoquant à la fois des traitements psychophysiques et cognitifs. Elle se base principalement sur des indices acoustiques *relatifs* ou *absolus* : un indice absolu donne une information sur les propriétés du son lors de sa première présentation ; un indice relatif ne donne que des informations sur les variations. La *sonie* et le *spectre fréquentiel* semblent être des indices relatifs tandis que les *différences binaurales* et le *rapport énergétique entre le champ direct et le champ réverbéré du son* (dans le cas d'émission sonore dans un endroit clos) semblent être des indices absolus.

La pertinence de ces indices semble pourtant médiocre : la perception de la distance chez les êtres humains est en effet peu précise ; bien moins précise que la localisation en azimuth et en site par exemple. Dans le cas général, et comme par analogie à la vision, cette précision baisse avec l'augmentation de la distance.

C'est peut-être justement à cause de ce manque de pertinence des indices acoustiques que leur efficacité semble pouvoir être autant influencée par d'autres facteurs, cognitifs et visuels. En effet, la connaissance du signal sonore, la connaissance de la salle de tests et de la position des sources sonores jouent un rôle important. De plus, l'adaptation aux conditions expérimentales, l'entraînement, les consignes données aux sujets lors des tests, les différentes distances de référence éventuellement employées, l'ordre de présentation des stimuli, etc. peuvent également influencer la perception de la distance.

Nous pourrions ainsi conclure que les indices sonores, les indices visuels et les paramètres cognitifs interagissent – collaborent ou rivalisent – afin de lever l'ambiguïté d'un phénomène perceptif multimodal qui, en l'occurrence, est la perception de la distance auditive.

## 2.4 PERCEPTION MULTIMODALE

### 2.4.1 Notions principales, domaines de recherche

La perception d'un événement présent dans notre environnement naturel, dans le cas général, ne se base pas sur la perception d'informations issues d'une seule modalité sensorielle, mais sur la perception d'informations hétéromodales, provenant de plusieurs modalités<sup>1</sup>.

Les modalités sensorielles fonctionnent ainsi rarement de façon isolée. Le plus souvent, elles sont actives simultanément et apportent des informations transférables de l'une à l'autre. On parle de transfert multimodal lorsqu'une information reçue à travers une modalité est utilisée correctement par une autre modalité. Tel est le cas quand un objet perçu visuellement est ensuite reconnu par le toucher ou quand une source sonore est localisée visuellement. Les transferts intermodaux sont très importants dans le fonctionnement de notre système perceptif car ils permettent des économies d'apprentissage, et assurent une connaissance cohérente et unifiée du monde.

Deux aspects peuvent être distingués dans la perception multimodale : l'intégration multimodale et les interactions sensorielles.

- L'intégration multimodale peut être définie comme la synthèse des informations provenant de deux ou de plusieurs modalités sensorielles, effectuée de telle manière que l'information qui en résulte ne pourrait pas être obtenue de chacune de ces modalités sensorielles séparément. Elle implique que, à un certain niveau de traitement, les informations ne sont pas représentées indépendamment, mais sont intégrées dans un seul processus de perception multimodale des événements.

L'intégration multimodale des informations se manifeste dans deux types de situations. Dans l'une, les modalités activées accèdent chacune à une propriété distincte des objets. Par exemple, l'œil peut percevoir la forme d'un réveil mécanique, la main son poids et l'oreille ses émissions sonores. Les modalités perceptives fonctionnent dans ce cas de manière complémentaire et l'intégration multimodale consiste à mettre en relation ces différentes perceptions pour former un objet unique multidimensionnel et multimodal. Dans l'autre type de situation, les systèmes perceptifs mobilisés informent sur la même propriété du même objet. Tel est le cas quand la taille d'un objet est évaluée visuellement et tactilement, ou quand la localisation spatiale d'une cible est appréhendée visuellement et auditivement. Pour que la connaissance soit cohérente, les deux modalités doivent alors normalement assigner la même valeur à la propriété perçue [Richelle et al., 1994].

---

<sup>1</sup> Nous présentons ici les significations des termes relatifs aux différentes opérations des modalités perceptives, utilisés dans cette section : par *intramodale*, on entend une opération effectuée dans une même modalité ; par *intermodale*, une opération effectuée entre deux modalités (c'est le cas de l'interaction) ; par *multimodale*, on entend une opération où plusieurs modalités sont impliquées ; par *amodale*, une opération qui n'est pas associée à une seule modalité mais à plusieurs (dans ce sens, amodale est similaire à multimodale). Par ailleurs, on parle de stimuli *hétéromodaux* quand il s'agit de stimuli de différentes modalités ; de stimuli *bimodaux* quand il s'agit de stimuli où deux modalités sont impliquées ; et enfin, de stimuli *unimodaux* quand ils ne concernent qu'une seule modalité.

- L'interaction entre deux modalités sensorielles peut être définie comme la modification d'un certain concept dans une modalité, à cause de la présence d'une stimulation dans une autre modalité. Ceci suggère que l'évaluation perceptive concerne une seule modalité, cette modalité est considérée comme la modalité *principale*. La deuxième modalité qui n'est pas celle directement étudiée mais peut influencer cette évaluation, est la modalité *accessoire* [Gribenski, 1978].

On remarque qu'entre les concepts d'intégration multimodale et d'interaction sensorielle, une différenciation importante se situe au niveau de l'orientation de l'attention : ainsi nous nous intéresserons, soit à l'évaluation d'un aspect multimodal (ou amodal) d'un événement, par exemple la *distance d'une personne* que l'on voit et que l'on entend, soit à un seul des aspects unimodaux de l'événement, malgré la présence d'informations issues d'autres modalités sensorielles, par exemple la *distance visuelle d'une personne* que l'on voit mais que l'on entend également.

La recherche sur la perception multimodale est effectuée dans une variété de contextes expérimentaux. Les disciplines scientifiques traitant du sujet varient de la recherche fondamentale en psychologie à la recherche appliquée dans des domaines bien spécifiques.

La perception multimodale est principalement étudiée dans le cadre de la psychologie cognitive et de la psychologie expérimentale. La psychologie cognitive traite tout ce qui concerne la représentation (l'information stockée dans la mémoire) et le traitement de la connaissance relative à la perception multimodale. La psychologie expérimentale vise plutôt à comprendre les processus, à étudier les conditions d'apparition et à mesurer l'ampleur des interactions sensorielles.

Un domaine typique de recherche appliquée sur la perception multimodale est celui de la perception audiovisuelle de la parole. Les études s'intéressent tant à l'apport de l'information visuelle sur l'intelligibilité de la parole sous des conditions acoustiques non favorables [Sumby et Pollack, 1954], [Dodd et Campbell, 1987] qu'à la création d'illusions perceptives observées lors de la présentation de stimuli hétéromodaux contradictoires (effet Mc Gurk) [McGurk et MacDonald, 1976]. Il s'agit aussi d'étudier les conditions sous lesquelles les indices visuels et auditifs de la parole sont intégrés dans les environnements artificiels pour créer l'illusion d'un seul événement perceptif. Il est évident qu'avec le développement des systèmes de télécommunications, des interfaces multimédia et des applications de la réalité virtuelle – où il est nécessaire d'accompagner la parole synthétique avec des animations de 'talking heads' (tête de présentateur) – ce domaine connaît un intérêt de plus en plus croissant.

La perception multimodale est également étudiée dans le cadre de l'évaluation de la qualité de systèmes audiovisuels (par exemple la télévision et le cinéma). Ces études s'intéressent à l'influence de différentes qualités vidéo et audio sur l'appréciation globale de la qualité audiovisuelle. Quelques-uns de ces travaux récents seront présentés dans les sections suivantes ; des discussions et références supplémentaires peuvent être trouvées dans [Woszczyk et al. 1995].

Une série de travaux traitant du sujet des interaction sensorielles entre vision et audition, mais aussi des paradigmes d'intégration intermodale, sera présentée dans la section suivante. Les différents types d'interactions seront présentés et les conditions de leur apparition sur le plan physique – mais aussi cognitif – seront discutées.

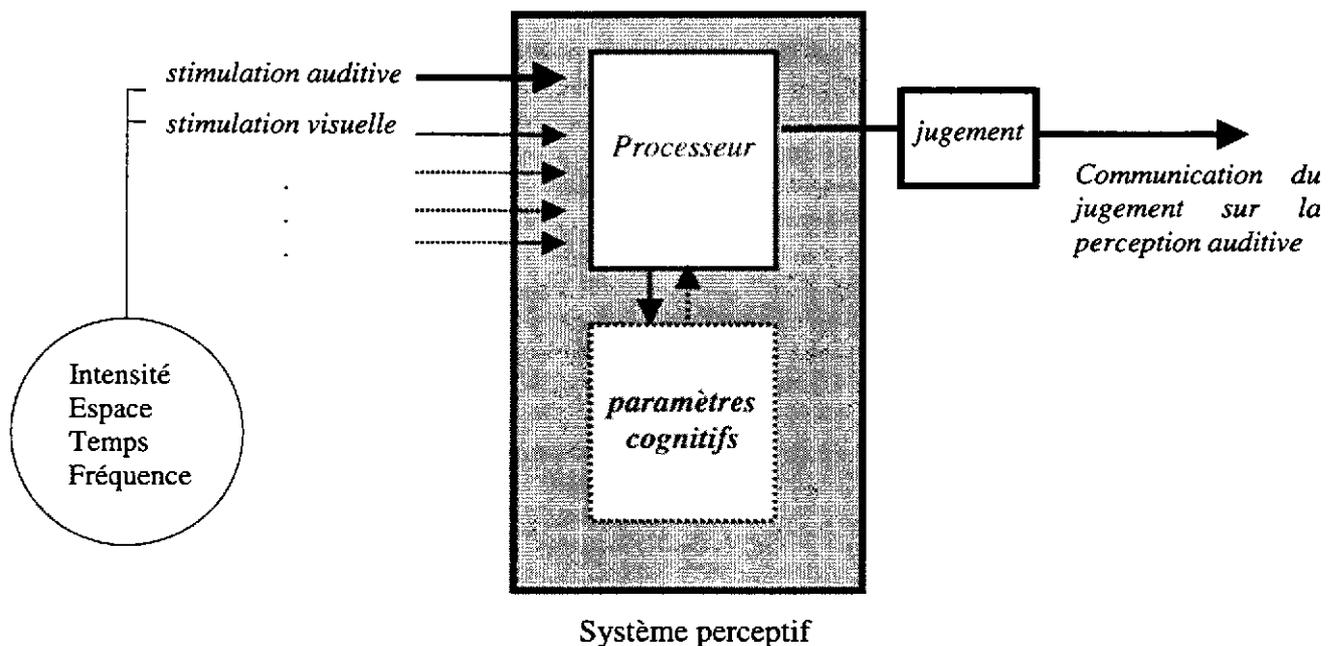
## 2.4.2 Interactions sensorielles entre vision et audition

### 2.4.2.1 Introduction

Dans le cadre de ce travail, nous nous intéresserons à évaluer l'influence des informations visuelles sur la perception auditive. C'est donc la notion d'interaction sensorielle qui nous concerne plus particulièrement, considérée ici comme l'une des difficultés de l'évaluation subjective. Dans l'ensemble de nos investigations, la modalité auditive est la modalité principale et la modalité visuelle est la modalité accessoire.

Plusieurs théories de la perception ont été proposées par les psychologues. L'approche dominante est celle de l'approche cognitive. Elle propose de considérer la perception comme un système de traitement de l'information dont les processus perceptifs se déroulent dans le temps, ils impliquent des transformations des données d'entrée et leur confrontation avec des informations mémorisées (représentations). La distinction 'sensation-perception' a une valeur importante dans cette approche : la première fait référence aux traitements sensoriels des informations physiques, indépendamment de leur signification, et la seconde est plus directement concernée par les significations des objets et donc par l'interprétation des informations sensorielles [Richelle et al., 1994].

Le but de ce travail n'étant pas de discuter les diverses théories de la perception, l'approche cognitive est adoptée pour représenter sur la **figure 2.8** un modèle schématique de la perception. Est représenté plus particulièrement, le processus des interactions sensorielles entre audition et vision.



**Figure 2.8** : Schéma des interactions sensorielles entre audition et vision où la perception est considérée comme un système de traitement d'information. La modalité auditive est la modalité principale et la modalité visuelle la modalité accessoire. Chacune des stimulations physiques peut être décrite à l'aide de quatre dimensions physiques fondamentales : son intensité, sa distribution spatiale, temporelle et fréquentielle. Les signaux entrants sont traités et confrontés aux informations mémorisées (représentées ici sous le titre plus général de paramètres cognitifs) afin d'arriver à un jugement et par la suite à la communication de ce jugement.

Au même moment qu'une stimulation auditive est reçue par notre système perceptif, d'autres stimulations provenant d'autres modalités sensorielles – et plus particulièrement de la modalité visuelle – sont également reçues. Ces informations sont traitées par notre système perceptif afin d'arriver à un jugement sur la perception auditive. On peut qualifier ce traitement de perceptif mais aussi de cognitif parce que des paramètres autres que les paramètres physiques des stimulations entrantes sont pris en compte dans le processus. Il s'agit de paramètres cognitifs, c'est à dire de paramètres liés à la mémoire (représentations mentales, significations des stimuli), aux connaissances, à l'attention, etc., qui sont activés, d'une part directement par les informations extraites de la stimulation, et d'autre part par le contexte de la situation [Casalis et al., 1991]. On reviendra sur ces notions importantes à la fin de cette section (§ 2.4.3).

Arriver à un jugement et communiquer ce jugement sont deux opérations distinctes. Ainsi, le produit final de ce processus – la communication du jugement sur la perception auditive – n'est pas non plus indépendant de la modalité sensorielle chargée de le communiquer. Ainsi, si ce jugement se base sur des descriptions verbales, il dépend des représentations lexicales chez les personnes qui jugent ; s'il s'effectue à l'aide d'échelles perceptives, il dépend de l'utilisation qui en est faite ; s'il est communiqué à l'aide d'une tâche tactile, il dépend des capacités et particularités psychomotrices des sujets, etc. Il s'agit en effet d'un des problèmes généraux de l'appréciation subjective qui ne sera pas discuté davantage ici (cf. § 2.2.6).

Les stimulations auditives et visuelles reçues par notre système perceptif peuvent être décrites par quatre grandeurs physiques fondamentales : leur intensité, leurs position et géométrie dans l'espace, leur distribution temporelle et leur distribution fréquentielle. Les interactions apparaissent généralement entre attributs équivalents des deux modalités : nous avons ainsi des interactions spatiales, temporelles, et aussi des interactions qui concernent les intensités et la fréquence des deux modalités.

La grande majorité des études traite les problèmes des interactions spatiales et temporelles. Ceci est dû au fait que dans ce cas les deux modalités sont appelées à nous informer sur la même propriété d'un objet. Pour l'intensité et la fréquence, où les modalités activées accèdent chacune à une propriété distincte des objets, ce sont surtout les 'correspondances' ou 'équivalences' intermodales qui sont recherchées [Marks, 1987]. Quelques uns de ces travaux sont présentés à la suite; l'importance est donnée sur les interactions spatiales, comme c'est aussi le cas pour la littérature spécialisée. Les paramètres physiques et cognitifs qui influencent les interactions sont relevés ; ces paramètres sont ensuite récapitulés.

#### 2.4.2.2 Interactions spatiales

Les interactions spatiales entre vision et audition concernent l'influence de la localisation visuelle sur la localisation auditive et vice versa. Par localisation d'une source, on entend sa position sur le plan horizontal (azimut), sur le plan vertical (site) ainsi que sa distance dans la direction définie par les deux premiers paramètres. La grande majorité des travaux sur les interactions spatiales concerne la localisation en azimut. Quelques travaux sporadiques traitant indirectement ou directement la localisation en profondeur (influence de la perception visuelle sur la perception de la distance auditive) sont présentés dans la partie précédente (§ 2.3).

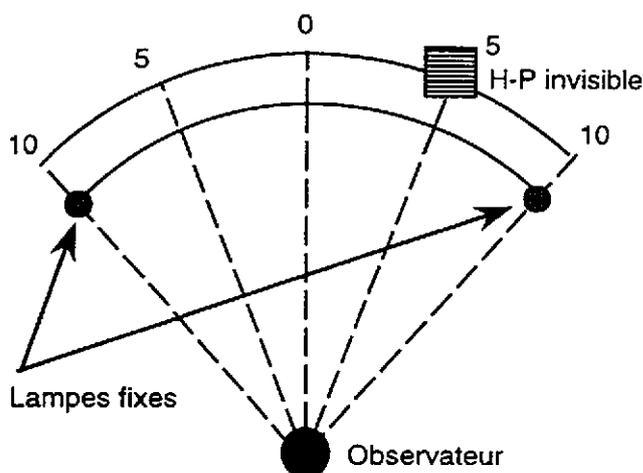
Au début du siècle, KLEMM [1909] a trouvé que la localisation auditive peut être influencée par une stimulation visuelle. Si, dans l'obscurité, une source sonore est exactement en face du sujet, une tache lumineuse à gauche semble la faire se déplacer légèrement vers la gauche. Autrement dit, un stimulus

lumineux semble attirer une source sonore vers lui. Toujours selon Klemm, une source sonore et une source lumineuse, pour être jugées distinctes, doivent être plus éloignées l'une de l'autre que deux sources sonores ou deux sources lumineuses.

Ont suivi les expériences de THOMAS [1941] où les stimulations visuelles étaient des taches lumineuses qui, comme la stimulation sonore, pouvaient clignoter avec divers motifs rythmiques. Son expérience était montée comme indiqué sur la **figure 2.9**.

Thomas a remarqué de la part de ses observateurs, une tendance à juger la localisation sonore vers une tache lumineuse clignotante plutôt que vers une tache lumineuse continue, même dans le cas où la source sonore était également continue. Il a aussi remarqué une faible tendance à juger la localisation sonore décalée vers une tache lumineuse qui clignotait en rythme avec la source sonore plutôt que vers une tache lumineuse qui clignotait de façon asynchrone.

Les phénomènes qui ont influencé la localisation sonore étaient bien sûr la *position* même de la source lumineuse et sa *simultanéité* avec la source sonore, il s'agissait donc de deux caractéristiques visuelles physiques. Cependant, l'*action* ressort comme un paramètre déterminant dans ces expériences. Si dans le champ visuel des observateurs, quelque chose change ou bouge, ceux-ci montrent une tendance à juger la localisation sonore décalée vers ce lieu d'agitation. Ce phénomène est lié à la propriété des signaux sonores de nature impulsive (montées et descentes rapides de l'amplitude) à attirer notre attention, voire même à nous interpeller.



**Figure 2.9 :** Expérience de Thomas [1941] : les stimuli visuels étaient des lampes fixes, le haut-parleur était invisible et pouvait prendre l'une des 5 positions sur un trajet semi-circulaire.

Quelques années plus tard, JACKSON [1953] a effectué deux expériences où il a, à son tour, remarqué l'influence de la cadence temporelle des stimuli sonore et visuel sur la localisation sonore. Deux autres paramètres apparaissent dans son travail : *l'ampleur de la discordance spatiale* et *la connaissance préalable*.

La facilité avec laquelle les observateurs jugent la localisation sonore décalée vers la localisation visuelle dépend largement de la discordance spatiale entre les stimuli. Quand celle-ci augmente, les

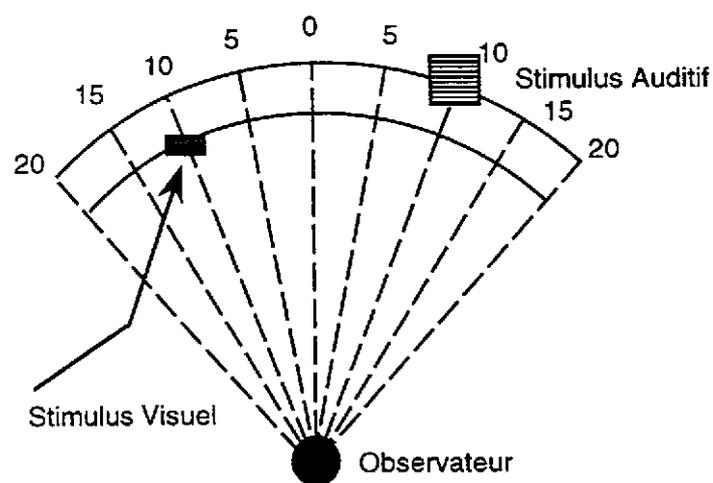
observateurs associent plus difficilement les deux stimuli et jugent plus correctement la localisation sonore.

La connaissance préalable dépend de l'expérience relative à l'apparition simultanée des stimuli hétéromodaux dans le passé. Jackson a utilisé deux couples de stimuli dans ses expériences, chacun étant caractérisé par un niveau différent de 'connaissance préalable' : le premier était le son d'une sonnette dissimulée et une série de sonnettes 'visuelles' accompagnées de la lumière d'une lampe ; le second était le sifflement d'une bouilloire et une série de bouilloires 'visuelles' qui pouvaient émettre un jet de vapeur silencieux. Il a remarqué que le niveau de connaissance préalable relative aux stimuli influence fortement les interactions spatiales : l'influence de la localisation visuelle sur la localisation auditive apparaît sous des conditions de séparation spatiale bien plus importante pour le cas des bouilloires accompagnées de jet de vapeur, que pour le cas des sonnettes accompagnées de lumière (haut et bas niveaux d'association cognitive respectivement). Toutefois, lorsque les séparations spatiales augmentent encore, les interactions cessent d'exister, même pour le cas le plus 'crédible' des bouilloires. Enfin, les résultats de Jackson sont accompagnés de mesures de temps de réaction (TR) qui servent d'indice de confiance sur les réponses des sujets. Ainsi, le TR dans le cas des bouilloires est inférieur à celui des sonnettes, et il augmente dans les deux cas avec la séparation spatiale.

WARREN [1979] a diminué l'intensité de l'éclairage (le stimulus visuel était un haut-parleur) de la salle de tests afin de savoir si l'influence de la localisation visuelle sur la localisation sonore allait diminuer également. Les résultats n'ont pas été probants.

En 1981, WARREN [1981] a encore réalisé une expérience d'incompatibilité spatiale pour tester l'influence de deux facteurs sur les interactions audiovisuelles : *le niveau de cohérence* des stimuli et *le contenu des informations préliminaires* données aux sujets concernant la discordance spatiale.

Il a testé trois niveaux de cohérence : aux niveaux haut et moyen, les observateurs voyaient sur un écran TV un présentateur qui lisait un texte, et sa voix était diffusée par un haut-parleur. Au niveau de haute cohérence, le stimulus sonore était synchrone au stimulus visuel, tandis qu'au niveau moyen le stimulus sonore arrivait avec un retard de 150 ms. Pour le niveau bas de cohérence, l'écran TV restait allumé mais sans image, avec une bande adhésive collée à l'endroit où, dans les autres conditions, se trouvait la bouche de l'orateur. Des séparations angulaires constantes de 10° ou 20° étaient utilisées (figure 2.10).



**Figure 2.10** : Expérience de Warren [1981] : le stimulus visuel était un écran TV, le stimulus auditif un haut-parleur invisible. Trois niveaux de cohérence étaient testés ainsi que deux types d'informations préliminaires (séparations angulaires de 20°).

Warren a trouvé que l'influence de la localisation visuelle sur la localisation auditive est bien plus grande que l'influence réciproque. Il a conclu que plus les stimuli sont cohérents, plus les observateurs se trompent sur leur jugement de localisation. La séparation spatiale des stimuli doit augmenter pour que les sujets commencent à les dissocier. La simultanéité des stimuli joue aussi un rôle important : quand les stimuli sont concordants et simultanés, les interactions spatiales sont d'autant plus évidentes. Enfin, il a observé que le contenu des informations préliminaires données aux observateurs peut aussi influencer les résultats : lorsque les observateurs ne sont pas informés de la possibilité d'une discordance éventuelle, ils sont plus facilement trompés sur la localisation sonore que s'ils en sont avertis.

RAGOT et al. [1988] ont montré l'influence de l'*orientation d'attention* des observateurs sur les interactions spatiales entre vision et audition, les stimuli étant toujours bimodaux. L'attention des observateurs peut être orientée par l'expérimentateur uniquement sur le stimulus visuel, uniquement sur le stimulus auditif, ou enfin, sur le stimulus bimodal. Il a été suggéré que la dominance visuelle trouvée par Colavita [1974] devait être attribuée au fait que l'orientation d'attention des observateurs dans ses expériences était bimodale, c'est à dire partagée entre les deux modalités. Le fait que l'information visuelle semble être accessible plus rapidement que l'information auditive, a conduit à une dominance visuelle qui n'est pas observée dans le travail de Ragot et al. quand l'attention des sujets est orientée vers la seule modalité auditive. Les instructions données aux sujets par rapport à la modalité sensorielle à laquelle ils doivent porter leur attention, jouent donc un rôle important sur la dominance ou non de la modalité visuelle, et par conséquent sur les interactions audiovisuelles.

EASTON [1983] a montré que, lorsque le mouvement de la tête des observateurs est permis, ceux-ci peuvent mieux désigner la localisation de la source sonore, et que par conséquent, ils sont moins influencés par la perception visuelle.

Une expérience de *mouvement apparent auditif* où des informations visuelles interviennent dans le phénomène, est présentée dans ce travail. Par mouvement apparent auditif, on entend une illusion perceptive de mouvement d'une source sonore sur un trajet qui est décrit par l'activation alternative et rapide d'une série de sources immobiles, placées sur ce trajet.

LAKATOS [1993] a voulu étudier la faculté de percevoir la direction du mouvement apparent auditif sur un trajet circulaire, dans un plan horizontal autour de la tête d'un auditeur. Il a mis en évidence une difficulté de la part des sujets à percevoir la direction de ce mouvement, sur le trajet frontal, lorsque les deux hémisphères (frontal et postérieur) étaient parcourus successivement. Les sujets ne rencontraient pas cette même difficulté pour percevoir la direction du mouvement sur le trajet postérieur. Les expériences ont montré qu'une seule stimulation lumineuse placée au centre de l'hémisphère frontal et activée en même temps que la stimulation sonore correspondante, favorise fortement la détection de la direction du mouvement sur ce trajet, malgré le fait qu'elle ne porte pas d'information sur le sens du mouvement.

Lacatos a conclu que des informations visuelles semblent être nécessaires pour la résolution du mouvement auditif sur l'hémisphère frontal quand les deux hémisphères, frontal et postérieur, sont stimulés. Il suggère que la représentation mentale (la manière dont les informations sont stockées) pour l'hémisphère frontal est amodale, c'est à dire qu'elle ne dépend pas d'une seule modalité sensorielle, tandis qu'elle est uniquement auditive pour l'hémisphère postérieur.

Deux hypothèses principales ont été proposées pour expliquer les effets facilitatoire ou perturbateur de la vision sur la localisation sonore (dans les cas de coïncidence spatiale ou de discordance spatiale

respectivement). La première hypothèse considère que la vision fournit un cadre de référence spatial pour les jugements auditifs, et suppose que la vision sert à organiser l'espace auditif [Warren, 1970]. La deuxième hypothèse se base sur le fait que le libre mouvement des yeux, même dans l'obscurité, semble faciliter la localisation sonore ; elle suggère ainsi que la vision facilite la mémoire spatiale [Jones, 1975], [Jones et Kabanoff, 1975].

Enfin, d'autres travaux comme ceux de Shelton et Searle [1980] montrent que ces deux processus ne doivent pas être considérés comme mutuellement exclusifs. Il semble en effet qu'ils soient tous les deux impliqués dans l'influence visuelle et que, au moins pour le cas des environnements aux contextes complets (full-context environnements), ils doivent être considérés comme des mécanismes complémentaires.

### 2.4.2.3 Autres influences entre vision et audition

#### *i) Interactions temporelles*

Plusieurs études ont traité le sujet de la synchronisation – ou de la désynchronisation – subjective entre son et image. Le résultat principal est qu'il existe une relative grande tolérance des sujets vis-à-vis des décalages intersensoriels par rapport à la résolution temporelle de chacune des modalités, visuelle et auditive. De plus, une nette asymétrie est observée sur la sensibilité que montrent les êtres humains par rapport aux deux situations suivantes : le son précédant l'image ou l'image précédant le son. Les retards du son par rapport à l'image sont ainsi bien plus acceptables que les retards de l'image par rapport au son. L'interprétation la plus couramment donnée à cette asymétrie est le retard naturel du son par rapport à l'image, dû à une vitesse de propagation bien inférieure pour le son.

Les retards acceptables dépendent largement du type de stimuli utilisés : pour les stimuli de nature impulsive (comme un marteau qui frappe par exemple), les retards sont moins acceptables que pour les stimuli de nature plus continue (comme par exemple la parole) [Dixon et Spitz, 1980], [Cavé et al., 1992] ; les réactions peuvent dépendre également des consignes données aux sujets [Kohlrausch et de Par, 1999].

#### *ii) Influence de l'éclairage sur la sensibilité auditive*

Des interactions dues aux informations provenant de la perception visuelle, sans être perçues auditivement, peuvent influencer la perception auditive. C'est le cas de l'influence de l'éclairage sur la perception de la hauteur du son et du niveau sonore.

Il a en effet été observé que des stimulations visuelles peuvent influencer la sensibilité auditive. Le seuil de perception d'intensité sonore et le seuil de discrimination d'intensité sonore semblent être abaissés avec la présence d'éclairs lumineux, cet effet disparaît dans le temps [Gribenski, 1978].

HARTMAN [1934] a voulu tester la discrimination de la hauteur et de l'intensité sonore sous différentes conditions d'éclairage. Il a trouvé que sous des conditions de bon éclairage, la performance des sujets était faiblement améliorée par rapport à la pénombre ou à l'obscurité. La discrimination dans l'obscurité était meilleure que dans la pénombre. On pourrait conclure que l'acuité auditive semble être favorisée lorsqu'il n'y a pas d'ambiguïté dans le champ visuel des sujets (bien voir ou ne pas voir du tout).

### iii) Equivalences intermodales

MARKS [1987] a étudié des *similarités ou équivalences intermodales*.

Les stimuli sonores jouaient sur deux facteurs :

- Hauteur Sonore : son bas ou haut (*Pitch : low-high*)
- Niveau Sonore : niveau faible ou fort (*Loudness : soft-loud*)

et les stimuli visuels jouaient sur trois caractères :

- Luminosité : sombre ou lumineux (*Lightness : dark-light*)
- Clarté : foncé ou clair (*Brightness : dim-bright*)
- Forme : arrondie ou aiguë (*Shape : round-sharp*)

Chaque test comprenait quatre stimuli provenant du couplage d'un facteur auditif avec un caractère visuel. Les observateurs ont montré plus de facilité à discriminer la variation sur un facteur auditif lorsque le stimulus auditif était accompagné d'un stimulus visuel correspondant. Ainsi, il a trouvé des équivalences pour les couples : hauteur sonore avec luminosité, hauteur sonore avec clarté des couleurs, niveau sonore avec luminosité, et hauteur sonore avec forme.

## 2.4.3 Méthodes expérimentales et facteurs qui influencent les interactions audiovisuelles

La méthode la plus souvent utilisée pour étudier les interactions spatiales entre vision et audition est la méthode de la discordance spatiale. Elle consiste à effectuer des tests bimodaux en mettant en conflit spatial plus ou moins grand les deux modes de perception, et à tester la localisation issue de l'une des modalités sous des conditions de localisation conflictuelles avec l'autre modalité : les sujets doivent ainsi localiser visuellement un objet en présence de stimuli auditifs (tests de conflit visuel) ou localiser auditivement la source sonore en présence de stimuli visuels (tests de conflit auditif). Des tests de contrôle, uniquement visuels et uniquement auditifs, sont également effectués.

La méthode d'analyse utilisée est dans la plupart des cas l'Analyse de Variance, avec comme variable dépendante l'erreur moyenne de localisation en degrés. Elle a l'avantage de permettre d'étudier simultanément les effets de plusieurs paramètres sur les résultats ainsi que leurs interactions. Les paramètres les plus souvent étudiés sont : l'influence de la vision sur l'audition et vice-versa, l'influence de l'ampleur de la discordance spatiale, le niveau de cohérence cognitive et la cadence relative des stimuli, l'orientation d'attention des sujets, d'autres consignes relatives à la nature des stimuli, etc.

Ces paramètres jouent un rôle primordial sur les interactions entre vision et audition. Ils déterminent leur apparition et influencent l'ampleur des interactions. Les plus importants sont regroupés ci-dessous, lorsque cela est possible, ils sont distingués en facteurs physiques, cognitifs ou mixtes.

i) La *discordance spatiale* : l'ampleur des interactions diminue avec l'augmentation de la discordance spatiale entre les stimuli sonores et les stimuli auditifs. Les interactions sont favorisées avec la coïncidence spatiale. Il s'agit d'un facteur de nature physique [Jackson, 1953].

ii) La *cadence relative* des stimuli audiovisuels. Les retards temporels entre stimuli hétéromodaux affaiblissent fortement les interactions ; ces dernières bénéficient beaucoup de la coïncidence temporelle (simultanéité, synchronisme). C'est un facteur de nature physique, mais l'asymétrie

observée par rapport aux retards acceptables (son avant ou après l'image) suggère que les attentes – autrement dit des facteurs cognitifs – y jouent un rôle [Dixon et Spitz, 1980], [Warren, 1981], [Cavé et al., 1992].

iii) *La simultanéité du rythme* ou la *cohérence du mouvement visuel et auditif*. Lorsqu'un changement apparaît sur le plan visuel et auditif et que ce changement est cohérent dans les deux modalités, les interactions audiovisuelles sont considérablement favorisées. C'est le cas par exemple de flashes lumineux qui suivent un même motif rythmique que la stimulation sonore qui les accompagne [Thomas, 1941], ou celui d'un message sonore de parole correspondant au mouvement des lèvres du présentateur que l'on voit.

iv) *L'intensité relative* des stimuli audiovisuels. Il est généralement admis dans le domaine de l'audiovisuel que l'intensité sonore correspond à la taille de l'image et non pas à sa luminosité par exemple [Woszczyk et al., 1995], [Bech et al., 1995]. C'est un facteur qui semblerait être partiellement lié à la connaissance, les objets visuels de grande taille émettent habituellement des sons de haut niveau sonore.

v) *Le mouvement des yeux, de la tête*. L'influence de ces paramètres sur les interactions est complexe : le libre mouvement des yeux vers la direction du son, même sous des conditions d'obscurité, contribue à la facilitation visuelle de la localisation sonore. Aussi, le mouvement des yeux peut améliorer la localisation sonore dans le cas de stimuli concordants et peut la détériorer dans le cas de stimuli discordants. L'hypothèse proposée pour expliquer ces effets est que le mouvement des yeux aide à maintenir la mémoire spatiale et à organiser ainsi la mémoire auditive [Jones et Kabanoff, 1975].

vi) *La connaissance préalable* ou l'*expérience vécue*. Les interactions sensorielles dépendent en grande partie de nos représentations cognitives du monde, c'est à dire des informations mémorisées tout au long de notre vie. Ces représentations se forment à l'aide de l'apparition simultanée de stimuli hétéromodaux dans le passé. Le plus souvent ces stimuli apparaissent ensemble, plus forte est la représentation qui en résulte et plus les interactions sont favorisées. Ainsi, le son d'une bouilloire est rapidement associé à l'image d'une bouilloire, le son d'une voiture à l'image d'une voiture, ou encore – avec une représentation moins forte – à un contexte urbain [Jackson, 1953], [Warren, 1981].

La connaissance préalable englobe un sens quelque peu plus large que l'expérience vécue, puisqu'elle concerne aussi la connaissance apprise. Nous pouvons par exemple développer des attentes par rapport à un événement par le simple fait d'en avoir entendu parlé. Il est par ailleurs connu que la réputation d'une salle peut influencer l'évaluation de sa qualité acoustique. La connaissance préalable, l'expérience vécue, les attentes, la réputation etc. sont des paramètres purement cognitifs.

vii) *L'entraînement* ou l'*adaptation* aux conditions expérimentales. Le processus perceptif n'étant pas indépendant du temps, il s'agit de toute sorte d'expérience acquise pendant les tests. Le caractère très répétitif de ces derniers associé à la nature relative – et donc influençable – de nos jugements, peuvent fortifier les interactions parce que l'on s'habitue même à des conditions expérimentales peu crédibles : on peut par exemple apprendre à associer un son et une image qui, au début étaient considérés comme complètement décorrélés, on peut s'adapter à un décalage spatial au point de ne plus l'apercevoir, s'habituer à un changement constant dans notre environnement et ne plus y porter attention, etc. Ces phénomènes peuvent persister un certain temps après les expériences, et puis avec l'exposition normale, disparaissent ensuite [Château, 1998].

viii) *L'orientation d'attention*. L'attention est en général un facteur très important pour le processus d'évaluation qualitative, et particulièrement important pour des évaluations impliquant une exposition multimodale. Notre système de traitement d'information a une capacité limitée. Le rôle de l'attention est justement de sélectionner les entrées qui doivent être analysées par notre organisme. Il est ainsi généralement admis que la modalité (ou la tâche) vers laquelle on focalise notre attention, est moins influencée par des stimulations provenant des autres modalités.

Dans le cas des expériences sur les interactions sensorielles, l'attention peut être orientée volontairement par l'expérimentateur vers l'une ou l'autre modalité sensorielle. L'orientation d'attention vers l'audition par exemple, résulte en une influence bien moins forte de la vision sur l'audition [Ragot et al., 1988]. Dans le cas de l'orientation d'attention bimodale – c'est à dire partagée entre deux modalités – c'est la modalité visuelle qui, pour des tâches spatiales, semble être la plus favorisée [Colavita, 1974].

L'attention peut aussi être orientée ou réorientée involontairement<sup>1</sup>, c'est à dire non pas par l'expérimentateur, mais par les conditions d'exposition elles-mêmes.

Un changement d'attention peut être provoqué par des propriétés physiques et surtout temporelles des stimuli : un signal à caractère impulsif, un mouvement, un changement large et rapide dans notre environnement, attirent involontairement notre attention vers la modalité sensorielle concernée. Cette modalité se trouve favorisée par rapport aux autres, son traitement s'effectue prioritairement et peut influencer le traitement des informations hétéromodales [Thomas, 1941].

Par ailleurs, les mêmes propriétés physiques peuvent dans des conditions d'exposition normales, conduire à une sensation d'alerte (signaux d'avertissement par exemple, [Vogel, 1999]). Même si la sensation d'alerte est rarement rencontrée dans des conditions habituelles de laboratoire, les signaux d'une telle nature peuvent nous interpeller et attirer notre attention. Ils cessent d'attirer notre attention s'il sont attendus ou prévisibles (cf. § 4.9.4).

Un changement d'attention peut aussi être provoqué par des propriétés des stimuli que l'on peut qualifier de cognitives. Il s'agit de tout ce qui peut attirer notre intérêt dans le sens large du terme. Ainsi, dans le cas général, un stimulus complexe est plus intéressant qu'un stimulus simple : par exemple, une image vidéo est plus intéressante que le clignotement d'une lampe. Un message de parole dans une langue que l'on comprend est plus intéressant et a plus de sens qu'un message dans une langue étrangère. Plus le sens que porte un signal est riche, plus il est considéré intéressant ou important, et plus il attire notre attention. La modalité sensorielle concernée peut alors se trouver dans une position favorable du point de vue attentionnel et influencer les autres modalités. [Woszcyk et al., 1995].

Ce phénomène pourrait par ailleurs expliquer les résultats trouvés par certains chercheurs vis-à-vis de signaux sonores de la parole. Ces signaux, parmi d'autres signaux sonores utilisés, semblent souvent être les seuls à ne pas être influencés par la modalité visuelle [Violon et Lavandier, 1999]. Or, la parole est le signal sur lequel un être humain a appris à porter son attention, c'est son propre son. Le message qu'elle porte a donc en général beaucoup plus de sens que les autres signaux, il nous intéresse davantage, on s'y trouve impliqués, même involontairement. L'attention ainsi attirée vers la modalité auditive empêche les influences de nature visuelle de se manifester.

Pour conclure, disons qu'afin d'éviter des biais d'une telle nature cognitive dans les résultats, il est très important de contrôler la signification des stimuli en essayant de respecter un équilibre entre l'intérêt, la complexité et le sens que ces stimuli portent dans les deux modalités.

ix) Les notions du *contexte* et de *l'activité* sont bien entendu indispensables à la discussion. Les êtres humains peuvent être impliqués de différentes manières dans les événements qui les entourent. Ils peuvent ainsi être 'spectateurs' ou 'acteurs'.

---

<sup>1</sup> Dans les sciences cognitives, il est admis que le processus d'attention a deux composantes fondamentales. Premièrement, des mécanismes d'alerte qui détectent de larges changements des grandeurs perceptibles dans l'environnement. Deuxièmement, les mécanismes qui ont pour but de sélectionner les entrées qui nécessitent des réponses de notre organisme, et ensuite de les analyser et d'y répondre [Glass et Holyoak, 1986].

Dans le premier cas, on considère que le but principal est le plaisir, comme c'est le cas dans une salle de concert ou une exposition photographique par exemple. Il s'agit principalement de la dimension hédonique.

Dans le deuxième cas, il s'agit principalement de vivre ou 'd'utiliser' notre environnement (la dimension hédonique peut aussi exister, mais ne reflète pas le but principal). Ainsi, rien que le fait de vivre dans une ville par exemple, nous rend acteurs dans ce contexte. De plus, on peut y être impliqué de manière passive ou active : entendre ou même écouter le bruit de la circulation par exemple, a une connotation passive, tandis que participer à la création de ce bruit nous rend forcément actifs. Enfin, en mode passif, il sera différent d'entendre le bruit d'un bus quand on essaie de dormir ou d'écouter la radio, ou si l'on est en train de l'attendre à son arrêt.

Il est en effet généralement accepté que la perception des différents phénomènes peut être influencée par l'activité ou l'intérêt des sujets<sup>1</sup>. Aussi, le degré d'implication dans une situation peut influencer considérablement notre perception : pour prendre un autre exemple que celui du bus, il suffit de considérer la manière dont on réagit au 'son' de notre poste de télévision par rapport au 'bruit' du poste du voisin, aux travaux chez-nous par rapport aux travaux dans la rue, etc. Ce degré d'implication est enfin un paramètre qui peut être aussi lié à l'attention et peut de manière conjointe influencer les interactions.

x) *Les informations préliminaires aux tests* données aux sujets, concernant l'existence éventuelle d'une discordance sensorielle : les sujets savent-ils que les stimuli sonore et visuel peuvent être dus à deux événements distincts, ou pensent-ils qu'il s'agit de différentes manifestations d'un même événement ? Dans le deuxième cas, l'apparition des interactions est plus sûre et leur ampleur sera évidemment plus grande [Warren, 1981].

Dans tous les cas – c'est à dire avec ou sans consignes qui concernent la nature des stimuli – l'hypothèse que se fait l'observateur au sujet de leur origine est très importante. Les stimuli sont représentés mentalement par un événement unique ou par deux événements indépendants. "L'hypothèse d'événement unique" que les êtres humains ont tendance à essayer de concrétiser, dépend des facteurs mentionnés ci-dessus, mais principalement de l'ampleur de la discordance, de la connaissance préalable et des informations préliminaires.

Mémoire, connaissance, expérience, représentations ou associations mentales, attentes, réputation, apprentissage, adaptation, entraînement, orientation d'attention, intérêt, degré d'implication, complexité, sens, compatibilité, cohérence, consignes, activité, contexte, sont tous des paramètres importants dans le processus des interactions sensorielles, ils déterminent en effet l'apparition même et l'ampleur des ces dernières. Ils doivent absolument être compris et contrôlés, et sinon, pris activement en compte afin de favoriser (ou de défavoriser) les interactions entre les sens.

En conclusion, disons que les mots clés pour l'apparition des interactions sont les notions similaires de cohérence, compatibilité, équilibre, etc. entre les stimuli hétéromodaux : ainsi, compatibilité spatiale, compatibilité temporelle, cohérence du rythme et du mouvement, équilibre des intensités, compatibilité cognitive, équilibre de l'intérêt et de l'attention, sont toutes des notions que doivent incorporer les études ou les applications dont le but repose sur les interactions sensorielles.

---

<sup>1</sup> Des études existent d'ailleurs sur le sujet, notamment sur la perception de la durée des événements en fonction de l'activité [Marmaras, 1995].

## **2.5 DISCUSSION A PARTIR DE L'ETUDE BIBLIOGRAPHIQUE - OBJECTIFS**

Dans cette discussion est présentée une brève revue des conclusions des recherches bibliographiques dans les domaines : de l'acoustique des salles, des interactions sensorielles entre vision et audition et de la perception de la distance auditive. La synthèse de ces conclusions nous permettra de définir le cadre dans lequel s'inscrit le présent travail et de présenter les approches expérimentales adoptées. Nous justifierons le choix de ces méthodes ainsi que le choix de l'attribut de la distance auditive pour les expériences de la première partie de cette étude. Nos objectifs deviendront clairs au fur et à mesure de cette discussion et seront présentés de manière récapitulative à la fin.

Nous avons pu constater à travers les études portant sur la qualité acoustique des salles, que cette qualité constitue un espace perceptif multidimensionnel qui peut être décrit à l'aide d'un ensemble limité d'attributs perceptifs, traduisant les principales sensations auditives créées chez les auditeurs lors d'une représentation musicale. Des études supplémentaires ont été menées pour permettre de corrélérer ces sensations à des grandeurs physiques capables de décrire de manière objective la qualité acoustique.

Plusieurs corrélations entre facteurs perceptifs et indices acoustiques ont été proposées, elles peuvent être distinguées en quatre groupes correspondant aux quatre dimensions physiques fondamentales (et par conséquent psychophysiques) : la dimension de l'intensité, la dimension temporelle, la dimension spatiale et la dimension fréquentielle. Les corrélations obtenues pour les sensations liées à l'intensité et à la dimension temporelle sont plutôt satisfaisantes, tandis que celles obtenues pour les sensations liées à la dimension d'espace et à la dimension fréquentielle semblent pouvoir bénéficier d'améliorations. De telles améliorations semblent être plus nécessaires pour la dimension spatiale du son car les aspects spatiaux du son sont très demandés, ils apparaissent en effet très bien placés dans la hiérarchisation des préférences des auditeurs.

L'origine des faibles corrélations entre attributs perceptifs et grandeurs physiques se trouve principalement dans le processus complexe de l'appréciation subjective. Ainsi, les préférences interindividuelles, l'expérience d'écoute, les attentes, la mémoire, et encore d'autres facteurs cognitifs et même psychologiques, peuvent influencer cette appréciation. Un problème d'origine différente est celui des informations provenant des modalités sensorielles autres que la modalité auditive et qui, sans être directement étudiées, peuvent influencer l'évaluation des attributs auditifs. Ceci est d'autant plus vrai pour la perception visuelle dans les salles de concert, puisque ces salles sont conçues dans le but de plaire à leurs auditeurs-spectateurs non seulement auditivement, mais aussi visuellement. La différence des informations visuelles par rapport aux autres limitations liées à l'appréciation subjective, se situe dans le fait qu'il s'agit du seul facteur issu d'une stimulation physique, et donc reproductible d'une part, mesurable de l'autre, et ceci de manière objective. Or, l'étude de son influence a été relativement négligée dans les travaux du domaine de l'acoustique des salles.

Ces considérations nous amènent à rappeler l'objectif principal de ce travail : il s'agit d'étudier l'influence des informations visuelles sur l'appréciation de la qualité acoustique des salles, dans le but d'améliorer les corrélations entre facteurs perceptifs et indices acoustiques, ainsi qu'aider à la compréhension du processus complexe de l'appréciation subjective

Il est intéressant de voir à partir de la recherche bibliographique quels sont les attributs auditifs sur lesquels des informations visuelles semblent plus susceptibles d'intervenir.

D'une part, certains travaux du domaine de l'acoustique des salles optant pour des expériences en conditions réelles de concert, sans étudier le sujet spécifiquement, ont permis d'observer l'influence des informations visuelles sur l'appréciation de la qualité acoustique. Les paramètres visuels qui ont

changé durant ces expériences étaient, la distance entre l'auditeur et la scène, et l'emplacement de l'auditeur par rapport aux parois de la salle. Les facteurs auditifs influencés étaient la distance auditive de la source sonore (ou l'éloignement, la proximité, la présence etc.), l'intimité, l'enveloppement, mais aussi l'impression générale, et dans l'une des études, la puissance sonore. Nous remarquons que les trois premiers de ces facteurs sont principalement liés à la dimension spatiale du son, ce qui nous incite à supposer que c'est surtout les attributs perceptifs spatiaux qui sont les plus susceptibles d'être influencés par la modalité visuelle.

D'autre part, le domaine de la psychologie expérimentale est riche de preuves sur la collaboration des modalités visuelle et auditive, notamment dans la dimension spatiale. L'intégration d'informations sensorielles en provenance de ces deux modalités ainsi que de nos représentations mentales du monde, constituent en effet un cadre de référence unique sur lequel se base la localisation spatiale. La perception de la direction et de la profondeur se fait ainsi dans un espace perceptif unique. Par ailleurs, si l'on considère que traditionnellement, la vision l'emporte sur l'audition lors des tâches spatiales, on comprend l'importance de l'étude particulière de l'influence visuelle sur les attributs spatiaux du son.

Nos expériences visent ainsi à étudier de manière plus approfondie l'influence des aspects visuels des salles sur les facteurs perceptifs liés à la dimension spatiale du son. Néanmoins, les autres attributs liés à la dimension d'intensité, aux dimensions temporelle et fréquentielle ne sont pas négligés. Ils contribuent en effet à la qualité acoustique globale et constituent un ensemble perceptif avec les aspects spatiaux. De plus, l'étude simultanée des facteurs spatiaux et non-spatiaux permettra de vérifier l'hypothèse avancée selon laquelle les premiers ont tendance à être plus influencés par la perception visuelle.

Pour répondre à ces objectifs, nous avons choisi d'employer deux méthodes expérimentales issues de deux disciplines scientifiques différentes. Nous les présentons brièvement et justifions leur choix :

D'une part, nous adopterons l'approche méthodologique traditionnellement utilisée dans le cadre des travaux sur les interactions sensorielles. Il s'agit de la méthode visant à créer des situations intermodales conflictuelles et à étudier l'influence d'une modalité sur l'autre. La principale raison de ce choix est évidente : il s'agit en effet de la méthode spécialement conçue pour étudier les interactions sensorielles, il serait donc dommage de ne pas profiter de son apport tant d'un point de vue méthodologique que théorique. Deux autres raisons concourent : il s'agit d'une méthode qui permet l'étude d'un seul facteur à la fois – dans notre cas nous nous intéresserons à un facteur auditif spatial – et il s'agit aussi d'une méthode flexible qui permet de modifier et ainsi d'étudier plusieurs aspects de la procédure expérimentale; les conditions d'apparition des interactions spatiales pourront ainsi être approfondies. Le facteur que nous avons choisi d'étudier est la distance auditive des sources sonores, les raisons sont rapportées dans la suite de cette partie.

D'autre part, nous adopterons une méthode similaire à celle utilisée dans les travaux traditionnels de la qualité acoustique des salles. Il s'agit de la méthode des questionnaires portant sur un ensemble d'attributs perceptifs descriptifs de la qualité acoustique des salles; elle sera adaptée pour incorporer des indices visuels et étudier ainsi leur influence sur les attributs auditifs. Ce choix est motivé :

- par le fait que l'emploi de la méthode classique de l'acoustique des salles nous permettra de comparer nos résultats aux résultats typiques du domaine ; par ailleurs, il s'agira de tester si cette méthode peut prendre en compte des influences intermodales, autrement dit, de tester si l'espace auditif multidimensionnel typiquement décrit par ces méthodes peut être considéré comme un espace perceptif multidimensionnel et aussi multisensoriel ;
- par l'étude simultanée des attributs issus des quatre dimensions fondamentales physiques (intensité, temps, espace et fréquence) qui nous permettra de valider notre hypothèse par rapport à la nature des attributs qui sont plus susceptibles d'être influencés par la perception visuelle, c'est à dire les attributs liés à la dimension spatiale.

Pour la première partie de ces expériences, adoptant la méthode empruntée au domaine des interactions sensorielles, nous avons choisi parmi les attributs spatiaux du son celui de la distance auditive des sources sonores. Les raisons de ce choix, issues principalement de notre étude bibliographique, sont présentées à la suite.

En ce qui concerne la distance auditive:

- il s'agit d'un attribut spatial beaucoup moins étudié que l'enveloppement ou la largeur apparente de source ;
- il existe des indications concernant l'influence de la distance visuelle sur différents attributs sonores dans les études de l'acoustique des salles ;
- elle présente de fortes corrélations avec l'intimité, attribut perceptif d'importance établie ;
- elle présente moins d'ambiguïtés sémantiques que les autres attributs spatiaux du son, ce qui permettra de réduire les biais introduits dans nos résultats<sup>1</sup> par les différentes représentations cognitives et lexicales ;
- la simulation visuelle de la distance visuelle est relativement facile, tant du point de vue de la conception que de celui de la pratique.

Nos objectifs présentés de manière déductive ci-dessus, ainsi que les raisons qui nous ont incité à entreprendre nos travaux, sont présentés et regroupés à la suite :

Notre objectif principal est d'étudier l'influence des informations visuelles sur l'appréciation de la qualité acoustique des salles. Cette étude devrait constituer une étape nécessaire vers une amélioration des corrélations des facteurs perceptifs avec les indices objectifs, améliorations qui peuvent s'avérer utiles dans la conception ou la modification des salles ainsi que dans la collaboration entre architectes et acousticiens.

Dans un plan plus général, cette étude devrait permettre une meilleure compréhension du processus de l'évaluation subjective, une évaluation qui implique le plus souvent des informations issues de plusieurs modalités sensorielles. L'influence des facteurs visuels bien sûr, mais aussi celle des facteurs attentionnels et cognitifs particulièrement importants dans la perception multimodale, seront ainsi largement discutées.

Enfin, ce travail a pour but d'étudier et de mettre en place une méthodologie cohérente, permettant d'étudier les interactions sensorielles, tant dans le cadre de la simulation des environnements réels (tels que les salles de spectacle) que dans le cadre des environnements artificiels (tels que les environnements synthétiques créés pour les applications de télécommunications ou de multimédia).

---

<sup>1</sup> Il semble exister un consensus bien plus fort que pour les autres attributs spatiaux sur ce que signifie la distance d'une source sonore. On est en effet plus habitué à juger la distance d'un son que son enveloppement ou son ampleur spatiale. Ceci est probablement dû au fait que l'évaluation de la distance auditive étant plus liée à la notion du danger – on a plus souvent *besoin* d'estimer la distance d'un son – elle est plus essentielle, donc plus courante, elle présente ainsi moins d'ambiguïtés sémantiques.

## 3 METHODE

### 3.1 INTRODUCTION

Les deux méthodes expérimentales utilisées dans le cadre de ce travail sont empruntées à deux domaines scientifiques bien distincts entre eux : le domaine des interactions audiovisuelles dans le cadre de la psychologie expérimentale et celui de l'acoustique des salles. Lorsque nécessaire, ces méthodes ont été adaptées pour tenir compte des particularités de la présente étude.

Ainsi, pour étudier l'influence de la distance visuelle sur la distance auditive, une des méthodes bien établies de la psychologie expérimentale a été utilisée (cf. § 2.4) : des stimuli sonores<sup>1</sup> représentant différentes distances auditives ont été couplés avec des stimuli visuels qui – à leur tour – représentaient différentes distances visuelles. Ces stimuli bimodaux ont été présentés à des sujets qui devaient en juger la distance auditive. Les stimuli de notre travail présentent la particularité d'être plus complexes que ceux que l'on rencontre dans les études fondamentales des interactions spatiales (où il s'agit le plus souvent de stimuli simples : sons purs, bruits, flashes lumineux), mais les différences se limitent ici : le souci de contrebalancer les éventuels effets d'ordre dans les réponses est maintenu et la même analyse statistique est employée.

Ensuite, en ce qui concerne l'étude de l'influence visuelle sur un ensemble de facteurs acoustiques, caractéristiques de la qualité acoustique des salles, la méthode employée est similaire à celle utilisée dans les travaux traditionnels du domaine (cf. § 2.2) : plusieurs simulations du champ sonore d'une salle sont présentées aux sujets qui sont priés de répondre à l'aide d'un questionnaire. Le questionnaire est composé d'échelles linéaires correspondant à des attributs perceptifs descriptifs de la qualité acoustique des salles. La seule différence avec les travaux classiques est que les sons ont été diffusés sous des conditions visuelles variées, quatre simulations du champ visuel des spectateurs-auditeurs ont été créées à ce propos. Le même type d'analyse que celui utilisé habituellement est employé pour le dépouillement des résultats (analyse multidimensionnelle).

La description détaillée de chacune des deux méthodes expérimentales telles qu'elles ont été appliquées ici, est donnée aux chapitres correspondants (§ 4.2 et § 5.2.2).

Les expériences ont nécessité de simuler une salle tant de manière sonore que de manière visuelle : la méthode de simulation sonore employée (synthèse de scènes sonores à l'aide du *Spatialisateur*), la méthode de simulation visuelle (projection des vues stéréoscopiques) ainsi que la salle de tests et le dispositif expérimental, sont présentés dans ce chapitre. Les outils statistiques principaux (ANOVA et ACP) utilisés dans le cadre de ce travail sont présentés ensuite.

---

<sup>1</sup> Une distinction doit être faite entre les différentes notions de stimuli : du point de vue de l'acoustique des salles, on entend par stimuli, les sons anéchoïques utilisés pour exciter les salles et traduire ainsi la transformation acoustique apportée par ces dernières sur les signaux sonores. Du point de vue des expériences dans ce travail, on entend par stimuli, les sons présentés aux sujets (il s'agit de sons déjà modifiés par l'acoustique des salles en question, ici, par les différentes configurations du *Spatialisateur*).

## 3.2 LA SIMULATION SONORE

La synthèse des scènes sonores utilisées dans le cadre de ce travail a été réalisée grâce au *Spatialisateur*, un processeur d'acoustique virtuelle développé par l'association Espaces Nouveaux et l'Ircam. Le *Spatialisateur* reçoit les signaux sonores des sources instrumentales ou de synthèse et leur ajoute, en temps réel, des effets de spatialisation. Il permet de reproduire et de contrôler la localisation des sources sonores en trois dimensions et de reconstituer un effet de salle artificiel (réverbération).

Après une brève présentation des principes généraux relatifs à la synthèse de scènes sonores, la méthode de création de ces scènes à l'aide du *Spatialisateur* est présentée. Le module du traitement de signal de ce dernier est détaillé et les principes de la synthèse binaurale et de la diffusion transaurale utilisés dans ces expériences sont décrits. Enfin, l'interface de contrôle direct des attributs perceptifs offerts par cet outil est présenté. Les atouts du *Spatialisateur* par rapport à la présente application sont discutés, ils permettent de mettre en évidence l'intérêt de son utilisation.

### 3.2.1 La synthèse des scènes sonores

On entend par scène sonore complexe, une scène composée de plusieurs sources sonores situées en différents emplacements d'un même espace, existant ou virtuel. La création des scènes sonores implique des informations concernant les aspects temporels, fréquentiels et spatiaux des signaux à restituer ou à synthétiser.

Toutes les informations temporelles et fréquentielles nécessaires sont inscrites sur la réponse impulsionnelle de la salle, c'est à dire sur la réponse de la salle à une impulsion Dirac<sup>1</sup>. La méthode de prise de son (monophonique, stéréophonique avec un couple de microphones rapprochés, par tête artificielle, etc.) détermine la finesse des informations spatiales obtenues.

Diverses méthodes existent pour obtenir la réponse impulsionnelle (monophonique ou stéréophonique) d'une salle :

- La méthode la plus simple – et à condition évidemment que la salle en question existe – est de mesurer la réponse impulsionnelle sur place. Alternativement, la réponse impulsionnelle peut être mesurée dans des maquettes réduites construites à ce propos. Cette dernière méthode présente des limitations, notamment par rapport à la simulation des matériaux de la salle et des caractéristiques de directivité des transducteurs ; par ailleurs, la taille de ces derniers limite les mesures en fréquence vers l'aigu.
- Une méthode plus complète consiste à utiliser la modélisation numérique, c'est à dire à employer des algorithmes d'acoustique prévisionnelle qui reposent sur des propriétés de propagation des salles. Une fois les paramètres relatifs à la salle et aux transducteurs fournis (paramètres architecturaux, position et directivité des transducteurs), la réponse impulsionnelle peut être calculée au prix d'une simplification de phénomènes physiques comme la diffraction. La modélisation numérique permet d'obtenir directement les paramètres acoustiques propres à la salle, tels que le temps de réverbération, la répartition fréquentielle de l'énergie, etc.

---

<sup>1</sup> Pour des emplacements et des directivités donnés de la source et du récepteur.

La création des scènes sonores s'effectue alors en adoptant l'une des deux approches suivantes:

- soit par la convolution de la réponse impulsionnelle (mesurée ou calculée) avec le signal sonore monophonique et de préférence anéchoïque. Les scènes sonores ainsi créées peuvent être très fidèles à la réalité, mais dans les limites imposées par la directivité des transducteurs de mesure (directivités du haut-parleur et du microphone de mesures, différentes des directivités des sources réelles de musique et du récepteur humain) et aussi le nombre de sources que l'on peut recréer.
- soit par un logiciel de traitement de signal qui réalise la synthèse d'un effet de salle générique. Les paramètres de traitement de signal (paramètres TNS) nécessaires au logiciel sont déduits par l'exploitation de la réponse impulsionnelle. En effet, il s'agit plus de la *simulation d'une qualité sonore* caractérisée par une réponse impulsionnelle, et moins de la *reproduction d'une scène sonore* réelle.

Clarifions ici ces deux derniers termes employés :

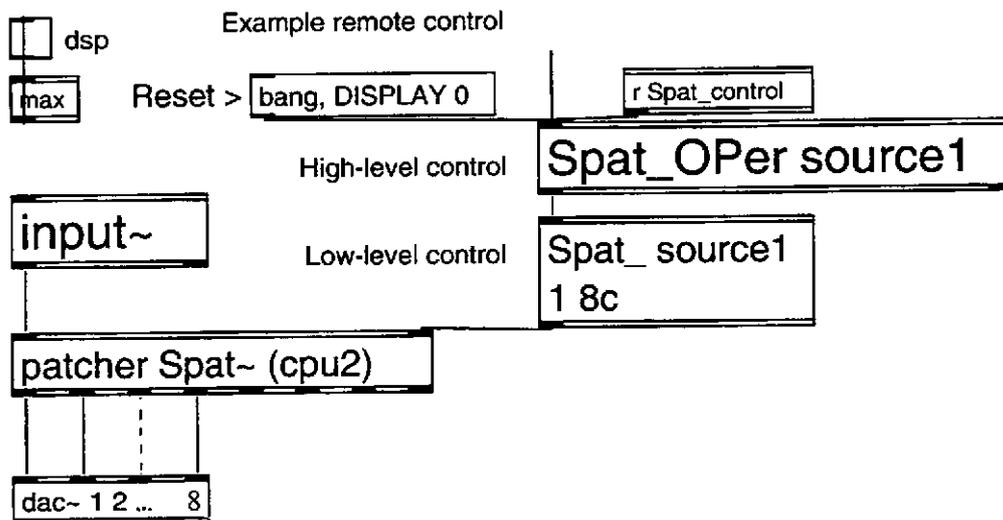
Par *reproduction* d'une scène sonore, on parle de la procédure passant par la mesure de la réponse binaurale (à l'aide d'une tête artificielle) – par souci de souplesse, plutôt que d'enregistrer directement les sources musicales dans la salle, on ne mesure le plus souvent que la réponse impulsionnelle – vient ensuite la convolution en laboratoire avec les signaux anéchoïques. Il s'agit donc formellement de *reproduction*, la première approche mentionnée ci dessus.

Par *simulation* d'une qualité sonore, on entend le passage par une procédure d'analyse et de synthèse de la qualité acoustique : l'analyse de la réponse impulsionnelle permet de déduire les paramètres (objectifs ou perceptifs) que l'on s'intéresse à simuler ; suit la synthèse d'un effet de salle contrôlé par ces paramètres. C'est donc uniquement la similitude de ces derniers qui est assurée et non pas la similitude d'autres caractéristiques acoustiques de la salle. C'est la méthode de création de champs sonores utilisée dans ce travail; elle est détaillée dans la suite.

### 3.2.2 Le module de traitement de signal du Spatialisateur

Le *Spatialisateur* est un logiciel de création de scènes sonores développé dans l'environnement de traitement de signal FTS/ Max. Le type de programmation Max est uniquement graphique. Il est fondé sur la notion de « patch » qui est un ensemble de modules symbolisant les opérations élémentaires effectuées sur des signaux ou des messages de contrôle. On voit sur la **figure 3.1** la structure du *Spatialisateur* à une seule source. Il est composé d'un module de traitement du signal (la partie « signal ») et d'un ensemble d'interfaces permettant de fournir les informations nécessaires pour gérer les opérations de la partie signal (la partie « contrôle ») [Jot et Warusfel, 1997].

## 1 source -&gt; 2 or 4 speakers



**Figure 3.1 :** Structure du *Spatialisate*ur à une source. La partie signal est représentée par le module générant les signaux d'entrée « input~ », le module de traitement du signal « Spat~ » et le convertisseur N/A « dac~ 1 2...8 ». La partie contrôle est représentée par deux types de « patch » : le « Spat\_ » qui est le contrôle bas-niveau et le « Spat\_OPer » qui est le contrôle haut-niveau.

La partie signal est composée du patcher « input~ », du « patcher Spat~ » et du « dac~ 1 2...8 » :

« input~ » est le module générant les signaux d'entrée (fichiers sons sur disque dur, convertisseur A/N ou des signaux test générés par le *Spatialisate*ur en temps réel),

« Spat~ » est le module de traitement du signal. Il fournit les paramètres TNS et simule l'effet de salle,

« dac~ 1 2...8 » est un convertisseur N/A, il peut recevoir et fournir de 1 à 8 canaux. Dans le cas de notre étude, le mode de reproduction utilisé (binaural ou transaural) nécessitait uniquement deux sorties.

La partie « contrôle » est composée du patch « Spat\_ » et du « Spat\_OPer » :

le « Spat\_ » est le contrôle bas-niveau ; il concerne la description simplifiée de la distribution de l'énergie (4 sections temporelles, 3 bandes fréquentielles),

le « Spat\_OPer » est le contrôle haut-niveau. Les paramètres perceptifs sont traduits sous forme de distribution simplifiée de l'énergie et pilotent ainsi directement le module de traitement de signal « Spat~ ».

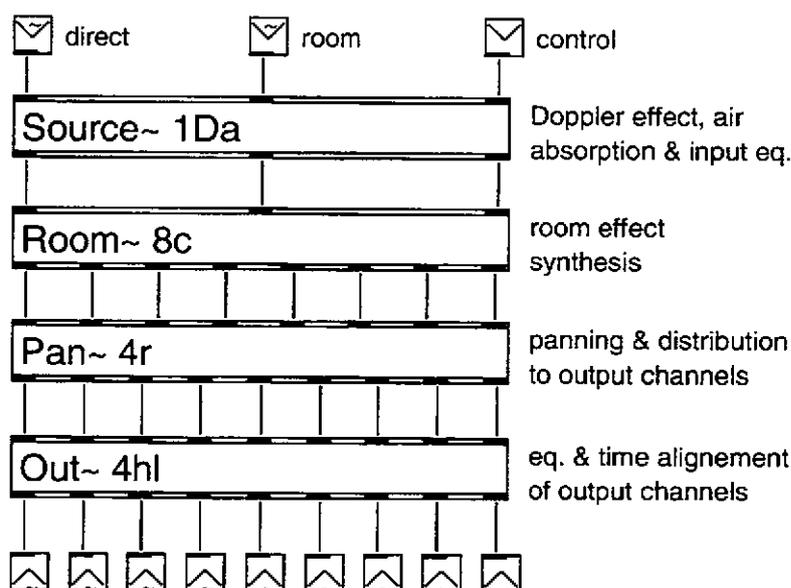
Le module de traitement du signal « Spat~ » est composé de quatre parties : *Source*, *Room*, *Pan* et *Out* (cf. **figure 3.2**). Chacune de ces parties peut être configurée séparément en appelant des versions prédéfinies, disponibles dans la bibliothèque *Spat*, une bibliothèque qui contient des modules de traitement de signal et d'interfaces de contrôle pour un traitement spatial temps-réel des sons. Les modules élémentaires sont des réverbérateurs artificiels, des correcteurs spectraux, des 'panpots' (panoramic potentiometers), etc.

### 3.2.2.1 Synthèse de la distribution temps / fréquence de l'effet de salle

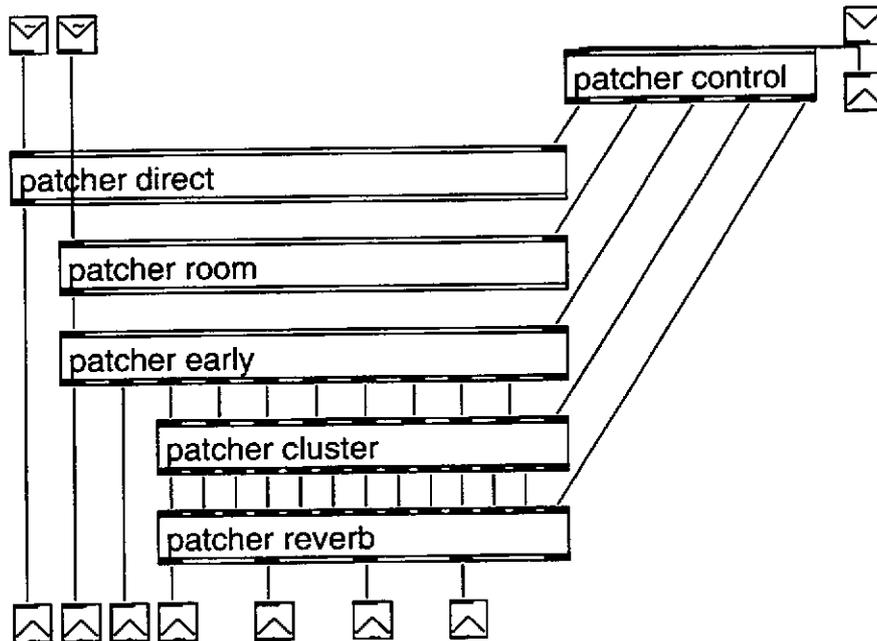
La partie *Source* contient des filtres passe-bas et des filtres de retards pour reproduire les effets de l'absorption de l'air et les effets Doppler, ainsi que des égaliseurs spectraux pour permettre des corrections supplémentaires concernant la nature du signal d'entrée qui est supposé être anéchoïque.

La partie *Room* réalise la distribution temporelle et fréquentielle de l'effet de salle en se basant sur une description simplifiée de la réponse impulsionnelle en quatre plages temporelles (*direct*, *early*, *cluster* et *reverb* correspondent respectivement aux intervalles temporels 0-20, 20-40, 40-100 et >100 ms).

*Room* est un réverbérateur artificiel qui reçoit deux signaux en entrée : un signal révélateur du spectre de la source dans une direction de référence de la source, par exemple la direction du pavillon d'un haut-parleur ou d'un instrument à vent, et un deuxième signal révélateur du spectre de puissance de l'instrument (moyenne spatiale de l'émission de l'instrument), destiné à alimenter l'effet de salle. *Room* est composé de modules élémentaires de réverbération qui permettent la construction d'une variété d'architectures de traitement afin de répondre à des contraintes en temps de calcul et à des besoins en densité spectrale ou temporelle de la réverbération synthétique. Une telle architecture de traitement est présentée à la **figure 3.3** : ses modules simulent les différentes parties de la réponse impulsionnelle, représentées par des 'patchers' correspondants : *direct*, *room*, *early*, *cluster* et *reverb*.



**Figure 3.2** : Le module de traitement de signal (*Spat*) est composé de quatre parties : *Source*, *Room*, *Pan* et *Out*.

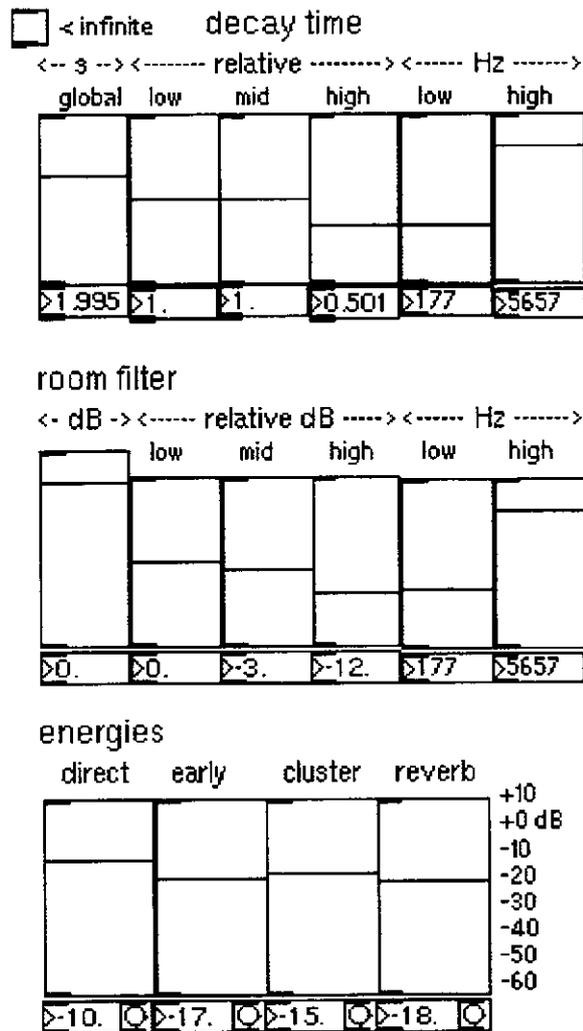


**Figure 3.3 :** Structure interne de *Room* qui réalise la distribution temporelle et fréquentielle de l'effet de salle.

*Direct* et *room* effectuent des filtrages pour prendre en compte la directivité de la source sonore, *direct* fournit le son direct et *room* alimente *early*. *Early* fournit les premières réflexions (gauche et droite). Son signal est réparti dans 8 lignes à retard qui alimentent le *cluster*. Enfin, le *cluster* alimente le module de *réverb* qui réalise la partie exponentiellement décroissante de l'effet de salle. *Room* fournit ainsi 7 sorties : une pour le son direct, deux pour les premières réflexions et quatre pour l'effet de salle tardif (*cluster* et *reverb*).

Le *Spatialisateur* offre une interface de contrôle 'bas-niveau' basée sur la description simplifiée de la réponse impulsionnelle ; elle est présentée à la **figure 3.4**. La première partie permet de gérer le temps de réverbération global et son profil fréquentiel. La deuxième partie effectue un contrôle fréquentiel en trois bandes pour chaque section de l'effet de salle (sur la figure 3.3, on ne voit la partie de contrôle fréquentiel que pour une seule section). Les fréquences de transition pour ces deux parties sont contrôlables. Enfin, la troisième partie de l'interface contrôle l'énergie de chacune des 4 sections de la réponse impulsionnelle (*direct*, *early*, *cluster* et *reverb*).

source1



**Figure 3.4 :** Structure de la partie contrôle bas-niveau du *Spatialisateur*. On observe de haut en bas l'interface de contrôle du temps de la réverbération, l'interface de contrôle fréquentiel de l'effet de salle (pour une seule section de la réponse impulsionnelle) et l'interface de contrôle de l'énergie de chacune des 4 sections de la réponse impulsionnelle (pour les deux premières interfaces, les valeurs de temps de réverbération et de l'énergie dans les trois bandes fréquentielles sont relatives à des valeurs globales).

### 3.2.2.2 Synthèse de la distribution spatiale

La quatrième partie du module de traitement de signal (*Spat*) est optionnelle.

La dimension *spatiale* des scènes sonores est un facteur très important de la synthèse sonore, il s'agit en effet d'un des facteurs clé pour améliorer l'*aspect naturel*, la *lisibilité* d'une scène sonore, ou encore la *télé-présence* dans les applications de télécommunications (téléconférences) dans une scène

virtuelle [Jot, 1996]. Le *Spatialisateur* donne la possibilité de reproduire et contrôler la localisation des sources sonores en trois dimensions et de reconstituer un effet de salle artificiel (réverbération).

Deux catégories d'indices sont nécessaires pour la reproduction spatiale des scènes sonores : les indices qui concernent la localisation des sources sonores et ceux qui concernent l'effet de salle. Ils sont reconstruits artificiellement à l'aide d'un système de traitement de signal, le module *Pan* du *Spatialisateur*.

Les 7 signaux de la sortie de *Room* sont : un pour le son direct, deux pour les premières réflexions, quatre pour le *cluster* et la réverbération tardive. La synthèse des indices spatiaux pour ces différents signaux ne requiert pas le même type de traitement ni le même niveau d'exigence en terme de précision de localisation. Ils sont alors traités par le module *Pan* comme suit :

- Une localisation précise est nécessaire pour le signal correspondant au son direct.
- Les premières réflexions sont réparties en deux groupes, localisés de part et d'autre du son direct (typiquement à  $\pm 30^\circ$ ) et le 'suivent' lorsque la source se déplace. Ce choix est imposé par l'architecture du *Spatialisateur*. Il est justifié par les études perceptives qui montrent l'importance particulière des premières réflexions parvenant suivant l'angle solide entourant la source, contrairement aux réflexions arrières qui auront tendance à s'intégrer à l'effet de salle [Lavandier, 1989]. Ce choix répond également au souci de contrôler de manière convaincante l'enveloppement sonore.
- La seule contrainte pour les quatre signaux créant l'effet de salle tardif (*cluster* et *reverb*) est qu'ils doivent être perçus de manière diffuse, non-localisable, ce qui veut dire que la corrélation spatiale des signaux doit être près de zéro. Pour certains modes de reproduction (binaural, transaural), il convient de corrélérer ces signaux en basses fréquences afin de restituer un IACC (cf. § 2.2.4.3) similaire à celui observable entre les deux oreilles d'une tête placée en champ diffus.

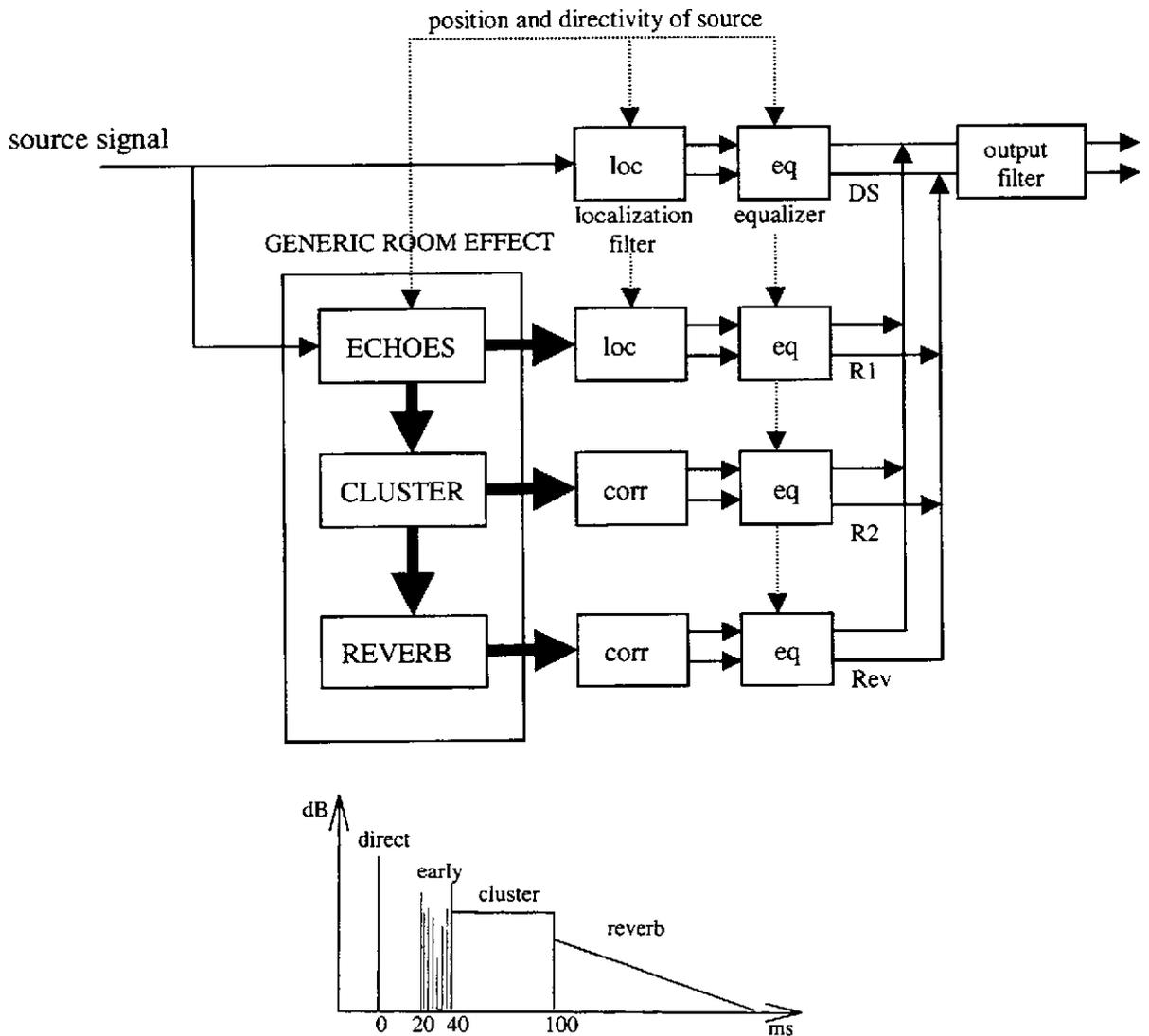
Le module *Pan* fournit autant de canaux de sortie que nécessaire pour la technique d'encodage spatial. Il s'agit ainsi d'une matrice  $7 \times P$  (où  $P$  est le nombre de sorties du *Pan*) qui permet de convertir le format «3/4» issu des 7 sorties de *Room* (1 pour le son direct, 2 pour les premières réflexions et 4 pour la réverbération tardive), au format d'encodage directionnel spécifié par l'utilisateur. Ces différents formats sont :

- la technique de *pondération d'intensité* entre paires de haut-parleurs voisins (« discrete surround ») utilisée pour synthétiser un effet de localisation dans le plan horizontal et destinée à la reproduction sur un dispositif composé de  $N$  haut-parleurs.
- la technique *Ambisonics* qui consiste à simuler électroniquement l'effet d'un dispositif de prise de son composé d'un système de microphones encodant le champ acoustique sur la base des 4 premières harmoniques sphériques, équivalentes respectivement aux sensibilités microphoniques, omnidirectionnelle pour la première et bidirectionnelles dans les directions  $x$ ,  $y$  et  $z$  pour les trois suivantes (prise de son « soundfield »).
- la technique de *stéréophonie* conventionnelle.
- la technique *binaurale* d'encodage spatial du signal en 3-D sur casque d'écoute, qui peut ensuite être décodé pour la reproduction *transaurale* sur haut-parleurs. Il s'agit de la technique utilisée ici, elle est détaillée dans une section spécifique (cf. 3.1.3).

Il s'agit de la partie *Out* qui est en fait utilisée pour décoder et adapter la sortie du module *Pan* à la géométrie d'implantation et à la réponse acoustique du système de haut-parleurs. Elle est chargée d'effectuer l'égalisation spectrale et temporelle (sur chaque canal de sortie), le décodage d'un signal

B-format (technique *Ambisonics*) pour reproduction sur haut-parleurs ou casque audio, ou la conversion du format binaural en un format adapté pour la diffusion transaurale.

La **figure 3.5** montre la structure d'un algorithme DSP (parties *Room, Pan* et *Out* du Spatialisateur) qui permet la reproduction d'une scène sonore sur deux canaux (binaurale, transaurale ou stéréophonique). La distribution temporelle simplifiée de l'énergie adoptée par le Spatialisateur est aussi présentée. On observe ainsi que les directions d'arrivées du son direct (*source signal*) et des premières réflexions (*Echoes*) sont contrôlées à l'aide des filtres de localisation (*loc*). Supposant des descriptions statistiques des distributions temporelles et spatiales des réflexions tardives et de la réverbération, les signaux au niveau de *Cluster* et de *Reverb* sont complètement décorrélés. Leurs aspects directifs peuvent ensuite être reproduits par le contrôle du coefficient de «cross-corrélation» (*IACC*) ; il s'agit d'une opération de re-corrélation importante, surtout pour le rendu des basses fréquences. Après une étape de correction (*eq*), les signaux sont utilisables pour une diffusion stéréophonique conventionnelle ou pour un passage par des filtres de sortie (*output filter*) contrôlables par l'ordinateur et permettant de permuter entre deux modes de diffusion : binaurale ou transaurale.



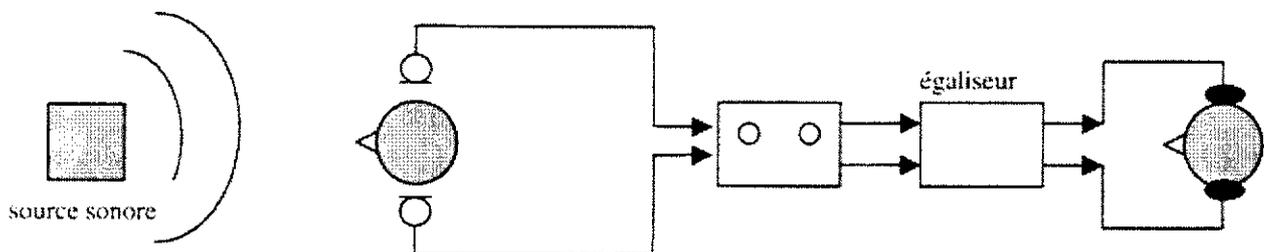
**Figure 3.5 :** Structure d'un algorithme DSP pour une reproduction sur deux canaux (reproduction binaurale, transaurale ou stéréophonique) et distribution temporelle simplifiée de l'énergie dans une salle adoptée par le Spatialisateur [Jot et al., 1993].

### 3.2.3 Synthèse binaurale et reproduction transaurale

Nous avons vu que le *Spatialisateur* offre la possibilité d'utiliser différentes méthodes d'encodage directionnel et différents dispositifs de reproduction. Un cas particulier de *Pan* est l'encodage directionnel en format binaural. La synthèse binaurale et la restitution transaurale sur haut-parleurs, sont des techniques utilisées pour la reproduction des scènes sonores dans le cadre de ce travail ; elles sont présentées ci-après.

#### 3.2.3.1 La synthèse binaurale

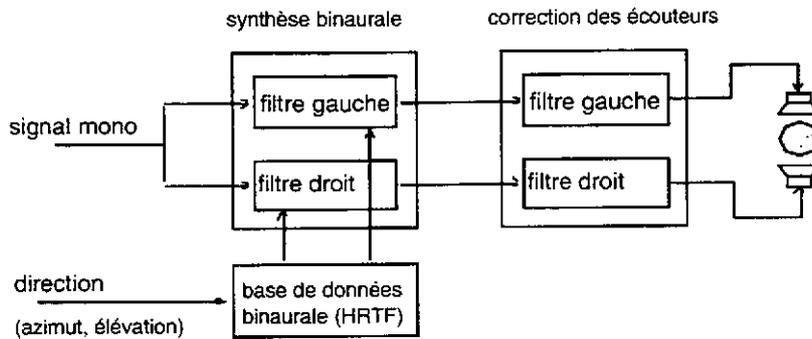
Le principe de l'encodage directionnel en format binaural est présenté sur la **figure 3.6**. Il s'agit de mesurer le signal de pression acoustique au niveau des tympans des oreilles d'un auditeur présent dans une scène sonore, et de la reproduire ensuite sur casque d'écoute, en passant par une étape d'égalisation afin de compenser la sensibilité du casque.



**Figure 3.6 :** L'encodage directionnel en format binaural. Les signaux sont enregistrés au niveau des oreilles des auditeurs et, après une étape d'égalisation, sont diffusés sur casque d'écoute.

De manière naturelle, les indices acoustiques responsables de notre perception directionnelle sont ainsi codés dans la prise de son et conduisent à une restitution très convaincante. En revanche, il existe des limitations : pour une restitution possédant le moins possible d'ambiguïté de localisation, il faudrait que les enregistrements soient réalisés au niveau des oreilles de la même personne qui va ensuite écouter sur casque. De plus, les mesures ne sont valables que pour une seule position de la tête de l'auditeur et pour un environnement sonore donné (l'environnement sonore lors de la séance d'enregistrement).

Ce même principe d'encodage directionnel se trouve à l'origine de la synthèse binaurale. Il s'agit d'une technique qui permet, à partir d'un enregistrement monophonique, de reproduire une image sonore tridimensionnelle par l'intermédiaire d'un couple de filtres linéaires (HRTF). Ces filtres encodent les indices spatiaux pour chacune de nos oreilles et permettent de reconstruire ainsi les signaux sonores qu'elles auraient reçus sous des conditions d'écoute naturelle, c'est à dire en présence d'une source sonore réelle (**figure 3.7**).



**Figure 3.7 :** Principe de la synthèse binaurale permettant de simuler des conditions d'écoute en champ libre sur casque d'écoute.

Une HRTF (Head-Related Transfer Function) est une fonction de transfert caractérisant la transformation subie par le champ acoustique en présence de l'auditeur, transformation produite par la diffraction du son sur le pavillon de l'oreille, sur la tête et sur le torse de ce dernier. Plusieurs couples de filtres numériques – chacun correspondant à une direction particulière de la source sonore virtuelle (azimut et site) – sont disponibles dans des bases de données HRTF (HRTF database). En fonction du signal utilisé comme référence, il existe différentes définitions :  $HRTF_{champ\ libre}$ ,  $HRTF_{frontale}$ ,  $HRTF_{champ\ diffus}$ , etc. ; elles sont définies ci-dessous pour un angle de source  $\theta$  [Larcher et al., 1998] :

$$HRTF_{\theta, champ\ libre} = \frac{\text{HRTF directionnelle (source à } \theta \text{ degrés)}}{\text{Pression acoustique à l'emplacement du milieu de la tête en l'absence de l'auditeur (micro omnidirectionnel)}}$$

$$HRTF_{\theta, frontale} = \frac{\text{HRTF directionnelle (source à } \theta \text{ degrés)}}{\text{HRTF directionnelle avec la source à } 0 \text{ degré}}$$

$$HRTF_{\theta, champ\ diffus} = \frac{\text{HRTF directionnelle (source à } \theta \text{ degrés)}}{\text{moyenne des HRTF directionnelles ou HRTF mesurée en champ diffus}}$$

Dans notre cas, c'est la  $HRTF_{champ\ diffus}$  qui a été utilisée. Elle a l'avantage de limiter l'instabilité du rapport (et par conséquent celle du filtre) puisque la fonction du dénominateur (la moyenne des HRTF directionnelles ou l'HRTF mesurée en champ diffus) présente moins de zéros que par exemple, une HRTF directionnelle. De plus, il a été constaté que la normalisation champs diffus limite les différences interindividuelles entre les HRTF.

Les bases de données HRTF sont obtenues à l'aide d'une tête artificielle immobile, positionnée dans un espace dans lequel est placé un haut-parleur successivement en plusieurs directions sur une sphère autour de la tête (la tête artificielle peut être remplacée par un sujet humain, de petits microphones étant insérés dans les canaux auditifs). Une mesure est faite au niveau de chaque oreille de la tête artificielle et ce, pour chacun des emplacements du haut-parleur. La synthèse binaurale peut alors être réalisée par la convolution du signal d'entrée monophonique avec le couple de HRTF correspondant à la direction de provenance souhaitée. Une étape de correction est aussi nécessaire pour compenser la réponse de la chaîne de restitution sonore (réponse du casque audio).

Les HRTF encodent les effets de diffraction au niveau de la tête de l'auditeur et dépendent donc de sa morphologie particulière. Ainsi, afin d'assurer une reproduction parfaite il faudrait en principe réaliser des mesures pour chaque auditeur, et enregistrer les HRTF qui lui sont propres. La variabilité interindividuelle des HRTF limite la qualité de reproduction, il en résulte typiquement une certaine difficulté à localiser le son dans l'hémisphère frontale, celui-ci est souvent perçu au dessus de la tête ou derrière, très près, voire même dans la tête (sensation intracrânienne).

### 3.2.3.2 La reproduction transaurale

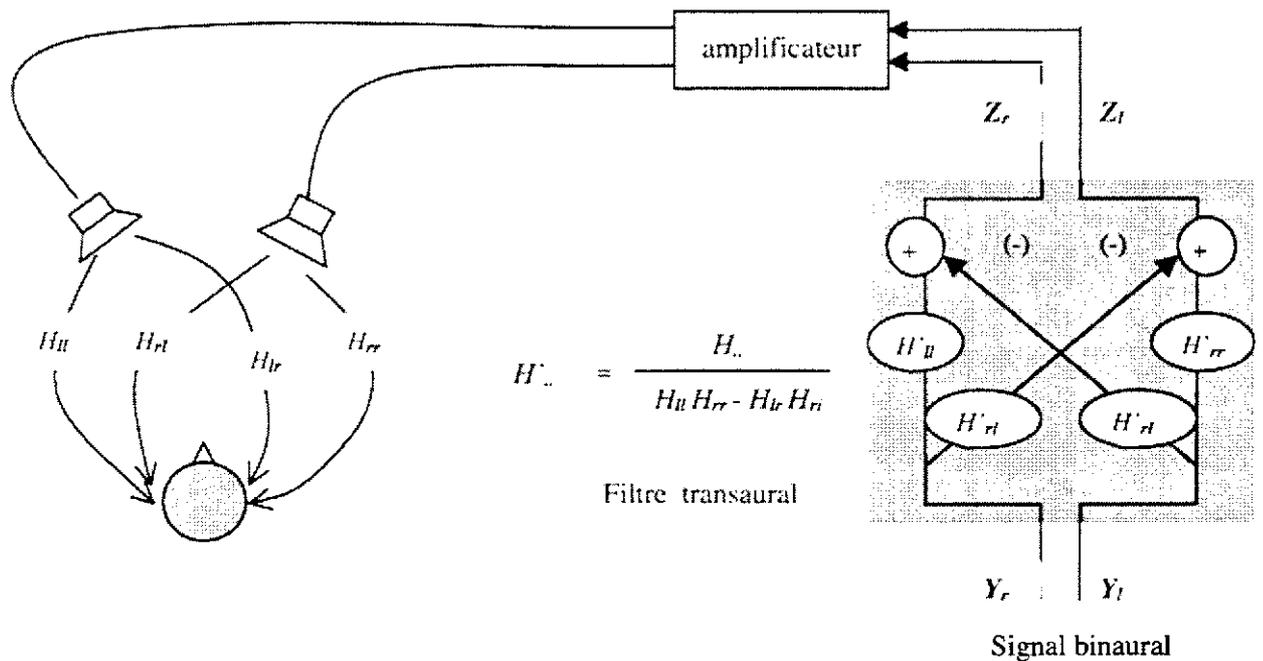
La reproduction transaurale consiste à reproduire les signaux binauraux au niveau des oreilles des auditeurs à l'aide d'un couple de haut-parleurs. Le format d'encodage directionnel reste donc binaural, c'est la technique et le dispositif de restitution qui diffèrent.

Le principe de la technique transaurale (cross-talk cancellation) est présenté sur la **figure 3.8**. Elle permet d'annuler les ondes acoustiques parvenant de chaque haut-parleur vers l'oreille opposée (signaux  $H_{lr}$  et  $H_{rl}$ ). Ainsi, si  $Y_l$  et  $Y_r$  représentent le signal binaural que l'on veut reconstruire, la propagation des signaux issus des haut-parleurs  $Z_l$  et  $Z_r$  est caractérisée par une matrice à quatre fonctions de transfert [Jot et al., 1995] :

$$\begin{bmatrix} Y_l \\ Y_r \end{bmatrix} = \begin{bmatrix} H_{ll} & H_{rl} \\ H_{lr} & H_{rr} \end{bmatrix} \begin{bmatrix} Z_l \\ Z_r \end{bmatrix}$$

Le signal binaural peut ainsi être décodé par une matrice inverse de fonctions de transfert (filtre transaural) afin d'obtenir les signaux des haut-parleurs :

$$\begin{bmatrix} Z_l \\ Z_r \end{bmatrix} = \frac{\begin{bmatrix} H_{rr} & -H_{rl} \\ -H_{lr} & H_{ll} \end{bmatrix}}{(H_{ll} \cdot H_{rr}) - (H_{lr} \cdot H_{rl})} \begin{bmatrix} Y_l \\ Y_r \end{bmatrix}$$



**Figure 3.8 :** Diffusion transaurale sur haut-parleurs. Les quatre fonctions de transfert acoustique et le filtre transaural (cross-talk canceller) servent à décoder le signal binaural (le système doit aussi comporter une étape de compensation des réponses des haut-parleurs).

En théorie, la technique transaurale ne nécessite ni une disposition symétrique des haut-parleurs, ni un environnement de reproduction anéchoïque. La seule contrainte est que les quatre fonctions de transfert soient mesurées dans des conditions *identiques* à celles de la reproduction, ce qui impose notamment une contrainte sur la position de la tête par rapport aux haut-parleurs. De plus, pour pouvoir reproduire l'écoute transaurale dans des lieux différents du lieu des mesures, mais aussi pour simplifier l'implémentation du filtre, il est souhaitable d'opter pour une disposition symétrique des haut-parleurs – où la tête de l'auditeur et les deux haut-parleurs se trouvent aux sommets d'un triangle équilatéral – et de réaliser la diffusion dans des conditions anéchoïques (ou du moins d'éviter les réflexions précoces fortes pouvant détruire l'effet du filtrage inverse).

Si les contraintes délicates vis-à-vis des emplacements des haut-parleurs et de la tête des auditeurs sont respectées, la technique transaurale permet une restitution sonore tridimensionnelle très convaincante – cependant un peu moins robuste pour la partie postérieure – tout en préservant une distinction nette entre le champ direct et le champ diffus (respectivement le son localisé et la réverbération tardive).

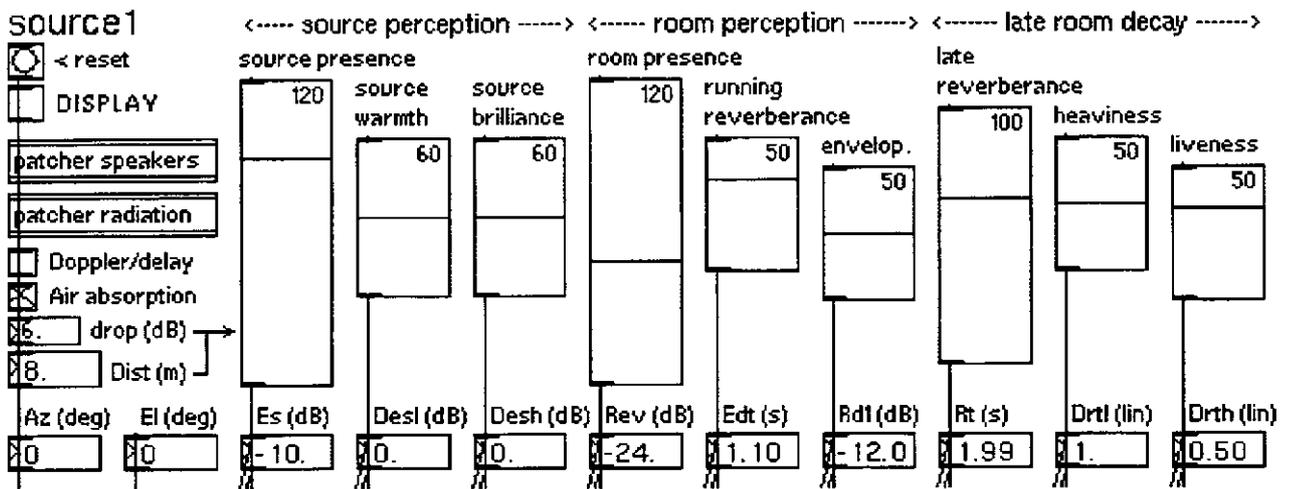
En pratique, la technique transaurale permet de réduire les inversions d'écoute avant / arrière ainsi que la localisation du son *dans* la tête (problèmes courants de la diffusion sur casque d'écoute, surtout lorsque l'on ne dispose pas des HRTF individuelles, ni de suiveur de position) [Blauert, 1996], [Begault, 1999]. Par rapport à la précision de localisation, une certaine ambiguïté existe sous forme de biais frontal, mais elle est moins gênante pour le réalisme de la scène virtuelle reconstituée que le biais arrière observé avec le casque. De plus, et du fait que cette technique ne nécessite pas de port de casque, elle offre une amélioration au niveau du « réalisme », ou de la crédibilité de la situation en

général, aspect important pour de nombreuses applications de simulation acoustique, sensibles à ce type de qualité.

### 3.2.4 L'interface de contrôle 'perceptif' du *Spatialisateur*

Au chapitre précédent, nous avons vu que des études psychoacoustiques ont permis de définir un vocabulaire perceptif pertinent pour décrire la qualité acoustique d'une salle. Ainsi, un nombre limité de facteurs perceptifs indépendants entre eux est suffisant pour décrire les impressions auditives liées à la perception d'une scène sonore. C'est sur ce principe que le *Spatialisateur* a été basé pour fournir une interface perceptuelle *haut-niveau* ; c'est-à-dire une interface perceptuelle donnant à tout utilisateur la possibilité d'intervenir directement sur les facteurs perceptifs, à l'aide de paramètres qui ont *un sens* pour lui, sans qu'il soit obligé de passer par la description physique ou architecturale du champ sonore, description a priori étrangère à son langage.

Pour l'implantation de l'interface haut-niveau, des résultats issus des recherches précédentes sur la caractérisation perceptuelle de la qualité acoustique des salles ont été incorporés dans le *Spatialisateur* [Jullien et al., 1992], [Jullien, 1995]. Ces recherches ont permis de définir un vocabulaire perceptif, descriptif de cette qualité, de révéler une indépendance de ces facteurs et de mettre en évidence des liens entre des attributs perceptifs et des grandeurs physiques (mesurables sur la réponse impulsionnelle de la salle). Neuf de ces facteurs figurent sur l'interface haut-niveau du *Spatialisateur*, qui donne ainsi la possibilité de les contrôler de manière indépendante à l'aide d'une interface graphique (curseurs) ou numérique (figure 3.9).



**Figure 3.9** : L'interface haut-niveau du *Spatialisateur* (higher-level user interface). Elle permet de gérer les attributs perceptifs et de reproduire ainsi l'effet de salle désiré. Elle permet aussi de définir des paramètres de configuration comme l'absorption de l'air en fonction de la distance, la position de la source (en azimut et en site).

L'interface haut-niveau contrôle simultanément les différents 'sous-modules' (sub-modules) d'un processeur *Spat* (module du traitement de signal) et permet de spécifier un effet de salle désiré à l'aide d'un ensemble de paramètres dont la définition ne dépend pas du format ou du dispositif de reproduction. Parmi ces paramètres, on retrouve l'azimut, le site et la distance de la source par rapport à l'auditeur, ainsi que les attributs perceptifs mentionnés auparavant.

Les curseurs de l'interface graphique de la **figure 3.9** varient selon une échelle basée sur la sensibilité moyenne des auditeurs par rapport au facteur qu'ils contrôlent. Chaque facteur perceptif est lié à un indice acoustique mesurable qui caractérise la transformation du signal sonore lors de sa propagation dans la salle. Certains de ces indices correspondent à des indices établis du domaine de l'acoustique des salles (comme *envelopment* ou *early decay time*), mais en fait, le choix des indices a été réalisé dans un but différent : celui de fournir une caractérisation exhaustive de la qualité acoustique d'une salle à l'aide d'un nombre minimum d'indices *indépendants* [Lavandier, 1989], [Jullien et al., 1992], [Jullien, 1995].

Les attributs perceptifs de l'interface haut-niveau du *Spatialisateur* appartiennent à trois groupes : un groupe qui concerne la perception de la source, un groupe qui concerne l'interaction source / salle et un groupe qui concerne la perception de la salle [Jot, 1997].

Les attributs perceptifs de chacun des groupes sont présentés ci-après :

i) Perception de la source (*source perception*)

- Présence de la source (*source presence*) : elle décrit la sensation auditive de la distance de la source<sup>1</sup>.
- Chaleur (*warmth*) et brillance (*brilliance*) : ces attributs décrivent la couleur tonale associée à la source respectivement dans les basses et dans les hautes fréquences.

ii) Interaction source / salle (*source / room interaction*)

- Présence de la salle (*room presence*) : force sonore de l'effet de salle.
- Réverbérance courante (*running reverberance*) : par opposition à la réverbérance tardive qui se réfère à la sensation de 'traînage' perçue lors des transitoires à la fin des phrases musicales, la réverbérance courante traduit la persistance de cette sensation pendant la phrase musicale ou le message de parole.
- Enveloppement (*envelopment*) : sensation de l'étendue de la source.

iii) Perception de la salle (*room perception*)

- Réverbérance tardive (*late reverberance*) : sensation de 'traînage' des transitoires à la fin des phrases musicales.
- Lourdeur (*heaviness*) et vivacité (*liveness*) : variation de la réverbérance respectivement aux basses et hautes fréquences.

La présentation détaillée des paramètres de contrôle et de la syntaxe haut-niveau est donnée dans le tableau 3.1. La définition objective des descripteurs perceptifs y est également présentée. Enfin, le modèle simplifié de la distribution temporelle de l'énergie dans une salle adoptée par le *Spatialisateur* permet de préciser les notations utilisées (figure 3.10).

---

<sup>1</sup> Le terme de « présence de la source » a été préféré à celui de « distance de la source », pour éviter une trop grande connotation à la géométrie (la tentation étant de définir cette distance en mètres par exemple). Or, les propriétés des salles peuvent, pour une même distance du spectateur par rapport à l'orchestre ou au soliste, procurer des sensations de « présence » de cette source, très différentes et parfois en « inadéquation » avec la distance géométrique. Il s'agit donc bien de propriétés acoustiques et non pas architecturales.

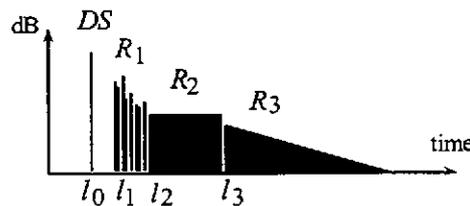
## Descripteurs de localisation

Nom	Unité	Différence juste perceptible	Indices objectifs
Azimut	°	3° à 10°	dépend du format d'enregistrement
Élévation (site)	°	10° à 30°	dépend du format d'enregistrement

## Descripteurs de l'acoustique des salles

Nom	Indice objectif	Unité	Différence juste perceptible	Indices objectifs
Présence de la source	Es	dB	1dB	= DS + R1
Chaleur	Desl	dB	1dB	= ( DS <sub>BF</sub> + R1 <sub>BF</sub> ) / Es
Brillance	Desh	dB	1dB	= ( DS <sub>HF</sub> + R1 <sub>HF</sub> ) / Es
Présence de la salle	Rev	dB	1dB	= C*R3 + [ (1-C)*R3 ] <sup>2</sup> / (8*Es) si (1-C)*R3/Es = R3 - 2*Es sinon, où C = 10 <sup>-1.2/Rt</sup>
Réverbérance courrante	Edt	sec	10%	= 0.4 + Rt * (1 - 0.667*log10[1 + (Es+R2)/R3]) si Es+R2 < 30.622 * R3 = 0.6 / log10[1 + (Es+R2)/R3 ] sinon
Enveloppement	Rd1	dB	1.4 dB	= ( 0.3*R1 + 0.05*R2 ) / Es
Réverbérance tardive	Rt	sec	10%	(temps de réverbération TR <sub>60dB</sub> )
Lourdeur	Drtl	lin	20%	= Rt <sub>BF</sub> / Rt
Vivacité	Drth	lin	10%	= Rt <sub>HF</sub> / Rt

**Tableau 3.1** : Présentation détaillée des paramètres de contrôle et de la syntaxe haut-niveau du *Spatialisateur* (paramètres de localisation et paramètres liés à l'acoustique de la salle). L'unité, la différence juste perceptible et la définition objective de chaque attribut y sont présentées.



**Figure 3.10** : Le modèle simplifié de la distribution temporelle de l'énergie adoptée par le *Spatialisateur*. Elle est divisée en quatre sections DS, R1, R2 et R3 définies par les limites temporelles  $l_1$ ,  $l_2$ ,  $l_3$  qui correspondent respectivement à 20, 40 et 100 ms par rapport à l'arrivée du son direct en  $l_0$ .

### 3.2.5 Le choix de la méthode de synthèse sonore (Spatialisateur)

La création des scènes sonores à l'aide du *Spatialisateur* ne représente pas la méthode la plus généralement répandue.

Une autre méthode – que l'on qualifierait de traditionnelle – consisterait à effectuer la convolution du signal d'entrée désiré avec des réponses impulsionnelles mesurées en différents emplacements de la salle dont on souhaite simuler la qualité acoustique. Par exemple, pour simuler des scènes sonores qui donnent des impressions de distance auditive croissante, des réponses impulsionnelles en des emplacements de distance croissante devraient être mesurées et convoluées ensuite avec le signal. Cette méthode aurait l'avantage de garantir a priori le réalisme des scènes sonores présentées aux sujets.

Or, le but principal de l'étude n'est pas de reproduire exactement la qualité acoustique des théâtres simulés. Il s'agit plutôt de créer des scènes audiovisuelles que nous voulons étudier sur les plans auditif et visuel. Plus particulièrement, nous nous intéressons à l'étude de l'influence de la perception visuelle sur la perception auditive, or, il semble plus facile d'observer l'influence visuelle sur une échelle perceptive auditive régulière. L'objectif de la création des champs sonores est ainsi de donner des impressions auditives successives et, dans la mesure du possible, réparties de manière homogène sur une échelle perceptive. Ainsi, pour tester la distance auditive par exemple, les scènes sonores simulées doivent fournir des sensations de distance croissante – et surtout pour assurer l'ordre de succession – les différences perceptives entre chaque scène doivent être à peu près équivalentes.

Si on voulait néanmoins utiliser la méthode mentionnée ci-dessus (mesures de réponses impulsionnelles et convolution), il s'avère que certaines salles ont une tendance à homogénéiser ou au contraire à amplifier les différences de sensation auditive de distance [Kahle, 1995]. Le choix de la salle serait donc difficile, de même que le serait celui de la discrétisation en distance géométrique pour chaque réponse impulsionnelle (faudrait-il par exemple mesurer ces réponses à des distances successives équivalentes à 4, 8 et 12 mètres de la scène ou bien à des distances équidistantes sur une échelle logarithmique ?).

Par ailleurs, les réponses impulsionnelles mesurées dans les salles varient simultanément sur de multiples aspects de la perception auditive, ce qui compromettrait les étapes d'interprétation des résultats.

Une autre solution est de s'appuyer sur les connaissances préalables concernant la perception auditive, et d'utiliser par exemple un indice acoustique mesurable pour traduire l'impression de la distance auditive. Ce faisant, le choix de l'ensemble des stimuli devient plus facile et leur répartition régulière sur l'échelle perceptive est plus assurée. De plus, l'effet de l'influence visuelle devient plus observable sur une telle structure perceptive régulière.

Ces considérations nous ont conduit à concevoir des stimuli auditifs contrôlés directement sur la base des facteurs perceptifs tels qu'ils sont proposés par le *Spatialisateur*. De plus, sachant que l'interface perceptive du *Spatialisateur* a été conçue dans l'objectif de permettre une description indépendante des principales sensations auditives liées à la perception des scènes sonores, il devient évident qu'il est plus aisé de concevoir des stimuli ne variant que sur un seul paramètre à la fois.

Le principe de simulation des stimuli auditifs à l'aide d'un contrôle direct des attributs perceptifs peut évidemment s'étendre à un ensemble d'attributs qui caractérisent la qualité acoustique des salles. Ainsi, de manière similaire à la sensation de la distance auditive, prise ici comme un exemple, des variations perceptives sur d'autres paramètres auditifs peuvent être simulées de manière facile et fidèle. C'est le cas pour la deuxième partie de ce travail (chapitre 5) où les perceptions

d'enveloppement, de réverbérance et de vivacité sont également modifiées à l'aide de l'interface perceptive du *Spatialisateur*.

Nous résumons à la fin de cette section, la démarche qui est utilisée pour créer les stimuli auditifs dans le cadre de ce travail : la possibilité de relier les facteurs perceptifs à des indices objectifs mesurables à partir de la réponse impulsionnelle d'une salle, a permis de régler le *Spatialisateur* de manière à simuler une situation réelle : la réponse impulsionnelle mesurée dans la salle en question (pour toutes les expériences principales, dans le théâtre des Champs-Élysées).

Cette simulation est basée sur une procédure d'analyse et de synthèse [Jot et al., 1997] : l'analyse temps-fréquence de la réponse impulsionnelle mesurée dans la salle, permet d'extraire les grandeurs énergétiques et temporelles à partir desquelles s'expriment les indices objectifs (ceux qui sont répertoriés dans le tableau 3.1). Le *Spatialisateur* a ainsi permis de reproduire un effet de salle réel dont les principales caractéristiques acoustiques sont respectées.

A la suite, des variations sur cet effet de salle ont été effectuées en modifiant un ou plusieurs paramètres perceptifs ; les nouveaux effets obtenus ne sont pas forcément représentatifs de ce qui peut se passer réellement dans le théâtre des Champs-Élysées d'une place à l'autre.

La synthèse de l'effet de salle à partir de la réponse impulsionnelle mesurée dans la salle représentée à l'écran, présente l'avantage d'assurer le réalisme (ou du moins la vraisemblance) de la situation présentée aux sujets. Aussi, les variations, par l'intermédiaire de l'interface perceptive, permettent de maintenir une certaine cohérence au sein de la série de stimuli, et de construire un test dans lequel les effets auditifs sont contrôlables, voire prévisibles.

### 3.3 LA SIMULATION VISUELLE

La projection de grandes images stéréoscopiques, donc tridimensionnelles, a été réalisée pour ces expériences afin de simuler au mieux les conditions visuelles présentes dans de vraies salles de spectacle. Les images représentaient des vues des scènes du Théâtre de Louvrais à Pontoise et du Théâtre des Champs Elysées à Paris qui ont été visités à ce propos. Les techniques de prise de vue et de projection des images stéréoscopiques sont brièvement présentées ci-après ; la description des vues utilisées dans les expériences vient ensuite.

#### 3.3.1 Prises de vues et restitution du relief

La cause principale de la perception du relief, c'est à dire de la faculté des êtres humains à pouvoir apprécier en profondeur les distances relatives, est la vision binoculaire. Ainsi, pour la stéréoscopie, comme pour la stéréophonie, deux «signaux» sont nécessaires afin de restituer la dimension spatiale, chacun correspondant au signal capturé par l'un de nos deux récepteurs sensoriels.

Deux images représentant le même objet, légèrement différentes l'une de l'autre, sont captées simultanément par nos deux yeux ; leur différence correspond à une mesure de la distance en profondeur de l'objet : celle-ci est d'autant plus grande que l'objet visualisé est proche. L'appréciation des distances relatives est en effet beaucoup moins précise pour les objets éloignés ; la vision binoculaire naturelle ne jouant plus aucun rôle à partir d'environ deux cents mètres.

C'est au niveau du cerveau que les deux images sont ensuite fusionnées pour donner une image en relief. Il s'agit d'un traitement inconscient prenant en compte principalement les différences entre les deux images, mais aussi des informations pouvant aider à la restitution du 'relief monoculaire', c'est à dire de différents repères (par exemple la taille implicite des objets connus), la perspective, les textures, les ombres, le masquage d'objets situés en arrière plan par des objets situés à proximité, etc. [Cahen, 1990].

Les images présentées dans nos expériences ont été obtenues par stéréoscopie conventionnelle : il s'agit d'obtenir sur émulsion photographique deux prises des vues d'une même scène, légèrement décalées, chacune correspondant à la vue de chaque œil. Deux prises de vues successives ont ainsi été réalisées pour chaque image – l'une correspondant à la vue de l'œil droit et l'autre à celle de l'œil gauche– avec une distance d'écartement de 65 mm, distance qui correspond à la distance interoculaire moyenne (base stéréoscopique). Entre les deux prises de vues, l'appareil photographique a été glissé sur une « réglette », un appareil qui permet d'assurer le déplacement horizontal et la précision de l'écartement.

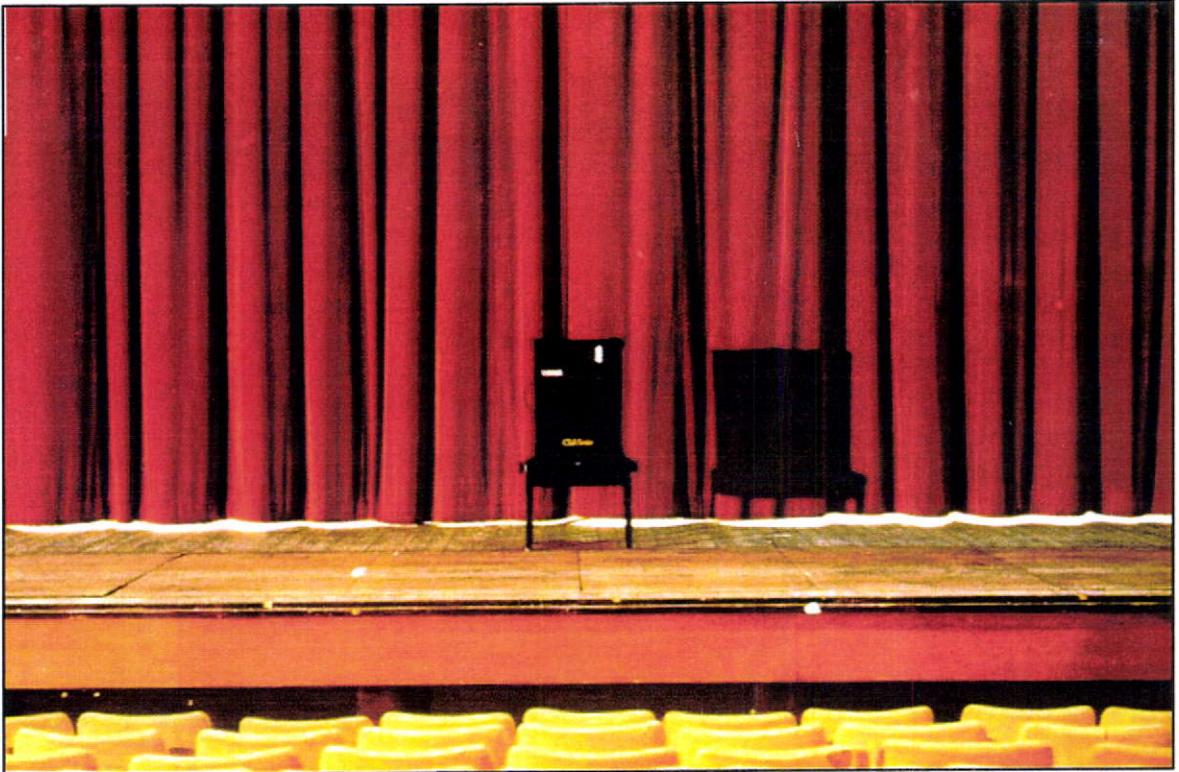
Le développement des diapositives est suivi de l'étape délicate d'«alignement» des vues (assurant la bonne superposition des vues), elle a été effectuée à l'aide d'un second appareil stéréoscopique spécifique, la «monteuse».

La projection en lumière polarisée a été réalisée ensuite afin de permettre de fournir à chaque œil la vue qui lui correspond. Elle consiste à placer un polariseur devant chaque objectif de projection, et à regarder l'image résultante avec des lunettes polarisantes. La projection en relief exige deux projecteurs, ils sont placés l'un sur l'autre. Enfin, surtout pour éviter le bruit des projecteurs auprès des sujets, la projection a été réalisée depuis l'arrière de l'écran (par rétro-projection) ; ce qui a nécessité l'utilisation d'un écran translucide qui permet de respecter la polarisation. La projection a été réalisée dans des conditions d'obscurité complète. La taille de l'image sur l'écran était de 100 cm x 150 cm ; la distance à laquelle les sujets-spectateurs étaient assis était de 160 cm par rapport à l'écran.

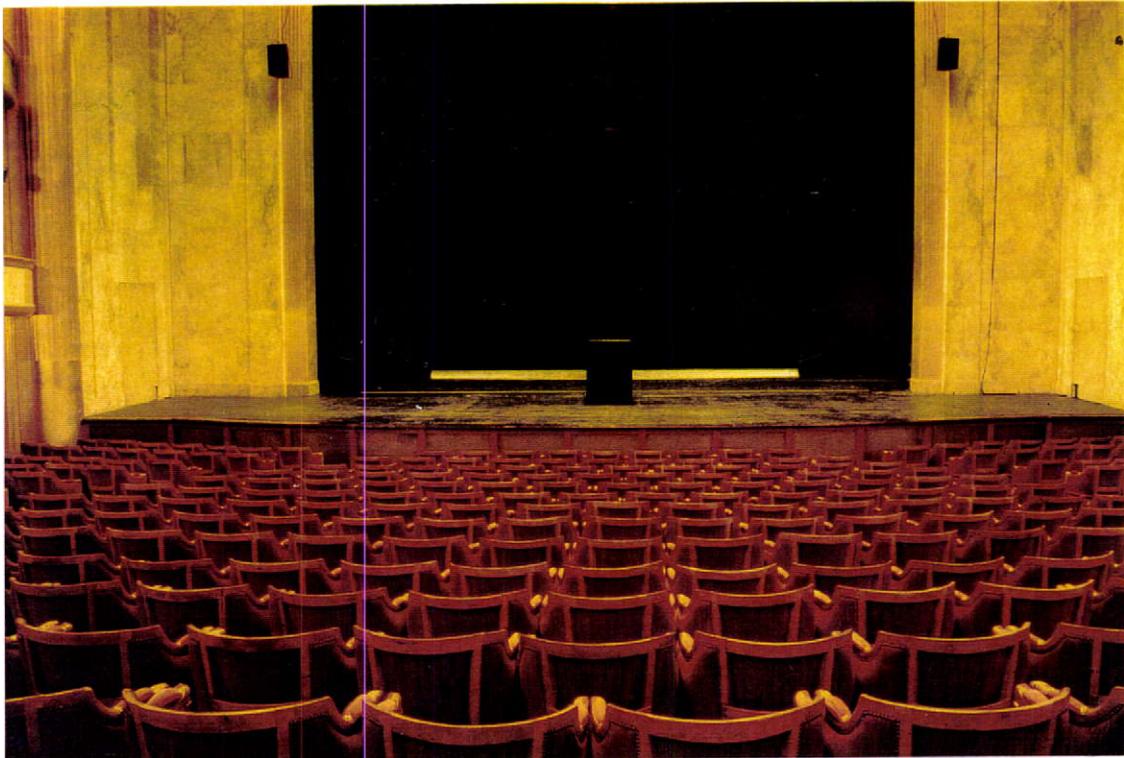
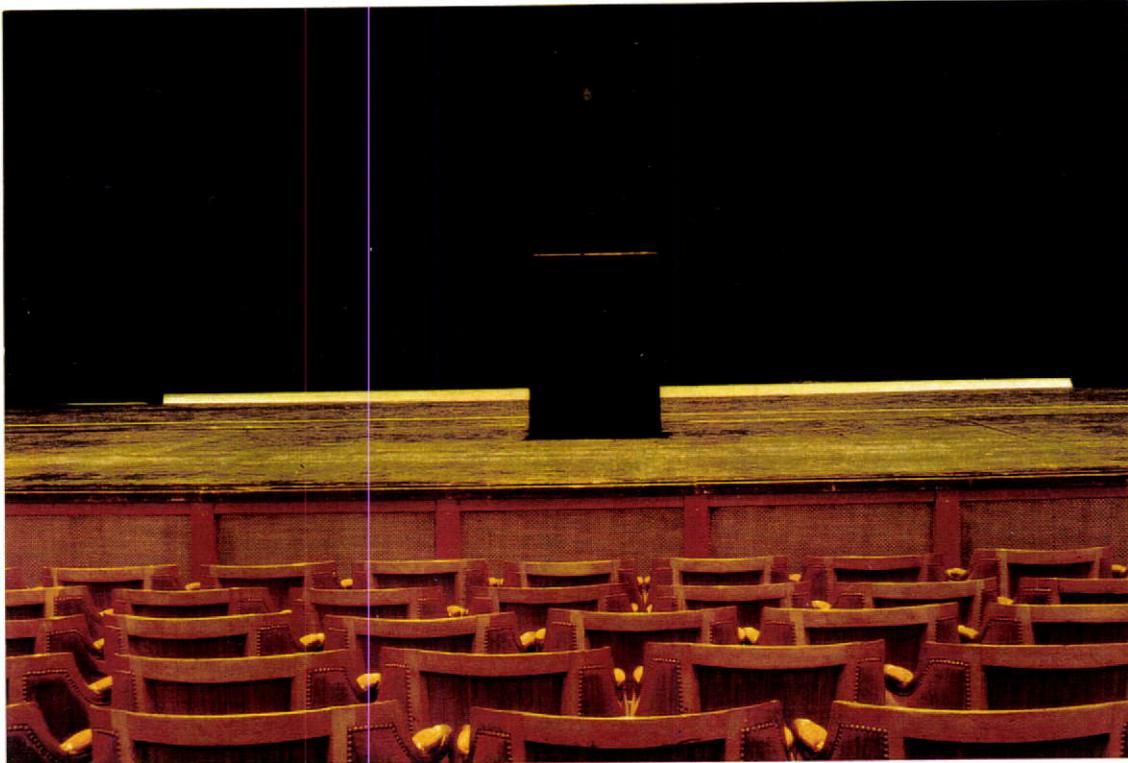
### 3.3.2 Les vues

Les vues utilisées dans ces expériences ont été prises dans le Théâtre de Louvrais à Pontoise et dans le Théâtre des Champs Elysées à Paris. Elles ont été prises depuis différents emplacements des spectateurs par rapport à la scène, au milieu de laquelle dans les deux cas, était placé un haut-parleur.

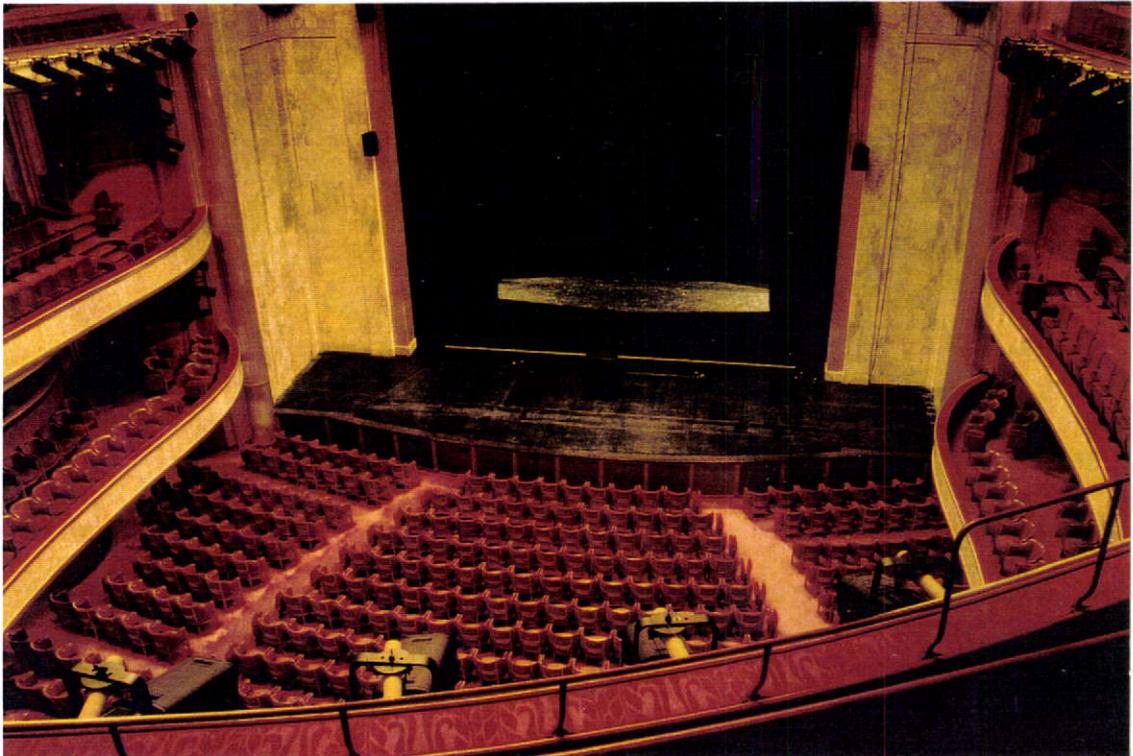
Les vues obtenues dans le Théâtre de Louvrais ont été utilisées uniquement pour l'expérience préliminaire. Les vues stéréoscopiques V1 et V2 ont été prises au niveau des yeux des spectateurs, sur l'axe central de la salle et à des distances respectives de 6 m et 18 m du bord de la scène (distance focale de l'objectif: 28 mm). Le haut parleur (taille de face: 62 cm x 46 cm) était placé sur un tabouret (à 45 cm de hauteur) à une distance de 3 m par rapport au bord de la scène (cf. **figure 3.11**).



**Figure 3.11** : Un aperçu des images V1 et V2 utilisées dans l'expérience préliminaire. En projection, les images sont en couleur, tridimensionnelles et grandes (100 cm x 150 cm).

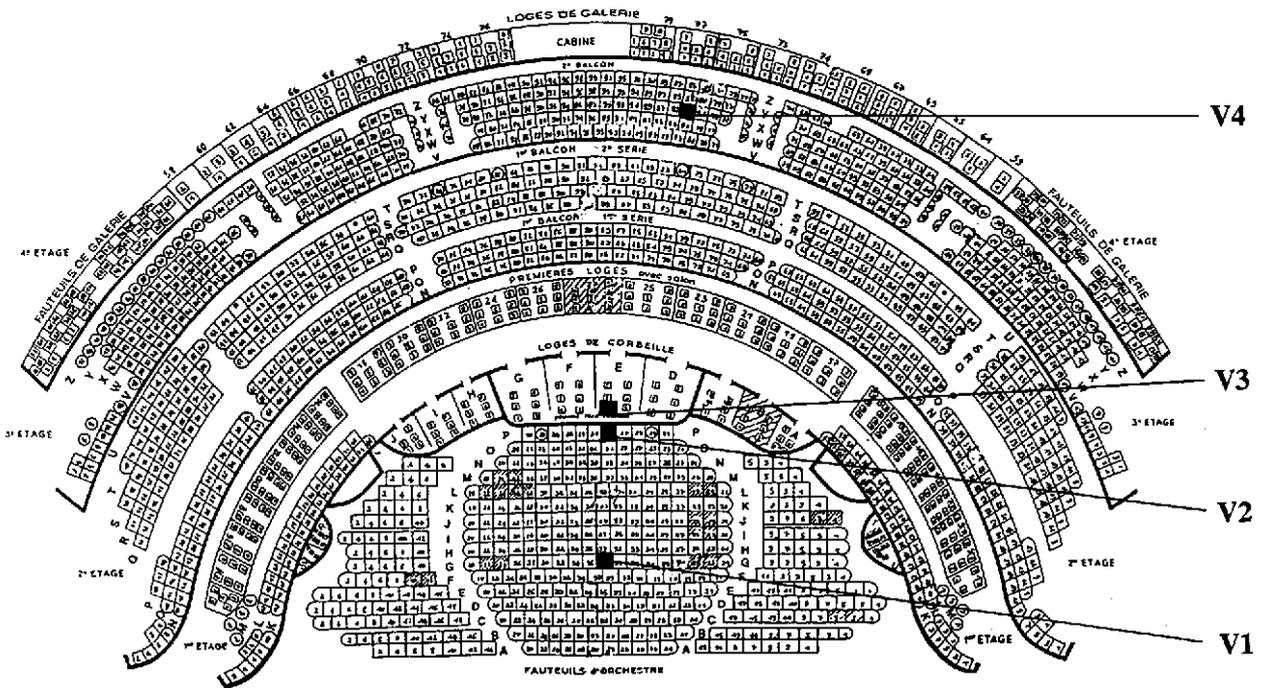


**Figure 3.13** : Un aperçu des images V1 (place G37, distance focale  $df=50\text{mm}$ ) et V2 (place P29,  $df=28\text{mm}$ ) utilisées dans les expériences. Les images projetées étaient en couleur, tridimensionnelles et grandes (100 cm x 150 cm).



**Figure 3.13 (suite) :** Un aperçu des images V3 (place E1,  $df=28mm$ ) et V4 (place X83,  $df=28mm$ ) utilisées dans les expériences. Les images projetées étaient en couleur, tridimensionnelles et grandes (100 cm x 150 cm).

Les images présentées dans la grande majorité des expériences ont été prises dans le Théâtre des Champs Elysées à Paris, salle de 1900 personnes inaugurée initialement en 1913. Il s'agit d'une salle d'un volume total d'environ 17 000 m<sup>3</sup> de forme générale circulaire et relativement compacte (aucun spectateur n'est à plus de 30 m du chef d'orchestre). Le plan développé de ce théâtre et les emplacements depuis lesquels les vues V1, V2, V3 et V4 ont été obtenues, sont présentés à la figure 3.12. La figure 3.13 donne un aperçu de ces quatre images.



**Figure 3.12 :** Plan développé du Théâtre des Champs Elysées (les balcons sont en réalité superposés). Les emplacements G37, P29, E1 et X83 depuis lesquels les prises de vues V1, V2, V3 et V4 ont été respectivement obtenues, sont noircis.

Il est évident qu'à la vue de ces photos, les trois premières ont été prises à des distances croissantes sur l'axe centrale de la salle. Comme il n'était pas possible que ces trois vues soient prises à la même hauteur au dessus de la scène, les vues V1 et V2 ont été prises au niveau de l'orchestre tandis que la vue V3 est prise au niveau des loges de corbeille ; cette dernière représente donc une vue très légèrement surélevée. La vue V4 est encore plus éloignée mais excentrée et bien plus élevée. Elle n'a pas été utilisée pour les expériences sur la distance auditive, puisqu'en même temps que la distance, d'autres paramètres visuels changent dans cette photo. Par ailleurs, ces quatre vues ont fait l'objet d'un test préliminaire 'visuel' dont une partie des résultats est présentée en annexe C.

La focale de l'objectif photographique étant différente de la focale de l'objectif humain (et parfois même – à cause de limitations techniques – différente entre diverses prises de vues), les distances perçues ne peuvent pas être représentées par des distances réelles. Il serait ainsi difficile de dire qu'une distance sur une image correspond à une distance réelle précise. De plus, il est connu qu'en

stéréoscopie, il est impossible de juger les distances visuelles de manière absolue en l'absence notamment d'un objet dans l'image dont la taille réelle soit connue<sup>1</sup>.

Les distances visuelles ainsi perçues ne pouvant être évaluées que de manière relative, un moyen simple de les comparer est de mesurer la taille d'un objet commun présent dans les différentes images : conformément aux lois de perspective, la taille d'un objet est inversement proportionnelle à la distance à laquelle on l'observe. Le haut-parleur présent au milieu de la scène peut alors servir de référence ; sa hauteur a été mesurée sur l'écran de projection et les distances relatives obtenues ont été utilisées pour l'analyse des résultats. Ces rapports de distances sont les suivants :  $V2/V1 = 3.23$ ,  $V3/V2 = 1.24$ ,  $V4/V3 = 1.27$ ,  $V3/V1 = 4.00$  (les hauteurs du haut-parleur mesurées à l'écran étant en cm  $V1 : V2 : V3 : V4 \sim 24.2 : 7.5 : 6 : 4.7$ ).

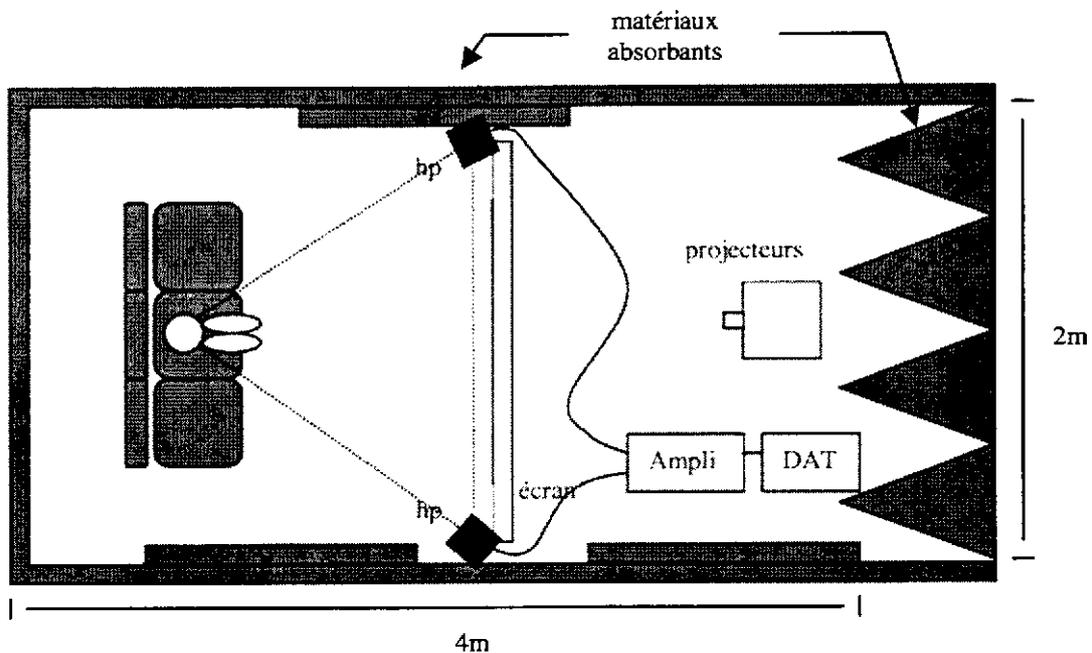
---

<sup>1</sup> Ce qui nous rappelle la faiblesse du système auditif pour juger de façon absolue la distance auditive d'un signal inconnu.

### 3.4 SALLES ET MATERIEL DE SIMULATION.

Toutes les expériences effectuées dans le cadre de ce travail ont eu lieu dans la chambre assourdie de l'IUT Génie Civil de l'Université de Cergy-Pontoise entre mars 1996 et février 1998. Une seule expérience de caractère préliminaire, sans simulation visuelle, a eu lieu dans la salle d'écoute sèche du Laboratoire d'Acoustique Musicale de l'Université Paris VI en février 1997.

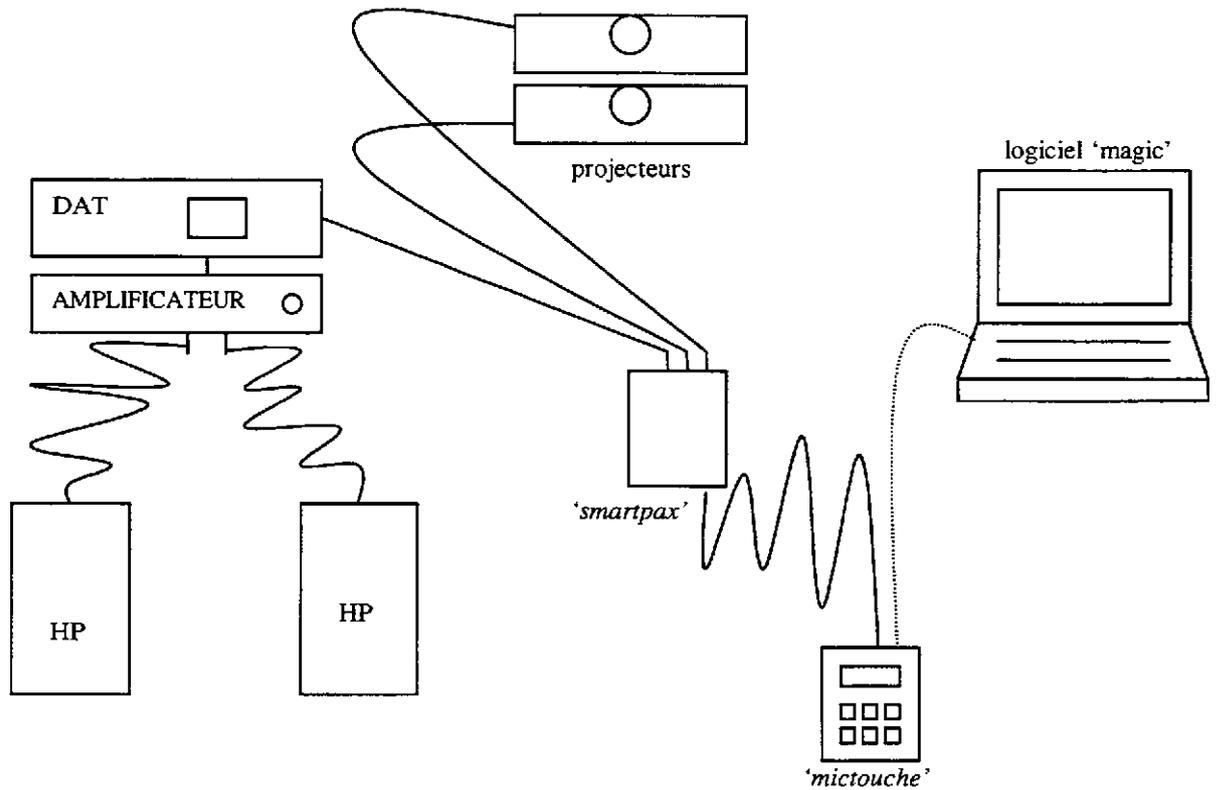
La salle de l'Université de Cergy-Pontoise et le dispositif expérimental sont décrits à la suite. Il s'agit d'une salle rectangulaire, aux dimensions intérieures approchées de 4,2 m x 2 m, la hauteur est de 4 m (Figure 3.14). Des panneaux acoustiques de laine de roche recouvrent les trois surfaces verticales; des dièdres volumineux de laine de roche sont posés sur la quatrième surface verticale ainsi qu'au plafond. Le sol est revêtu de moquette noire. Les murs autour de la place des sujets sont également en noir. Les sujets sont assis dans des sièges du même type que ceux de la salle de spectacle visualisée, à une distance de 1,6 m de l'écran qui lui a des dimensions de 175 cm x 115 cm (taille de l'image 150 cm x 100 cm).



**Figure 3.14 :** Plan schématique de la salle de tests vue de dessus (dimensions : longueur 4,2 m, largeur 2 m, hauteur 4 m). La totalité des surfaces intérieures est recouverte de matériaux absorbants. La salle présente une asymétrie de traitement acoustique sur les parois latérales (double épaisseur de panneaux acoustiques répartis de façon aléatoire). Les deux haut-parleurs et la tête des auditeurs se trouvent aux sommets d'un triangle à peu près équilatéral.

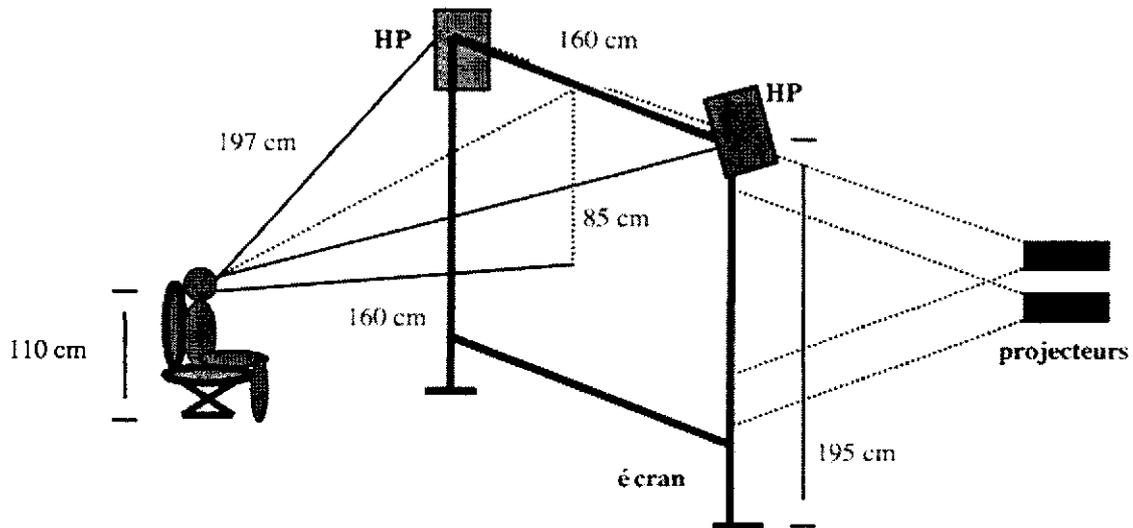
Le dispositif expérimental est présenté à la figure 3.15 (les détails techniques sont présentés en annexe A) La chaîne sonore est composée d'un lecteur DAT, d'un amplificateur et de deux enceintes. Le dispositif visuel est composé de deux projecteurs, de filtres polarisants, d'un écran translucide pour la rétro-projection et de lunettes – polarisantes également – que portaient les sujets (seuls les projecteurs superposés, sont représentés sur la figure).

La gestion des images et des sons était réalisée à l'aide d'un dispositif de contrôle (*DATATON AC PAX Adaptor*) qui permettait de commander à la fois les deux projecteurs et le DAT. Les séquences des stimuli étaient programmées par le logiciel *magic* et chargées sur un automate (*mictouche*) avant chaque session des tests. L'automate commandait le *Smartpax* qui à son tour gérait les projecteurs et le DAT. Le système permettait également de noter et d'enregistrer les réponses.



**Figure 3.15 :** Dispositif expérimental : chaîne sonore et dispositif visuel commandés par un système de contrôle programmable (*magic, smartpax, mictouche*).

Nous avons vu dans la section précédente concernant la simulation sonore, que la diffusion transaurale du son impose des contraintes par rapport à la position de la tête des auditeurs ; il faut en effet que les deux enceintes et la tête des auditeurs forment un triangle équilatéral. Ceci a été respecté dans la mesure du possible, sachant que des limitations supplémentaires ont été imposées par les dimensions de la salle et surtout par la simulation visuelle (dimensions fixes de l'écran, grande taille de l'image projetée, contraintes de la projection 3-D). De plus, les haut-parleurs ne devaient pas masquer l'écran, ni être trop visibles ; ils ont donc été surélevés. Des réglages fins d'emplacement de la tête des auditeurs ont été effectués pour chaque sujet afin que la restitution sonore soit convaincante (cf. § 4.2.1 'Réglage subjectif du dispositif transaural'). Les distances moyennes auxquelles se trouvaient les sujets par rapport aux enceintes, l'écran, etc., sont indiquées sur la **figure 3.16**.



**Figure 3.16 :** Dispositif stéréophonique utilisé et distances moyennes auxquelles se trouvaient les sujets par rapport aux enceintes, l'écran, etc. Les deux enceintes et la tête des auditeurs doivent normalement former un triangle équilatéral, égal à la base stéréophonique (ici de 160 cm de côté). Cette règle fut respectée dans la mesure du possible, mais cela n'a pas toujours été le cas. Par ailleurs, des réglages avant les tests ont permis de vérifier la bonne restitution spatiale du son pour chaque sujet.

## 3.5 METHODES D'ANALYSE

Les analyses statistiques utilisées dans le cadre de ce travail sont essentiellement l'Analyse de Variance (ANOVA) et l'Analyse en Composantes Principales (ACP). Les principes généraux et les étapes importantes de calcul de ces deux analyses sont présentés dans ce chapitre. Des exemples de tableaux et de graphiques issus de ces analyses permettent à la fin de chaque section de faciliter la compréhension des résultats.

### 3.5.1 Analyse de la Variance

#### 3.5.1.1 Introduction

Le principe de l'Analyse de Variance (ANOVA) est basé sur la décomposition de la variance totale d'un ensemble d'observations en variances partielles et sur leurs comparaisons entre elles. Elle consiste essentiellement à calculer et à tester le rapport des variances *entre* les populations que l'on veut comparer, avec les variances *à l'intérieur* de chacune de ces populations.

L'Analyse de Variance s'applique dans les cas où la variable dépendante – la variable à expliquer – est *métrique* (représentée par un chiffre) alors que les variables indépendantes – les facteurs explicatifs – sont *nominales* [Lagarde, 1983].

Il s'agit d'une analyse qui permet de tester simultanément l'influence de deux ou de plusieurs facteurs sur les résultats et d'évaluer les interactions entre ces facteurs. C'est une analyse puissante, capable de révéler des tendances subtiles à condition qu'elles soient observées de manière systématique, c'est à dire qu'elles soient présentes dans la grande majorité des observations. Enfin, il s'agit d'une analyse robuste au non-respect des hypothèses de départ [Bray et Maxwell, 1985]. Ces propriétés lui confèrent une place privilégiée parmi les méthodes d'analyses statistiques.

Afin de montrer les principes fondamentaux de l'Analyse de Variance (hypothèses, décomposition de la variance, test de significativité), le cas d'une analyse à un seul facteur (ou variable explicative) sera d'abord présenté. Sa généralisation aux cas complexes de l'analyse à plusieurs facteurs avec interactions et de l'analyse de variance à mesures répétées – cas particuliers utilisés dans le cadre de ce travail – permettra ensuite de mettre davantage en évidence la technique et l'intérêt de la méthode.

#### 3.5.1.2 Analyse de la variance à un facteur

L'analyse de la variance à un facteur permet de découvrir si un seul facteur peut expliquer ou non les variations constatées dans les observations de la variable étudiée. Supposons que l'on soit en présence de  $k$  populations, chacune correspondant à l'un des  $k$  différents *niveaux* (ou *modalités*) du facteur en question. On extrait indépendamment un échantillon aléatoire de chacune des  $k$  populations. Si on note par  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$  les moyennes de cette variable pour chacun de ces échantillons, on cherche à tester l'hypothèse :

$$H_0 : \bar{x}_1 = \bar{x}_2 = \dots = \bar{x}_k$$

contre l'hypothèse :

$H_1$  : les moyennes ne sont pas toutes égales (au moins une de ces égalités n'est pas vérifiée).

La véracité de l'hypothèse  $H_0$  signifie que le facteur n'a aucun effet particulier sur la variable étudiée. L'ensemble des observations peut alors être considéré comme appartenant à une seule population. En revanche, la véracité de l'hypothèse  $H_1$  montre que le facteur a un effet significatif et quantifiable sur les résultats [Martel et Nadeau, 1988].

Nous décrivons ci-après comment l'Analyse de Variance procède pour tester ces hypothèses : considérons un grand nombre  $n$  d'observations (par exemple les réponses  $x$  des sujets sur un attribut auditif des sons présentés lors d'une expérience). Supposons que l'on cherche à évaluer l'effet d'un seul facteur (par exemple les différents stimuli sonores présentés). Supposons aussi que ce facteur ait  $k$  modalités (par exemple  $k$  sons différents  $C_1, C_2, \dots, C_k$ ) qui ont respectivement fait l'objet de  $n_1, n_2, \dots, n_k$  observations (avec bien sûr  $n_1+n_2 + \dots+n_k = n$ ) et dont les moyennes relatives sont  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ . Comme nous l'avons vu auparavant, la procédure consiste à tester si ces moyennes sont significativement différentes de  $\bar{x}$ , ce qui reviendrait à dire que le type du son présenté a un effet sur le jugement de l'attribut auditif. On peut représenter ces observations par un tableau de la forme du **tableau 3.2**.

Populations	Observations				Moyennes
1	$x_{11}$	$x_{12}$	...	$x_{1n_1}$	$\bar{x}_1$
2	$x_{21}$	$x_{22}$	...	$x_{2n_2}$	$\bar{x}_2$
.	.	.	...	.	.
.	.	.	...	.	.
.	.	.	...	.	.
k	$x_{k1}$	$x_{k2}$	...	$x_{kn_k}$	$\bar{x}_k$
					$\bar{x}$

**Tableau 3.2** : Classification des observations utilisées dans une Analyse de Variance à un facteur [Martel et Nadeau, 1988].

La variance totale  $V_{tot}$  des  $n$  observations est par définition égale à la moyenne des carrés des écarts de ces  $n$  observations par rapport à la moyenne générale  $\bar{x}$ . Or, le théorème de Huygens permet de décomposer cette variance en des variances partielles associées chacune à de différentes modalités du facteur testé. La variance totale peut alors être écrite sous la forme :

$$V_{tot} = \underbrace{1/n \sum (x - \bar{x}_1)^2 + n_1/n (\bar{x}_1 - \bar{x})^2}_{\text{part de variance due à la modalité 1}} + \underbrace{1/n \sum (x - \bar{x}_2)^2 + n_2/n (\bar{x}_2 - \bar{x})^2 + \dots}_{\text{part de variance due à la modalité 2}} + \underbrace{1/n \sum (x - \bar{x}_k)^2 + n_k/n (\bar{x}_k - \bar{x})^2}_{\text{part de variance due à la modalité k}}$$

Cette même équation peut s'écrire :

$$V_{tot} = \underbrace{1/n \sum (x - \bar{x}_1)^2 + 1/n \sum (x - \bar{x}_2)^2 + \dots + 1/n \sum (x - \bar{x}_k)^2}_{\text{variance intra}} + \underbrace{n_1/n (\bar{x}_1 - \bar{x})^2 + n_2/n (\bar{x}_2 - \bar{x})^2 + \dots + n_k/n (\bar{x}_k - \bar{x})^2}_{\text{variance inter}}$$

On observe que la variance totale  $V_{\text{tot}}$ , qui correspond à SCT (voir plus loin), est la somme de deux variances [Lagarde, 1983] :

- La variance *inter* (*between*) obtenue à partir des écarts de la moyenne *particulière* de chaque modalité avec la moyenne globale. Elle représente la variation des observations qui est due au facteur, elle est en quelque sorte une mesure de son effet qui se trouve ainsi isolé du reste. Elle est mesurée par l'intermédiaire de la somme SCF des carrés dus au facteur [Martel et Nadeau, 1988] :

$$SCF = \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{X}_i - \bar{X})^2$$

- La variance *intra* (*within*) obtenue à partir des écarts de chaque observation avec la moyenne *particulière* de sa modalité. Elle représente la variation des observations qui n'est pas due au facteur testé et qui peut être expliquée, soit par d'autres facteurs (qui n'ont pas été pris en compte ou qui n'ont pas pu être contrôlés), soit par le hasard pur. C'est le *résidu*, il est mesuré par l'intermédiaire de la somme SCE des carrés dus à l'erreur :

$$SCE = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

Le caractère plus ou moins significatif de l'effet du facteur explicatif (ici du son utilisé) dépend du rapport entre ces deux quantités. Pour les comparer, il faut faire intervenir le nombre de *degrés de liberté* : Si  $k$  est le nombre de modalités du facteur, le nombre de degrés de liberté pour la variance due au facteur est égal à  $k-1$  et le nombre de degrés de liberté pour la variance due à l'erreur est égale à  $n-k$ . On forme ensuite le rapport :

$$F = \frac{SCF / (k - 1)}{SCE / (n - k)}$$

Si les populations sont normales et ont la même variance, et si les échantillons tirés de ces populations sont indépendants, le rapport  $F$  suit une distribution de Fisher à  $k-1$  et  $n-k$  degrés de liberté [Martel et Nadeau, 1988].

On procède alors à un *test unilatéral à droite* en comparant le  $F$  obtenu au  $F$  théorique, ce dernier correspondant à un effet nul donné par une table de Fisher-Snedecor. Au niveau  $p$ , il s'agit de trouver pour  $k-1$  degrés de liberté au numérateur et  $n-k$  au dénominateur, la valeur  $F_p$  telle que  $P(F < F_p) = p$ . Cette table donne en général deux valeurs de  $F_p$  correspondant respectivement à 95% ( $p=0.05$ ) et 99% ( $p=0.01$ ) de seuil de signification ou niveau de certitude. Dans la pratique, c'est le logiciel statistique qui calcule les rapports de variances  $F$  et estime les seuils. Si le rapport obtenu  $F$  est supérieur à  $F_p$  ( $p=0.05$ ), il correspond à un seuil de signification supérieur à 95% ( $p < 0.05$ ). On peut alors affirmer – au niveau de certitude obtenu – que le facteur a un effet. La différence des moyennes partielles  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p$  avec la moyenne générale  $\bar{x}$  est une bonne mesure de cet effet.

Si l'effet d'un facteur se révèle significatif, les différences entre les modalités peuvent être testées avec une série de tests supplémentaires offerts par l'ANOVA, les comparaisons Post-Hoc. Ces tests permettent de tester la différences entre les moyennes de deux ensembles d'observations, en prenant en compte justement, le fait qu'il s'agit de moyennes et non pas de deux simples observations [Winer, 1962], [Milliken et Johnson, 1984].



Un des tests Post-Hoc les plus conservatifs est le test Tukey HSD (Honesty Significance Difference), utilisé ici, il est basé sur la distribution 'studentized range'. Une valeur  $t_s$  est calculée pour chaque paire de moyennes :  $t_s = (M_i - M_j) / \sqrt{(MSE / n_h)}$ , où  $M_i - M_j$  est la différence entre la  $i^{\text{ème}}$  et la  $j^{\text{ème}}$  moyenne, MSE est la moyenne des carrés des erreurs (Mean Square Error) et  $n_h$  est la moyenne harmonique des tailles d'échantillons des groupes  $i$  et  $j$ . La valeur critique de  $t_s$  utilisée pour tester la significativité est déterminée par la distribution 'studentized range', elle est la valeur de  $t_s$  qui ne doit pas être dépassé par la valeur calculée pour que l'hypothèse nulle soit rejetée.

### 3.5.1.3 Analyse de la variance à deux facteurs avec interaction

Le principe précédent se généralise pour les cas de deux facteurs ou plus. En principe, nous pouvons considérer qu'il n'y a pas d'influence ou d'interaction entre les différents facteurs. Cela signifie que l'effet quelconque d'un facteur sur les observations est indépendant de la modalité de l'autre facteur. Or, en réalité, ce n'est pas toujours le cas : l'effet d'un facteur sur la variable dépendante peut très bien dépendre de la modalité de l'autre facteur.

Ainsi, si par exemple la variable étudiée  $x$  est la distance auditive d'un ensemble de sons, les deux facteurs pour lesquels on cherche à évaluer l'effet sont, d'une part les différents stimuli sonores utilisés, et d'autre part les conditions visuelles sous lesquelles sont présentés ces sons : le fait que les conditions visuelles puissent avoir un effet différent sur un stimulus particulier va entraîner une interaction entre ces deux facteurs.

Dans le cas de l'analyse de la variance à deux facteurs, on est en présence de  $(k \times n)$  populations : chacune des  $k$  modalités du premier facteur définit une population, et chacune des  $n$  modalités du second facteur définit aussi une population. Les facteurs peuvent ainsi avoir des effets conjoints qui, pour les étudier, nécessitent de disposer de deux observations ou plus par population. Nous devons alors avoir un échantillon de  $k \times n \times r$  observations (avec  $r \geq 2$ ). Le tableau d'observations peut ainsi prendre la forme du **tableau 3.3**. Les lignes du tableau représentent les  $k$  modalités du facteur  $F_1$  et les colonnes les  $n$  modalités du facteur  $F_2$ .

La variance totale est décomposée en quatre parties : une variance due au facteur 1 ( $SCF_1$ ), une variance due au facteur 2 ( $SCF_2$ ), une variance due à l'interaction entre les deux facteurs (SCI) et une variance due à l'erreur (SCE).

Les sommes de carrés correspondant à chacune des ces variances sont ainsi :

- La somme de carrés totale due à la variance totale

$$SCT = \sum_{i=1}^k \sum_{j=1}^n \sum_{h=1}^r (x_{ijh} - \bar{x})^2$$

- La somme de carrés due au premier facteur ( $F_1$ )

$$SCF_1 = \sum_{i=1}^k nr (\bar{x}_{i..} - \bar{x})^2$$

- La somme de carrés due au deuxième facteur ( $F_2$ )

$$SCF_2 = \sum_{j=1}^n kr (\bar{x}_{.j.} - \bar{x})^2$$

- La somme de carrés due à l'interaction

$$SCI = \sum_{i=1}^k \sum_{j=1}^n r (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2$$

où  $x_{ij}$  est la moyenne des observations à l'intersection de la  $i$ -ème ligne et de la  $j$ -ème colonne.

- La somme de carrés due à la variance liée à l'erreur

$$SCE = SCT - SCF_1 - SCF_2 - SCI$$

Colonnes	1	...	$j$	...	$n$	Moyenne
Lignes						
1	$x_{111}$	...	$x_{1j1}$	...	$x_{1n1}$	$\bar{x}_{1.}$
	.	...	.	...	.	
	.	...	.	...	.	
	$x_{11h}$	...	$x_{1jh}$	...	$x_{1nh}$	
	.	...	.	...	.	
	$x_{11r}$	...	$x_{1jr}$	...	$x_{1nr}$	
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
$i$	$x_{i11}$	...	$x_{ij1}$	...	$x_{in1}$	$\bar{x}_{i.}$
	.	...	.	...	.	
	.	...	.	...	.	
	$x_{i1h}$	...	$x_{ijh}$	...	$x_{inh}$	
	.	...	.	...	.	
	$x_{i1r}$	...	$x_{ijr}$	...	$x_{inr}$	
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
$k$	$x_{k11}$	...	$x_{kj1}$	...	$x_{kn1}$	$\bar{x}_{k.}$
	.	...	.	...	.	
	.	...	.	...	.	
	$x_{k1h}$	...	$x_{kjh}$	...	$x_{knh}$	
	.	...	.	...	.	
	$x_{k1r}$	...	$x_{kjr}$	...	$x_{knr}$	
Moyenne	$\bar{x}_{.1}$	...	$\bar{x}_{.j}$	...	$\bar{x}_{.n}$	$\bar{x}$

**Tableau 3.3 :** Classification des observations utilisées dans une Analyse de Variance à deux facteurs avec interaction. Les lignes du tableau représentent les  $k$  modalités du facteur  $F_1$  et les colonnes les  $n$  modalités du facteur  $F_2$ . Pour tester l'interaction nous devons disposer de  $r$  observations par population ( $r \geq 2$ ), [Martel et Nadeau, 1988].

Les estimateurs des variances sont obtenus ensuite en faisant intervenir les nombres de degrés de liberté et la comparaison des variances s'effectue en formant les rapports suivants :

$$F_1 = \frac{SCF_1 / (k - 1)}{SCE / [kn(r - 1)]}$$

$$F_2 = \frac{SCF_2 / (n - 1)}{SCE / [kn(r - 1)]}$$

$$F_{\text{interaction}} = \frac{SCI / [(k - 1)(n - 1)]}{SCE / [kn(r - 1)]}$$

En conséquence, dans l'Analyse de Variance à deux facteurs avec interaction, on est amené à effectuer trois tests d'hypothèses différentes : un test qui concerne l'effet du premier facteur, un test qui concerne l'effet du second facteur et un test qui concerne les effets d'interaction entre les deux facteurs.

Si la variable  $x$  que l'on étudie dans les  $kn$  populations est normale, le rapport  $F_1$  suit une distribution de Fisher à  $k-1$  et  $kn(r-1)$  degrés de liberté, et le rapport  $F_2$  suit une distribution de Fisher à  $n-1$  et  $kn(r-1)$  degrés de liberté. Le rapport  $F_{\text{interaction}}$  suit une distribution à  $(k-1)(n-1)$  et  $kn(r-1)$  degrés de liberté.

Comme dans le cas à un seul facteur, on procède ensuite à des *tests unilatéraux à droite* en comparant les  $F$  obtenus aux  $F$  théoriques, les  $F$  théoriques correspondant à des effets nuls donnés par les tables de Fisher-Snedecor pour les niveaux de certitude choisis. Si les rapports obtenus sont supérieurs à ces valeurs critiques, on peut affirmer – au niveau de certitude obtenu – que les facteurs correspondants ont un effet. En général, le seuil de signification est de 95 %, seuil qui correspond à une probabilité d'erreur  $p = 0.05$  (si  $p$  est inférieure à 0.05, l'effet est considéré significatif, si  $p$  est supérieur, l'effet n'est pas significatif et l'hypothèse  $H_0$  est retenue). Evidemment, plus les valeurs  $F_1$ ,  $F_2$  et  $F_{\text{interaction}}$  sont grandes, plus les probabilités  $p$  d'erreur sont petites et plus les effets correspondants sont significatifs.

### 3.5.1.4 Analyse de Variance à Mesures Répétées

Il faut souligner que l'Analyse de Variance permet de traiter une grande variété de problèmes, d'où des versions adaptées pour différentes applications. Dans cette section, on présente l'*Analyse de Variance à Mesures Répétées* (Repeated Measures ANOVA) utilisée dans le cadre de ce travail. Elle s'applique dans les cas particuliers où un même groupe de sujets a pu tester tous les traitements apportés à un produit ou à une situation. On entend par « traitements » les modifications engendrées par les différentes modalités des facteurs dont on cherche à évaluer l'effet.

Prenons ici un exemple type n'ayant pas de rapport direct avec nos expériences mais qui permet de démontrer dans quels cas particuliers s'applique cette méthode : supposons que la variable dépendante  $x$  est un indice biologique mesuré chez un groupe de patients et que les deux facteurs pour lesquels on cherche à évaluer l'effet sont, d'une part différents médicaments, et de l'autre, la durée du traitement. Il faudrait que tous les patients aient pris tous les médicaments, pour toutes les durées testées (et dans

des ordres différents) pour avoir le droit (et l'obligation) d'entrer dans le champ d'application d'une Analyse de Variance à Mesures Répétées. Notre équivalent ici pourrait par exemple être l'estimation de la distance auditive comme variable dépendante, testée sous différentes conditions de vision et d'écoute.

La différence entre l'ANOVA à Mesures Répétées et l'ANOVA classique décrite auparavant, se situe au niveau du terme d'erreur. Plutôt que d'utiliser comme terme d'erreur le résidu, issu de la soustraction sur la variance totale de deux types de variances (variances dues aux facteurs et variances dues aux interactions entre les facteurs), l'ANOVA à Mesures Répétées considère le nombre d'observations  $r$  par population (cf. **tableau 3.3**) comme un « nouveau facteur » et utilise comme terme d'erreur la variance due à l'interaction de ce « facteur » supplémentaire avec l'effet à tester. Ainsi, pour tester par exemple l'effet d'un facteur quelconque  $F_1$  qui comporte  $k$  modalités, l'analyse compose le rapport :

$$F_1 = \frac{SCF_1 / (k - 1)}{SCI_1 / [(k - 1)(r - 1)]}$$

où  $SCF_1$  est la somme des carrés due au facteur  $F_1$ ,  $SCI_1$  est la somme des carrés due à l'interaction entre le facteur  $F_1$  et le facteur supplémentaire à  $r$  modalités, issu des  $r$  observations par population.

Pour évaluer des effets dus par exemple à l'interaction des facteurs  $F_1$  et  $F_2$ , l'analyse procède de manière similaire et forme le rapport :

$$F_{\text{interaction}} = \frac{SCI / [(k - 1)(n - 1)]}{SCI_{12} / [(k - 1)(n - 1)(r - 1)]}$$

où  $SCI$  est la somme des carrés due à l'interaction des facteurs  $F_1$  et  $F_2$  avec  $k$  et  $n$  modalités respectivement,  $SCI_{12}$  est la somme des carrés due à la triple interaction entre le facteur  $F_1$ , le facteur  $F_2$  et le facteur issu des  $r$  observations par population.

Il est à noter que l'effet isolé de ce facteur supplémentaire correspondant aux différentes observations par population n'est pas d'un grand intérêt et qu'il n'est généralement pas calculé. Sa création consiste en une étape intermédiaire de l'analyse, ce sont seulement ses interactions qui sont calculées et utilisées comme termes d'erreur pour estimer les autres effets.

L'utilisation de tels termes d'erreur élaborés, par l'intermédiaire de leur interaction directe sur les effets testés, conduit à une analyse plus fine et mieux contrôlée parce qu'il s'agit d'une analyse qui prend en compte les différences individuelles [Bernstein, 1988]. Considérons ainsi, comme c'est le cas dans nos expériences, que les  $r$  observations par population correspondent aux réponses obtenues pour  $r$  sujets ayant participé à une étude. L'interaction du facteur issu des sujets avec le facteur dont on veut tester l'effet est une mesure de l'homogénéité des réponses des sujets. Si cette interaction est grande, cela signifie que les sujets ont répondu très différemment par rapport à ce facteur ; son effet se montrera diminué (puisque l'interaction se trouve au dénominateur). Au contraire si l'interaction s'avère faible, cela signifie que les sujets ont répondu de manière similaire ou cohérente ; l'effet du facteur en question se révélera alors élevé.

### Forme du tableau de l'Analyse de la Variance à Mesures Répétées :

Il est commode de regrouper les principales étapes de l'Analyse de la Variance dans un tableau qui, dans le cas d'analyses à Mesures Répétées, a la forme du **tableau 3.4**.

Effet : Sources de variation	Effet : Degrés de liberté	Effet : Moyennes des carrés	Erreur : Degrés de liberté	Erreur : Moyennes des carrés	F	p
Facteur 1	k-1	$SCF_1 / (k-1)$	$(k-1)(r-1)$	$SCI_1 / (k-1)(r-1)$	$F_1$	p
Facteur 2	n-1	$SCF_2 / (n-1)$	$(n-1)(r-1)$	$SCI_2 / (n-1)(r-1)$	$F_2$	p
Interaction	$(k-1)(n-1)$	$SCI / (k-1)(n-1)$	$(k-1)(n-1)(r-1)$	$SCI_{12} / (k-1)(n-1)(r-1)$	$F_{interaction}$	p

**Tableau 3.4 :** Tableau de l'Analyse de la Variance à Mesures Répétées. F représente le rapport des moyennes des carrés effet / erreur.  $SCF_1$  et  $SCF_2$  sont les sommes des carrés dues respectivement aux facteurs 1 et 2 (avec k et n modalités respectivement);  $SCI_1$  et  $SCI_2$  sont les sommes des carrés dues aux interactions entre le facteur issu des r observations par population (facteur intermédiaire à l'analyse) et les facteurs  $F_1$  et  $F_2$  respectivement ;  $SCI_{12}$  est la somme des carrés due à l'interaction entre le facteur  $F_1$ , le facteur  $F_2$  et le facteur issu des r observations par population.

## 3.5.2 L'Analyse en Composantes Principales

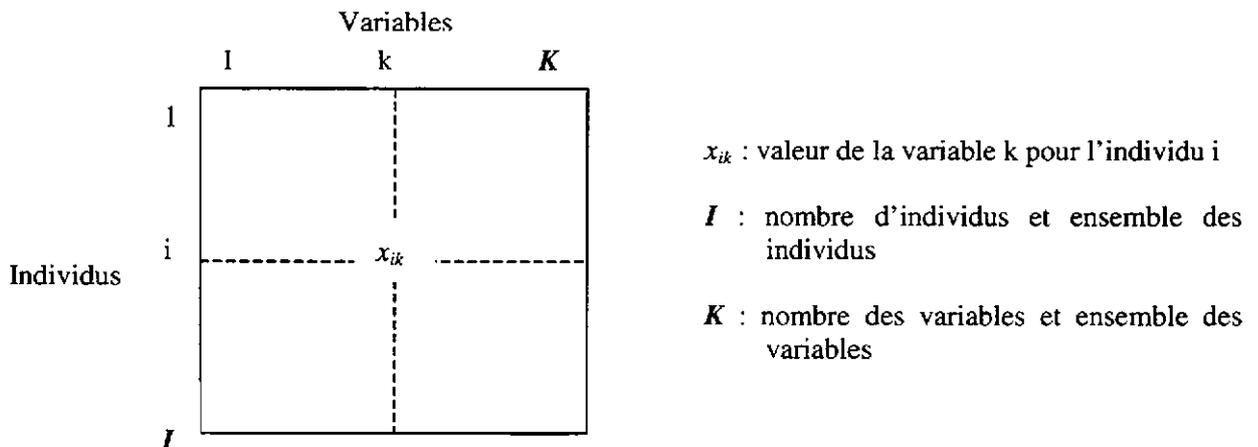
### 3.5.2.1 Introduction

L'Analyse en Composantes Principales ou ACP fait partie des méthodes d'analyse de données appelées multidimensionnelles, par opposition aux méthodes de statistique descriptive qui ne traitent qu'une ou deux variables à la fois. Les méthodes multidimensionnelles permettent la confrontation entre de nombreuses informations, elles aident à en extraire les tendances les plus marquantes, à les hiérarchiser ainsi qu'à éliminer les effets marginaux ou ponctuels qui perturbent la perception globale des faits [Escofier et Pages, 1988].

Parmi les méthodes d'analyse des données, les analyses factorielles tiennent une place primordiale; due en partie à des représentations géométriques des données qui transforment en distances euclidiennes des proximités statistiques entre éléments. Elles se basent toutes sur le même principe de double analyse d'un nuage de points, de recherche des axes d'inertie de ce nuage et de la double projection sur ces axes.

### 3.5.2.2 Tableau et transformation des données - Objectifs

L'Analyse en Composantes Principales (ACP) est une analyse factorielle qui s'applique à des tableaux à deux dimensions croisant des individus et des variables quantitatives. Selon un usage bien établi, les lignes du tableau représentent les individus et les colonnes représentent les variables. A l'intersection de la ligne i et de la colonne k se trouve la valeur de la variable k pour l'individu i. La **figure 3.17** illustre ces notions et complète les notations.



**Figure 3.17** : Le tableau des données. Les colonnes représentent les variables et les lignes représentent les individus [Escofier et Pages, 1988].

Les principales étapes de calculs de l'ACP et le type d'information qui peut être obtenu sont présentés ci-après :

La première opération effectuée sur le tableau initial des données consiste à le transformer, c'est à dire à le *centrer* et à le *réduire*. *Centrer* le tableau consiste à soustraire à chaque valeur  $x_{ik}$  la moyenne de la variable en cause  $\bar{x}_k$ . Le tableau obtenu a alors le terme général :

$$x_{ik} - \bar{x}_k$$

*Réduire* le tableau consiste à exprimer la valeur centrée en nombre d'écart-types  $s_k$  de la variable en cause. Le tableau obtenu a alors comme terme général :

$$(x_{ik} - \bar{x}_k)/s_k$$

Ce faisant, l'unité de mesure pour la variable k devient l'écart-type  $s_k$  de cette variable. Cette opération permet de s'affranchir d'unités de mesure arbitraires. Toutes les variables présentent la même variabilité et de ce fait la même influence dans le calcul des distances entre individus.

Les questions que l'on se pose sur les individus et celles que l'on se pose sur les variables ne sont pas de même nature : Ainsi, à propos des individus on essaie d'évaluer leur *ressemblance*, tandis qu'à propos des variables on cherche à évaluer leurs *liaisons*. Le tableau des données, une fois centré et réduit peut être considéré, soit à travers ses lignes comme un nuage d'individus, soit à travers ses colonnes comme un nuage de variables.

### 3.5.2.3 Le nuage des individus

Deux individus se ressemblent d'autant plus qu'ils possèdent des valeurs proches pour l'ensemble des variables. En ACP, la distance  $d(i,l)$  entre deux individus  $i$  et  $l$  est définie par :

$$d^2(i,l) = \sum (x_{ik} - x_{lk})^2, k \in K$$

Géométriquement, la notion de ressemblance entre deux individus est représentée par la distance euclidienne usuelle.

L'ensemble des distances interindividuelles constitue ce que l'on appelle la forme du *nuage des individus*  $N_I$ . Réaliser un bilan de distances revient ainsi à étudier la forme du nuage, c'est à dire à y déceler une *partition des points* ou des *directions d'allongement*. Dès que  $K$  (le nombre de variables) est supérieur à 3, l'étude directe du nuage  $N_I$  est impossible du fait de la limitation à trois dimensions de notre sens visuel. D'où l'intérêt de l'ACP – et des analyses factorielles en général – qui fournissent des images planes approchant le mieux possible – au sens du critère de l'inertie projetée maximum – un nuage de points situé dans un espace de grande dimension.

En pratique, on recherche une suite  $\{ u_s ; s = 1, \dots, S \}$  de directions privilégiées appelées axes factoriels qui, prises deux à deux, définissent des plans factoriels sur lesquels on projette le nuage  $N_I$ . Chaque direction choisie rend maximale l'inertie par rapport à l'origine  $O$  de la projection de  $N_I$  sur  $u_s$ . Ensuite, on impose à chaque nouvelle direction d'être orthogonale aux directions déjà trouvées. Le plan engendré par les deux premiers axes  $u_1$  et  $u_2$  rend maximale l'inertie projetée sur ce plan. Il en est de même pour le sous-espace engendré par les trois premiers axes, etc.

### 3.5.2.4 Le nuage des variables, les composantes principales

Nous avons vu qu'à propos des variables, on cherche à évaluer leurs *liaisons*. Elles sont mesurées par le coefficient de corrélation linéaire, soit :

$$\text{corrélacion}(j, k) = \frac{1}{I} \sum_{i \in I} \left( \frac{x_{ij} - \bar{x}_j}{s_j} \right) \left( \frac{x_{ik} - \bar{x}_k}{s_k} \right)$$

avec  $\bar{x}_j, \bar{x}_k$  les moyennes et  $s_j, s_k$  les écart-types des variables  $j$  et  $k$  respectivement. Les variables sont centrées et réduites. Une variable peut être représentée comme un vecteur de l'espace vectoriel à  $I$  dimensions, dont chaque dimension représente un individu. L'ensemble des extrémités des vecteurs représentant les variables constitue le *nuage des variables*  $N_K$ . Chaque vecteur représentant une variable a pour norme 1. Soit :

$$\|\text{variable } k\|^2 = \sum_{i=1}^I \left( \frac{x_{ik} - \bar{x}_k}{s_k} \right)^2 = 1$$

Le nuage est donc situé sur une sphère de rayon 1 (on dit aussi hypersphère pour rappeler que les vecteurs variables appartiennent à un espace vectoriel supérieur à 3). Le cosinus de l'angle formé par les vecteurs représentant les deux variables  $j$  et  $k$  est obtenu en calculant le produit scalaire entre ces deux vecteurs et est égal au coefficient de corrélation entre ces deux variables :

$$\text{cosinus}(j, k) = \sum_i \frac{1}{I} \left( \frac{x_{ij} - \bar{x}_j}{s_j} \right) \left( \frac{x_{ik} - \bar{x}_k}{s_k} \right) = \text{corrélacion}(j, k)$$

L'interprétation du coefficient de corrélation comme un cosinus est une propriété très importante puisqu'elle donne un support géométrique, donc visuel, au coefficient de corrélation. De plus, la longueur des vecteurs représentant les variables étant égale à 1, la coordonnée de la projection d'une variable sur une autre s'interprète comme un coefficient de corrélation.

L'étude directe des angles entre les vecteurs étant impossible du fait de la dimension des vecteurs des variables supérieure à 3, l'intérêt de l'ACP est de fournir des variables synthétiques qui construisent un résumé de l'ensemble des variables initiales, et sont la base d'une représentation plane approchée des variables et de leurs angles.

Pour obtenir ces variables synthétiques  $\{v_s; s = 1, \dots, S\}$ , l'ACP applique au nuage des variables la même démarche qu'au nuage des individus. Ainsi, c'est encore le critère de l'inertie projetée maximale qui doit être satisfait dans le choix des axes, mais cette fois ce sont les angles entre les vecteurs représentant les variables qui doivent être peu déformés par les projections et non pas les distances entre les points du nuage.

De même que pour les projections entre variables initiales, la projection d'une variable sur une variable synthétique  $v_j$  est égale à son coefficient de corrélation avec cette variable. En effet, le plan de deux variables synthétiques  $(v_1, v_2)$  en maximisant l'inertie à l'origine du nuage projeté, rend maximale la somme des cosinus carrés des angles entre les vecteurs et leur projection, ajuste les vecteurs et déforme donc le moins possible leurs angles. Cette opération équivaut à rechercher la combinaison linéaire la plus liée à l'ensemble des variables. En ce sens  $v_j$  est la variable qui synthétise le mieux l'ensemble des variables initiales. Les axes factoriels étant orthogonaux deux à deux, on met en évidence une suite de variables synthétiques, *les composantes principales*, non corrélées entre elles, qui résument l'ensemble des variables initiales. Une composante principale peut être considérée ainsi comme la représentante (la synthèse) d'un groupe de variables liées entre elles.

### 3.5.2.5 Relations de dualité

Nous avons vu que le tableau initial une fois centré et réduit peut être considéré, à travers ses lignes comme un nuage d'individus, ou à travers ses colonnes comme un nuage de variables. Cette double représentation du même tableau conduit à des relations fortes appelées relations de dualité :

- i) *L'inertie totale* des deux nuages est la même et égale au nombre de variables (lorsque celles ci sont réduites).

$$\text{Inertie totale de } N_I = \text{Inertie totale de } N_K = \frac{1}{I} \sum_i \sum_k \left( \frac{x_{ik} - \bar{x}_k}{s_k} \right)^2 = K$$

La projection de chacun des deux nuages sur une suite d'axes orthogonaux correspond à une décomposition de l'inertie totale. Les deux décompositions sont identiques : les inerties des nuages projetés sur les axes factoriels de même rang sont égales.

- ii) *Facteurs* : l'ensemble des projections de tous les points du nuage d'individus sur son premier axe factoriel est appelé premier facteur sur les individus et constitue une nouvelle variable. Cette variable se confond avec la première composante principale obtenue dans la projection du nuage des variables. De la même façon, les projections des individus sur les autres axes factoriels s'identifient aux composantes principales de même rang.

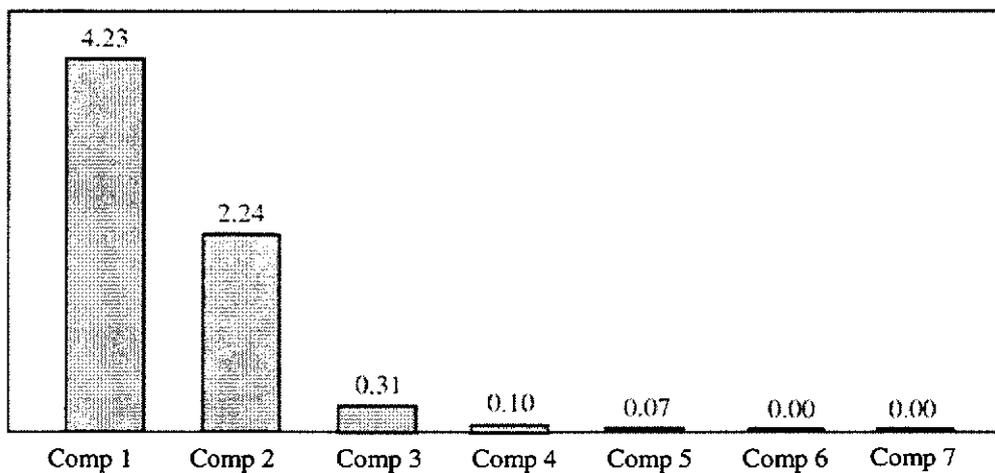
De manière analogue, la projection des  $K$  variables sur le premier axe factoriel  $v_1$  de leur nuage, définit des valeurs pour chacune des  $K$  variables, et donc un individu nouveau, qui caractérise le facteur. Cette notion d'individu "type" est réciproque à celle de la composante principale, bien que beaucoup moins classique.

### 3.5.2.6 Valeurs propres, qualité de représentation, représentation simultanée

#### Valeurs propres

L'inertie associée à un facteur est appelée *valeur propre*. Les facteurs sont appréhendés dans l'ordre décroissant de leurs valeurs propres (ou de leur inertie) ; le facteur d'ordre  $S$  ( $S > 1$ ) traduit les tendances résiduelles non prises en compte par les facteurs précédents. Plus la valeur propre d'un facteur est grande plus elle résume des variables et plus le facteur risque d'être intéressant.

Le *diagramme des valeurs propres* (figure 3.18) permet d'étudier l'allure de cette décroissance afin de retenir les facteurs les plus importants sur la base du principe suivant : si deux facteurs sont associés à des valeurs propres presque égales, ils représentent la même part de variabilité et il n'y a pas lieu a priori de retenir l'un et pas l'autre dans l'interprétation. Réciproquement, une forte décroissance entre deux valeurs propres successives incite à retenir dans l'interprétation les facteurs précédant cette décroissance. Les facteurs suivants ne représentent que le bruit inévitable qui accompagne toute observation de nature statistique.



**Figure 3.18** : Exemple numérique arbitraire de diagramme des valeurs propres d'une analyse à six variables. Les trois premiers facteurs représentent presque 97%  $[(4,23+2,24+0,31)/7=0,97]$  de la variabilité des réponses (ou de l'inertie totale qui vaut ici 7) et sont donc les facteurs retenus.

Le pourcentage d'inertie extrait par un facteur est le rapport entre l'inertie associée au facteur (par exemple la valeur propre) et l'inertie totale du nuage étudié, cette dernière étant égale au nombre de variables. Il est souvent utilisé sous la forme cumulée qui indique le pourcentage d'inertie extrait par les  $S$  premiers facteurs. Ainsi, pour l'exemple numérique arbitraire de la figure 3.18, le pourcentage d'inertie pour chaque composante est donné dans le **tableau 3.5**.

<i>Tableau des inerties</i>						
<i>Composante 1</i>	<i>Composante 2</i>	<i>Composante 3</i>	<i>Composante 4</i>	<i>Composante 5</i>	<i>Composante 6</i>	<i>Composante 7</i>
60%	32%	4,5%	1,5%	1%	≈0%	≈0%
(60%)	(92%)	(97%)	(98,5%)	(98%)	(=100%)	(100%)

**Tableau 3.5 :** Tableau des inerties associées aux composantes principales (inertie indépendamment calculée pour chacune des composantes et inertie cumulée).

### *Qualité de représentation*

Les axes factoriels fournissent des images approchées d'un nuage de points. Il est donc nécessaire de pouvoir mesurer la qualité de l'approximation tant pour chacun des points que pour l'ensemble du nuage. La qualité de la représentation d'un individu ou d'une variable par un axe est donnée par le rapport :

$$\frac{\text{Inertie de la projection de l'élément (individu ou variable) sur l'axe}}{\text{Inertie totale de l'élément}}$$

Cette définition se généralise au cas d'un plan. En outre, du fait de l'orthogonalité des axes factoriels, la qualité de représentation d'un élément par un plan (défini par exemple par les axes  $s$  et  $t$ ) est la somme des qualités de représentation de cet élément par l'axe  $s$  et l'axe  $t$ .

En ce qui concerne la qualité de la représentation d'un nuage par un axe, la définition précédente se généralise également à l'ensemble d'un nuage par le rapport :

$$\frac{\text{Inertie de la projection du nuage sur l'axe}}{\text{Inertie totale du nuage}}$$

Cet indicateur, appelé pourcentage d'inertie associée à un axe, mesure l'importance d'un axe factoriel dans la variabilité des données. Comme dans le cas d'un seul élément, ces pourcentages peuvent être cumulés sur plusieurs axes, on parle alors du pourcentage d'inertie extrait par un plan ou par les  $S$  premiers axes<sup>1</sup> (voir tableaux 3.5 et 3.6).

Dans le cadre de ce travail, c'est la qualité de la représentation du nuage des individus par un axe – ou par les  $S$  premiers axes pour la variance cumulative – et la qualité de la représentation des variables par un axe, qui nous intéressent. Des tableaux du type du **tableau 3.6** sont ainsi présentés aux sections des résultats des différentes expériences. Les qualités de la représentation sont présentées uniquement pour les  $S$  premières composantes principales regroupant la plus grande partie de la variance.

<sup>1</sup> Ce qui correspond à la représentation géométrique de la variance expliquée par un plan ou les  $S$  premiers axes.

i)		Composante 1	Composante 2	Composante 3
Proportion de Variance		0.60	0.32	0.045
Variance Cumulative		0.60	0.92	0.97

ii)		Composante 1	Composante 2	Composante 3
Variables				
<i>force</i>		0.40	0.35	-0.29
<i>précision</i>		0.47		0.34
<i>réverbérance</i>		-0.45	0.24	0.16
<i>vivacité</i>		-0.20	0.58	0.45
<i>distance</i>		-0.47	0.10	0.28
<i>enveloppement</i>		-0.11	0.61	0.51
<i>impression générale</i>		0.39	0.33	0.48

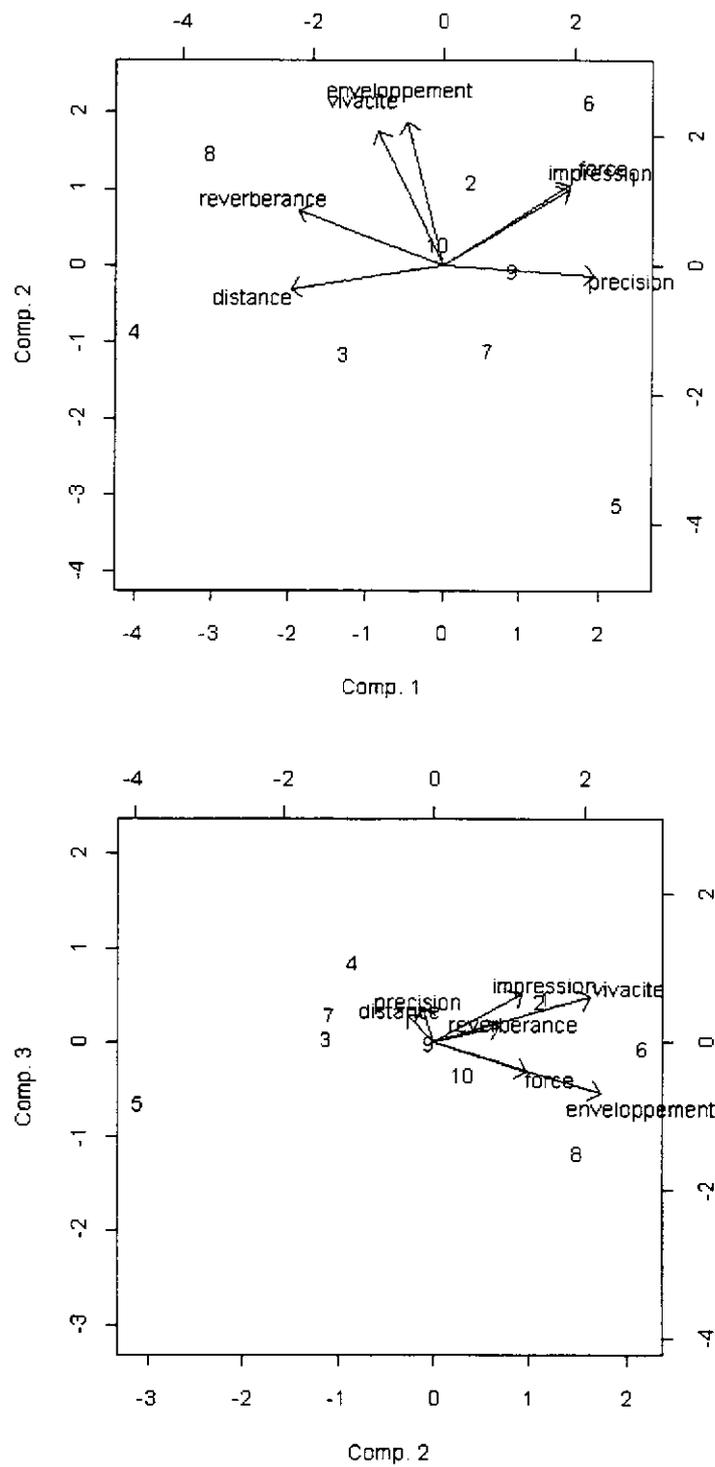
**Tableau 3.6 :** Exemple numérique des résultats d'une ACP (*partie unimodale* de l'expérience principale, § 5). Ici, les trois premières composantes principales expliquent 97% de la variance ( $S = 3$ ) et sont donc retenues. Nous observons : i) la qualité de la représentation du nuage des individus par les trois premières composantes principales (séparées et cumulatives), et ii) la qualité de représentation des variables initiales sur ces mêmes composantes.

### Représentation graphique simultanée

Les facteurs issus d'une analyse peuvent être étudiés séparément ou deux par deux à l'aide des plans factoriels. Du fait des relations de dualité entre les représentations des individus et des variables, les interprétations des axes factoriels doivent être menées simultanément. Cette nécessité d'interprétation conjointe conduit souvent à la superposition des deux nuages, le nuage des individus et le nuage des variables, ce qui entraîne toutefois le problème de la représentation sur un même graphique de points de nature différente, évoluant dans des espaces différents. La représentation simultanée peut être utilisée sans problème à condition de respecter les deux règles suivantes :

- Il ne faut pas interpréter la position d'un individu par rapport à une seule variable et réciproquement.
- Ce n'est pas la proximité entre un individu et un ensemble de points représentant des variables qui est importante, mais l'éloignement de l'individu dans la direction de cet ensemble de variables.

Dans l'exemple des plans factoriels 1-2 et 2-3 de la **figure 3.19** (correspondant aux résultats du tableau 3.6), nous pouvons observer par la longueur des vecteurs des variables qu'elles sont plutôt bien représentées sur le plan 1-2. En revanche, sur le plan 2-3 la plupart d'entre elles sont mal représentées (la longueur des vecteurs est très petite). On observe également que les variables *précision* et *distance* sont très corrélées au premier facteur (composante principale 1) tandis que les variables *enveloppement* et *vivacité* sont très corrélées au deuxième facteur (composante principale 2). A propos des individus, nous pouvons observer que les individus 8, 2 et 6 varient de manière significative le long de l'axe de la variable *distance* (les projections sont visiblement croissantes) tandis qu'ils varient beaucoup moins par rapport à *l'enveloppement* par exemple.



**Figure 3.19 :** Exemple de représentation simultanée de deux nuages sur le premier et sur le deuxième plans factoriels. Les variables sont représentées par des vecteurs et les individus par des chiffres.



## 4 EXPERIENCES SUR LA DISTANCE AUDITIVE

### 4.1 INTRODUCTION

L'influence de la distance visuelle sur la distance auditive a été étudiée à l'aide d'une série d'expériences : une expérience préliminaire et six expériences principales présentées dans le tableau ci-après (**tableau 4.1**).

L'objectif de l'expérience préliminaire était de tester un certain nombre de choix expérimentaux (méthode de diffusion sonore, nombre de présentations des stimuli, etc.); elle a également permis l'observation des premiers effets indicatifs de l'influence visuelle sur la distance auditive.

L'objectif des expériences principales était multiple :

- définir une méthode de simulation qui permette de créer des conditions visuelles et sonores les plus 'crédibles' possible,
- tester et établir une procédure expérimentale adaptée pour étudier les interactions audiovisuelles,
- permettre d'observer l'influence des informations visuelles sur la perception de la distance auditive et essayer de la quantifier, c'est-à-dire tenter de lier les effets perçus à des distances visuelles relatives.

Les différentes études sur les choix expérimentaux effectuées à l'aide des expériences 1, 2, 3 et 4 (comme la méthode de diffusion du son, la nature du signal et les configurations sonores, le nombre de présentations des stimuli, la méthode de projection des images, etc.) sont mis en œuvre dans le cadre des deux dernières expériences 5 et 6, expériences qui servent également pour la validation des effets observés.

Ces dernières expériences adoptent deux méthodes de présentation des stimuli : une méthode avec *vues statiques* et une méthode avec *vues changeantes* (cf. § 4.2.6). Leur étude parallèle permet de simuler et de comparer des contextes visuels différents : des conditions visuelles présentes dans les salles de concert d'une part, et d'autre part, des conditions visuelles évolutives telles qu'en environnements virtuels.

Etant donné le grand nombre de paramètres expérimentaux à tester, le nombre de sujets qui ont participé à certaines expériences est parfois limité. Ces tests permettent néanmoins d'observer les effets marquants. Les deux derniers tests – qui ont bénéficié d'un nombre plus élevé de sujets – ont par ailleurs permis de valider ces résultats.

Les expériences sur la distance auditive présentent plusieurs aspects communs en ce qui concerne la procédure expérimentale<sup>1</sup>. Cette procédure générale est présentée au paragraphe suivant, les aspects spécifiques de chaque test sont ensuite décrits dans les paragraphes correspondants.

Type et N° de test	Objectif	Son (stimuli, signal, diffusion)	Images, mode de projection	Nombre de présentations de stimuli	Méthode de présentation des vues	Nb. de sujets
Préliminaire	Choix expérimentaux	A1-A6, S.Vega, dif. binaurale	Théâtre de Louvrais, 3-D	3	statiques	12
Principal auditif 1	Etude des signaux et configurations	A1-A11, S.Vega et opéra, transaurale	---	1 et 2 (6 et 8 sujets)	---	14
Principal 2	Etude du nombre de présentations	A1-A6, opéra, dif. transaurale	(V0,V1,V3), TCE, 3-D	7	statiques	12
Principal 3	Comparaison 3-D et 2-D	A1-A6, opéra, dif. transaurale	(V0,V1,V3), TCE, 2-D	2	statiques	16
Principal 4	Comparaison 3-D et 2-D	A1-A6, opéra, dif. transaurale	(V1,V2,V3), TCE, 2-D	2	changeantes	16
Principal 5	Validation. Influence visuelle.	A1-A6, opéra, dif. transaurale	(V0,V1,V3), TCE, 3-D	2	statiques	28
Principal 6	Validation. Influence visuelle.	A1-A6, opéra, dif. transaurale	(V1,V2,V3), TCE, 3-D	2	changeantes	28

**Tableau 4.1** : Liste des expériences sur la distance auditive (TCE : Théâtre des Champs Elysées).

<sup>1</sup> Mais aussi en ce qui concerne la méthode de simulation et les méthodes d'analyse.

## 4.2 PROCEDURE EXPERIMENTALE GENERALE

La méthodologie utilisée dans les expériences portant sur la perception de la distance est largement empruntée à la psychologie expérimentale et notamment au domaine traitant des interactions sensorielles. Les stimuli auditifs simulant différentes distances auditives sont accompagnés d'images qui – elles – représentent différentes distances visuelles. La tâche des sujets consiste à *juger la distance auditive* et à répondre oralement à l'aide d'une échelle numérique représentant cet attribut.

### 4.2.1 Réglage subjectif du dispositif transaural.

Dans le cas de tests impliquant une simulation sonore transaurale, une vérification subjective du réglage du dispositif stéréophonique a été effectuée pour chacun des sujets : 4 sons provenant de différentes directions ont été présentés afin de contrôler la bonne position de la tête des auditeurs. Ainsi – considérant que la direction frontale correspond à un angle de  $0^\circ$  – des sons des directions de  $-30^\circ$ ,  $-80^\circ$ ,  $+30^\circ$  et  $+80^\circ$  ont été présentés aux sujets. Lorsque ces derniers les entendaient provenir (à peu de choses près) de ces directions, les tests pouvaient commencer ; dans le cas contraire, des réglages fins étaient effectués au niveau de la position du siège jusqu'à ce que ces effets spatiaux soient perçus. Tous les sujets sans exception ont pu les percevoir.

Au terme de cette séance de réglage, les sujets étaient priés de ne plus bouger la tête pendant la durée des expériences, cette consigne fut respectée. Tous les sons émis lors des tests étaient frontaux (direction  $0^\circ$ ).

### 4.2.2 Consignes

Les consignes pour les tests étaient données oralement (une feuille de consignes écrites a été fournie au tout premier test préliminaire, mais l'expérience a montré que les consignes orales étaient mieux appréciées et plus efficaces). Elles se résument par les indications suivantes :

*« Vous allez écouter des sons et voir des images spécialement conçus pour simuler les conditions sonores et visuelles d'une salle de concert. Il s'agit de la musique de... et de la salle de... »*

*Vous avez deux tâches à réaliser :*

- *D'abord, utilisez le son et l'image afin d'essayer de vous imaginer au mieux dans une vraie salle de concert.*
- *Ensuite, essayez de juger la distance auditive apparente, c'est à dire : avez vous l'impression que la source sonore est proche ou éloignée par rapport à vous. Vous répondrez oralement, de préférence dans une échelle arbitraire comprise entre 1 et 10 ; 1 représente la distance auditive la plus courte et 10 la distance auditive la plus grande.*

*Vous avez le droit de vous tromper : c'est à dire que si par exemple, vous avez utilisé la note 9 pour un son très lointain et que le son suivant vous semble encore plus lointain, vous pouvez utiliser une note supérieure, par exemple 11 ou 12. Egalement, pour les sons très proches vous pouvez si vous le souhaitez utiliser par exemple les notes 0 ou -1.*

*Avant les tests, vous allez écouter les deux sons extrêmes, c'est-à-dire le son le plus proche et le son le plus lointain, ceci autant de fois que vous le souhaitez afin de saisir la différence et d'avoir une idée de la tâche demandée.*

*Soyez détendus, ne prenez pas trop de temps pour répondre, donnez simplement votre impression telle que vous la sentez. Il n'y a pas de réponse bonne ou mauvaise, toutes les réponses sont d'une valeur identique. Ce qui est important pour nous est de voir comment les gens perçoivent les sons, il ne s'agit pas de tester leurs capacités. »*

Une attention particulière a été portée lors de ces consignes pour ne pas mentionner des mot ou des phrases évoquant la simulation : écran, photos, casque, etc.

Des entretiens informels eurent lieu à la fin de chaque test concernant la tâche auditive, le rôle des images, mais aussi l'expérience d'écoute des sujets, leur culture, éducation, etc.

### 4.2.3 Référence

Deux stimuli auditifs extrêmes A1 et A6 sont présentés au début de chaque session pour référence. Les sujets pouvaient les écouter autant de fois qu'ils le souhaitaient afin de pouvoir discerner la différence et de se sentir prêts à passer aux tests. L'expérience a montré que deux présentations du couple – du type ordre A1-A6-A1-A6 – étaient suffisantes pour la plupart des sujets.

La référence servait également à orienter l'attention des sujets vers la perception auditive.

Cette présentation des stimuli extrêmes A1 et A6 avant les tests ne peut pas être considérée comme un entraînement pré-experimental, d'une part parce qu'elle n'impliquait que deux des six sons utilisés, et surtout, parce que la différence perceptive entre eux était relativement grande (la plus grande possible lors les tests) et ne représentait donc pas la finesse de jugement demandée de la part des sujets pour la suite.

### 4.2.4 Sujets

Une partie des sujets qui ont participé aux tests étaient des volontaires. Dans certains cas, ces mêmes sujets sont revenus après un certain temps (quelques semaines) et ont participé à un deuxième test. De plus, un petit nombre d'entre eux avait une expérience d'écoute. L'un des sujets était âgé de 80 ans, les autres étaient âgés entre 20 et 50 ans.

La majeure partie des sujets, et en tous cas ceux qui ont participé aux deux derniers tests principaux, étaient des étudiants de l'IUT et de l'IUP de Neuville sur Oise (génie électrique, génie civil). Ils étaient âgés de 19 à 26 ans et 40F/heure leur étaient offerts pour leur participation. Ils n'avaient aucune expérience des tests d'écoute préalable, et les audiogrammes réalisés n'ont pas montré de déficience auditive notable.

### 4.2.5 Stimuli sonores et visuels

Dans toutes les expériences sur la distance, six stimuli auditifs ont été utilisés ; ils sont appelés A1, A2, A3, A4, A5 et A6 et ils ont été créés afin de représenter des *distances auditives croissantes*.

En outre, trois images ont été utilisées, appelés V1, V2 et V3 ; elles représentent également des *distances visuelles croissantes* par rapport à la scène du théâtre visualisé. ( $V1 = V_{\text{proche}}$ ,  $V2 = V_{\text{moyen}}$  et  $V3 = V_{\text{loin}}$ ). De plus, une condition visuelle de contrôle V0 ( $V_{\text{contrôle}}$ ) a été créée : aucune image n'était projetée sur l'écran, seulement une lumière diffuse y était perçue. Un filtre de couleur orange pâle était placé sur l'un des projecteurs, l'autre restant éteint, de manière à ce que la luminosité de cette condition visuelle soit équivalente à la luminosité des conditions visuelles figuratives.

Lors des tests en *vues statiques* (voir paragraphe suivant), ce sont uniquement les conditions visuelles V0, V1 et V3 qui étaient utilisées. Lors des tests en *vues changeantes*, ce sont les conditions visuelles V1, V2 et V3 qui étaient employées.

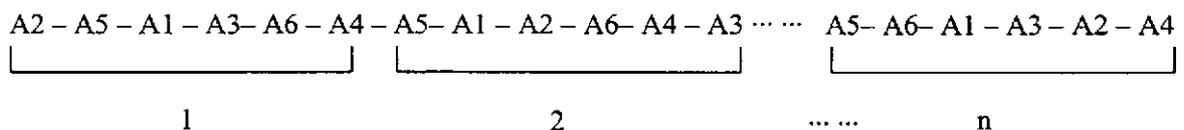
### 4.2.6 Types de tests, séquences et nombre de présentations

Deux types de tests ont été réalisés : les tests avec *vues statiques* et les tests avec *vues changeantes*. Pour chaque type de test des séquences spécifiques ont été créées.

- Tests avec vues statiques :

Lors des tests avec vues statiques (tests *statiques*), la première image (ou condition visuelle de contrôle) restait fixe sur l'écran pendant que la séquence sonore (série de stimuli auditifs) était présentée aux sujets. Quand la première présentation d'une séquence sonore était terminée, la deuxième image remplaçait la première. Cette même séquence était alors présentée pour la deuxième fois. En tout, la séquence était présentée trois fois à chaque sujet pour chacune des trois conditions visuelles (deux images et une condition de contrôle).

Les séquences utilisées pour les tests *en vues statiques* étaient des séquences sonores : les six stimuli auditifs sont présentés  $n$  fois dans un ordre pseudo-aléatoire créant ainsi des séquences de  $n \times 6$  stimuli (**figure 4.1**). Pour le test préliminaires,  $n$  était égal à 3 ; Pour les tests principaux,  $n$  était égal à 7 ou à 2.



**Figure 4.1:** Séquence sonore typique pour les tests *en vues statiques* ( $6 \times n$  stimuli sonores). Des séquences de ce type étaient présentées trois fois à chaque sujet, une fois pour chaque condition visuelle.



D'autre part, maintenir la tête immobile par rapport aux haut-parleurs est une condition importante pour la bonne restitution des champs sonores en diffusion transaurale (cf. § 3.2.2). De plus, il a été supposé que le fait de ne pas trop bouger la tête et de garder les yeux fixés sur l'écran, pouvait aider les sujets à mieux s'imaginer dans la salle simulée.

Les questionnaires portant sur la qualité acoustique des salles utilisent dans leur grande majorité des échelles sémantiques bipolaires (du type fort-faible, confus-précis, etc.). Une telle échelle a été choisie pour ces expériences ; les sujets avaient pour consigne de juger la distance auditive et de répondre sur une échelle comprise entre 1 et 10<sup>1</sup>. Cette échelle était suggérée aux sujets, mais elle n'était pas imposée (cf. *Consignes* à la même section), elle a été choisie de façon à ce que la plupart des personnes se sentent à l'aise pour l'utiliser.

Toutes les réponses ont été normalisées afin de minimiser les problèmes liés aux échelles propres que les sujets pouvaient utiliser (par exemple un sujet a répondu entre 1 et 6, un autre entre 5 et 11, etc.). Pour chaque personne, une valeur moyenne des réponses était imposée à 0, ainsi qu'un écart-type égal à 1. Ainsi, si  $x_{11}, x_{21}, x_{31}, \dots, x_{n1}$  sont les  $n$  réponses du sujet 1, et  $\bar{x}_1$  et  $\sigma_1$  sont respectivement la moyenne et l'écart-type des ces réponses, la valeur normalisée de la valeur  $x_{21}$  par exemple, est donnée par :

$$x_{21\text{norm}} = (x_{21} - \bar{x}_1) / \sigma_1$$

Ensuite, les valeurs obtenues sont recentrées autour d'une valeur moyenne  $\bar{x}$  et d'un écart-type  $\sigma$ , ce sont des valeurs arbitraires et communes pour chaque sujet. Par exemple, la nouvelle valeur  $x_{21\text{nouv}}$  pour la valeur initiale normalisée  $x_{21\text{norm}}$ , est donnée par :

$$x_{21\text{nouv}} = (x_{21\text{norm}} + \bar{x}) \cdot \sigma$$

La valeur moyenne utilisée était de 4.5 (elle correspond à peu près à la moyenne globale de l'ensemble des réponses) et l'écart-type imposé était de 3. De cette manière, les valeurs exploitables obtenues étaient ramenées à une échelle qui correspond à peu près à celle qui était suggérée au début des tests.

Des entretiens informels ont été réalisés avec les sujets après la fin des tests. Leurs réponses, ainsi que les remarques spontanées des sujets, servent à l'interprétation des résultats.

## 4.2.9 Durée des tests

Les sujets disposaient d'autant de temps qu'ils le souhaitent pour répondre. A titre d'exemple, pour répondre à 36 stimuli (par exemple, 3 séquences de 2 x 6 stimuli pour les tests en vues statiques, ou 2 séquences de 18 stimuli pour les tests à vues changeantes, le nombre  $n$  de présentations des stimuli étant de 2 dans les deux cas), la durée moyenne des tests était de l'ordre de 25 à 30 min. ; il n'y avait pas de grande variabilité sur cette durée.

Il faut noter ici que le temps nécessaire pour passer d'un son à l'autre (sur une cassette DAT) était relativement long, ce qui pénalisait la durée globale des tests. Un intervalle d'environ 8 à 10 s était nécessaire entre la fin de l'émission d'un son et le début du son suivant. Ce retard avait pourtant l'avantage d'éviter des réponses successives trop immédiates et donc des réponses qui, par comparaison, seraient complètement dépendantes du son qui a précédé.

<sup>1</sup> Exceptionnellement pour le test préliminaire l'échelle proposée était de 0 à 9.

## 4.3 EXPERIENCE PRELIMINAIRE

- *Test à écoute binaurale - Vues statiques*
- *3 présentations de stimuli*
- *Signal sonore : «Tom's dinner » de Suzanne Vega  
(Stimuli sonores : A1, A2, A3, A4, A5 et A6).*
- *Vues : Théâtre de Louvrais à Pontoise  
(Stimuli visuels : condition visuelle de contrôle V0, images 3-D : V1 et V3).*
- *12 sujets*

### 4.3.1 Objectif

L'expérience préliminaire a été réalisée afin de clarifier un certain nombre de choix expérimentaux (nombre des présentations des stimuli, mode de diffusion sonore, échelle de réponses, etc.). Elle a permis également d'observer les premiers effets indicatifs de l'influence visuelle sur la distance auditive [Nathanail et al., 1996 ; 1997a ; 1997b ; 1997c].

### 4.3.2 Méthode

1. *Stimuli sonores.* Le champ sonore est créé par le *Spatialisateur*, un processeur d'acoustique virtuelle permettant de reconstituer un effet de salle artificiel (cf. § 3.2). Le *Spatialisateur* a permis de simuler deux réponses d'une salle de conférence de l'Université de Cergy<sup>1</sup> et, à partir de ces deux réponses, de créer quatre autres réponses en modifiant essentiellement le niveau sonore du champ direct tandis que le niveau du champ réverbéré restait constant (cf. § 3.2.5). Pour cette expérience préliminaire, ces réponses ont été créées de manière arbitraire<sup>2</sup> afin que – après avoir été appliquées à un extrait de chanson (“Tom's Dinner” de Suzanne Vega) – elles puissent donner subjectivement l'impression d'une distance auditive croissante : A1, A2, A3, A4, A5 et A6. Le son est enregistré et diffusé sur DAT (Digital Audio Tape) via un casque audio.

---

<sup>1</sup> Les réponses impulsionnelles du théâtre de Louvrais n'ont pas pu être obtenues.

<sup>2</sup> Les deux réponses impulsionnelles mesurées dans la salle de conférence de l'Université de Cergy et simulées par le *Spatialisateur* ont été appelées A3 et A4 : elles représentaient des distances moyennes dans cette salle (le signal A3 était mesuré à environ 8 m du haut-parleur placé sur la scène, plus proche que le signal A4 qui lui était mesuré à environ 12 m). Les signaux A2 et A1 ont été créés à partir de A3, respectivement en augmentant de +3 dB et +6 dB l'énergie du son direct (OD). A5 et A6 ont été créés à partir de A4, respectivement en abaissant de -3 dB l'énergie du son direct (OD) et en diminuant de -3 dB le rapport entre son direct et son réverbéré (Es). Un filtrage progressif des hautes fréquences a été effectué avec l'augmentation de la distance. Enfin, R1, R2 et R3 ont été maintenus constants (cf. § 3.2, tableau 3.1, figure 3.10).

2. *Conditions visuelles.* Le champ visuel est créé par la projection de grandes images tridimensionnelles, obtenues par polarisation de la lumière et projetées depuis l'arrière sur grand écran dans des conditions d'obscurité complète (cf. § 3.3). Les photographies ont été prises dans le *Théâtre de Pontoise*, en des emplacements de différentes distances par rapport à la scène, au milieu de laquelle était placé un haut-parleur. Les vues V1 et V3 utilisées dans les tests, sont prises aux distances respectives de 6 m et 18 m<sup>1</sup>. Pour la condition visuelle de contrôle V0, aucune image n'était projetée sur l'écran mais une lumière diffuse y était perçue (cf. § 4.2). Durant les tests, les sujets étaient assis dans des sièges du même type que ceux de la salle de spectacle visualisée, à une distance de 1,6 m de l'écran.

3. *Sujets, consignes.* Douze adultes ont participé à cette expérience. La tâche des sujets consistait à juger la distance auditive et à répondre oralement dans une échelle comprise entre 0 et 9 (0 représentant la distance auditive la plus courte et 9 la plus grande). Ils avaient pour consigne d'utiliser le son et l'image afin de s'imaginer dans la "salle", mais de ne juger que la distance auditive. Les stimuli extrêmes (A1 et A6) sont présentés au début de chaque session pour référence. Les consignes et l'écoute des sons de référence sont décrits en détail dans la section précédente (cf. § 4.2.2 et 4.2.3).

4. *Procédure expérimentale* Les six stimuli auditifs sont présentés trois fois (présentations P1, P2 et P3) dans un ordre pseudo-aléatoire créant ainsi des séquences de dix-huit stimuli (cf. figure 4.1 pour  $n=3$ , § 4.2). Afin de compenser les effets d'ordre de présentation des stimuli, quatre séquences sont créées, chaque séquence a été présentée à quatre sujets différents (12 sujets). Les quatre séquences utilisées sont présentées en annexe B.

Chaque sujet subit la même séquence trois fois pour chacune des trois conditions visuelles :

- Condition visuelle de contrôle V0 ou  $V_{\text{contrôle}}$ .
- Vue proche de la scène V1 ou  $V_{\text{proche}}$ .
- Vue lointaine de la scène V3 ou  $V_{\text{loin}}$ .

Chaque sujet participait ainsi en tout à trois tests de 18 stimuli chacun ( $18 \times 3 = 54$  réponses). En premier lieu, tous les sujets participaient à la condition visuelle V0, puis une moitié participait à la condition visuelle V1 et l'autre à la condition visuelle V3; chaque moitié subissait enfin la condition restante.

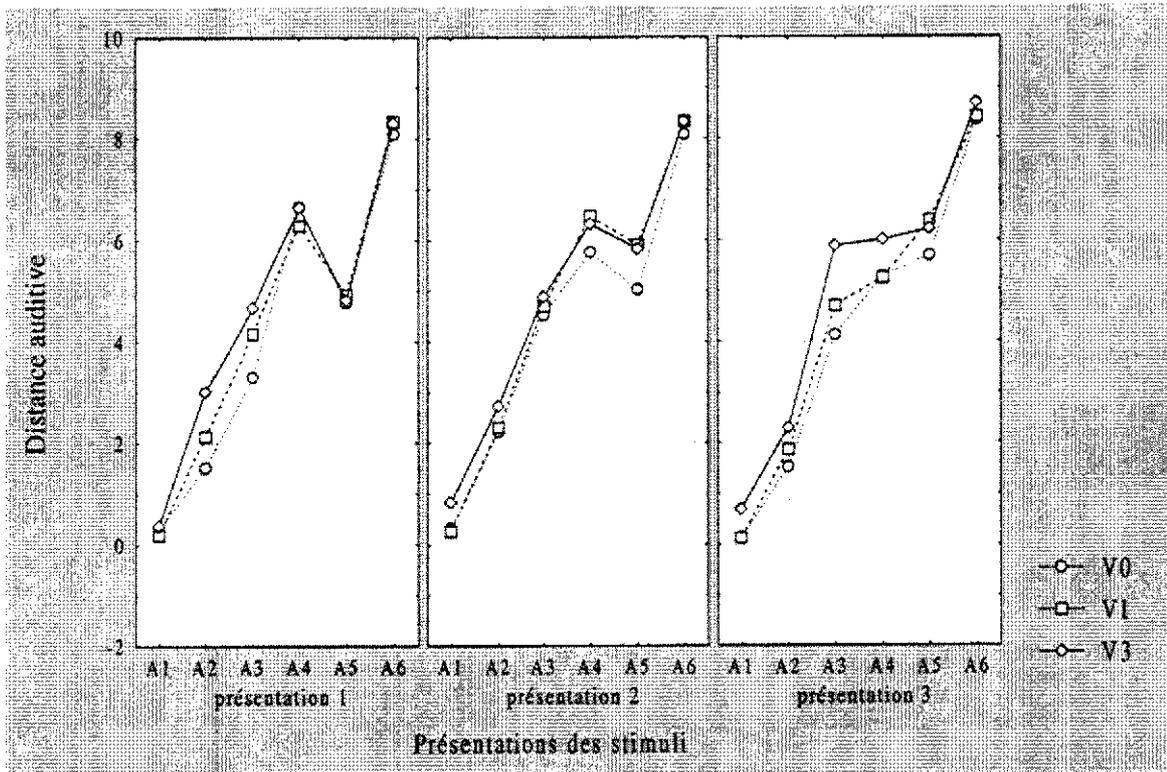
La *durée* moyenne des tests était de 35 à 45 min. (12 à 15 min. par condition visuelle).

### 4.3.3 Résultats

Les réponses des sujets sur la distance auditive sont normalisées comme décrit au § 4.2.8 ( $\bar{x} = 4,5$  et  $\sigma = 3$ ). Les moyennes et les écart-types des réponses sur les 12 sujets pour chacun des stimuli auditifs, les conditions visuelles et les présentations de stimuli sont présentés à l'annexe B.

La *distance apparente auditive* pour chacun des stimuli auditifs et en fonction des conditions visuelles V0, V1 et V3 et de 3 présentations des stimuli, est présentée sur la **figure 4.3**.

<sup>1</sup> La distance focale était de 28mm afin d'assurer un large champ visuel.



**Figure 4.3 :** Distance apparente auditive des stimuli sonores en fonction des conditions visuelles V0, V1 et V3 ( $V0 = V_{\text{contrôle}}$ ,  $V1 = V_{\text{proche}}$  et  $V3 = V_{\text{loin}}$ ). Les trois graphes correspondent aux trois présentations de stimuli P1, P2 et P3.

Nous remarquons que les stimuli auditifs, créés de manière à donner des impressions successives des distances auditives, semblent avoir été perçus effectivement comme tels : A1 est perçu plus proche que A2, A2 est perçu plus proche que A3, etc. Une inversion perceptive est pourtant observée entre les stimuli A4 et A5 (A5 semble plus proche que A4) ; elle pourrait être liée à la création quelque peu arbitraire des champs sonores pour cette expérience préliminaire. Un regard plus attentif des résultats pour la 3<sup>ème</sup> présentation des stimuli montre pourtant que cette inversion semble peu pertinente, en effet, A4 est perçu plus proche que A5 à la 3<sup>ème</sup> présentation.

Les conditions visuelles semblent avoir un effet sur la perception de la distance auditive, surtout à la 3<sup>ème</sup> présentation des stimuli : les stimuli semblent être jugés plus lointains quand ils sont présentés sous la condition visuelle V3 et plus proches sous la condition visuelle V0.

Afin de tester la validité de ces observations, une *Analyse de Variance* (ANOVA) à trois facteurs avec des *mesures répétées* est réalisée sur les réponses des sujets. L'ANOVA permet de décomposer la variabilité totale d'une variable dépendante, ici la *distance auditive*, en contributions partielles des différents facteurs. Les trois facteurs de cette expérience étaient : les *conditions visuelles* (3 niveaux), les *présentations des stimuli* (3 niveaux) et les *stimuli auditifs* (6 niveaux). Le **tableau 4.2** présente les résultats de cette analyse.

On observe que l'analyse révèle effectivement un effet principal des *stimuli auditifs* [ $F(5,55) = 248.77$ ,  $p < 0.01$ ]. Des comparaisons Post Hoc (Tukey HSD test) entre les stimuli auditifs révèlent des différences significatives (au niveau de significativité de 0.05) entre tous les stimuli (sauf entre les

stimuli A4 et A5). Les différentes distances auditives des stimuli sont donc, d'une manière générale, bien distinguées entre elles.

Table de tous les effets						
1 : Conditions Visuelles, 2 : Présentations de stimuli, 3 : Stimuli Auditifs						
	df Effet	MS Effet	df Erreur	MS Erreur	F	Niveau-p
1	2	19.80	22	5.34	3.70	* 0.040
2	2	3.28	22	5.55	0.59	0.561
3	5	877.72	55	3.52	248.77	* 0.000
12	4	0.69	44	2.40	0.29	0.882
13	10	1.61	110	1.01	1.58	0.118
23	10	6.40	110	2.78	2.30	* 0.017
123	20	0.82	220	1.05	0.78	0.727

**Tableau 4.2:** Tableau des résultats de l'Analyse de la Variance pour l'expérience préliminaire. Les trois facteurs sont : les *conditions visuelles* (facteur 1), les *présentations des stimuli* (facteur 2) et les *stimuli auditifs* (facteur 3). 12, 13, 23 et 123 représentent les interactions entre les facteurs respectifs : 12 est par exemple l'interaction entre les facteurs 1 et 2, 123 l'interaction entre les facteurs 1, 2 et 3, etc. Les effets significatifs sont signalés par un astérisque (\*).

Les *conditions visuelles* semblent avoir également un effet significatif [ $F(2,22) = 3.71$ ,  $p < 0.05$ ] ; ceci indique que les situations visuelles influencent les réponses auditives. Afin d'étudier les différences entre conditions visuelles, des comparaisons Post Hoc (Tukey HSD test) ont été réalisées : elles révèlent une différence significative entre les conditions V0 et V3 [ $p < 0.05$ ] mais ne montrent pas de telles différences entre les conditions V0 et V1 ou V1 et V3<sup>1</sup>.

La présentation n'a pas d'effet principal mais son interaction avec les stimuli auditifs est significative [ $F(10,110) = 2.30$ ,  $p < 0.05$ ]. Ce dernier résultat montre que les stimuli auditifs ne sont pas jugés de la même façon dans les trois présentations. Cet effet pourrait être partiellement dû à l'inversion initiale des stimuli A4 et A5 qui se rétablit à la 3<sup>ème</sup> présentation, mais il nous incite de toute manière à étudier le rôle des présentations de stimuli.

La **figure 4.4** représente les réponses moyennes de distance auditive pour chacun des stimuli auditifs, chacune des présentations et des conditions visuelles. On observe que les réponses augmentent de

<sup>1</sup> Les différences entre stimuli pris indépendamment, par exemple entre les stimuli A4V0 et A4V3, ou même entre ces mêmes stimuli pour toutes les présentations, ne sont pas significatives à elles seules (une telle différence pourrait par exemple être testée avec un t-test entre deux échantillons). L'ANOVA permet de tester l'effet *global* entre V0 et V3, c'est à dire pour l'ensemble des stimuli auditifs (qui dans ce cas est significatif). Il serait donc inutile, et cela dégraderait la lisibilité de la figure 4.3, de reporter les écart-types dans ce chapitre (puisque dans presque tous les cas, les courbes se trouveront *dans* les écart types). Ceci est le cas pour les autres figures de ce chapitre, les écart-types pour certains des résultats sont donc présentés en annexe B.

manière générale à la 2<sup>ème</sup> présentation des stimuli. Ceci est encore plus visible sur la **figure 4.5** qui montre les réponses moyennes de distance auditive sur les stimuli auditifs pour chacune des présentations et des conditions visuelles (il s'agit en fait d'une «moyenne» des 6 graphes de la **figure 4.4**).

Ce dernier graphe nous permet d'observer que les différences dues aux conditions visuelles semblent plus importantes pour la 3<sup>ème</sup> présentation des stimuli (P3). Une analyse prenant en compte uniquement ces dernières réponses (3<sup>ème</sup> présentation), montre à nouveau un effet principal des stimuli auditifs [ $F(5,55) = 136,76, p < 0.001$ ], et – effectivement – un effet principal plus significatif des conditions visuelles [ $F(2,22) = 5,67, p < 0.05$ ].

### 4.3.4 Discussion

Les différences observées sur la perception de la distance auditive entre les conditions visuelles V0 (condition sans image) et V3 (vue lointaine de la scène), semblent montrer que la simple présence d'une image tridimensionnelle favorise la perception de profondeur de l'espace sonore.

Pourtant, cet effet est peut être plus prononcé ici, au moins au début de chaque séance, à cause de l'utilisation du casque audio qui lui, donne généralement un espace sonore « restreint » et difficilement « extériorisé » de la tête des auditeurs. En outre, comme la condition de contrôle V0 est toujours présentée au début des tests, il est possible que les sujets répondent prudemment, dans l'expectative, et sous-estiment alors la distance auditive, résultat en accord avec d'autres travaux sur la perception de la distance (cf. § 2.2). Une telle conclusion sur l'effet « favorable » ou « facilitateur » de l'image sur la perception de la distance auditive nécessite alors une étude plus approfondie, avec une diffusion sonore améliorée et des ordres de présentation des conditions visuelles aléatoires.

Nous focalisons ensuite notre intérêt sur trois autres aspects de l'expérience : l'utilisation du *Spatialisateur* pour la simulation sonore, la diffusion binaurale sur casque audio et l'influence du nombre de présentations de stimuli.

L'utilisation du *Spatialisateur* pour la simulation des champs sonores semble adaptée pour ce type d'expérience. Celui-ci permet en effet de créer un effet de salle homogène et de donner des impressions de distances auditives « crédibles » et « discernables » entre elles (ou au moins jugées comme si elles l'étaient). Un soin particulier doit pourtant être porté sur la conception des stimuli auditifs, ceux-ci ne doivent pas être basés uniquement sur des évaluations subjectives mais également sur des indices acoustiques pertinents liés à la distance auditive (indices issus de l'interface de contrôle bas ou haut-niveau du *Spatialisateur*). La 1<sup>ère</sup> expérience étudie justement cet aspect de la simulation sonore.

Les questions posées aux sujets après les tests concernaient le sentiment « d'être dans la salle », la provenance du son et les indices sur lesquels les sujets se basent pour juger la distance auditive. Les remarques intéressantes à faire sont les suivantes :

- la moitié des sujets environ a trouvé que le son « venait du casque », ou bien qu'il était situé « au dessus » ou « derrière » la tête.

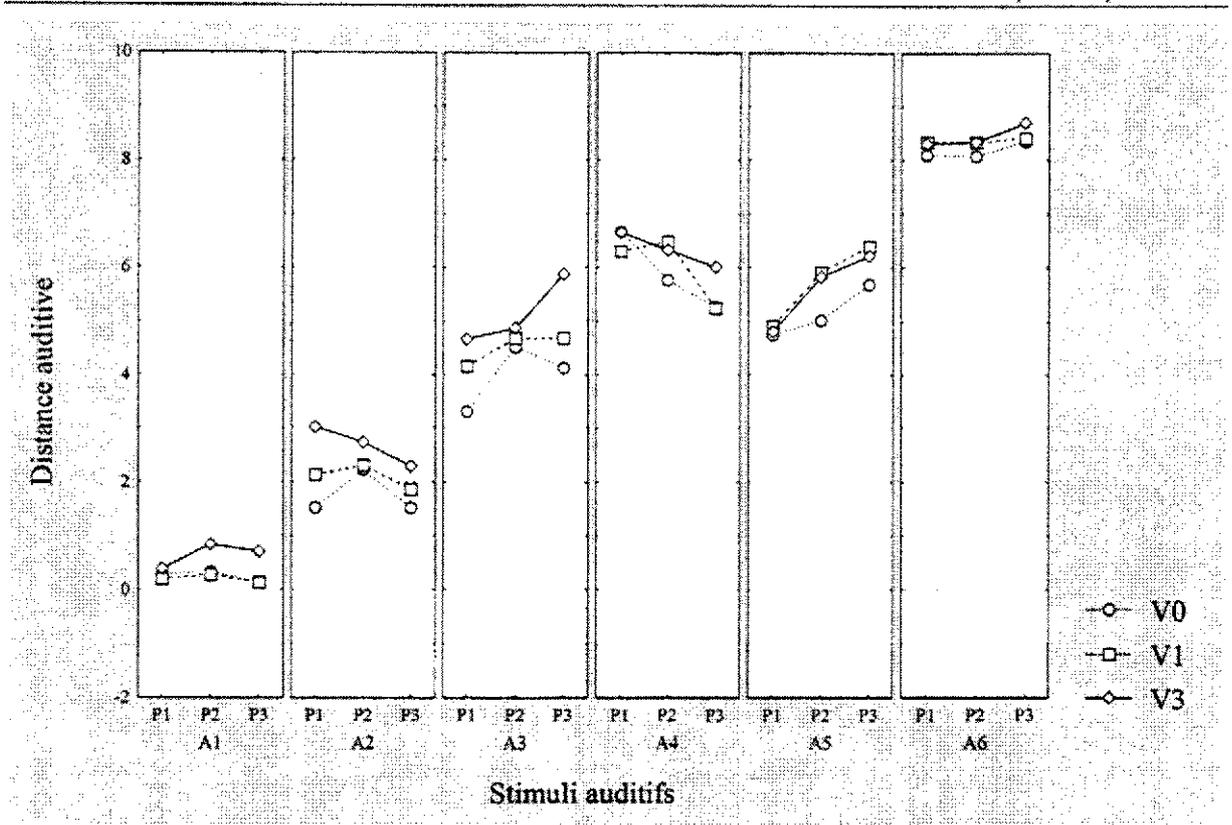


Figure 4.4 : Distance apparente auditive de chacun des stimuli sonores en fonction des conditions visuelles V0, V1 et V3 et des présentations de stimuli P1, P2 et P3.

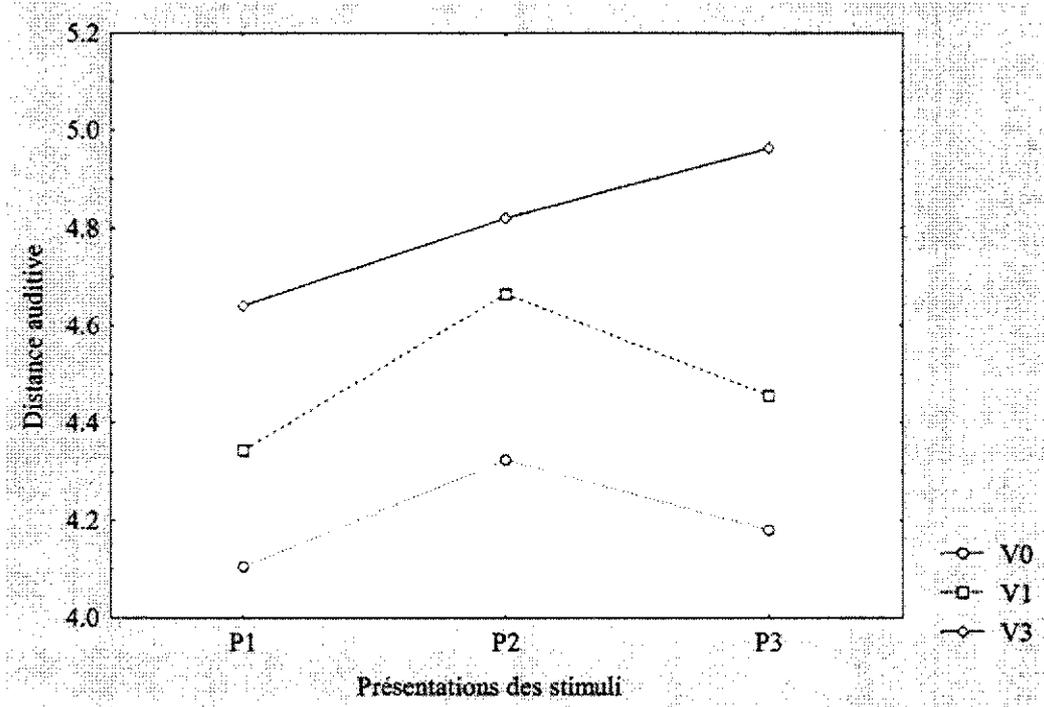


Figure 4.5 : Distance apparente auditive pour les trois présentations de stimuli en fonction des trois conditions visuelles. Les réponses sont moyennées sur les sujets et les six stimuli auditifs.

- Certains sujets ont remarqué que la présence de l'image les a aidés à sentir le son venir d'en face, ou bien, sans mentionner l'image ils ont dit que, au bout d'un certain temps, ils arrivaient à avoir l'impression que le son venait de devant, ou tout simplement, le fait que le son venait du casque ne les gênait plus.
- Par rapport au sentiment «d'être dans la salle», les réponses étaient très variées. Nous émettons toutefois des doutes pour le cas des réponses positives : nous considérons que, comme les sujets *savaient* qu'il s'agissait d'une simulation, ils jugeaient justement la qualité de la simulation – donc du son et de l'image par exemple – et non pas le degré de l'illusion que cette simulation pouvait engendrer. La question n'était peut-être ni suffisamment claire ni assez intéressante.

Ces remarques des sujets nous ont amenés à considérer qu'une amélioration de la technique de diffusion du son était nécessaire. Elle permettrait aux sujets de mieux « se sentir dans la salle », et surtout de percevoir le son à l'extérieur de la tête. Par ailleurs, si en plus de ces réactions des sujets, nous considérons le simple fait que les expériences portaient sur un attribut spatial du son – pour lequel une représentation spatiale la plus fidèle possible du son est essentielle<sup>1</sup> – nous comprenons l'importance de l'amélioration du dispositif sonore. Par conséquent, toutes les expériences qui ont suivi cette expérience préliminaire ont utilisé la méthode de diffusion du son transaurale, sur haut-parleurs (cf. § 3.2.2).

L'importance du nombre des présentations de stimuli a également été révélée lors de cette expérience. En effet, la perception de la distance auditive semble changer avec le temps [Blauert, 1996]. Ceci est peut-être dû à l'entraînement progressif des sujets pendant les tests, à leur adaptation avec les conditions expérimentales, avec les stimuli sonores, etc., mais aussi à d'autres paramètres cognitifs ou sensoriels en partie discutés au chapitre concernant la perception de la distance (cf. § 2.2.3).

Ainsi, il semblerait qu'au début des tests, les sujets hésitent encore à utiliser des notes élevées et la plupart des distances sont légèrement sous-estimées. (cf. [Coleman, 1962], [Mershon et Bowers, 1979]). Dès la 2<sup>ème</sup> présentation, cet effet semble disparaître ; en effet, les moyennes obtenues sur les distances auditives de six stimuli sonores pour la 2<sup>ème</sup> et la 3<sup>ème</sup> présentation ne montrent pas de différences notables.

Il semble donc clair que cet effet demande une étude plus approfondie : l'influence de l'ordre de présentation des stimuli (1<sup>ère</sup>, 2<sup>ème</sup>, 3<sup>ème</sup>, etc.) – et même du nombre de stimuli jugés depuis le début des tests – sur l'appréciation de la distance auditive, est l'objet de l'expérience 2. Elle y sera discutée plus en détail.

### 4.3.5 Conclusion

En conclusion, nous retenons quatre points importants issus de cette expérience préliminaire :

Le *Spatialisateur* semble être un outil bien adapté pour créer les impressions auditives nécessaires pour ces expériences. De nouveaux stimuli sonores mieux contrôlés sur le plan physique doivent pourtant être créés afin d'aider à établir des corrélations avec les aspects perceptifs. La 1<sup>ère</sup> expérience a pour but l'étude des nouveaux stimuli sonores basés sur des indices acoustiques pertinents.

---

<sup>1</sup> Il serait maladroit par exemple d'étudier l'influence des facteurs visuels sur un son pauvre en effets spatiaux, et donc affaibli sur la dimension spatiale, chose qui pourrait conduire à de fausses conclusions en faveur de la perception visuelle

La reproduction du son transaurale sur haut-parleurs est nécessaire afin d'améliorer la « crédibilité » de la simulation, mais aussi afin de donner au son la dimension spatiale nécessaire au jugement d'un attribut auditif spatial comme la distance.

La perception de la distance auditive semble changer avec le temps : les réponses sont légèrement sous-estimées dans la 1<sup>ère</sup> présentation de stimuli. L'importance du nombre de présentations de stimuli a ainsi été révélée et sera étudiée plus en détail dans la suite (expérience 2).

Enfin, les résultats fournissent un premier indice de l'influence des conditions visuelles sur la perception de la distance. L'étude de cette influence étant le but principal de ce travail, elle sera également étudiée en profondeur dans la suite.

## 4.4 EXPERIENCE 1 : ETUDE DES STIMULI AUDITIFS

- *Test à écoute transaurale sans vue*
- *1 ou 2 présentations de stimuli*
- *Signaux sonores :*
  - «Tom's dinner » de Suzanne Vega et*
  - «La Somnambule» de Bellini, chant interprété par Phyllis Bryn-Jullson.*
  - (Stimuli sonores : A1, A2, A3, A4, A5, A6, A7, A8, A9, A10 et A11)*
- *14 sujets*

### 4.4.1 Objectif

Le but de la 1<sup>ère</sup> expérience est de repérer les indices acoustiques pertinents pour la perception de la distance. Les tests réalisés ne concernaient que la perception auditive sans vue ; les stimuli variaient selon trois indices acoustiques afin de déterminer celui qui permettrait d'engendrer les variations les plus conséquentes sur la sensation perceptive de la distance. Deux signaux sonores ont également été utilisés afin de repérer le mieux adapté aux expériences qui suivront.

L'expérience 1 est la première expérience qui utilise la diffusion sonore transaurale sur haut-parleurs.

### 4.4.2 Méthode

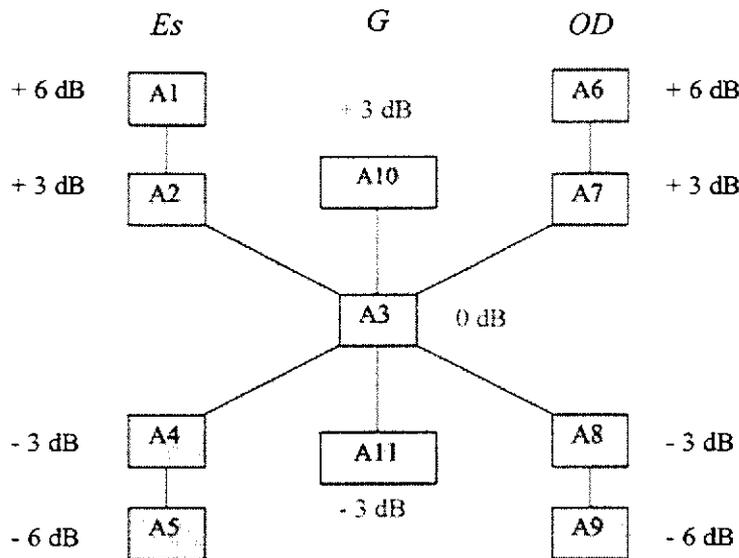
1. *Stimuli sonores.* Onze configurations de salle ont été créées à l'aide du *Spatialisateur* (cf. § 3.2.5), afin de donner différentes impressions de distance auditive. Le *Spatialisateur* a permis avant tout de simuler une réponse impulsionnelle mesurée au Théâtre des Champs Elysées à Paris. Le but n'était pas de reproduire exactement le champ sonore mesuré mais de s'inspirer de cette mesure – obtenue pour un auditeur placé au milieu du théâtre – afin de recréer une réponse qui correspondrait à ce type de salle. A partir de cette réponse de référence (stimulus auditif A3), dix autres réponses ont été créées, sans rapport avec ce qui peut se passer réellement dans ce théâtre lorsque l'on s'approche ou s'éloigne. Elles variaient sur les trois paramètres acoustiques :

- Energie précoce Es
- Niveau sonore du son direct OD
- Niveau sonore global G.

Variation de l'énergie précoce Es revient à varier le rapport d'énergies Es/Er entre le son direct et le son réverbéré (Er reste stable). Es appartient à l'interface haut-niveau du *Spatialisateur* tandis que OD et G appartiennent à l'interface bas-niveau (cf. § 3.2.3). Par ailleurs, il faut préciser ici que ces indices ne sont pas indépendants : Es fait intervenir le niveau du son direct OD ; de même, une variation de Es ou OD entraîne une variation de G. La salle des Champs-Elysées qui a servi de base pour construire la situation de référence est relativement sèche. Ainsi, le niveau des premières réflexions (R1) est très inférieur à celui du son direct, seules les configurations A8 et A9 rendent OD et R1 du même ordre de

grandeur . Ce qui veut dire qu'il faut s'attendre à peu de différence entre le pilotage par OD ou par Es (sauf pour A8 et A9).

La **figure 4.6** présente ces 11 configurations. Trois catégories de distances perceptives *croissantes* se distinguent, chacune correspondant à un indice acoustique : A1, A2, A3, A4 et A5 pour Es, A6, A7, A3, A8 et A9 pour OD et A10, A3 et A11 pour G. Notons que la configuration de référence A3 se trouve au milieu de chacune de ces catégories.



**Figure 4.6:** Les 11 stimuli auditifs du premier test principale. Ils varient, selon l'énergie précoce Es (A1, A2, A3, A4 et A5), selon le niveau du son direct OD (A6, A7, A3, A8 et A9) et selon le niveau global G (A10, A3 et A11) ; les stimuli sont réalisés afin de donner des distances perceptives *croissantes*.

Une différence de 3 dB sur l'indice objectif servant à piloter chaque catégorie, existe entre chaque couple de stimuli successifs. La description détaillée des 11 configurations – interfaces de contrôle *bas* et *haut-niveaux* du *Spatialisateur* – est présentée en annexe B. Les niveaux sonores mesurés à la place de la tête des auditeurs sont également présentés en annexe B.

Les 11 configurations ont été appliquées aux signaux sonores de base qui étaient des voix féminines chantées : il s'agissait de «Tom's dinner» de Suzanne Vega et de «La Somnambule» de Bellini, chanté par Phyllis Bryn-Jullson. La diffusion sonore était transaurale sur haut-parleurs.

**2. Sujets, consignes.** Quatorze adultes en tout ont participé à cette expérience. Avant chaque session de tests, le réglage subjectif du dispositif transaural a eu lieu afin de contrôler la bonne position de la tête des auditeurs (cf. § 4.2.1). Les consignes données aux sujets étaient de juger la distance auditive des stimuli et de répondre oralement dans une échelle comprise entre 1 et 10 (cf. § 4.2.2). Pour que les sujets se fassent une idée des sons, la présentation de 3 des sons choisis comme référence a eu lieu avant les expériences : le son A1 représentait « un des sons les plus proches », A3 « un des sons de distance moyenne » et A6 « un des sons les plus lointains » (cf. § 4.2.3).

**3. Procédure expérimentale.** Pour six des sujets, les sons n'étaient présentés qu'une seule fois pour chacun des signaux sonores (11 configurations x 1 fois x 2 signaux sonores = 22 notes). Le test eut lieu après une petite séance d'entraînement où 6 stimuli – que les sujets ne jugeaient pas – étaient présentés. Pour les huit autres sujets, les 11 sons étaient présentés deux fois, une fois pour chacun des

signaux sonores (11 configurations  $\times$  2 fois  $\times$  2 signaux sonores = 44 notes). Quatre ordres aléatoires de présentation des stimuli ont été utilisés. Dans tous les cas, la moitié des sujets jugeait d'abord les configurations convoluées avec la chanson de S. Vega et ensuite celles convoluées avec «La Somnambule» chanté par P. Bryn-Jullson. L'ordre était inversé pour la deuxième moitié des sujets.

Ces tests ont eu lieu dans la salle d'écoute 'sèche' du Laboratoire d'Acoustique Musicale, de l'Université Paris VI.

### 4.4.3 Résultats

La normalisation des réponses sur la distance auditive était effectuée comme décrit au § 4.2.8 (avec  $\bar{x} = 4,5$  et  $\sigma = 3$ ). Les moyennes et les écart-types des réponses de tous les sujets sont présentés à la suite, pour chacun des signaux sonores (S. Vega ou P. Bryn-Jullson) et pour chaque stimulus auditif isolé d'après l'indices acoustique selon lequel ils varie :

S. Vega						Es					OD					G		
	A1	A2	A3	A4	A5	A6	A7	A3	A8	A9	A10	A3	A11					
<i>m</i>	0.90	2.33	4.94	6.38	8.89	1.03	2.23	4.94	2.23	5.99	2.72	4.94	6.01					
<i>et</i>	0.89	0.49	1.64	1.19	0.88	0.92	0.74	1.64	0.74	1.02	0.89	1.64	1.43					

**Tableau 4.3 :** Valeurs moyennes (*m*) et écart-types (*et*) de distances auditives des stimuli utilisant «Tom's dinner» de S. Vega. Les trois tableaux correspondent respectivement aux stimuli variant sur Es, OD et G.

P. Bryn-Jullson						Es					OD					G		
	A1	A2	A3	A4	A5	A6	A7	A3	A8	A9	A10	A3	A11					
<i>m</i>	0.92	2.54	4.40	6.70	8.51	0.57	2.16	4.40	5.78	7.77	3.52	4.40	6.63					
<i>et</i>	1.28	1.12	1.65	1.00	1.33	0.87	0.98	1.65	1.35	0.76	0.93	1.65	1.00					

**Tableau 4.4 :** Valeurs moyennes (*m*) et écart-types (*et*) de distances auditives des stimuli utilisant «La Somnambule» chanté par P. Bryn-Jullson. Les trois tableaux correspondent respectivement aux stimuli variant sur Es, OD et G.

Les figures 4.7 et 4.8 donnent pour chacun des signaux sonores, les graphiques des *distances apparentes auditives* issues de ces moyennes.

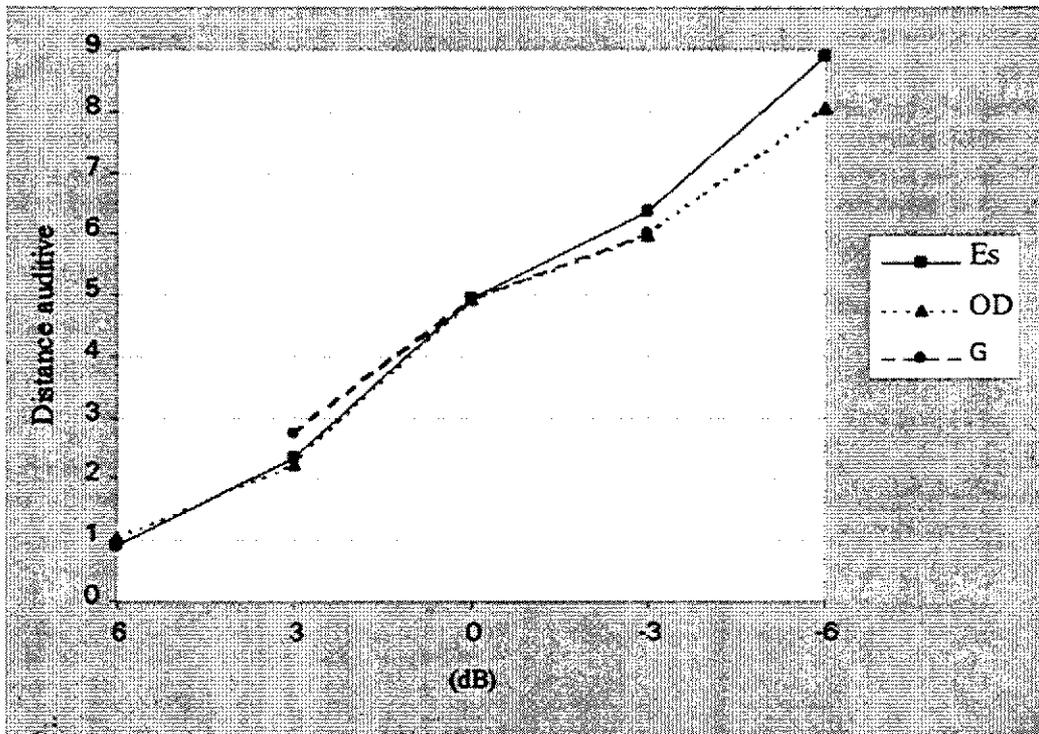


Figure 4.7 : Valeurs moyennes de la distance apparente auditive des 11 stimuli sonores pour le signal sonore de S. Vega. 5 stimuli varient selon l'énergie précoce Es (A1, A2, A3, A4 et A5), 5 varient selon le niveau du son direct OD (A6, A7, A3, A8 et A9) et 3 selon le niveau global G (A10, A3 et A11).

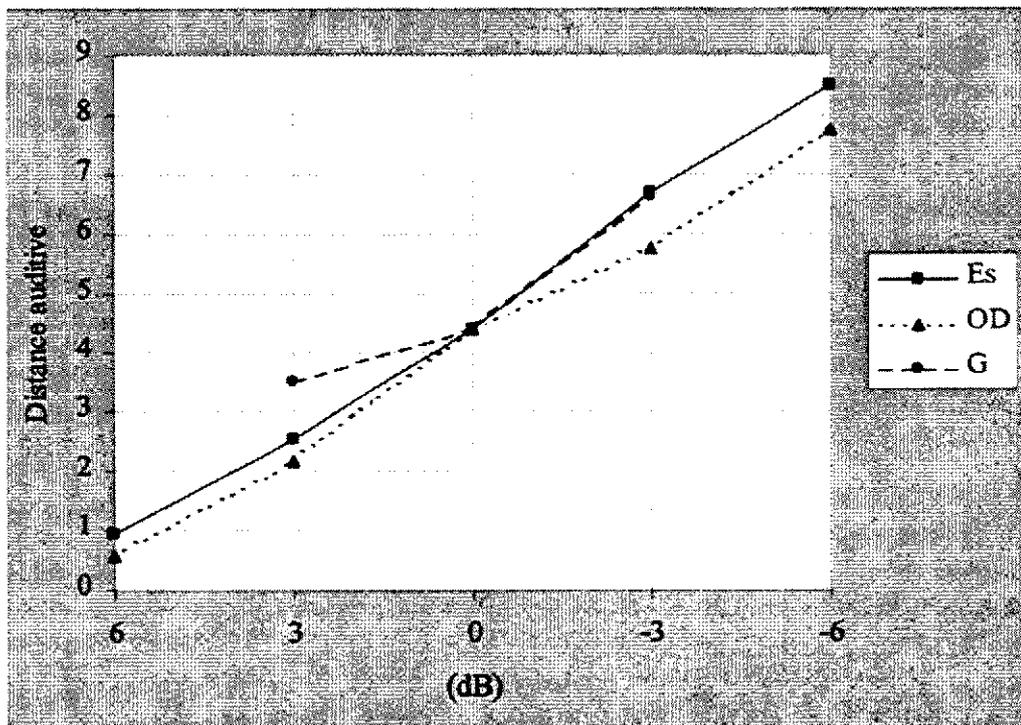
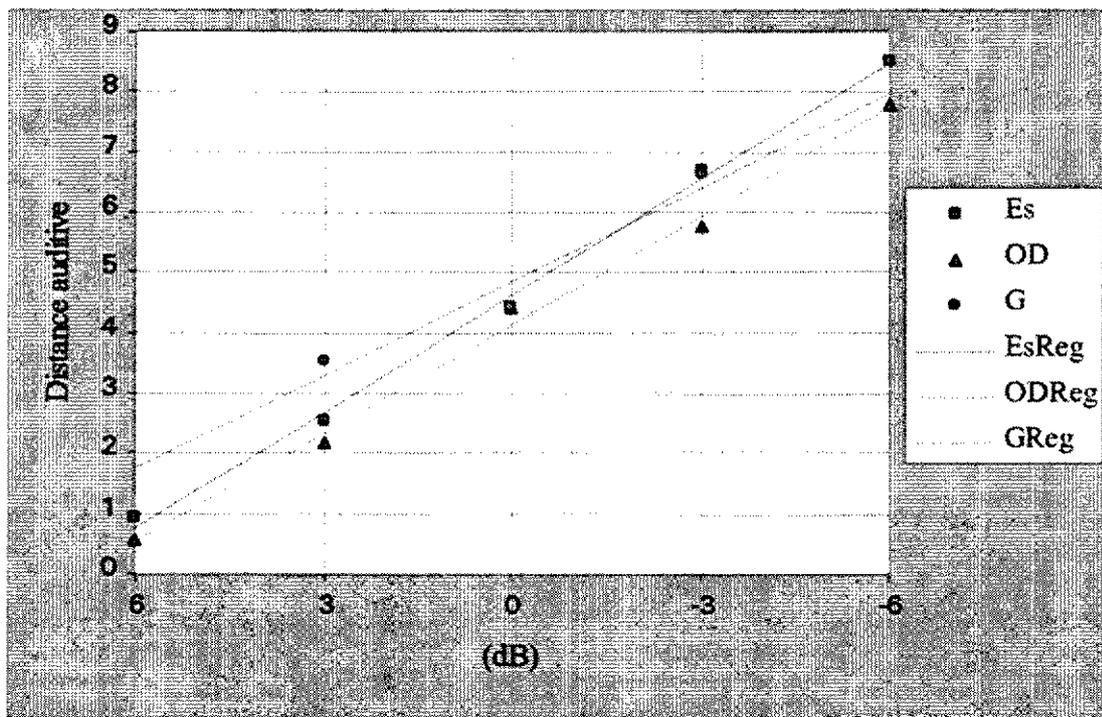


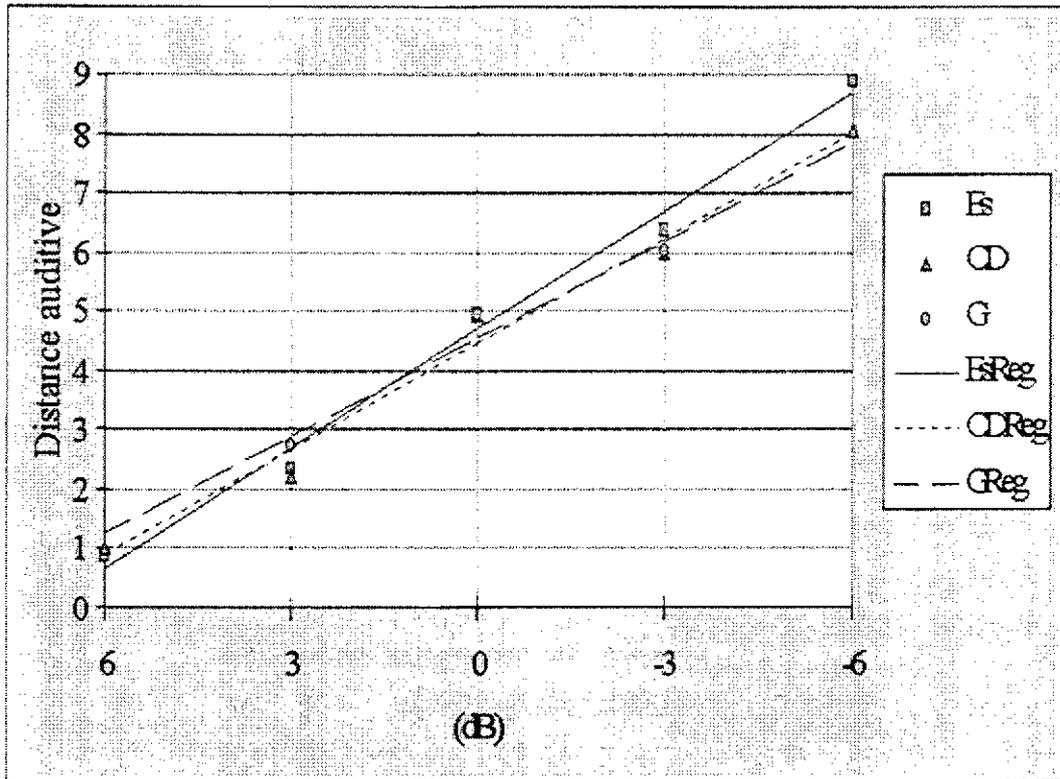
Figure 4.8 : Valeurs moyennes de la distance apparente auditive des 11 stimuli sonores pour le signal sonore de P. Bryn-Jullson. 5 stimuli varient selon l'énergie précoce Es (A1, A2, A3, A4 et A5), 5 varient selon le niveau du son direct OD (A6, A7, A3, A8 et A9) et 3 selon le niveau global G (A10, A3 et A11).

#### 4.4.4 Discussion

Les stimuli qui ont été créés pour cette expérience engendrent des réponses qui varient linéairement en fonction des indices acoustiques Es, OD et G (eux mêmes représentés sur une échelle logarithmique). Afin d'étudier lequel des trois catégories (ou séries de stimuli) correspondant à chacun de ces indices est perçu de la manière la plus significative, les valeurs sont comparées par rapport à une tendance linéaire. Une droite de régression, calculée selon la méthode des moindres carrés, est ainsi ajustée pour chacune de ces séries. Les écarts (erreurs) des valeurs moyennes obtenues sur la distance auditive par rapport à cette droite sont également calculés. Les figures 4.9 et 4.10 montrent ces valeurs moyennes de distance perçue, ainsi que les droites de régression linéaires ; les erreurs sont présentées à la suite.



**Figure 4.9 :** Valeurs moyennes de distance auditive et droites de régression linéaire correspondant à «Tom's dinner» de S. Vega (les valeurs 6, 3, 0, -3 et -6 figurant sous l'axe des abscisses sont les écarts en dB des stimuli par rapport au stimulus de référence A3, lui même correspondant au 0).



**Figure 4.10 :** Valeurs moyennes de distance auditive et droites de régression linéaire correspondant à «La Somnambule» chanté par P. Bryn-Jullson (les valeurs 6, 3, 0, -3 et -6 figurant sous l'axe des abscisses sont les écarts en dB des stimuli par rapport au stimulus de référence A3, lui-même correspondant au 0).

Les erreurs des valeurs moyennes sur la distance auditive par rapport aux droites de régression linéaire sont présentées au **tableau 4.5** ci-dessous pour chacun des signaux sonores (cf. annexe B) :

Erreur globale <i>S Vega</i>		Erreur globale <i>P. Bryn-Jullson</i>	
Es erreur	0,092	Es erreur	0,030
OD erreur	0,127	OD erreur	0,032
G erreur	0,108	G erreur	0,149

**Tableau 4.5 :** Erreurs moyennes sur la distance auditive

On remarque que les valeurs des erreurs sur les indices Es et OD pour «La Somnambule» de Bellini chanté par P. Bryn-Jullson, sont bien inférieures à celles obtenues pour la chanson de S. Vega.

De plus, dans les deux cas, l'énergie précoce Es semble être l'indice le mieux corrélé avec les impressions perceptives sur la distance auditive. Ce résultat n'est pas étonnant, dans la mesure où l'indice Es fait partie de l'interface de contrôle haut-niveau du *Spatialisateur*, interface dont la

conception est basée sur des tests du même type que le présent test (dont le but est de mieux correspondre aux sensations subjectives). Es correspond en effet au facteur de l'interface de contrôle haut-niveau appelé « présence » de la source (cf. § 3.2.3).

Enfin, il faut noter ici que lors des entretiens informels après les tests, certains sujets ont trouvé la chanson de S. Vega « artificielle », de volume « trop fort » ou même « plus facile à juger » que le signal de l'opéra. Il y avait même pour certains, une petite tendance à juger les stimuli distinctement : ceux variant sur OD étaient parfois repérables et étaient plus souvent caractérisés d'« artificiels ». En fait, la chanson de S. Vega n'est pas un enregistrement purement anéchoïque, il est simplement supposé « exempt de réverbération audible ». Ce fait a été repéré par les sujets qui avaient une certaine habitude d'écoute. Les stimuli sonores créés à partir de l'enregistrement anéchoïque de «La Somnambule» de Bellini ont, en général, été jugés plus homogènes et considérés comme étant plus crédibles.

Ce dernier signal, «La Somnambule» de Bellini, sera donc retenu pour les tests suivants ; les configurations du *Spatialisateur* seront variées sur l'indice de l'énergie précoce Es.

Une dernière remarque générale doit être faite, concernant la diffusion sonore transaurale qui a été employée pour la première fois dans le cadre de ce travail. Lors du réglage subjectif du dispositif transaural réalisé afin de contrôler la bonne position de la tête des auditeurs, tous les sujets ont entendu le son venir des cotés (pour les azimuts  $+80^\circ$ ,  $+30^\circ$ ,  $-30^\circ$  et  $-80^\circ$ ) (cf. § 4.2.1). Le son était dans tous les cas localisé à l'extérieur de la tête, et le son frontal (azimut  $0^\circ$ ) donnait l'impression de venir de devant dans la majeure partie des cas.

#### **4.4.5 Conclusion**

La première expérience a été réalisée afin de repérer le signal sonore et l'indice acoustique le mieux corrélé à l'impression subjective de distance auditive. Elle a montré que la chanson d'opéra de Bellini «La Somnambule» semble plus adaptée que «Tom's dinner» de S. Vega. De plus, le rapport d'énergies Es/Er entre son direct et son réverbéré semble l'indice le plus représentatif des sensations subjectives de distance auditive ; les deux autres indices testés étaient l'énergie de l'onde direct OD et le niveau sonore global G. Ce résultat n'est pas étonnant puisque « Es » correspond au facteur « présence de la source » de l'interface de contrôle haut-niveau du *Spatialisateur*, une interface conçue à partir de tests perceptifs. Ces choix du signal et de l'indice acoustique sont également confirmés par les réactions des sujets, obtenues après chaque session de tests.

En outre, la diffusion transaurale du son sur haut-parleurs, utilisée pour la première fois dans ces tests, semble être une méthode nettement meilleure que la diffusion binaurale. Le son est localisé à l'extérieur de la tête et aucune inversion devant-derrrière n'est apparue. La dimension spatiale du son semble globalement favorisée. Enfin, le port du casque audio, susceptible d'engendrer des problèmes de « crédibilité » ou de « réalisme », est évité.

## 4.5 EXPERIENCE 2 : EVOLUTION DES REPONSES

- *Test à écoute transaurale - Vues statiques*
- *7 présentations de stimuli*
- *Signal sonore : «La Somnambule» de Bellini interprété par Phyllis Bryn-Jullson  
(Stimuli sonores : A1, A2, A3, A4, A5 et A6).*
- *Vues : « Théâtre des Champs Elysées » à Paris  
(Stimuli visuels : condition visuelle de contrôle V0, images 3-D : V1 et V3).*
- *12 sujets*

### 4.5.1 Objectif

La deuxième expérience a été réalisée afin d'étudier l'effet des différentes présentations des stimuli – ou plus généralement l'effet de l'entraînement ou de l'adaptation des sujets – sur la perception de la distance auditive. L'intérêt d'une telle expérience a été révélé lors de l'expérience préliminaire où l'appréciation de la distance variait entre les différentes présentations de stimuli, et en particulier, elle augmentait entre la première et la deuxième présentation.

Cette expérience est bimodale – les stimuli sonores sont accompagnés par des images – elle devrait donc permettre également d'observer quelques effets de l'influence visuelle sur la distance auditive.

### 4.5.2 Méthode

1. *Stimuli sonores.* Les 6 stimuli sonores utilisés dans cette expérience varient sur l'énergie précoce Es : A1, A2, A3, A4, A5 et A6 (il s'agit des 5 stimuli de l'expérience 1 et de 5 stimuli supplémentaires créés à ce propos) (cf. § 3.2.5 et § 4.4.2). Leur description détaillée est présentée à l'annexe B.

2. *Conditions visuelles.* Les vues V1 et V3 utilisées dans les tests furent prises dans le « Théâtre des Champs Elysées » à Paris, en des emplacements de différentes distances par rapport à la scène, au milieu de laquelle était placé un haut-parleur (cf. § 3.3.2). Pour la condition visuelle de contrôle V0 aucune image n'était projetée sur l'écran, seule une lumière diffuse y était perçue (cf. § 4.2.5). Les sujets étaient assis à une distance de 1,6 m de l'écran.

3. *Sujets, consignes.* Douze adultes ont participé à cette expérience. La tâche des sujets consistait à juger la distance auditive et à répondre oralement dans une échelle comprise entre 1 et 10 (1 représentant la distance auditive la plus courte et 10 la plus grande). Ils avaient pour consigne d'utiliser le son et l'image afin de s'imaginer dans la "salle", mais de ne juger que la distance auditive (cf. § 4.2.2).

Le réglage subjectif du dispositif transaural et la présentation des sons de référence (stimuli extrêmes A1 et A6) eurent lieu avant chaque session (cf. § 4.2.1 et § 4.2.3).

4. *Procédure expérimentale* : Les six stimuli auditifs sont présentés sept fois dans un ordre pseudo-aléatoire, créant ainsi des séquences de quarante-deux stimuli (cf. figure 4.1 pour  $n=7$ ). Afin de compenser les effets d'ordre de présentation des stimuli, deux séquences sont créées, chaque séquence a été présentée à six sujets (sur les 12 sujets). Les deux séquences utilisées sont présentées à l'annexe B.

Chaque sujet subit la même séquence trois fois, une fois pour chacune des trois conditions visuelles :

- Condition visuelle de contrôle V0 ou  $V_{\text{contrôle}}$
- Vue proche de la scène V1 ou  $V_{\text{proche}}$
- Vue lointaine de la scène V3 ou  $V_{\text{loin}}$

Chaque sujet participe ainsi en tout à trois tests de 42 stimuli chacun ( $42 \times 3 = 126$  stimuli en tout). Six ordres de présentation des conditions visuelles ont été utilisés : V0-V1-V3, V1-V3-V0, V3-V0-V1, V3-V1-V0, V1-V0-V3 et V0-V3-V1 (respectivement présentés à 3, 3, 2, 2, 1 et 1 sujets).

La *durée* moyenne des tests variait entre 1 heure 15 min. et 1 heure et 45 min. (25 à 35 min. par condition visuelle) ; les sujets se reposaient entre les différentes sessions du test.

### 4.5.3 Résultats

Les réponses des 12 sujets sur la distance apparente auditive sont normalisées comme décrit au § 4.2.8 ( $\bar{x} = 4,5$  et  $\sigma = 3$ ). Les résultats de l'Analyse de la Variance (cf. § 3.6.1.4) sont présentés au **tableau 4.6**.

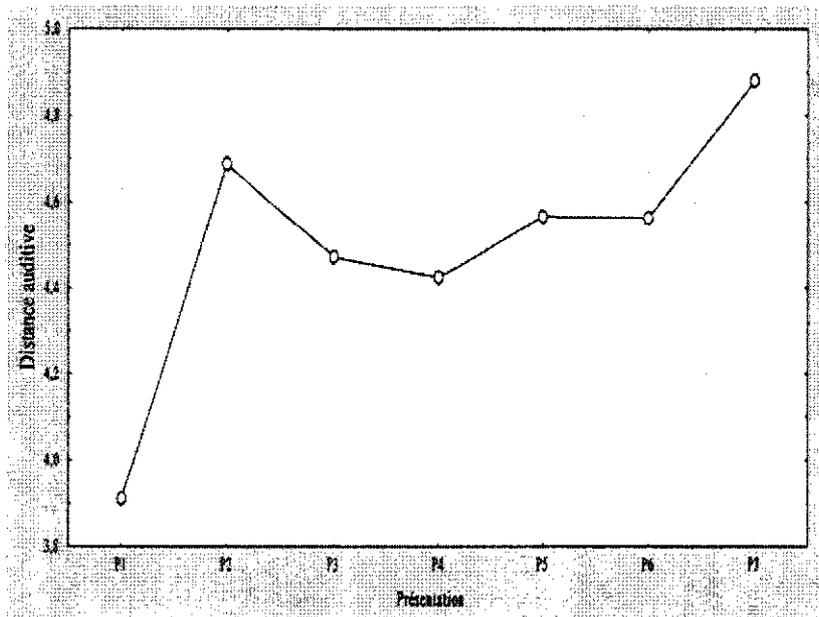
Nous observons un effet principal des stimuli auditifs [ $F(5,55) = 951.81$ ,  $p < 0.01$ ], un effet attendu puisque les stimuli ont été *créés* pour représenter de différentes impressions de distances auditives.

L'analyse ne révèle pas d'effet des conditions visuelles, ainsi elles ne semblent apparemment pas influencer la perception de la distance de manière significative, tout du moins sur l'ensemble des résultats. Pourtant, une analyse prenant en compte uniquement les réponses obtenues dans les conditions visuelles V1 et V3 – ce qui paraît normal puisqu'il s'agit justement d'étudier l'effet de la distance visuelle sur la distance auditive et pas vraiment l'effet de la condition V0 – révèle un effet significatif [ $F(1,11) = 5.49$ ,  $p < 0.05$ ]. Mais cela n'étant pas le but principal de cette expérience, ce résultat ne sera pas discuté davantage ici.

Table de tous les effets						
1 : Conditions Visuelles, 2 : Présentations de stimuli, 3 : Stimuli Auditifs						
	df Effet	MS Effet	df Erreur	MS Erreur	F	Niveau-p
1	2	11,01	22	4,13	2,67	0,092
2	6	19,48	66	2,45	7,97	* 0,000
3	5	2349,25	55	2,57	915,81	* 0,000
12	12	1,50	132	1,46	1,03	0,425
13	10	0,54	110	1,16	0,46	0,911
23	30	2,77	330	1,05	2,64	* 0,000
123	60	0,69	660	0,62	1,12	0,259

**Tableau 4.6** : Tableau des résultats de l'Analyse de la Variance pour l'expérience 2. Les trois facteurs sont : les *conditions visuelles* (facteur 1), les *présentations des stimuli* (facteur 2) et les *stimuli auditifs* (facteur 3). 12, 13, 23 et 123 représentent les interactions entre ces facteurs respectifs : 12 est par exemple l'interaction entre les facteurs 1 et 2, 123 l'interaction entre les facteurs 1, 2 et 3, etc. Les effets significatifs sont signalés par un astérisque (\*).

Enfin, un effet principal des *présentations des stimuli* est observé [ $F(6,66) = 7.97, p < 0.01$ ], cela signifie que la distance auditive est jugée différemment dans les différentes présentations. La **figure 4.11** montre la variation de la distance auditive avec les présentations des stimuli.



**Figure 4.11** : Distance apparente auditive pour les 7 présentations des stimuli (les valeurs sont moyennées sur les conditions visuelles et sur les stimuli auditifs).

On observe que les réponses à la 1<sup>ère</sup> présentation semblent être inférieures aux réponses des présentations suivantes. Afin de tester une telle hypothèse, des comparaisons Post-Hoc (Tukey HSD test) entre les 7 niveaux de présentations ont été réalisées. Elles sont présentées au **tableau 4.7** ; les moyennes des distances auditives pour ces 7 présentations y sont également présentées.

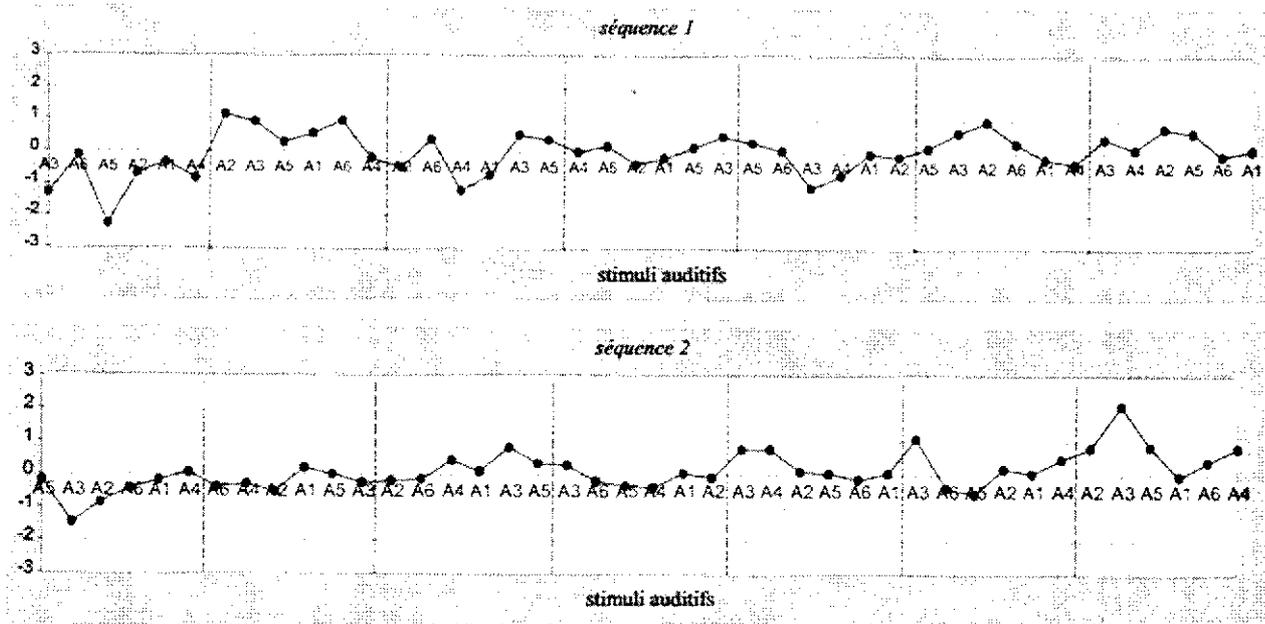
Probabilités pour des tests Post Hoc (Test Tukey HSD)							
Effet principal : Présentation							
	1	2	3	4	5	6	7
<b>Moyennes</b>	<b>3,91</b>	<b>4,69</b>	<b>4,47</b>	<b>4,42</b>	<b>4,56</b>	<b>4,56</b>	<b>4,88</b>
1		* 0,000	* 0,007	* 0,019	* 0,001	* 0,001	* 0,000
2	* 0,000		0,780	0,581	0,982	0,980	0,858
3	* 0,007	0,780		1,000	0,996	0,997	0,110
4	* 0,019	0,581	1,000		0,965	0,968	0,051
5	* 0,001	0,982	0,996	0,965		1	0,366
6	* 0,001	0,980	0,997	0,968	1		0,357
7	* 0,000	0,858	0,110	0,051	0,366	0,357	

**Tableau 4.7** : Moyennes des distances auditives pour les 7 présentations des stimuli et comparaisons Post-Hoc effectuées entre ces 7 niveaux. Les effets significatifs sont signalés par un astérisque (\*).

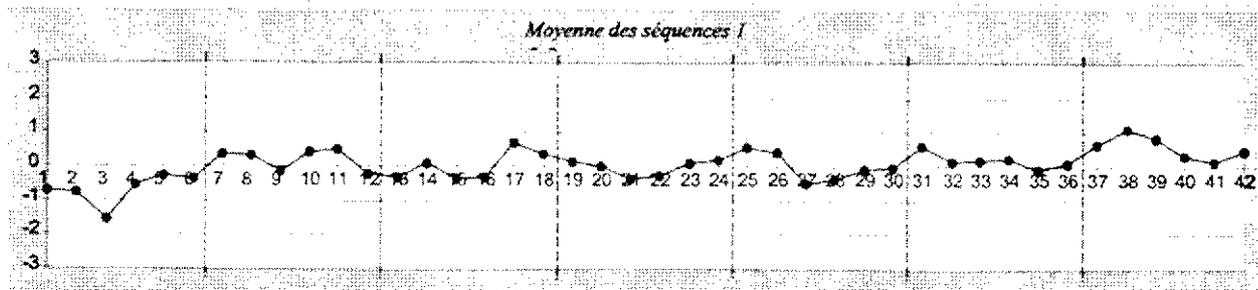
On observe que les réponses à la 1<sup>ère</sup> présentation sont effectivement inférieures aux autres de manière significative. En revanche, il n'y a pas de telle différence pour les présentations suivantes. Ce résultat semble suggérer que deux présentations de stimuli suffisent pour obtenir des jugements de la distance auditive relativement stables, les réponses ne semblent plus varier après.

Le même effet d'évolution des réponses a été testé par une autre technique : les séquences sont considérées comme des suites de 42 stimuli, sans tenir compte de leur découpage en 7 séries (les 7 présentations). Autrement dit, au lieu d'étudier les séquences comme l'enchaînement de 7 séries de stimuli comme ci-dessus, on les étudie comme une suite continue de stimuli.

Pour pouvoir observer cette évolution de stimuli de manière continue, nous devons nous affranchir du fait que les moyennes des réponses sont différentes pour les différents stimuli. Pour ce faire, nous sommes obligés de recourir à une opération nous permettant d'homogénéiser les résultats. Les séquences sont considérées indépendamment des conditions visuelles, et chaque stimulus est recentré autour de sa moyenne évaluée sur l'ensemble de la séquence. Par exemple, pour le premier stimulus de la première séquence, correspondant au stimulus auditif A3, la valeur présentée est l'écart entre la moyenne des réponses (des 6 sujets) et la moyenne globale obtenue pour ce stimulus auditif sur chacune des séquences utilisées ; la **figure 4.12** montre les résultats pour chacune des séquences utilisées, la **figure 4.13** présente la moyenne de ces deux séquences. Ces figures s'interprètent comme l'évolution des écarts entre les moyennes « instantanées » de chaque stimulus et les moyennes à long terme des ces stimuli sur l'ensemble d'une séquence.



**Figure 4.12 :** Evolution des réponses sur la distance auditive avec le temps pour la 1<sup>ère</sup> et la 2<sup>ème</sup> séquences utilisées dans le test (pour la séquence 1, A3 est le 1<sup>er</sup> stimulus présenté, A6 le 2<sup>ème</sup>, A5 le 3<sup>ème</sup>, etc.; pour la séquence 2, A5 est le 1<sup>er</sup> stimulus présenté, A3 le 2<sup>ème</sup>, A2 le 3<sup>ème</sup>, etc.). De chaque stimulus auditif, la moyenne globale a été soustraite, l'axe y = 0 correspond à cette moyenne globale.



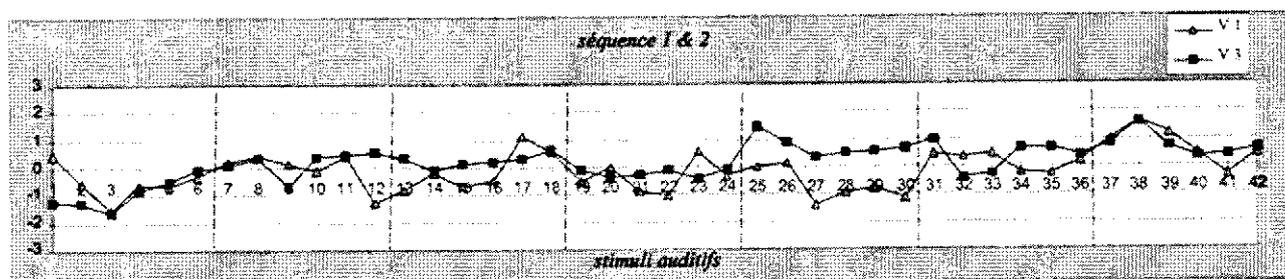
**Figure 4.13 :** Evolution des réponses sur la distance auditive avec le temps pour les deux séquences utilisées dans le test (1 correspond au 1<sup>er</sup> stimulus présenté, 2 au 2<sup>ème</sup>, 3 au 3<sup>ème</sup>, etc.). L'axe y = 0 correspond à la moyenne globale de ces stimuli.

Nous observons que dans les deux séquences, les valeurs jusqu'au sixième stimulus environ sont inférieures à la moyenne globale (représentée par l'axe y = 0). Les réponses sur la distance auditive semblent donc être effectivement inférieures au début des tests, indépendamment de la condition visuelle. Pourtant, au bout de six stimuli – et pour les 5 séries qui suivent – les réponses semblent, non pas se stabiliser – on remarque en effet que la variation est importante – mais cesser d'augmenter de manière systématique : les réponses semblent fluctuer autour de leur moyenne (axe y = 0).

Cependant, une exception à cette règle semble se produire à la fin de la séquence (ce qui correspond à la septième série de stimuli) : on remarque en effet que toutes les valeurs se trouvent supérieures à la moyenne (cf. figures 4.12 et 4.13). Bien que cet effet ne soit pas révélé significatif (cf. tableau 4.7), on ne peut pas passer outre : une hypothèse qui pourrait l'expliquer est que, vers la fin de chaque séquence, les sujets relâchent leur attention, « s'éloignent » du test et donnent des notes de manière

moins attentive. Des expériences supplémentaires devraient être menées pour étudier ce phénomène en profondeur ; mais ce qui nous importe ici, est de choisir un nombre (peu élevé) de présentations des stimuli, et non pas d'étudier l'évolution des réponses à long terme.

Enfin, et bien que l'étude de l'influence des images sur la perception auditive ne soit pas le but principal de cette expérience, la **figure 4.14** présente l'évolution des réponses avec le temps, et ce pour chacune des deux conditions visuelles V1 et V3 (la condition visuelle V0 a été omise pour une meilleure lisibilité du graphe).



**Figure 4.14 :** Evolution des réponses avec le temps pour chacune des conditions visuelles V1 (vue proche de la scène) et V3 (vue lointaine).

La lecture de ce graphique indique que les réponses qui correspondent à la condition visuelle V3 semblent, dans bien des cas, être supérieures à celles obtenues pour la condition visuelle V1 (il s'agit en effet d'une différence significative mentionnée auparavant [ $F(1,11) = 5.49, p < 0.05$ ]).

#### 4.5.4 Discussion

A la 1<sup>ère</sup> présentation des stimuli, les réponses sur la distance auditive sont globalement inférieures à celles obtenues aux autres présentations. Cette différence entre les présentations a été également observée à l'expérience préliminaire ainsi que dans des travaux antérieurs sur la perception de la distance, comme le travail de Coleman [1962] qui trouve que les réponses à la 1<sup>ère</sup> présentation des stimuli sont sérieusement sous-estimés (sauf pour un haut-parleur proche) et le travail de Mershon et King [1975] (cf. § 2.2.1).

On pourrait supposer que les sujets réservent les jugements de distances lointaines pour «plus tard», ils répondraient de manière réservée, dans l'attente des stimuli à venir. Quand ils commencent à s'habituer aux conditions des expériences et à sentir qu'ils maîtrisent les signaux sonores, ils commencent à faire des jugements de distances plus lointaines. Autrement dit, il semblerait qu'un certain temps soit nécessaire pour que les sujets commencent à juger la distance de manière plus cohérente. Ce temps semble correspondre à la durée de présentation des 6 à 7 premiers stimuli.

L'augmentation de la distance auditive dans les réponses dès la 2<sup>ème</sup> présentation des stimuli, est aussi observée à la **figure 4.15** qui montre les distances auditives obtenues pour chacun des stimuli auditifs en fonction de la présentation (cette augmentation est observée pour tous les stimuli, sauf pour le stimulus A4). Enfin, un autre effet semble être révélé par ce même graphique : à la 1<sup>ère</sup> présentation des stimuli, les réponses sur la distance auditive montrent une tendance à se regrouper vers les extrémités de l'échelle proposée aux sujets ; les notes au milieu de l'échelle semblent moins utilisées.

Cet effet de «polarisation» des réponses à la 1<sup>ère</sup> présentation pourrait être un indice du manque de confiance des jugements, d'un besoin de repères (cf. § 2.2.3) et par conséquent, d'une tendance à juger par rapport aux distances extrêmes de référence A1 et A6. Par ailleurs, nous avons vu auparavant que l'emploi des différentes distances de référence peut entraîner des jugements influencés justement vers ces mêmes distances [Cochran et al., 1968]. Et puis, ce phénomène pourrait être dû au fait qu'au début des expériences, les sujets ne peuvent discerner que deux 'catégories' de distance auditive : 'loin' et 'proche'. Leurs réponses sont alors concentrées autour de ces extrêmes. Avec l'entraînement, ils obtiennent plus de finesse dans leur jugement et utilisent toute l'échelle perceptive (cf. 'Affinement croissant de la différenciation perceptive' de E. Gibson [Casalis, 1991], § 2.2.6.1). Aussi, on observe que les réponses aux stimuli extrêmes A1 et A6 ne varient pas, du fait probablement qu'ils ont été présentés au début des tests comme références, ce qui va dans le sens d'une stabilisation rapide des réponses.

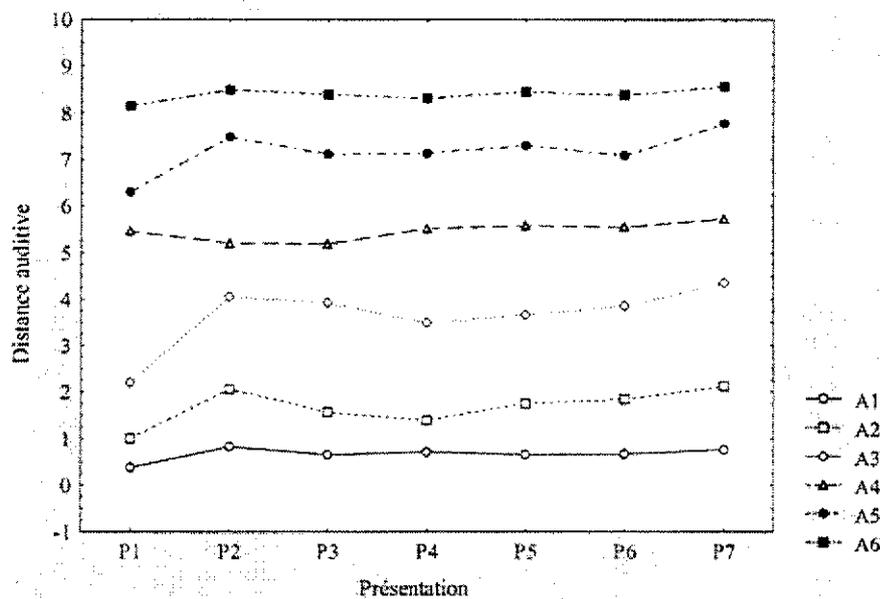


Figure 4.15 : Distance auditive apparente en fonction de la présentation pour chacun des stimuli auditifs.

### 4.5.5 Conclusion

Dans cette 2<sup>ème</sup> expérience, l'effet du nombre de présentations répétées des stimuli a été étudié. En d'autres termes, il s'agissait d'étudier l'évolution des réponses avec le temps et l'entraînement progressif des sujets.

Sept présentations de stimuli ont été utilisées. A la 1<sup>ère</sup> présentation des stimuli, les réponses sur la distance auditive sont inférieures à celles obtenues aux autres présentations. Dès la 2<sup>ème</sup> présentation les réponses augmentent et restent stables pour les présentations qui suivent (des variations relativement importantes peuvent exister, mais elles sont pratiquement autour de la même moyenne).

Les deux différentes méthodes utilisées pour tester ce résultat (étude de l'évolution des réponses en fonction des présentations ou en fonction des stimuli pris séparément) ont montré que 6 stimuli suffisent pour que les réponses commencent à se stabiliser. Cela veut dire qu'à partir de la 2<sup>ème</sup>

présentation de stimuli, une plus grande confiance peut être accordée aux résultats. En effet, il apparaît que même si 3, 5 ou 7 présentations étaient utilisées, les résultats seraient similaires.

Enfin, un effet de « polarisation » des réponses à la 1<sup>ère</sup> présentation vers les distances extrêmes de référence A1 et A6 est observé. C'est encore à partir de la 2<sup>ème</sup> présentation de stimuli que les sujets semblent ne plus être influencés par ces stimuli de référence, et utilisent davantage le milieu de l'échelle proposée.

Pour conclure, cette expérience semble montrer qu'un certain entraînement est nécessaire afin que les sujets commencent à juger la distance de manière plus correcte, c'est à dire sans sous-estimer les distances et en utilisant toute l'échelle de réponses proposée. Dans les expériences suivantes, deux présentations des stimuli seront adoptées afin de réaliser cet « entraînement » tout en respectant une durée acceptable.

## 4.6 EXPERIENCE 3 : INFLUENCE VISUELLE (VUES STATIQUES EN PROJECTION 2-D)

- *Test à écoute transaurale - Vues statiques*
- *2 présentations de stimuli*
- *Signal sonore : «La Somnambule» de Bellini interprété par Phyllis Bryn-Jullson  
(Stimuli sonores : A1, A2, A3, A4, A5 et A6).*
- *Vues : « Théâtre des Champs Elysées » à Paris  
(Stimuli visuels : condition visuelle de contrôle V0, images 2-D : V1 et V3).*
- *16 sujets*

### 4.6.1 Objectif

L'expérience 3 (comme aussi l'expérience 4) a été organisée afin d'étudier la dégradation éventuelle de la projection conventionnelle en 2-D par rapport à la projection des images en 3-D.

L'expérience 3 est identique à l'expérience 2 en vues statiques, la seule différence est que la projection des images est réalisée en 2-D et que le nombre de présentations des stimuli est de 2 au lieu de 7. Comme elle est aussi identique à l'expérience 5 – et considérant qu'il est peu commode pour la description de la méthode des tests finaux utilisés pour la validation des résultats (expérience 5) de devoir se référer à des tests destinés à étudier la méthode (expérience 3) – nous avons préféré le contraire : pour la description de la procédure de l'expérience 3, on se référera à l'expérience 5, en vues statiques (cf. § 4.8.2).

Ainsi, ce sont uniquement les résultats de cette expérience qui sont brièvement décrits.

### 4.6.2 Résultats

La **figure 4.16** donne les réponses sur la distance auditive des stimuli sonores pour chacune des conditions visuelles, séparément pour les deux présentations de stimuli. Le **tableau 4.8** présente les résultats de l'Analyse de la Variance.

Nous observons sur le graphique de faibles différences sur la perception de la distance, qui sont dues aux conditions visuelles, mais l'Analyse de Variance ne révèle pas d'effet significatif ( $[F(2,30) = 1.44, p=0.252]$ ).

Si l'intérêt porte maintenant uniquement sur les réponses obtenues sous les conditions visuelles V1 et V3 (puisque V0 ne représente pas une distance visuelle), nous remarquons que l'effet des conditions visuelles augmente mais reste non significatif ( $[F(2,30) = 3.06, p=0.10]$ ). On observe que ces différences sont plus grandes à la 2<sup>ème</sup> présentation des stimuli.

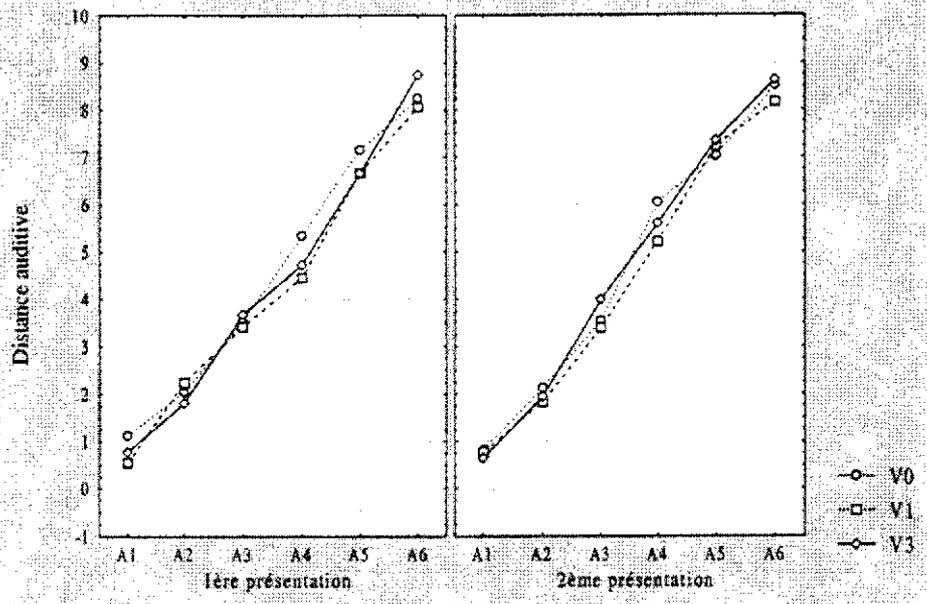


Figure 4.16 : Distance auditive des stimuli sonores pour chacune des conditions visuelles, séparément pour les deux présentations des stimuli.

Table de tous les effets						
1 : Conditions Visuelles, 2 : Présentations de stimuli, 3 : Stimuli Auditifs						
	df Effet	MS Effet	df Erreur	MS Erreur	F	Niveau-p
1	2	4,32	30	2,99	1,44	0,252
2	1	5,65	15	5,80	0,97	0,340
3	5	823,37	75	2,69	305,84	* 0,000
12	2	0,43	30	2,28	0,19	0,828
13	10	1,48	150	0,84	1,75	0,074
23	5	2,53	75	2,04	1,24	0,300
123	10	0,63	150	0,99	0,63	0,783

Tableau 4.8 : Tableau des résultats de l'Analyse de la Variance pour l'expérience 3 (expérience à vues statiques en projection 2-D). Les trois facteurs sont : les *conditions visuelles* avec 3 niveaux V0, V1 et V3 (facteur 1), les *présentations des stimuli* avec 2 niveaux P1 et P2 (facteur 2) et les *stimuli auditifs* avec 6 niveaux A1-A6 (facteur 3). 12, 13, 23 et 123 représentent les interactions entre les facteurs respectifs. Les effets significatifs sont signalés par un astérisque (\*).

### **4.6.3 Conclusion**

La simulation de profondeur des champs visuels à l'aide d'une projection conventionnelle (2-D) semble moins convaincante, ou encore, moins intéressante que celle obtenue à l'aide de la projection en 3-D ; les effets dus aux conditions visuelles et obtenus avec la projection 2-D ne sont pas significatifs. L'utilisation de la projection 3-D est donc importante – sinon nécessaire – pour le rendement d'une impression de distance visuelle.

## 4.7 EXPERIENCE 4 : INFLUENCE VISUELLE (VUES CHANGEANTES EN PROJECTION 2-D)

- *Test à écoute transaurale - Vues changeantes*
- *2 présentations de stimuli*
- *Signal sonore : «La Sonnambule» de Bellini interprété par Phyllis Bryn-Jullson  
(Stimuli sonores : A1, A2, A3, A4, A5 et A6).*
- *Vues : « Théâtre des Champs Elysées » à Paris  
(Stimuli visuels : images 2-D : V1, V2 et V3).*
- *16 sujets*

### 4.7.1 Objectif

L'expérience 4 (comme aussi l'expérience 3) a été organisée afin d'étudier la dégradation éventuelle de la projection conventionnelle en 2-D par rapport à la projection des images en 3-D.

L'expérience 4 est identique à l'expérience 6, à vues changeantes, la seule différence est que la projection des images est réalisée en 2-D. Pour la description de sa méthode, il faut donc se référer à l'expérience 6 (cf. § 4.9.2).

### 4.7.2 Résultats

La **figure 4.17** donne la distance auditive des stimuli sonores pour chacune des conditions visuelles, séparément pour les deux présentations des stimuli. Le **tableau 4.9** présente les résultats de l'Analyse de la Variance.

Ces résultats ne montrent pas de différences significatives dues aux conditions visuelles ( $[F(2,30) = 1.22, p = 0.311]$ ). L'interaction entre conditions visuelles et stimuli auditifs se révèle par contre significative, indiquant qu'en effet les conditions visuelles influencent la distance auditive perçue, mais que cette influence dépend fortement des stimuli. De faibles différences sont par ailleurs observées entre les conditions visuelles V2 et V3 à la 1<sup>ère</sup> présentation des stimuli, ces différences disparaissent à la 2<sup>ème</sup> présentation.

### 4.7.3 Conclusion

En conclusion, les mêmes résultats qu'à l'expérience 3 sont retrouvés, c'est à dire que la simulation de la distance des champs visuels à l'aide d'une projection conventionnelle (2-D) semble être moins convainquante que celle obtenue à l'aide de la projection en 3-D ; les effets principaux dus aux

conditions visuelles et obtenus avec la projection 2-D ne sont pas significatifs . L'utilisation de la projection 3-D semble donc nécessaire pour le rendement d'une impression visuelle de distance.

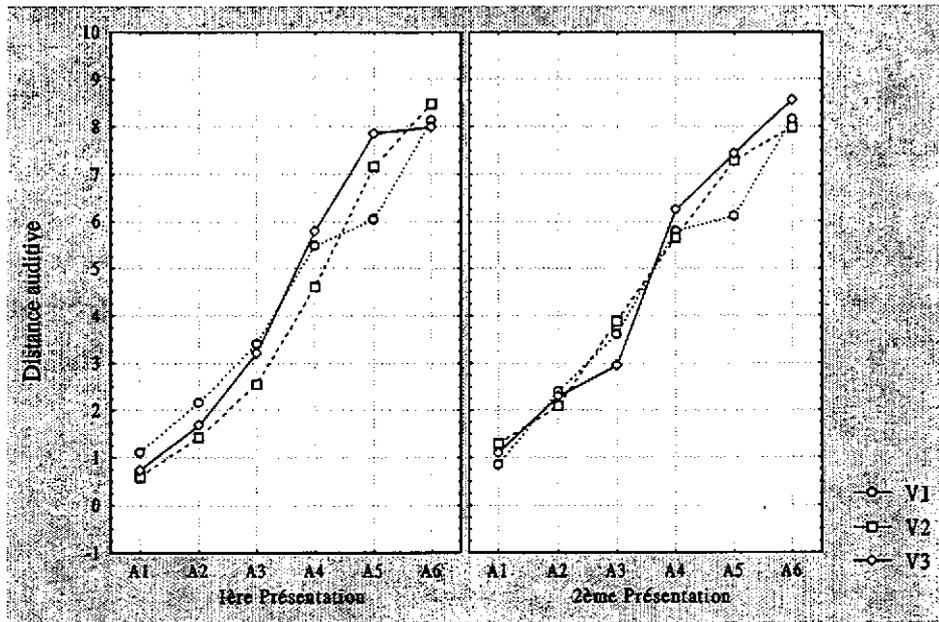


Figure 4.17 : Distance auditive des stimuli sonores pour chacune des conditions visuelles, séparément pour les deux présentations des stimuli.

Table de tous les effets						
1 : Conditions Visuelles, 2 : Présentations de stimuli, 3 : Stimuli Auditifs						
	df Effet	MS Effet	df Erreur	MS Erreur	F	Niveau-p
1	2	3,34	30	2,75	1,22	0,311
2	1	12,24	15	3,90	3,14	0,097
3	5	796,35	75	2,03	392,34	* 0,000
12	2	2,73	30	1,42	1,91	0,165
13	10	5,60	150	1,49	3,75	* 0,000
23	5	1,76	75	1,57	1,12	0,357
123	10	1,81	150	1,82	0,99	0,452

Tableau 4.9 : Tableau des résultats de l'Analyse de la Variance pour l'expérience 4 (expérience à vues changeantes en projection 2-D). Les trois facteurs sont : les conditions visuelles avec 3 niveaux V1, V2 et V3 (facteur 1), les présentations des stimuli avec 2 niveaux P1 et P2 (facteur 2) et les stimuli auditifs avec 6 niveaux A1-A6 (facteur 3). 12, 13, 23 et 123 représentent les interactions entre les facteurs respectifs. Les effets significatifs sont signalés par un astérisque (\*).

## 4.8 EXPERIENCE 5 : INFLUENCE VISUELLE (VUES STATIQUES EN PROJECTION 3-D)

- *Test à écoute transaurale - Vues statiques*
- *2 présentations de stimuli*
- *Signal sonore : « La Somnambule » de Bellini interprété par Phyllis Bryn-Jullson  
(Stimuli sonores : A1, A2, A3, A4, A5 et A6).*
- *Vues : « Théâtre des Champs Elysées » à Paris  
(Stimuli visuels : condition visuelle de contrôle V0, images 3-D : V1 et V3).*
- *28 sujets*

### 4.8.1 Objectif

La cinquième expérience a été réalisée afin d'étudier l'influence de la distance visuelle sur la distance auditive apparente et de valider les résultats obtenus aux expériences précédentes. Les différents choix expérimentaux effectués à l'aide des expériences précédentes (comme la méthode de diffusion du son, le type du signal, les configurations sonores, le nombre de présentations des stimuli, la méthode de projection des images, etc.) sont mis en œuvre dans le cadre de cette expérience.

La méthode utilisée, en *vues statiques* (cf. § 4.2.6), permet de se rapprocher des conditions visuelles dans lesquelles se trouvent les auditeurs-spectateurs de vraies salles de concert ; ces conditions restent en effet généralement inchangées pendant une représentation.

### 4.8.2 Méthode

1. *Stimuli sonores.* Les stimuli sonores utilisés pour cette expérience ont été choisis lors de l'expérience 1 ; ils sont d'ailleurs les mêmes que ceux utilisés aux expériences principales 2, 3, 4 et 6. Leur description détaillée est présentée dans l'annexe B (tableaux B.7 et B.8).

Les six stimuli ont été créés à partir d'une réponse impulsionnelle du « Théâtre des Champs Elysées » à Paris, comme l'explique le chapitre correspondant (cf. § 3.2.5). Ils varient sur l'énergie précoce  $E_s$  : une différence de 3 dB existe entre chaque stimulus successif ; A1, A2, A3, A4, A5 et A6 rendent ainsi une impression de distance auditive croissante.

Le signal sonore était « *La Somnambule* » de Bellini. Le son était enregistré et diffusé sur DAT (Digital Audio Tape). La diffusion du son était transaurale sur haut-parleurs.

2. *Conditions visuelles.* Les conditions visuelles de cette 5<sup>ème</sup> expérience principale sont également celles utilisées aux expériences 2 et 3 : le champ visuel est créé par la projection, depuis l'arrière de

l'écran, de grandes images tridimensionnelles<sup>1</sup> (taille de l'image=150 cm x 100 cm) dans des conditions d'obscurité complète (cf. § 3.3). Les vues V1 et V3 utilisées dans ces tests ont été prises dans le « Théâtre des Champs Elysées » à Paris, à différentes distances de la scène, au milieu de laquelle était placé un haut-parleur (cf. § 3.3.2). Pour la condition visuelle de contrôle V0 aucune image n'était projetée sur l'écran, seule une lumière diffuse y était perçue (cf. § 4.2.5). Les sujets sont assis à la distance de 1,6 m de l'écran de projection.

3. *Sujets, consignes.* Vingt-huit adultes ont participé à cette expérience (cf. § 4.2.2). La moitié d'entre eux a d'abord participé à la 6<sup>ème</sup> expérience principale en *vues changeantes*. Un petit nombre d'entre eux avait participé à certains des tests précédents avant ces deux tests finaux ; pour la plupart des sujets, il s'agissait de leur première expérience de test d'écoute.

Le réglage subjectif du dispositif transaural a lieu avant chaque session de tests ; la présentation des sons de référence vient ensuite (stimuli extrêmes A1 et A6) (cf. § 4.2.1 et 4.2.3).

Lors de la partie principale des tests, la tâche des sujets consiste à juger la distance auditive et de répondre oralement dans une échelle arbitraire comprise entre 1 et 10 (1 représentant la distance auditive la plus courte et 10 la plus grande). Ils ont pour consigne d'utiliser le son et l'image (dans le cas où il y a une image) afin de s'imaginer dans la "salle", mais de ne juger que la distance auditive (cf. § 4.2.2).

4. *Procédure expérimentale.* La méthodologie est largement empruntée à la psychologie expérimentale et notamment au domaine traitant des interactions sensorielles : les six stimuli auditifs sont présentés deux fois dans un ordre pseudo-aléatoire, créant ainsi des *séquences* de douze stimuli (cf. figure 4.1 pour n=2). Afin de compenser les effets d'ordre de présentation des stimuli, sept séquences ont été utilisées ; chaque séquence a été présentée à quatre des sujets sur les vingt huit. Ces séquences sont présentées à l'annexe B. La règle générale pour la création des séquences fut d'éviter l'apparition accidentelle d'un couple de sons dans un certain ordre bien plus de fois que pour les autres couples de sons (cf. 4.2.7).

Il s'agit d'un test en vues statiques, c'est à dire que les images restent stables lors la présentation des stimuli auditifs (cf. § 4.2.6). Le même son n'apparaît jamais deux fois à la suite. Chaque sujet subit ainsi la même séquence trois fois, une fois pour chacune des trois conditions visuelles :

- Condition visuelle de contrôle  $V_{\text{contrôle}}$  ou V0
- Vue proche de la scène  $V_{\text{proche}}$  ou V1
- Vue lointaine de la scène  $V_{\text{loin}}$  ou V3

Chaque sujet participe ainsi en tout à trois tests de 12 stimuli chacun ( $12 \times 3 = 36$  stimuli en tout). Six ordres de présentation des conditions visuelles ont été utilisés : V0-V1-V3, V1-V3-V0, V3- V0-V1, V3-V1-V0, V1-V0-V3 et V0-V3-V1 (respectivement présentés à 5, 5, 4, 5, 4 et 5 sujets).

La *durée* moyenne des tests était de 25 min. à 30 min. (8 à 10 min. par condition visuelle).

---

<sup>1</sup> Sauf pour l'expérience 3, où la projection réalisée était en 2-D.

### 4.8.3 Résultats

Les réponses des 28 sujets sur la distance apparente auditive sont normalisées de la manière décrite au § 4.2.8 ( $\bar{x} = 4,5$  et  $\sigma = 3$ ). Les moyennes et les écart-types des réponses sur la distance auditive pour les différentes conditions expérimentales (3 conditions visuelles x 2 présentations des stimuli x 6 stimuli auditifs = 36 réponses) sont présentés en annexe B.

Le **tableau 4.10** présente les résultats de l'Analyse de la Variance (cf. § 3.6.1.4), le **tableau 4.11** donne les valeurs moyennes pour chacun des stimuli auditifs et pour chaque condition visuelle. La **figure 4.18** donne une représentation graphique des ces résultats, séparément pour les deux présentations des stimuli.

Table de tous les effets						
1 : Conditions Visuelles, 2 : Présentations de stimuli, 3 : Stimuli Auditifs						
	df Effet	MS Effet	Df Erreur	MS Erreur	F	Niveau-p
1	2	4,21	54	2,19	1,92	0,157
2	1	30,83	27	1,50	20,62	* 0,000
3	5	1541,40	135	1,74	886,30	* 0,000
12	2	2,91	54	1,22	2,38	0,102
13	10	0,46	270	0,75	0,61	0,805
23	5	1,39	135	1,29	1,08	0,375
123	10	0,87	270	0,78	1,11	0,353

**Tableau 4.10** : Tableau des résultats de l'Analyse de la Variance pour la cinquième expérience en vues statiques. Les trois facteurs sont : les *conditions visuelles* avec 3 niveaux V0, V1 et V3 (facteur 1), les *présentations des stimuli* avec 2 niveaux P1 et P2 (facteur 2) et les *stimuli auditifs* avec 6 niveaux A1-A6 (facteur 3). 12, 13, 23 et 123 représentent les interactions entre les facteurs respectifs. Les effets significatifs sont signalés par un astérisque (\*)

	V0	V1	V3
A1	0.67	0.59	0.75
A2	1.88	1.67	1.89
A3	3.53	3.40	3.70
A4	5.24	5.30	5.59
A5	7.25	6.96	7.26
A6	8.48	8.34	8.40

**Tableau 4.11** : Expérience 5 en vues statiques : valeurs moyennes de la distance auditive pour chacun des stimuli et pour chacune des conditions visuelles. Les réponses sont moyennées sur les 28 sujets et sur les 2 présentations des stimuli.

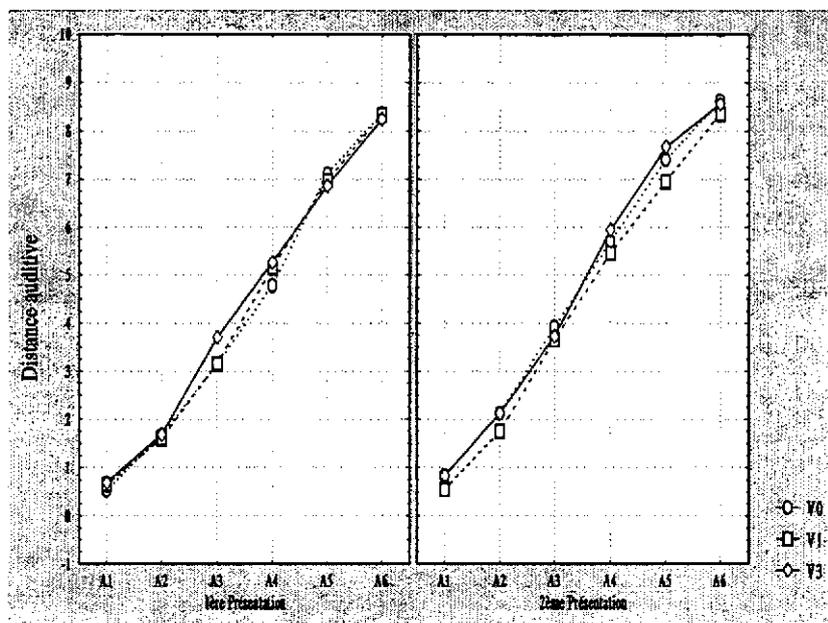


Figure 4.18 : Distance auditive apparente des stimuli auditifs en fonction des conditions visuelles séparément pour les deux présentations de stimuli.

L'Analyse de la Variance révèle un effet significatif des *stimuli auditifs*, un effet significatif de la *présentation* mais ne révèle pas d'effet des *conditions visuelles*. Les deux premiers effets étaient attendus, ils vérifient les résultats trouvés aux expériences précédentes. En ce qui concerne l'effet des conditions visuelles, on observe à l'aide du tableau 4.11 que c'est la condition visuelle de contrôle qui semble empêcher l'influence visuelle de se manifester (voir également l'analyse qui suit). En effet, les résultats montrent que les réponses de distance auditive sous la condition visuelle V1 sont inférieures aux réponses sous la condition visuelle V3. Or, rappelons que le but de cette expérience est d'étudier l'influence de la distance visuelle sur la distance auditive, et non pas l'influence d'une condition visuelle dont le contenu cognitif – ou le sens qu'elle évoque – ne serait pas contrôlé ; on entend par "contrôlée", une condition visuelle qui serait similaire aux autres conditions visuelles testées. Pour ces raisons, mais aussi afin de pouvoir comparer les résultats des deux dernières expériences sur la perception de la distance<sup>1</sup>, ce sont surtout les jugements de distance auditive sous les conditions visuelles V1 et V3 dont la comparaison nous intéresse.

Une analyse prenant en compte uniquement les réponses obtenues pour les *conditions visuelles* V1 et V3 a donc été réalisée. Les résultats sont présentés au **tableau 4.12**. Elle révèle des effets significatifs pour les trois facteurs expérimentaux : les stimuli sonores, les présentations des stimuli et les conditions visuelles.

<sup>1</sup> Comparaison de l'expérience 5 avec des vues statiques (vues V0, V1 et V3) et de l'expérience 6 en vues changeantes (vues V1, V2 et V3) : voir 'Discussion générale' (§ 4.10).

Table de tous les effets						
1 : Conditions Visuelles, 2 : Présentations de stimuli, 3 : Stimuli Auditifs						
	df Effet	MS Effet	df Erreur	MS Erreur	F	Niveau-p
1	1	8,31	27	1,45	5,73	* 0,024
2	1	12,48	27	1,45	8,63	* 0,007
3	5	1021,63	135	1,50	681,01	* 0,000
12	1	2,82	27	1,23	2,29	0,142
13	5	0,28	135	0,65	0,43	0,830
23	5	0,77	135	1,16	0,66	0,652
123	5	1,25	135	0,67	1,86	0,105

**Tableau 4.12** : Résultats de l'Analyse de la Variance pour la 5<sup>ème</sup> expérience en vues statiques. Les trois facteurs sont : les *conditions visuelles* avec 2 niveaux V1 et V3 (facteur 1), les *présentations des stimuli* avec 2 niveaux P1 et P2 (facteur 2) et les *stimuli auditifs* avec 6 niveaux A1-A6 (facteur 3). 12, 13, 23 et 123 représentent les interactions entre les facteurs respectifs. Les effets significatifs sont signalés par un astérisque (\*)

La **figure 4.19** montre les réponses de distance auditive moyennées sur les stimuli auditifs pour chacune des présentations et pour chaque condition visuelle.

La **figure 4.20** représente les valeurs moyennes de distance auditive pour chacun des stimuli auditifs, en fonction des présentations et des conditions visuelles (il s'agit en fait d'une «moyenne» des 6 graphes de la figure 4.19).

Ces deux figures nous permettent d'observer deux effets marquants :

- i) les réponses sur la distance auditive augmentent globalement à la deuxième présentation des stimuli.
- ii) les différences de jugement de la distance, dues aux conditions visuelles, augmentent également à la 2<sup>ème</sup> présentation des stimuli pour tous les stimuli auditifs (sauf pour le stimulus A3) :  $F(1,27) = 0.46$ ,  $p=0.5$  pour la première présentation seule, tandis que  $F(1,27) = 9.22$ ,  $p<0.01$  pour la seconde présentation.

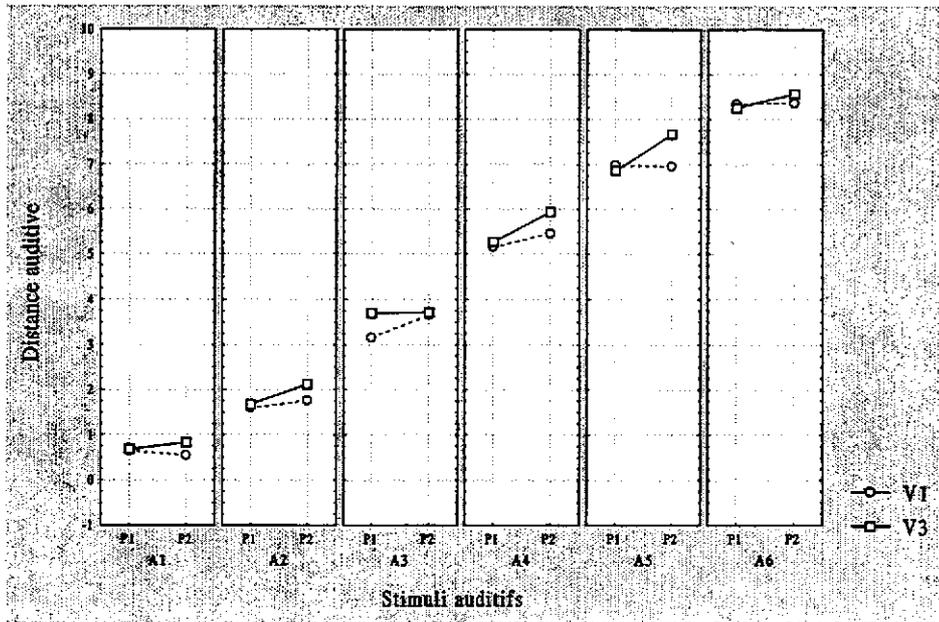


Figure 4.19 : Distance auditive apparente pour les présentations de stimuli P1 et P2 en fonction des conditions visuelles V1 et V3.

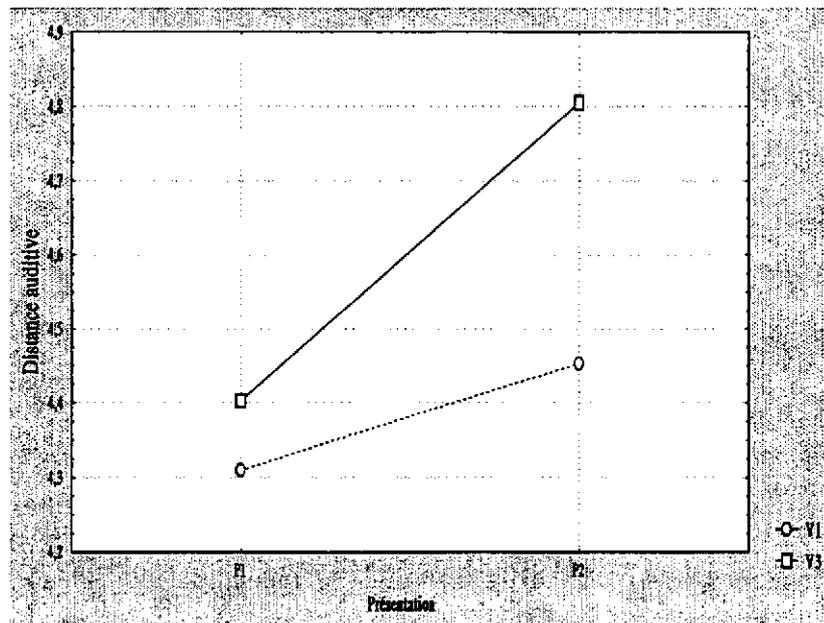


Figure 4.20 : Distance auditive apparente des stimuli auditifs en fonction des conditions visuelles V1 et V3 et des présentations de stimuli P1 et P2. Les réponses sont moyennées sur les sujets et sur les stimuli auditifs.

#### 4.8.4 Discussion

L'absence d'interaction entre les facteurs expérimentaux (stimuli auditifs, présentations des stimuli et conditions visuelles) nous permet de dissocier et de discuter séparément leurs effets :

##### i) Les stimuli auditifs

Nous constatons avant tout l'influence très importante des stimuli sonores sur les réponses ; il s'agit d'un effet tout à fait attendu puisque les stimuli ont été créés afin de représenter différentes impressions de distance auditive.

Par ailleurs, les stimuli auditifs semblent être perçus de manière très cohérente : les distances perceptives moyennes entre les stimuli successifs, données au **tableau 4.13**, révèlent une bonne distribution des réponses sur l'échelle proposée. Un faible effet de « polarisation » est observé, c'est à dire qu'il semble exister une tendance à donner des réponses plus proches des réponses données aux stimuli extrêmes. Les écarts de distances entre A1 et A2 et entre A5 et A6 semblent être en effet plus petits que les écarts entre les stimuli du milieu A2-A3, A3-A4 et A4-A5.

Ce phénomène pourrait être lié au « besoin de repères » pour la perception de la distance, il en résulte une tendance à juger par rapport aux distances de référence que sont les stimuli extrêmes A1 et A6 (cf. § 2.2.3). Il a en effet été constaté que l'utilisation de différentes distances de référence peut influencer les réponses et entraîner des jugements proches de ces distances de référence [Cochran et al., 1968].

	<i>A1-A2</i>	<i>A2-A3</i>	<i>A3-A4</i>	<i>A4-A5</i>	<i>A5-A6</i>
<i>V0</i>	1.21	1.65	1.71	2.01	1.23
<i>V1</i>	1.08	1.73	1.90	1.66	1.38
<i>V3</i>	1.14	1.81	1.89	1.67	1.14
<i>Global</i>	1.14	1.74	1.83	1.78	1.25

**Tableau 4.13 :** Ecarts de distances « perceptives » moyens entre les stimuli successifs pour chacune des conditions visuelles et pour l'ensemble des résultats

On peut ainsi conclure que la simulation de la distance auditive par le Spatialisateur – et aussi les choix expérimentaux effectués lors des expériences précédentes concernant le son (méthode de diffusion, signal et configurations sonores) – sont très satisfaisants et semblent bien adaptés pour ce type d'expérience.

##### ii) Les présentations des stimuli

Les réponses sur la distance auditive sont nettement inférieures à la 1<sup>ère</sup> présentation des stimuli ; on observe une augmentation globale significative à la 2<sup>ème</sup> présentation. C'est un effet valable pour l'ensemble des stimuli et des conditions visuelles (il est un peu moins prononcé pour les stimuli auditifs de référence A1 et A6 et pour la condition visuelle proche V1). Il s'agit d'un effet déjà observé aux tests précédents et étudié lors de l'expérience 2 : les jugements augmentaient à la

deuxième présentation et se stabilisaient ensuite. Le choix de deux présentations de stimuli dans ce test est justement issu des résultats de cette expérience 2 (cf. § 4.5).

L'augmentation des réponses sur la distance auditive avec le temps est un phénomène observé également par d'autres études sur la perception de la distance : Coleman [1962] trouve que les réponses à la 1<sup>ère</sup> présentation des stimuli sont sérieusement sous-estimées et Mershon et King [1975] introduisent la notion de «*tendance de distance spécifique*» (cf. § 2.2.1).

Au début des tests, les sujets semblent hésiter à donner des jugements de distance lointaine, ils ont tendance à réserver ces jugements pour des sons qui risqueraient éventuellement d'apparaître encore plus lointains. Ils répondent ainsi de manière réservée, dans l'expectative des stimuli à venir et en se référant peut-être à leur position (correspondant à une distance minimale). Dès qu'ils se sentent maîtriser les sons, et surtout connaître la plage de variation des signaux sonores, ils commencent à donner des jugements de distances plus lointaines en utilisant toute l'échelle proposée.

### iii) Influence des conditions visuelles

Premièrement, nous considérons que l'effet de la condition visuelle de contrôle sur les réponses ne peut pas vraiment être pris en compte, ni être comparé aux réponses sous une autre condition puisque le contenu cognitif de cette condition – sa signification – n'est pas contrôlé ; c'est à dire qu'elle n'est pas similaire à celle des autres conditions visuelles. En effet, une diapositive orange uniforme (condition visuelle de contrôle V0) n'évoque pas la même chose chez les sujets que les images de la scène d'un théâtre (conditions visuelles V1 et V3).

En ce qui concerne les résultats obtenus sous les conditions visuelles V1 et V3, ils montrent que l'effet des conditions visuelles est très faible à la 1<sup>ère</sup> présentation des stimuli, mais qu'il augmente et devient significatif à la 2<sup>ème</sup> présentation.

Nous pourrions supposer qu'au début des expériences, les sujets sont moins sûrs de leur capacité à juger la distance, ils sont très attentifs au son et se repèrent à l'aide des signaux de référence qui sont récents. Visuellement au début des tests, leur environnement immédiat (leur emplacement, la salle des tests, l'écran, la taille de l'image, la projection 3-D, etc.) attire leur attention plus que l'image projetée par elle-même. Cette concentration initiale de la part des sujets sur la tâche auditive et ce manque d'intérêt relatif aux images projetées, sont à l'origine des jugements peu influencés par les conditions visuelles lors de la 1<sup>ère</sup> présentation des stimuli.

Avec l'avancement des tests, les sujets se sentent plus confiants dans leurs réponses (un indice le traduit : les écart-types des réponses baissent à la 2<sup>ème</sup> présentation des stimuli). Ils répondent de façon plus cohérente, hésitent moins et font moins 'd'erreurs'. De plus, les sons de référence sont plus lointains dans le temps et – à cause des limites de la mémoire auditive – moins faciles à utiliser comme repères. En ce qui concerne les conditions visuelles, après avoir examiné rapidement leur environnement immédiat, les sujets semblent s'affranchir de la salle des tests et de l'écran, ils ont tendance à fixer les yeux sur les images projetées (chaque image reste projetée pendant environ 8 à 10 min.), ils imaginent alors plus facilement la salle simulée et se repèrent partiellement à l'aide des distances variables de la scène. Les images fournissent un cadre spatial de référence pour les jugements auditifs, et c'est à ce moment, à la 2<sup>ème</sup> présentation de stimuli et alors que les sujets semblent plus confiants dans leurs réponses, que l'on observe une influence significative de la distance visuelle sur les jugements de la distance auditive.

Les différences observées sur la perception de la distance auditive pour différentes distances visuelles sont du type 'offset', c'est à dire qu'elles sont similaires pour tous les stimuli auditifs et ne dépendent pas de la distance auditive de ces stimuli. Les différences sont petites mais significatives. Elles

apparaissent également robustes puisqu'elles ont été observées également dans tous les tests précédents (sauf à l'expérience 3 où la projection des images était conventionnelle en 2-D). Le **tableau 4.14** montre ces différences pour la 2<sup>ème</sup> présentation des stimuli.

V3-V1	A1	A2	A3	A4	A5	A6
2 <sup>ème</sup> présentation	0.28	0.36	0.07	0.48	0.71	0.20

**Tableau 4.14 :** Différences perceptives observées sur la perception de la distance auditive entre les conditions visuelles V3 (vue lointaine) et V1 (vue proche) à la 2<sup>ème</sup> présentation des stimuli sonores.

Nous observons par ailleurs, à l'aide de la figure 4.18 et du tableau 4.13, que la distance auditive est représentée par une relation linéaire entre les deux échelles (1.14 à 1.83 'unités perceptives' en global pour 3 dB de différence sur l'énergie précoce  $E_s$  des stimuli successifs). Grâce à cette linéarité et dans le souci d'une évaluation objective de l'effet dû à la perception visuelle, nous faisons la considération suivante : les 8 'unités perceptives' environ, observées entre les stimuli extrêmes A1 et A6, correspondent à une variation de 15 dB de l'indice  $E_s$  (figure 4.18). Une différence de 1 'unité perceptive' correspondrait donc à environ 2 dB de variation de cet indice, tandis que les différences d'une demie 'unité perceptive', observées au tableau 4.14, correspondraient à une variation de 1 dB de l'indice objectif ; cet écart correspond à la plus petite différence perceptible de l'indice  $E_s$  (cf. § 3.2.4). Le rapport perceptif entre les distances visuelles V3 et V1 qui a induit ces différences perceptives est de l'ordre de  $V3 : V1 = 6.66 : 1.65 = 4$  (distances visuelles estimées en fonction de la loi inverse de la taille du haut-parleur mesurée sur l'écran, cf. § 3.3.2).

## 4.9 EXPERIENCE 6 : INFLUENCE VISUELLE (VUES CHANGEANTES EN PROJECTION 3-D)

- *Test à écoute transaurale - Vues changeantes*
- *2 présentations de stimuli*
- *Signal sonore : « La Somnambule » de Bellini interprété par Phyllis Bryn-Jullson (Stimuli sonores : A1, A2, A3, A4, A5 et A6).*
- *Vues : « Théâtre des Champs Elysées » à Paris (Stimuli visuels : images 3-D : V1, V2 et V3).*
- *28 sujets*

### 4.9.1 Objectif

Comme la cinquième expérience, la sixième expérience – utilisant cette fois la méthode *en vues changeantes* (cf. § 4.2.6) – a été réalisée afin d'étudier l'influence de la distance visuelle sur la distance auditive apparente et de valider les effets observés aux expériences précédentes.

La méthode avec *vues changeantes* vis-à-vis de celle à *vues statiques*, se rapproche beaucoup moins de conditions visuelles de vraies salles de concert ; en effet, ces conditions restent généralement inchangées pour les auditeurs-spectateurs pendant une représentation. En revanche, elle se rapproche un peu plus des conditions visuelles des environnements virtuels où l'apparition d'une image peut être immédiate et le plus souvent de courte durée : comme c'est le cas pour la télévision ou les jeux vidéo, où souvent la différence est que ces changements sont dynamiques, c'est à dire qu'il y a du mouvement

L'objectif est donc d'étudier l'influence de la distance visuelle sur la distance auditive dans un autre contexte expérimental que celui de l'expérience précédente, et évidemment, de permettre la comparaison des résultats.

### 4.9.2 Méthode

Les seules différences en ce qui concerne la méthode entre les deux dernières expériences principales – celle *en vues statiques* et celle *en vues changeantes* – se trouvent au niveau du choix des stimuli visuels et de la procédure expérimentale. Les stimuli sonores, les sujets et les consignes sont les mêmes, ils ne sont pas rappelés ici (cf. § 4.8.2). Seules les conditions visuelles et la procédure expérimentale de ce test sont présentées.

*Conditions visuelles.* Les conditions visuelles de cette expérience sont les mêmes que celles utilisées à l'expérience 4 : le champ visuel est créé par la projection, depuis l'arrière de l'écran, de grandes images tridimensionnelles<sup>1</sup> (taille de l'image=150 cm x 100 cm) dans des conditions d'obscurité complète (cf. § 3.3). Les photographies V1, V2 et V3 utilisées dans ces tests ont été prises dans le « Théâtre des Champs Elysées » à Paris, à différentes distances de la scène, au milieu de laquelle était placé un haut-parleur (cf. § 3.3.2). Les sujets sont assis à la distance de 1,6 m de l'écran de projection.

*Procédure expérimentale.* Les six stimuli auditifs A1-A6 sont couplés avec les trois images V1, V2 et V3 pour créer des séquences pseudo-aléatoires de dix huit stimuli bimodaux. Deux de ces séquences sont présentées à chaque sujet qui a ainsi jugé 36 stimuli en tout (cf. figure 4.2). Afin de compenser les effets d'ordre de présentation des stimuli, sept séquences de 36 stimuli ont été utilisées, chaque séquence a été présentée à quatre sujets (sur les 28 sujets). Elles sont présentées à l'annexe B. La règle générale pour la création des séquences fut d'éviter l'apparition accidentelle d'un couple de sons dans un certain ordre bien plus de fois que pour les autres couples de sons (cf. 4.2.7).

Il s'agit d'un test *avec vues changeantes*, c'est à dire que les images changent avec les stimuli auditifs (cf. § 4.2.6). Ni le même son, ni la même image n'apparaissent deux fois à la suite. L'apparition du son et de l'image est synchrone.

La durée moyenne des tests était de 25 à 30 min..

### 4.9.3 Résultats

Les réponses des 28 sujets sur la distance apparente auditive sont normalisées comme décrit au § 4.2.8 ( $\bar{x} = 4,5$  et  $\sigma = 3$ ). Les moyennes et les écart-types des réponses sur la distance auditive pour les différentes conditions expérimentales (3 conditions visuelles x 2 présentations des stimuli x 6 stimuli auditifs = 36 réponses) sont présentées en annexe B.

Le **tableau 4.15** présente les résultats de l'Analyse de la Variance (cf. § 3.6.1.4), le **tableau 4.16** et la **figure 4.21** présentent les valeurs moyennes des distances auditives pour chacun des stimuli auditifs et pour chacune des conditions visuelles.

---

<sup>1</sup> Pour l'expérience 4, la projection des images était en 2-D.

Table de tous les effets						
1 : Conditions Visuelles, 2 : Présentations de stimuli, 3 : Stimuli Auditifs						
	df Effet	MS Effet	df Erreur	MS Erreur	F	Niveau-p
1	2	18,58	54	0,85	22,0	* 0,000
2	1	27,86	27	4,51	6,2	* 0,019
3	5	1507,79	135	1,28	1178,0	* 0,000
12	2	5,78	54	1,37	4,2	* 0,020
13	10	2,79	270	0,95	3,0	* 0,002
23	5	3,08	135	1,15	2,7	* 0,024
123	10	1,84	270	1,18	1,6	0,118

**Tableau 4.15 :** Résultats de l'Analyse de la Variance pour la sixième expérience en vues changeantes.

Les trois facteurs sont : les *conditions visuelles* avec 3 niveaux V1, V2 et V3 (facteur 1), les *présentations des stimuli* avec 2 niveaux P1 et P2 (facteur 2) et les *stimuli auditifs* avec 6 niveaux A1-A6 (facteur 3). 12, 13, 23 et 123 représentent les interactions entre les facteurs respectifs. Les effets significatifs sont signalés par un astérisque (\*)

	V1	V2	V3
A1	0.73	0.92	0.94
A2	1.72	1.87	1.80
A3	2.95	3.57	3.64
A4	4.79	5.38	5.87
A5	7.06	6.92	7.54
A6	8.26	8.44	8.53

**Tableau 4.16 :** Expérience 6 en vues changeantes : valeurs moyennes de la distance auditive pour chacun des stimuli et pour chaque condition visuelle (les réponses sont moyennées sur les 28 sujets et sur les 2 présentations des stimuli).

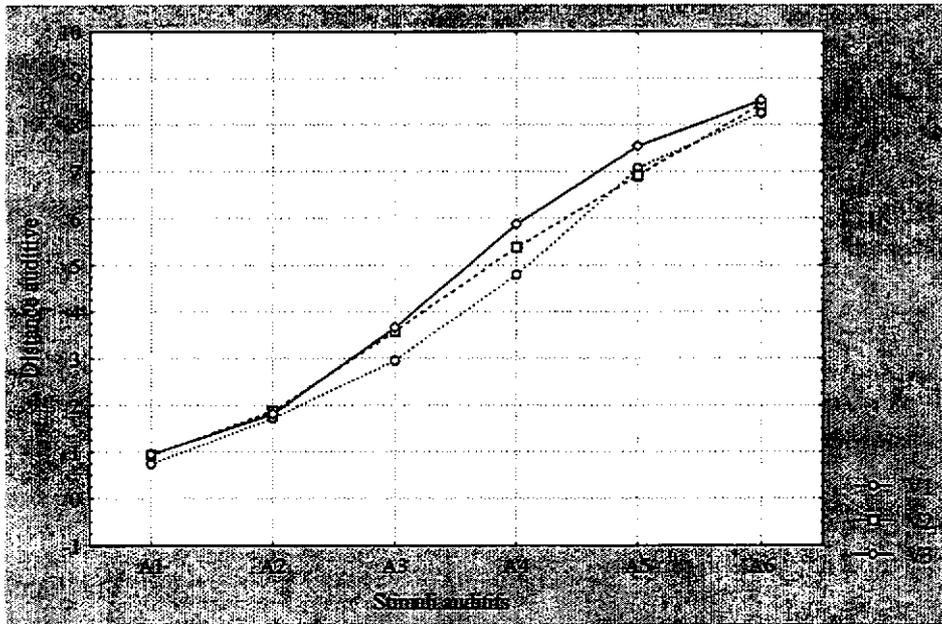


Figure 4.21 : Distance auditive apparente des stimuli auditifs en fonction des conditions visuelles.

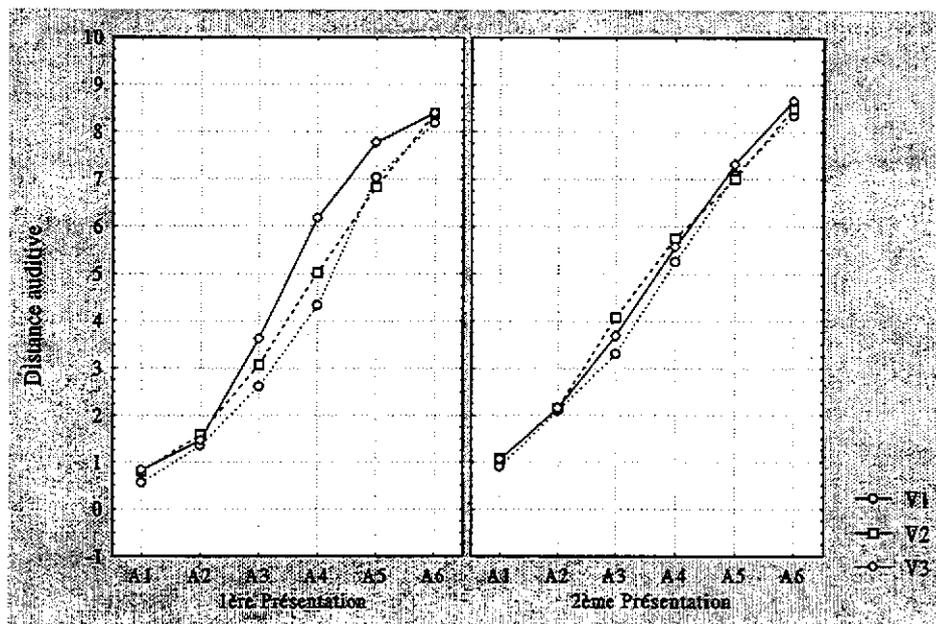
L'analyse de variance révèle des effets significatifs des stimuli auditifs, des conditions visuelles et des présentations des stimuli ; elle révèle aussi des effets d'interaction entre les trois facteurs pris deux par deux (interactions à deux facteurs).

En effet, on observe à la **figure 4.21** que la distance auditive varie avec les stimuli auditifs et semble également être influencée par les conditions visuelles. Le **tableau 4.17** présente les moyennes sur la distance auditive pour chacune des conditions visuelles et chacune des présentations des stimuli, elle présente également des comparaisons Post-Hoc (Tukey HSD) qui révèlent des différences significatives entre les niveaux de ces facteurs.

Test HSD Tukey						
Probabilités des Tests Post Hoc						
EFFET PRINCIPAL : <i>condition visuelle</i>				EFFET PRINCIPAL : <i>présentation</i>		
	V1	V2	V3		P1	P2
Moyennes	4,26	4,52	4,72	Moyennes	4,33	4,66
1		* 0,001	* 0,000	1		* 0,000
2	* 0,001		* 0,016	2	* 0,000	
3	* 0,000	* 0,016				

Tableau 4.17 : Moyennes et comparaisons Post-Hoc (Tukey HSD) des trois niveaux des *conditions visuelles* et des deux niveaux des *présentations de stimuli*.

Les réponses augmentent avec la distance visuelle (V1, V2 et V3 correspondant respectivement à des vues proche, moyenne et lointaine de la scène) ; elles augmentent aussi à la 2<sup>ème</sup> présentation des stimuli. L'existence des interactions entre la *présentation* et les deux autres facteurs, nous oblige à observer les résultats de manière plus analytique : la **figure 4.22** montre ainsi la distance auditive des stimuli auditifs pour chacune des conditions visuelles, séparément pour les deux présentations des stimuli.



**Figure 4.22** : Distance auditive des stimuli auditifs pour chacune des conditions visuelles, séparément pour les deux présentations des stimuli.

Ce graphique montre que les différences dues aux conditions visuelles sont observées principalement à la 1<sup>ère</sup> présentation des stimuli, tandis qu'à la 2<sup>ème</sup> présentation elles semblent diminuer fortement (voir aussi figures 4.23 et figure 4.24). Les stimuli auditifs du milieu semblent être les plus influencés.

Afin de pouvoir comparer les résultats des deux dernières expériences principales sur la perception de la distance<sup>1</sup>, une analyse prenant en compte uniquement les réponses obtenues pour les conditions visuelles V1 et V3 a été réalisée. Les résultats sont présentés au **tableau 4.18**. Cette analyse révèle des effets significatifs pour les stimuli sonores et les conditions visuelles. Elle ne révèle pas d'effet significatif pour les présentations, malgré le fait que les variations soient importantes (cf. **figure 4.23** et **figure 4.24**).

<sup>1</sup> Voir chapitre suivant : 'Discussion générale' (§ 4.10).

Table de tous les effets						
1 : Conditions Visuelles, 2 : Présentations de stimuli, 3 : Stimuli Auditifs						
	df Effet	MS Effet	df Erreur	MS Erreur	F	Niveau-p
1	1	36,96	27	0,93	39,82	* 0,000
2	1	11,37	27	3,11	3,66	0,067
3	5	1031,48	135	1,20	861,18	* 0,000
12	1	8,92	27	1,74	5,13	* 0,032
13	5	3,81	135	0,84	4,52	* 0,001
23	5	2,48	135	0,93	2,67	* 0,025
123	5	2,55	135	1,25	2,04	0,076

**Tableau 4.18 :** Tableau des résultats de l'Analyse de la Variance pour la sixième expérience en vues changeantes. Les trois facteurs sont : les *conditions visuelles* avec 2 niveaux V1 et V3 (facteur 1), les *présentations des stimuli* avec 2 niveaux P1 et P2 (facteur 2) et les *stimuli auditifs* avec 6 niveaux A1-A6 (facteur 3). 12, 13, 23 et 123 représentent les interactions entre les facteurs respectifs. Les effets significatifs sont signalés par un astérisque (\*).

La **figure 4.23** représente les réponses moyennes de la distance auditive pour chacun des stimuli auditifs, en fonction des présentations et des conditions visuelles. On observe encore que les différences dues aux conditions visuelles sont bien plus importantes pour la 1<sup>ère</sup> présentation des stimuli et sont surtout observées pour les stimuli du milieu A3, A4 et A5. Ceci est encore plus visible à la **figure 4.24** qui montre les réponses de distances auditives moyennées sur les stimuli auditifs, pour chacune des présentations et pour chaque condition visuelle (il s'agit en fait d'une «moyenne» des 6 graphes de la **figure 4.23**).

Ces deux figures nous permettent d'observer deux effets marquants :

- i) les réponses sur la distance auditive augmentent globalement à la deuxième présentation des stimuli ; ce sont surtout les réponses sous la condition visuelles V1 qui augmentent ;
- ii) les différences dues aux conditions visuelles sont beaucoup plus marquantes à la 1<sup>ère</sup> présentation des stimuli (surtout pour les stimuli auditifs A3, A4 et A5). En effet, une analyse prenant en compte seulement ces premières réponses révèle un effet des conditions visuelles :  $[F(1, 27) = 20.92, p < 0.01]$ , tandis qu'une analyse prenant en compte uniquement les réponses à la 2<sup>ème</sup> présentation révèle un effet moins important des conditions visuelles :  $[F(1,27) = 6.82, p < 0.05]$ <sup>1</sup>.

<sup>1</sup> Il s'agit pour cette deuxième présentation, d'une influence moins importante que celle observée en vues statiques (expérience 5) où  $[F(1,27) = 9.22, p < 0.01]$ .

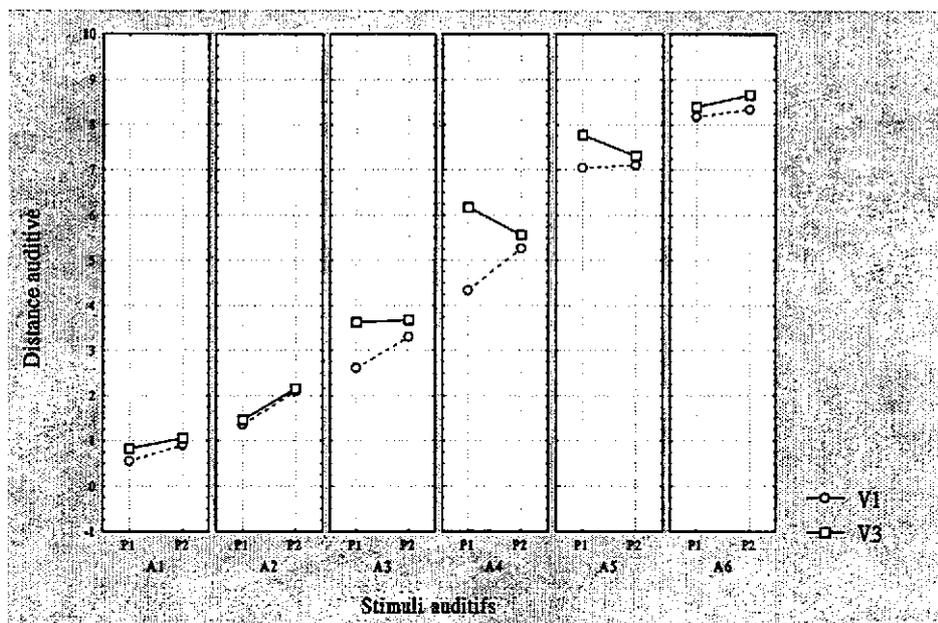


Figure 4.23 : Distance apparente auditive des stimuli auditifs en fonction des conditions visuelles V1 et V3 et des présentations de stimuli P1 et P2.

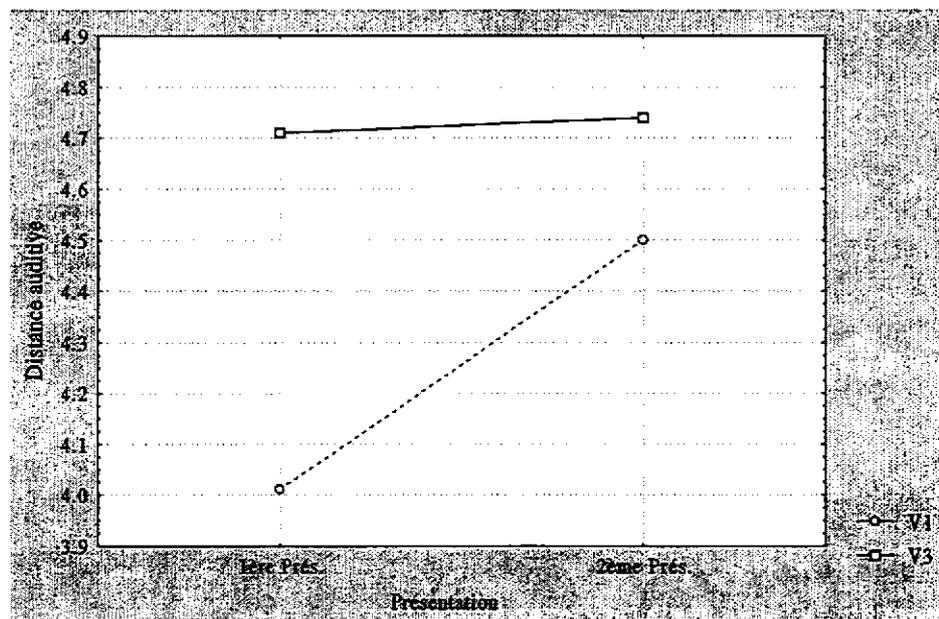


Figure 4.24 : Distance apparente auditive pour les présentations de stimuli P1 et P2 en fonction des conditions visuelles V1 et V3. Les réponses sont moyennées sur les sujets et sur les stimuli auditifs.

#### 4.9.4 Discussion

Dans l'expérience avec vues changeantes, les effets des trois mêmes facteurs expérimentaux étudiés dans l'expérience en vues statiques ont été testés : les effets des stimuli auditifs, des présentations des stimuli et des conditions visuelles ; ces effets se sont avérés significatifs. De plus, la présence d'interactions entre ces facteurs montre que les effets ne sont pas indépendants : l'effet de l'un de ces facteurs sur les résultats peut dépendre de la valeur (ou du niveau) d'un autre facteur. Ceci nous incite à étudier ces résultats de manière plus attentive et analytique.

En ce qui concerne les effets des stimuli auditifs et de la présentation des stimuli, les mêmes remarques que pour l'expérience 5 avec vues statiques, peuvent être faites (cf. § 1.2.4). Nous constatons ainsi l'influence importante des stimuli sonores sur les réponses de la distance auditive et la bonne répartition de ces réponses dans l'échelle proposée aux sujets (**tableau 4.19**). Nous constatons également que les réponses augmentent globalement à la 2<sup>ème</sup> présentation des stimuli (surtout sous les conditions visuelles V1 et V2 comme le montre la **figure 4.25**), cet effet était déjà observé dans les tests précédents et dans la littérature (cf. § 2.3.1).

	A1-A2	A2-A3	A3-A4	A4-A5	A5-A6
V1	0.99	1.23	1.84	2.28	1.19
V2	0.94	1.70	1.81	1.54	1.52
V3	0.86	1.84	2.23	1.67	0.99
<b>Global</b>	<b>0.93</b>	<b>1.59</b>	<b>1.96</b>	<b>1.83</b>	<b>1.23</b>

**Tableau 4.19** : Ecart de distances « perceptives » moyennes entre les stimuli successifs pour chacune des conditions visuelles et pour l'ensemble des résultats.

En ce qui concerne les effets des conditions visuelles, on observe un effet significatif pour la totalité des résultats : les mêmes distances auditives sont jugées plus lointaines quand elles sont accompagnées des images lointaines, et jugées plus proches quand elles sont accompagnées des images proches de la scène. L'effet est beaucoup plus marquant à la 1<sup>ère</sup> présentation des stimuli, il semble diminuer fortement, voire même disparaître (comme par exemple la différence entre les conditions V2 et V3) à la 2<sup>ème</sup> présentation (figure 4.25).

L'hypothèse que nous pourrions former est que, au début des expériences, les sujets sont surpris du changement constant des images, qui capturent ainsi inévitablement leur attention. Ceci crée une certaine incertitude par rapport à la tâche auditive : les sujets ont du mal à se concentrer et à juger les distances auditives. Ils utilisent ainsi inconsciemment les repères visuels disponibles (les images). Cet effet crée les grandes différences dues aux conditions visuelles à la 1<sup>ère</sup> présentation des stimuli et il est surtout observé pour les stimuli auditifs autres que les stimuli de référence A1 et A6.

Au bout d'un certain temps – chaque test dure en tout 25 à 30 min. – les sujets s'habituent au changement des images et ils cessent d'en être interpellés (il ne s'agit que de trois images interchangeables en tout). Ils se concentrent alors sur leur tâche auditive et arrivent progressivement à se reconstituer une échelle perceptuelle de la distance. Les images ne restent pas assez longtemps en place pour fournir un cadre spatial de référence (comme lors de l'expérience en vues statiques), leur influence semble alors diminuer fortement.

La présence d'effet dû à la perception visuelle au début de l'expérience et la disparition de cet effet au bout d'un certain temps, révèle un phénomène perceptif que l'on peut qualifier de 'transitoire'.

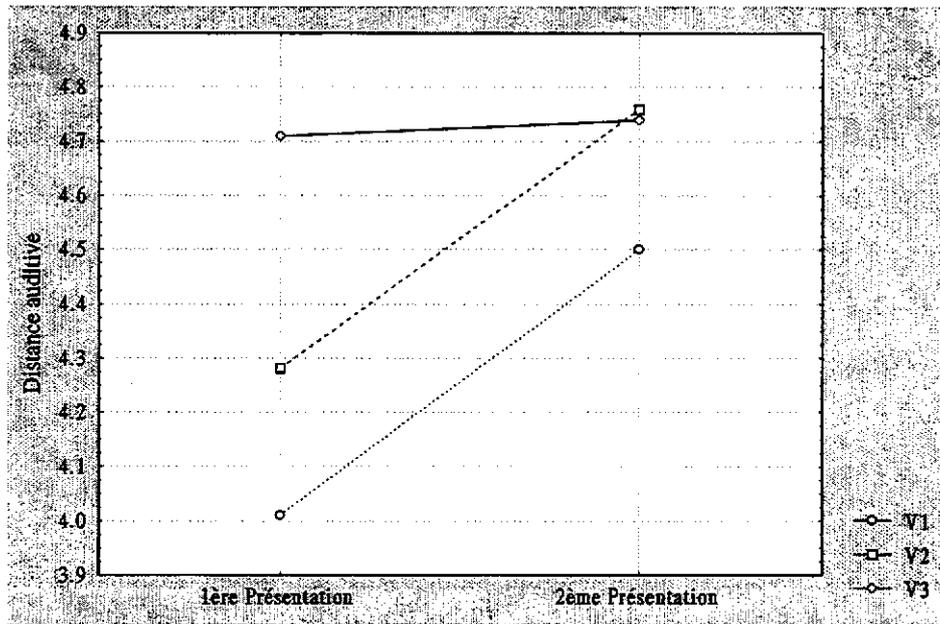


Figure 4.25 : Distance auditive apparente pour les présentations de stimuli P1 et P2 en fonction des conditions visuelles V1, V2 et V3. Les réponses sont moyennées sur les sujets et sur les stimuli auditifs.

## 4.10 DISCUSSION GENERALE

Nous avons noté que la perception de la distance des sources sonores est une faculté qui, chez les êtres humains, est peu précise et fortement influençable (cf. 2.3.3). Ce phénomène est lié entre autres à une constatation générale concernant le fonctionnement de l'ensemble de notre système perceptif : celui-ci perçoit des différences entre les grandeurs et non pas des grandeurs absolues. Afin d'effectuer ces comparaisons, le système perceptif a donc besoin de références, de repères. En général, les repères utilisés sont ceux issus de la modalité sensorielle en question, ici c'est la modalité auditive qui est testée. Nous avons voulu estimer si des repères issus d'autres modalités sensorielles – et notamment de la modalité visuelle – pouvaient être impliqués dans les jugements, ces modalités n'étant habituellement pas prises en compte dans les recherches pour l'évaluation de la qualité sonore.

Une série d'expériences a été menée dans le but d'étudier l'influence des conditions visuelles de salles de concert sur la perception de la distance auditive des sources sonores présentes. L'étude consistait essentiellement à simuler visuellement la scène d'un théâtre, à faire varier sa distance visuelle par rapport aux spectateurs-auditeurs, et à tester la perception de la distance auditive sous ces différentes conditions visuelles. Les aspects les plus intéressants des résultats de cette étude sont discutés à la suite.

Le résultat principal est qu'une augmentation de la perception de la distance auditive est observée avec l'augmentation de la distance visuelle. Cette influence visuelle est plus ou moins importante selon les expériences, mais apparaît pour toutes les expériences en projection des images 3-D, indiquant que la présence d'une image tridimensionnelle favorise la perception auditive spatiale. L'image semble reconstituer la profondeur de l'espace sonore et fournir un cadre spatial de référence pour la perception auditive. Elle semble également aider à l'organisation de l'espace sonore. Des entretiens informels avec les sujets à la fin des tests ont pu vérifier cette hypothèse : plusieurs d'entre eux ont indiqué qu'ils utilisaient les images pour 'placer' ou 'situer' le son dans l'espace, qu'ils 'se promenaient dans les images' pour juger la distance du son, ou simplement que les images les 'aidaient' dans leur tâche auditive.

L'influence de la perception visuelle est observable mais non significative pour les deux expériences en projection 2-D. Ce résultat permet de constater que c'est effectivement la *profondeur* des images 3-D qui est à l'origine de l'influence visuelle, et non plus ce que les images évoquent chez les sujets selon la taille du haut-parleur ou des détails architecturaux par exemple (les observateurs ont souvent une notion implicite de la taille des objets connus, et interprètent immédiatement celle-ci en termes de distance [Cahen, 1990]). De plus, ce manque d'effet en projection 2-D nous permet de vérifier que les consignes données aux sujets – de répondre en se basant seulement sur la modalité auditive – ont été respectées. Ceci parce que si par exemple, les sujets avaient compris qu'ils pouvaient également utiliser les images pour effectuer leur jugement, ils seraient influencés aussi par les images 2-D puisqu'elles donnent incontestablement l'impression de l'éloignement. L'influence visuelle se base par conséquent sur des caractéristiques physiques des stimuli (profondeur des images) et non plus sur des associations de mémoire (évoquées par la taille des objets connus du champ visuel, les consignes, etc.). Elle semble donc, dans notre cas, être davantage de nature perceptive que de nature cognitive.

Pour conclure sur les repères utilisés par les sujets dans le jugement de la distance auditive, nous observons qu'il s'agit de repères auditifs bien sûr (sons de référence, sons qui ont précédé le son en question), mais aussi – et dans un moindre degré – de repères visuels, en particulier la profondeur des images.

L'influence de la distance visuelle sur la perception de la distance auditive et la constatation que la modalité visuelle tient une place privilégiée dans la perception de l'espace (cf. § 2.4), nous amènent à

déduire qu'il faut porter une attention particulière aux conditions visuelles des expériences auditives, surtout lorsque ces dernières visent à tester des attributs spatiaux du son.

Dans nos expériences, nous avons essayé de contrôler le paramètre cognitif, au moins en ce qui concerne la conception et la comparaison des stimuli. Ainsi, pour chacune des expériences, le signal sonore utilisé était le même, les photos projetées provenaient d'une même salle de théâtre et étaient prises sur le même axe central de celle-ci. Aucun des sujets n'avait jamais visité la salle simulée (en tous cas aucun de ceux qui ont participé aux deux dernières expériences de validation) et le signal sonore faisait partie du répertoire de musique classique (considéré plus ou moins connu par tous). Seuls les résultats des conditions audiovisuelles sont comparés entre eux, et non pas ceux d'une condition uniquement auditive avec ceux d'une condition audiovisuelle. Ceci parce qu'une condition auditive et une condition audiovisuelle peuvent évoquer des impressions qualitativement différentes chez les sujets, ceux-ci peuvent alors leur attribuer des 'sens' différents. De plus, si l'on considère que l'expérience d'écoute chez les sujets était globalement équivalente (dans leur majorité des étudiants du même milieu socioculturel), nous pouvons supposer que les stimuli évoquaient des impressions – ou pour utiliser un terme plus juste, des représentations mentales – similaires chez les sujets.

Les consignes représentent un autre paramètre qui permet de contrôler – ou inversement – de biaiser les résultats. Ainsi, il est différent de demander aux sujets de juger la distance auditive ou la provenance sur le plan azimutal d'un son, ce n'est pas la même chose que de leur demander de juger 'l'agrément' ou le 'réalisme' d'un son. Plus l'attribut que l'on demande de juger est 'complexe' (il ne correspond pas par exemple à un attribut physique bien défini pour les sujets), plus nombreuses et plus difficilement prévisibles sont alors les impressions qu'il peut évoquer (puisque éminemment personnelles), et l'évaluation de cet attribut risque d'en être influencée d'autant.

Nous avons vu que l'influence de la perception visuelle sur la distance auditive est plus ou moins importante selon les expériences. Il est important de discuter ces différences, et pour ce faire, la confrontation des deux dernières expériences principales offre une possibilité intéressante.

Les expériences 5 et 6 adoptent deux différentes méthodes de présentation des stimuli visuels : l'une avec des vues statiques et l'autre avec des vues changeantes. Dans les deux cas, une augmentation de la distance auditive est observée avec la distance visuelle, mais la manière dont cet effet se manifeste dépend, d'une part de la méthode employée, et d'autre part des présentations de stimuli, c'est à dire de l'évolution des réponses avec le temps.

Ces effets sont présentés : pour la méthode en vues statiques aux figures 4.19 et 4.20 (expérience 5) et pour la méthode en vues changeantes aux figures 4.23 et 4.24 (expérience 6) ; ils sont résumés dans le tableau 4.20 ci-dessous.

Influence des conditions visuelles (XXX : forte, XX : moyenne, X : faible)	1 <sup>ère</sup> présentation	2 <sup>ème</sup> présentation.
Méthode à vues statiques (exp. 5)	néant	XX
Méthode à vues changeantes (exp. 6)	XXX	X

**Tableau 4.20 :** Comparaison des expériences 5 et 6 : tableau récapitulatif de l'influence des conditions visuelles (V1 et V3 uniquement) en fonction de la méthode de présentation des stimuli visuels et des présentations de stimuli.

On constate que l'effet visuel est presque inversé entre les deux expériences : avec la méthode à vues statiques, aucune influence de la distance visuelle n'est observée à la 1<sup>ère</sup> présentation ; cet effet augmente et devient significatif pour la 2<sup>ème</sup> présentation des stimuli. En revanche, pour la méthode à vues changeantes, l'effet visuel est maximal à la 1<sup>ère</sup> présentation et baisse fortement lors de la 2<sup>ème</sup> présentation des stimuli.

Nous considérons que ces résultats soulignent l'importance des effets dynamiques dans la perception. Si dans une situation de concert, on n'est généralement pas amené à bouger, cela pourrait néanmoins avoir une importance pour les applications audiovisuelles.

Par ailleurs, lors de l'interprétation de ces résultats aux sections précédentes (cf. § 4.8.4 et § 4.9.4), nous avons évoqué des processus cognitifs<sup>1</sup> tels que l'orientation d'attention, l'intérêt relatif aux deux modalités perceptives, l'adaptation aux conditions expérimentales, l'incertitude concernant les réponses, l'évolution de ces réponses avec le temps, etc.

Il est alors important de constater que, malgré les précautions prises initialement afin de contrôler le 'sens' ou les 'représentations mentales' évoqués par les stimuli sonores et visuels, d'autres influences de nature cognitive peuvent être introduites par le biais de la procédure expérimentale. Ces influences peuvent être importantes au point de faire apparaître ou disparaître un effet (dans notre cas, elles entraînent une inversion dans l'apparition d'effet visuel entre les deux méthodes).

En effet, les différentes procédures expérimentales (à vues statiques ou à vues changeantes) créent de différentes impressions chez les sujets, et de plus, ces impressions évoluent différemment avec le temps (entre la 1<sup>ère</sup> et la 2<sup>ème</sup> présentation). Etudier les résultats pas par pas en essayant toujours de prendre en compte ce que les différentes phases de la procédure ont pu évoqué chez les sujets, semble être le seul moyen de procéder à l'étude des phénomènes complexes de la perception audiovisuelle, en évitant de récupérer des effets qui sont des artefacts de la procédure.

En ce qui concerne l'ampleur de l'effet des conditions visuelles, nous observons que cet effet est bien plus grand pour l'expérience à vues changeantes que pour l'expérience à vues statiques ( $[F(1,27) = 39.82, p = 0.000]$  contre  $[F(1,27) = 5.73, p = 0.024]$ , cf. tableaux 4.12 et 4.18). Nous supposons qu'il s'agit de l'incertitude créée sur les réponses, bien plus élevée pour le cas des vues changeantes, qui a incité ce résultat. Les sujets ont en effet constaté après les expériences, que la méthode en vues changeantes les déstabilisait beaucoup plus dans leurs jugements, et rendait leur tâche difficile. Dans cette incertitude par rapport aux repères auditifs, des repères non auditifs – en l'occurrence visuels – ont pu alors être davantage utilisés, et avoir entraîné cette forte influence visuelle.

Les effets dus aux conditions visuelles sont en général robustes. Ils ont été observés dans une série d'expériences (expérience préliminaire, expériences 1 et 2), malgré des signaux sonores de nature différente utilisés (chanson de S. Vega et 'La Somnambule' de Bellini) et malgré les différentes méthodes de restitution sonore employées (diffusion binaurale ou transaurale). Ces effets ont été validés par les deux dernières expériences avec un plus grand nombre de sujets. Si par ailleurs, nous considérons que l'analyse de variance à mesures répétées permet un 'contrôle expérimental' des résultats – du fait que le terme d'erreur utilisé prend en compte les différences interindividuelles des sujets (cf. § 3.5.1) – nous pouvons être confiants sur la pertinence de la présence de l'effet visuel.

---

<sup>1</sup> Les fonctions qui constituent l'objet de la psychologie cognitive sont : la perception, l'attention, la mémoire, le langage, les activités intellectuelles, et le cas échéant, d'autres domaines d'étude qui sont directement apparentés à ceux-ci.

Il faut reconnaître que, malgré la nature significative des effets observés, les effets par eux mêmes sont faibles (cf. tableaux 4.11 et 4.16 par exemple). Tout d'abord, il faut tenir compte du fait qu'il s'agit de chercher le 'biais' visuel – or un 'biais' est par nature petit – et non plus de tester quelle aurait pu être l'influence visuelle sous d'autres conditions éventuellement bien plus favorables à son apparition. Ainsi, en principe tout effort a été fait pour éliminer l'effet visuel : des consignes qui orientaient clairement vers la modalité auditive, la référence auditive au début des tests, le choix d'un attribut à tester qui laisse le moins possible d'ambiguïtés sémantiques.

Aussi, il y a un autre paramètre qui a incontestablement influencé – voir conditionné – l'impact de la perception visuelle : il s'agit de la qualité de la simulation. Elle joue évidemment un rôle primordial sur les résultats. La même méthode de simulation concerne les expériences portant sur un ensemble des facteurs auditifs (cf. § 5) ; ce sujet sera donc discuté plus en profondeur lors la discussion finale de ce travail (cf. § 6.3). Nous nous contentons ici de constater que la simulation visuelle était relativement pauvre par rapport à la simulation auditive : pas de temps réel (et donc pas de changement à l'image), absence de présence humaine sur la scène et dans la salle, conditions visuelles peu semblables à une représentation d'opéra, etc. De plus, l'efficacité d'une simulation multimodale dépend non seulement de la qualité des deux simulations considérées séparément, mais aussi de l'équilibre entre ces qualités (cf. § 2.4). Or, nous considérons que la simulation auditive était bien plus réaliste. Tenant compte des faiblesses de notre simulation, nous faisons la considération suivante : si les effets visuels ont pu être observés sous des conditions de simulation visuelle – mais aussi bimodale – relativement pauvres, la question de leur pertinence en est a priori renforcée, puisque leur apparition est susceptible d'être bien plus accentuée sous des conditions visuelles de simulation améliorées, et surtout sous des conditions réelles de salles de spectacle.

## 4.11 CONCLUSION

Une série d'expériences a été réalisée afin d'étudier l'influence de la perception de la distance visuelle de la scène d'une salle de spectacle sur la perception de la distance auditive de la source sonore présente. Ces mêmes expériences ont également permis de tester différents aspects de la méthode expérimentale et de mettre en place un nombre de choix expérimentaux issus de ces recherches.

L'expérience préliminaire a servi d'expérience 'pilote' : l'influence de plusieurs facteurs expérimentaux, pour la plupart arbitrairement choisis, a pu être testée pour la première fois. Elle a permis de valider la simulation sonore de la distance auditive, ainsi que d'observer l'importance du nombre de présentations des stimuli. Aussi, des entretiens informels réalisés après les tests ont indiqué que la diffusion binaurale du son (format d'encodage binaural diffusé sur casque audio) était inadaptée pour la simulation spatiale du son (localisation du son 'dans la tête', 'au dessus de la tête', etc.) ; la diffusion transaurale du son a donc été choisie (format d'encodage binaural diffusé sur haut-parleurs) pour les expériences suivantes.

L'expérience suivante (expérience 1) a permis de choisir les stimuli sonores : l'énergie précoce s'est révélée être l'indice objectif le mieux corrélé – parmi les trois indices testés – avec la perception de la distance auditive. Elle a également permis de sélectionner – parmi les deux signaux testés – le signal sonore de l'opéra de Bellini 'La Somnambule'.

Le nombre de présentations des stimuli a été testé à l'aide de l'expérience 2. Les résultats ont montré que les réponses sur la distance auditive augmentent pour toutes les conditions visuelles entre la 1<sup>ère</sup> et la 2<sup>ème</sup> présentation, et semblent se stabiliser après.

Dans les expériences 3 et 4, la simulation visuelle était réalisée en 2-D (la projection en 3-D étant utilisée pour toutes les autres expériences). La méthode de présentation des conditions visuelles en vues statiques a été employée pour l'expérience 3, et celle en vues changeantes pour l'expérience 4. Les résultats n'ont pas montré d'effets significatifs des conditions visuelles sur la perception de la distance auditive, alors que des effets ont été observés sous des conditions visuelles réalisées en projection 3-D. L'effet est donc dû principalement au meilleur rendement de profondeur d'image obtenu avec ce type de projection. Ce résultat indique que l'influence des images est plutôt de nature perceptive que cognitive (c'est la profondeur des images qui donne l'impression de la distance plutôt que la taille des objets ou le nombre de fauteuils qui séparent les sujets de la scène par exemple). La projection en 3-D semble donc favoriser la perception de la distance auditive et l'effet visuel est donc dû au meilleur rendement de profondeur des images projetées en 3-D.

Toutes ces expériences – sauf celles en projection 2-D – ont permis de mettre en évidence les premiers indices de l'influence de la distance visuelle sur la perception de la distance auditive. L'effet visuel s'est en effet avéré significatif pour l'ensemble des résultats.

Les deux dernières expériences (expériences 5 et 6) ont opté pour un nombre plus élevé de sujets, et ont visé entre autres, la validation des effets observés aux expériences précédentes. Par ailleurs, les différents choix expérimentaux résultant des expériences précédentes (la méthode de diffusion sonore, le signal et les configurations sonores, le nombre de deux présentations des stimuli, la méthode de projection des images, etc.) ont été mis en œuvre dans le cadre de ces deux expériences.

Les expériences 5 et 6 ont adopté deux méthodes de présentation des stimuli : une méthode avec *vues statiques* et une méthode avec *vues changeantes* (cf. § 4.2.6) ; la première a permis de simuler des conditions visuelles présentes dans les salles de concert, la seconde se rapproche plus de conditions visuelles évolutives telles qu'en environnements virtuels.

Une augmentation des jugements de la distance auditive avec la distance visuelle a été observée d'une manière générale, elle vérifie les résultats des expériences précédentes et fortifie l'hypothèse selon laquelle la perception visuelle est susceptible de fournir un cadre spatial de référence pour les jugements auditifs. Elle semble également aider à l'organisation de l'espace sonore.

Toutefois, les effets visuels restent marginaux par rapport aux effets auditifs, de l'ordre de la plus petite différence perceptible auditivement (variation de 1 dB sur l'indice objectif Es pour un quadruplement de la distance visuelle).

L'apparition de l'effet visuel dépend largement de la méthode de présentation des stimuli visuels et du nombre des présentations des stimuli sonores. Des facteurs attentionnels et cognitifs, tels que l'orientation inconsciente de l'attention des sujets, l'intérêt relatif porté aux deux modalités, le besoin de davantage de repères créé par l'incertitude relative à la tâche demandée, etc. sont considérés comme à l'origine de ces différenciations.

Ceci nous amène à la conclusion que, malgré l'attention portée pour contrôler des associations cognitives évoquées par les stimuli sonores et visuels au niveau de la conception des expériences, d'autres influences de nature cognitive peuvent être introduites par la procédure expérimentale et peuvent s'avérer déterminantes pour les résultats. Des études menées de manière attentive, en essayant de contrôler les représentations mentales que les sujets peuvent se faire par rapport aux stimuli, mais aussi en prenant en compte les processus attentionnels souvent générés involontairement par la procédure expérimentale, permettraient d'obtenir des effets le moins biaisés possible par des facteurs cognitifs. Ces effets pourront alors être attribués aux interactions audiovisuelles spatiales que l'on qualifierait de pures, puisqu'elles se basent principalement sur la perception des grandeurs physiques.



## 5 EXPERIENCES SUR UN ENSEMBLE DE FACTEURS PERCEPTIFS

### 5.1 INTRODUCTION

Suite à l'étude approfondie de l'influence visuelle sur le facteur unique de la distance auditive, une étude plus large semble s'imposer, portant sur un ensemble de facteurs perceptifs correspondant aux principaux attributs auditifs présents dans les salles.

Cette suite se justifie du fait que l'espace perceptif décrivant la qualité acoustique des salles est un espace complexe, représenté par plusieurs dimensions perceptives et qu'il est, dans la grande majorité des cas, étudié en faisant appel à des méthodes multidimensionnelles. Par conséquent, une étude pertinente, visant à évaluer l'influence des conditions visuelles sur la qualité acoustique – telle que celle-ci est décrite habituellement – se doit d'adopter entre autres une approche du même type et étudier l'influence visuelle dans ce cas d'approche expérimentale traditionnelle.

Cette deuxième partie du travail utilise ainsi une méthodologie empruntée aux travaux classiques de la qualité acoustique des salles : celle de questionnaires portant sur un ensemble de facteurs perceptifs préalablement sélectionnés (cf. § 2.2.2). Ces facteurs correspondent à un jeu couramment utilisé pour étudier cette qualité. En plus de la distance auditive, il s'agit de : la puissance sonore, la réverbération, la précision d'écoute, la vivacité, l'enveloppement et l'impression générale. La méthode est adaptée de manière à incorporer des indices visuels dans les tests, afin d'étudier l'influence des informations visuelles sur la perception de la qualité acoustique. L'objectif est donc le même que celui des expériences précédentes, ce sont seulement l'approche méthodologique et le nombre de facteurs testés qui diffèrent.

En plus de l'objectif principal, l'étude simultanée de plusieurs facteurs vise à tester les hypothèses formulées précédemment sur la nature des facteurs auditifs qui risquent d'être les plus influencés par la perception visuelle, c'est à dire, les facteurs liés à la dimension spatiale du son (cf. § 2.5).

Enfin, les résultats de ces expériences en ce qui concerne le facteur de la distance auditive, permettront de tester si les effets observés dans la première partie de ce travail sur ce même facteur auditif, persistent sous des conditions d'expérimentation différentes. Il permettront ainsi de valider et de tester la robustesse de ces effets à travers les différentes approches méthodologiques.

Deux expériences préliminaires – l'une purement visuelle et l'autre purement auditive – ont été organisées, ainsi qu'une expérience principale divisée en deux parties. Les objectifs de chacune et les principaux aspects méthodologiques respectifs sont présentés dans le **tableau 5.1**.

Type de test	Facteurs testés	Objectif	Son (stimuli, signal, diffusion)	Images, mode de projection, mode de présentation	Nombre de présentations de stimuli	Sujets
préliminaire auditif	7 facteurs auditifs	Choix des stimuli sonores (configurations Spat)	A1-A14, opéra <sup>1</sup> , dif. transaurale	---	1	11
préliminaire visuel	4 facteurs visuels	Perception des images, Interprétation	---	(V1, V2, V3, V4), TCE <sup>2</sup> , 2-D et 3-D vues changeantes	1	20
principal (unimodal)	7 facteurs auditifs	Perception des stimuli sonores	A1-A10, opéra, dif. transaurale	---	1	32
principal (bimodal)	8 facteurs (7 auditifs + 1 bimodal)	Influence visuelle	A1-A10, opéra, dif. transaurale	(V1, V2, V3, V4), TCE, 3-D vues changeantes	1	32

**Tableau 5.1** : Liste des expériences sur un ensemble de facteurs perceptifs. Deux expériences préliminaires et une expérience principale ont été effectuées. L'expérience principale était composée d'une partie unimodale et d'une partie bimodale.

L'objectif de l'expérience préliminaire auditive – expérience où des stimuli auditifs sont présentés sans stimulation visuelle – était de choisir les stimuli sonores pour les tests principaux. Les 14 stimuli, dont la description est donnée en annexe C, variaient selon plusieurs indices acoustiques afin de créer des impressions sonores qui puissent, à leur tour, changer sur l'ensemble des facteurs perceptifs en question. Cette expérience ne sera pas décrite plus en détail ici puisque la plupart des stimuli testés ont été retenus (10 sur 14 pour l'expérience principale avec très peu de changements).

L'objectif de l'expérience préliminaire visuelle – expérience ne concernant que la perception visuelle sans stimulation auditive – est d'étudier certains aspects visuels des quatre images utilisées dans les tests principaux, dans le but de faciliter l'interprétation ultérieure des résultats. La procédure, ainsi qu'une brève description des résultats, sont présentées en annexe C.

L'expérience principale, décrite à la suite, est composée de deux parties : l'une est uniquement auditive, l'autre est bimodale.

<sup>1</sup> «La Sonnambule» de Bellini chanté par P. Bryn -Jullson.

<sup>2</sup> Théâtre des Champs Elysées.

## 5.2 EXPERIENCE PRINCIPALE

- *Test à écoute transaurale en deux parties :*  
*la partie auditive est sans vue, la partie bimodale est avec vues.*
- *Une seule présentation des stimuli*
- *Signal sonore : «La Somnambule» de Bellini, chant interprété par Phyllis Bryn-Jullson.*  
*(Stimuli sonores : A1 à A10)*
- *Vues : « Théâtre des Champs Elysées » à Paris*  
*(Stimuli visuels : images 3-D : V1, V2, V3 et V4).*
- *32 sujets*

Facteurs perceptifs testés : puissance sonore, précision d'écoute, réverbérance, vivacité, distance auditive, enveloppement, impression générale et compatibilité des stimuli hétéromodaux, visuels et auditifs<sup>1</sup>.

### 5.2.1 Objectif

L'expérience principale a été organisée dans le même but que les expériences présentées au chapitre précédent, c'est-à-dire pour étudier l'influence de la perception visuelle sur la perception auditive. La différence se situe dans le nombre de facteurs auditifs testés, et en partie, sur la méthode d'analyse des résultats. Les premières expériences portaient uniquement sur la perception de la distance auditive, tandis que les expériences décrites dans ce chapitre portent sur un ensemble de facteurs perceptifs, traditionnellement utilisés pour caractériser la qualité acoustique des salles. En outre, les résultats sont considérés dans leur ensemble et étudiés à l'aide d'analyses multidimensionnelles.

De plus, l'approche d'étude globale de l'espace perceptif adoptée ici, étant l'approche la plus répandue dans la caractérisation de la qualité acoustique des salles, il est intéressant d'étudier la pertinence des effets visuels sous ces conditions méthodologiques bien établies dans la pratique.

Enfin, les facteurs auditifs testés relèvent des quatre dimensions fondamentales liées à la perception : intensité, espace, temps et fréquence (cf. § 2.2.3), ce qui permettra de vérifier l'hypothèse selon laquelle les facteurs auditifs liés à la dimension d'espace sont plus susceptibles d'être influencés par la perception visuelle (cf. § 2.5).

---

<sup>1</sup> La compatibilité des stimuli a été testée uniquement dans la partie bimodale de l'expérience.

## 5.2.2 Méthode

### 1. Stimuli sonores

Dix configurations de salle créées à l'aide du "Spatialisateur" (cf. § 3.2.3) ont été utilisées pour cette expérience principale ; elles ont été choisies parmi quatorze configurations testées lors de l'expérience préliminaire qui était uniquement auditive.

Le stimulus de référence, noté ici A2, était le même que pour les expériences précédentes sur la distance, il a été créé afin de simuler une réponse impulsionnelle obtenue dans un emplacement du centre du Théâtre des Champs Elysées à Paris (cf. § 4.4.2). A partir de cette réponse de référence, treize autres réponses ont été créées pour l'expérience préliminaire auditive. Les dix réponses retenues<sup>1</sup> pour l'expérience principale variaient sur cinq facteurs perceptifs, la plupart d'entre eux faisant partie de l'interface haut-niveau du *Spatialisateur* (interface perceptive). Les réponses variaient aussi sur les cinq paramètres acoustiques objectifs correspondant à ces sensations subjectives. Les cinq facteurs perceptifs et les paramètres acoustiques correspondants sont les suivants, avec entre parenthèses, les noms utilisés en français<sup>2</sup> :

*Source presence* Es (présence de source)

*Liveness* Drth (vivacité)

G (Niveau sonore global)

*Envelopment* Rd1 (enveloppement)

*Late reverberance* Rt (réverbérance tardive)

Nous rappelons ici que Es est obtenue par l'intégration de l'énergie précoce de 0 à 40 ms, que Drth correspond à la variation fréquentielle du Tr en hautes fréquences (rapport entre Tr aux hautes fréquences et Tr aux fréquences moyennes), et que Rd1 est la somme pondérée de l'énergie des réflexions précoces (R1) et tardives (R2) rapportée à l'énergie précoce Es (la description de ces paramètres est présentée au § 3.2.4).

La figure 5.1 présente ces 10 configurations. Cinq séries se distinguent, chacune correspondant à des variations sur un indice acoustique :

A1, A2, A3 et A4 à '*source presence*' Es,

A2 et A5 à '*liveness*' Drth,

A6, A2 et A7 au niveau global G,

A8, A2 et A9 à '*envelopment*' Rd1,

A2 et A10 à '*late réverbérance*' Rt.

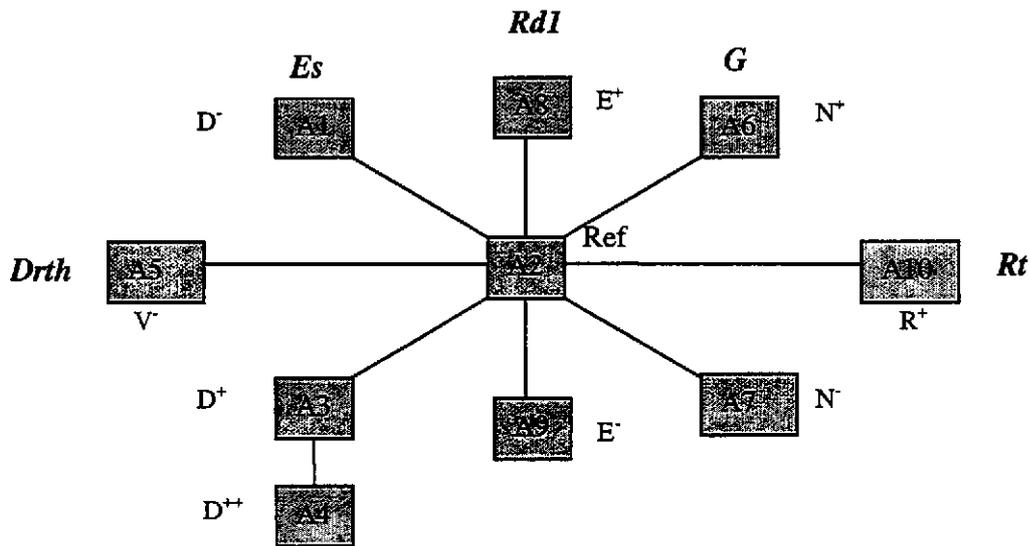
Notons que la configuration de référence A2 se trouve au milieu de chacune de ces séries. Les valeurs de tous les paramètres acoustiques des stimuli auditifs (haut et bas-niveau du *Spatialisateur*) sont

---

<sup>1</sup> Un seul changement au niveau de Rd1 pour le stimulus E<sup>+</sup> a été effectué : Rd1 = -6,5 (au lieu de -8,5 pour l'expérience préliminaire).

<sup>2</sup> A l'exception du 'niveau sonore global G' qui est un indice objectif contrôlé par l'interface bas-niveau (cf. § 3.2), et pour lequel le facteur perceptif correspondant est la 'puissance sonore'.

présentés dans l'annexe C. Les niveaux sonores mesurés à la place de la tête des auditeurs y sont également présentés.



**Figure 5.1:** Les 10 configurations du *Spatialisateur* pour l'expérience principale. Ils varient selon 'source presence' *Es* (A1, A2, A3 et A4), selon 'liveness' *Drth* (A2 et A5), selon le niveau global *G* (A6, A2 et A7), selon 'envelopment' *Rd1* (A8, A2 et A9), et selon 'late reverberance' *Rt* (A2 et A10).

Les configurations variant sur *Es* ont été créées afin de correspondre principalement à différentes impressions de distance auditive, celles variant sur le *Drth* correspondent à des différences perceptives de vivacité, le changement du niveau global *G* devrait influencer la perception de la puissance sonore, le *Rd1* l'enveloppement et le *Rt* la réverbérance.

Les configurations obtenues ont été appliquées au signal sonore d'entrée du *Spatialisateur* : «La Sonnambule» de Bellini chanté par Phyllis Bryn-Jullson (voix féminine chantée) afin de créer les 10 stimuli auditifs présentés aux sujets. La diffusion sonore était transaurale sur haut-parleurs.

Les niveaux sonores des stimuli des expériences sont fournis en annexe C, dans laquelle figure également une estimation du niveau sonore moyen entendu dans le théâtre des Champs Elysées.

## 2. Stimuli visuels

Le champ visuel est créé par la projection de grandes images tridimensionnelles (taille 150 cm x 100 cm), obtenues par polarisation de la lumière et projetées depuis l'arrière sur grand écran dans des conditions d'obscurité complète (cf. § 3.3).

Les photographies ont été prises dans le *Théâtre des Champs Elysées* à Paris, en des emplacements de différentes distances par rapport à la scène, au milieu de laquelle était placé un haut-parleur. Les vues V1, V2, V3 et V4 utilisées dans les tests représentent ainsi des distances visuelles *croissantes* par rapport à la scène du théâtre. Les images V1, V2 et V3 ont été prises sur l'axe central de la salle et à peu près à même hauteur (au niveau de l'orchestre) ; il s'agissait des mêmes images que celles utilisées lors des expériences principales sur la distance auditive. L'image V4 était une vue de la scène excentrée et prise plus en hauteur (3<sup>ème</sup> loge). Les vues V1, V2, V3 et V4 sont présentées sur la figure 3.13 (§ 3.3.2). Pendant les tests, les sujets étaient assis dans les sièges d'une vraie salle de spectacle, à une distance de 1,6 m de l'écran.

Pour la partie unimodale de l'expérience, aucune image n'était projetée sur l'écran, seule une lumière diffuse y était perçue (un filtre de couleur orange pâle était placé sur l'un des projecteurs, l'autre projecteur restait éteint).

### *3. Sujets, consignes*

Trente deux étudiants de l'IUT et de l'IUP de Neuville sur Oise (génie électrique, génie civil) ont participé à cette expérience. Ils étaient âgés de 19 à 26 ans et une rémunération de 40F/heure leur était offerte pour leur participation. Ils n'avaient aucune expérience préalable des tests d'écoute, et les audiogrammes réalisés avant les expériences n'ont pas montré de déficience auditive notable.

Avant chaque session de tests, le réglage subjectif du dispositif transaural eut lieu afin de contrôler la bonne position de la tête des sujets par rapport aux haut-parleurs (cf. § 4.2.1). Une courte session de démonstration de certains sons a suivi, afin de clarifier leur tâche et de leur permettre de distinguer des différences perceptives. Les sons variaient sur la vivacité, l'enveloppement et la réverbérance<sup>1</sup>, car les variations sur ces attributs étaient jugées plus difficiles à percevoir. Egalement, l'une des images (V4) du *Théâtre des Champs Elysées* a été montrée aux sujets en 3-D avant le début des tests, afin de vérifier la bonne restitution du relief et de leur donner un aperçu de la salle dans laquelle ils devaient s'imaginer.

Une courte présentation de l'objectif général de l'étude a eu lieu, c'est à dire une présentation de la notion de qualité acoustique des salles et de sa caractérisation à l'aide des facteurs perceptifs, avant que des consignes plus spécifiques aux expériences soient apportées. Ces consignes étaient données oralement et peuvent se résumer par :

*« Vous allez voir des images et écouter des sons spécialement conçus pour simuler les conditions visuelles et sonores d'une salle de spectacle. Il s'agit de la salle du « Théâtre des Champs Elysées » à Paris et de l'air d'opéra de Bellini « La Somnambule ».*

*Il vous est demandé de:*

*Utiliser le son et l'image (quand celle-ci est disponible) afin de vous imaginer au mieux dans la vraie salle.*

*Juger uniquement le son sur la base des facteurs perceptifs du questionnaire, vous répondrez par écrit, sur les échelles verbales représentant ces attributs. Les facteurs sont :*

*La puissance sonore : elle caractérise la force ou l'intensité d'un son, s'il est fort ou faible.*

*La précision d'écoute : elle concerne la netteté d'un son, s'il est net ou brouillé.*

*La réverbérance : elle représente la sensation que le son dure dans le temps, ceci après l'arrêt de son émission. Si le son dure, la réverbérance est grande, sinon elle est petite.*

*La vivacité : elle concerne la réverbération aux hautes fréquences. Est ce que la réverbération est plus ou moins riche en hautes fréquences ? C'est l'aspect 'brillant' ou, au contraire, 'pauvre' du son.*

*La distance auditive : elle décrit la sensation liée à la distance de la source sonore, si le son semble venir de près ou de loin.*

---

<sup>1</sup> Ref et V pour la vivacité ; E<sup>+</sup>, Ref et E<sup>-</sup> pour l'enveloppement ; Ref et R<sup>+</sup> pour la réverbérance.

*L'enveloppement ou impression d'espace : il décrit l'aspect spatial du son, le sentiment d'être entouré par le son. Si le son donne l'impression de venir de plusieurs côtés, l'enveloppement est fort, sinon il est faible.*

*L'impression générale concerne l'impression globale auditive de l'acoustique de la salle et non du signal musical ou du style d'interprétation par exemple.*

*La compatibilité des stimuli héréromodaux concerne le son et l'image : est ce qu'ils sont compatibles, est ce qu'ils 'collent' ensemble ? Evidemment, la compatibilité ne doit être jugée que lorsque les sons sont accompagnés des images. »*

*Soyez détendus, donnez simplement votre impression telle que vous la sentez. Il n'y a pas de réponses correctes ou fausses, elles sont toutes d'une valeur identique. Ce qui nous intéresse est de voir comment vous percevez les sons, il ne s'agit pas de tester vos capacités.*

Le questionnaire donné aux sujets est présenté à la **figure 5.2**. Les facteurs choisis correspondent à un jeu traditionnellement utilisé pour caractériser la qualité acoustique des salles (voir par exemple le questionnaire proposé par Barron [1988] ; figure 2.2). Les sujets expriment leur impression sur des échelles de jugements absolus (échelles différentielles sémantiques), la cotation des catégories est verbale [Bonnet, 1986].

Expérience Principale (février-mars 1998)		Nom :	No :
<b>SON 1</b>		No séquence : 1	
<i>puissance sonore</i>			
<i>précision d'écoute</i>			
<i>réverbérance</i>			
<i>vivacité</i>			
<i>distance auditive</i>			
<i>enveloppement sonore, impression d'espace sonore</i>			
<i>impression générale</i>			
<i>compatibilité des stimuli bimodaux</i>			

**Figure 5.2:** Questionnaire utilisé pour l'expérience principale. Les sujets répondent à la '*compatibilité des stimuli bimodaux*' uniquement lorsque les sons sont accompagnés des images (partie bimodale de l'expérience).

4. Procédure expérimentale.

Après le réglage subjectif du dispositif transaural et la démonstration des sons variant sur les attributs auditifs les plus difficiles à juger, la partie principale des tests a lieu. Des séquences comprenant les 50 stimuli sont présentées aux sujets : les 10 stimuli sonores sont présentés au début de chaque séquence dans un ordre aléatoire, ils ne sont pas accompagnés d'image. Une diapositive uniforme (non figurative) de couleur orange est projetée sur l'écran, représentant la seule source lumineuse de la salle de tests. Ensuite, ces mêmes stimuli sonores sont couplés avec les quatre images pour produire 40 stimuli bimodaux (10 stimuli auditifs x 4 images) qui sont également présentés avec des ordres aléatoires en deux 'blocs' de 20 stimuli. Huit séquences de 50 stimuli ont ainsi été créées afin de compenser les effets d'ordre.

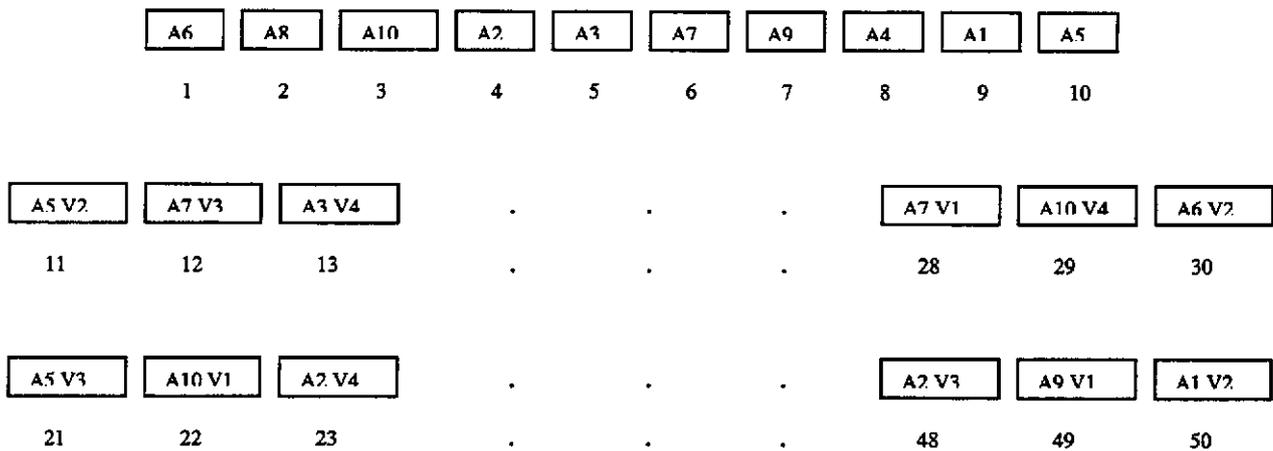


Figure 5.3 : Séquence type de 50 stimuli, présentée lors des expériences portant sur un ensemble de facteurs auditifs. Les dix stimuli auditifs sont d'abord présentés sans image (partie unimodale de l'expérience) ; ensuite, ils sont couplés aux quatre images de la figure 3.13 pour produire 40 stimuli bimodaux (partie bimodale). Huit séquences aléatoires de ce type ont été créées afin de compenser les effets d'ordre.

Les sujets ont donné par écrit (en cochant par une croix ou un trait) leurs jugements sur les échelles verbales du questionnaire. Chacun a répondu à 50 questionnaires, un pour chacun des 50 stimuli. Les questionnaires comprenant des échelles différentielles sur 7 facteurs perceptifs (6 seulement pour les 10 premiers sons), le nombre total des réponses obtenues par sujet était de 340 (60 réponses sur la partie unimodale des tests et 280 réponses sur la partie bimodale).

Les sujets pouvaient se reposer (pour environ 5 min.) entre le 'bloc' des 10 stimuli auditifs et le premier 'bloc' des 20 stimuli bimodaux, ainsi qu'entre les deux 'blocs' de stimuli bimodaux (pour 15 min. environ). La durée totale des tests était de 2 heures approximativement. Tous les sujets ont consacré ce temps pour terminer le test en une seule fois.

### 5.2.3 Résultats

Les réponses des sujets sur les échelles verbales du questionnaire ont été traduites en chiffres. Ainsi, pour la puissance sonore par exemple, 'très faible' correspondait à la note 1, 'faible' à la note 2, 'assez faible' à la note 3, etc. ; des notes intermédiaires ont également été utilisées (2,5 ou 2,8 par exemple). La puissance sonore, la précision d'écoute, la réverbérance, la distance auditive et l'impression générale correspondaient ainsi à des échelles numériques de 1 à 6, tandis que la vivacité, l'enveloppement et la compatibilité des stimuli à des échelles de 1 à 4. Cela est dû au fait que la sensibilité des sujets pour ces différentes échelles perceptives n'est pas identique : plus la sensibilité de la perception est fine, plus les sujets peuvent discriminer un nombre important de degrés sur l'échelle donnée. Il s'agit d'une approche généralement adoptée par les études en acoustique [Lavandier, 1989], [Kahle, 1995].

L'Analyse en Composantes Principales (ACP) a été utilisée pour étudier l'espace perceptif dans son ensemble (étude globale de l'ensemble des facteurs perceptifs), et l'Analyse de la Variance (ANOVA) pour étudier des résultats obtenus séparément pour chaque facteur.

La normalisation des réponses fait partie de l'ACP (cf. § 3.6.2), elle est donc réalisée par le logiciel utilisé (S-PLUS). Pour le cas de l'ANOVA (logiciel *Statistica*), la normalisation des réponses est effectuée comme précédemment pour les expériences sur la distance auditive (cf. § 4.2.8).

Cinquante réponses étaient manquantes sur l'ensemble. Le pourcentage correspondant sur les 10880 réponses au total qui devaient normalement être obtenues (32 sujets x 340 réponses) était ainsi inférieur à 0,5 %. A ces réponses manquantes sont attribuées des valeurs égales à la moyenne des réponses de tous les autres sujets pour le même stimulus.

#### 5.2.3.1 La partie unimodale

Les résultats de l'ACP pour la partie unimodale de l'expérience sont présentés en premier lieu : le **tableau 5.2** donne la proportion de la variance expliquée par les trois premières composantes principales, la proportion cumulative de cette variance ainsi que la qualité de représentation (pourcentage d'inertie) des facteurs-variables sur les axes synthétiques ; la **figure 5.4** présente les plans factoriels 1:2 et 2:3 (entre les composantes 1 et 2 et entre 2 et 3 respectivement).

	Composante 1	Composante 2	Composante 3
Proportion de Variance	0.60	0.32	0.045
Variance Cumulative	0.60	0.92	0.97

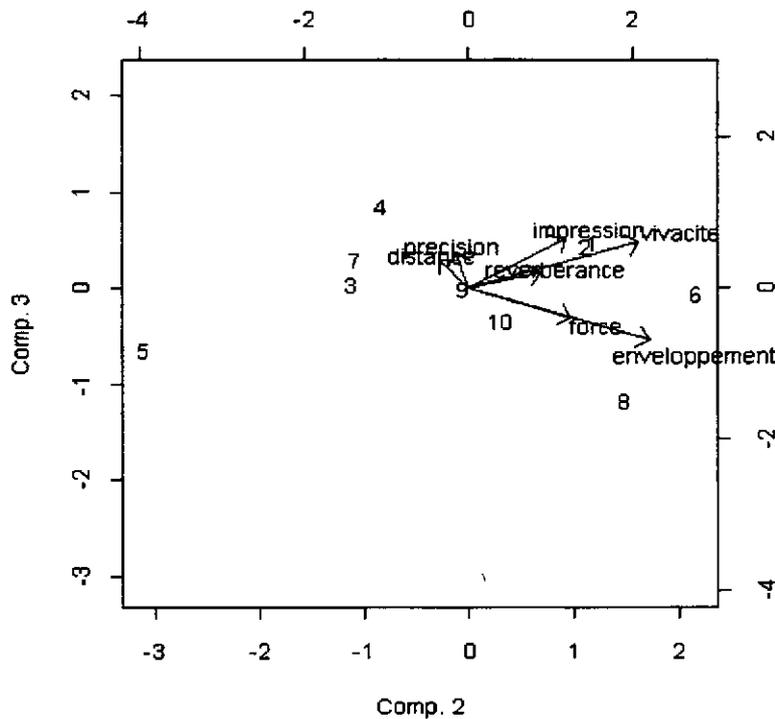
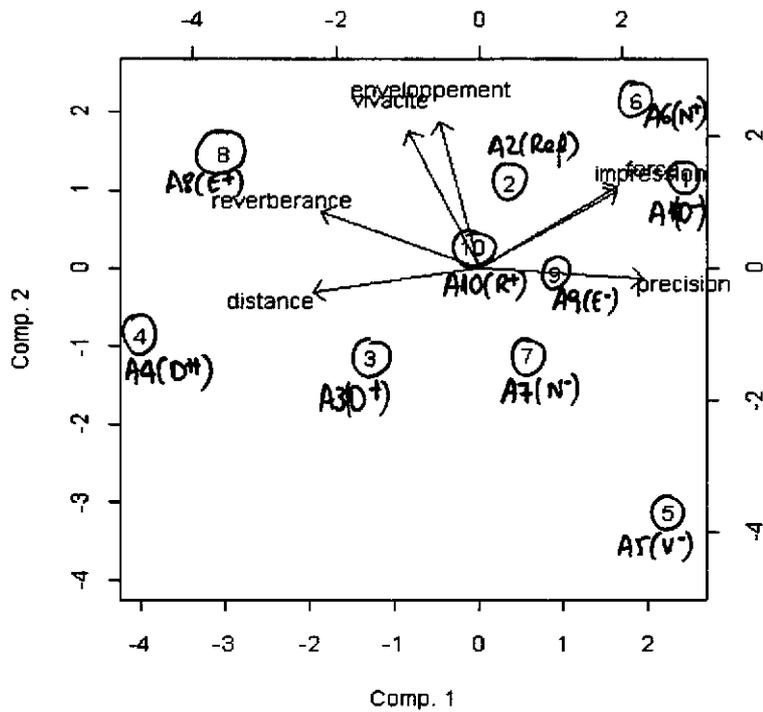
Variables	Composante 1	Composante 2	Composante 3
force	0.40	0.35	-0.29
précision	0.47		0.34
réverbérance	-0.45	0.24	0.16
vivacité	-0.20	0.58	0.45
distance	-0.47	0.10	0.28
enveloppement	-0.11	0.61	0.51
impression générale	0.39	0.33	0.48

**Tableau 5.2 :** i) Proportion de variance et proportion cumulative expliquées par les trois premières composantes principales pour la partie unimodale de l'expérience principale (10 stimuli sonores présentés sans vue). ii) Qualité de représentation des variables sur ces trois composantes principales.

Nous observons que le premier plan factoriel (composantes 1 et 2) représente 92 % de la variance totale des réponses. C'est donc sur les deux premières composantes principales que nous focaliserons notre intérêt : la première composante semble grouper les facteurs liés à la précision, elle est liée positivement à la force sonore et inversement à la distance et à la réverbérance. La deuxième composante regroupe l'enveloppement et la vivacité ; elle est liée positivement à la force sonore et à la réverbérance. Pour résumer, on peut dire que le premier axe est corrélé aux sensations liées à la source (distance, précision), tandis que le deuxième est corrélé à un faisceau de sensations liées à la salle (enveloppement, vivacité, réverbérance).

L'impression générale est partagée entre les deux axes, elle est liée à une perception simultanée d'une grande présence de la source et d'un effet de salle notable (enveloppement, vivacité).

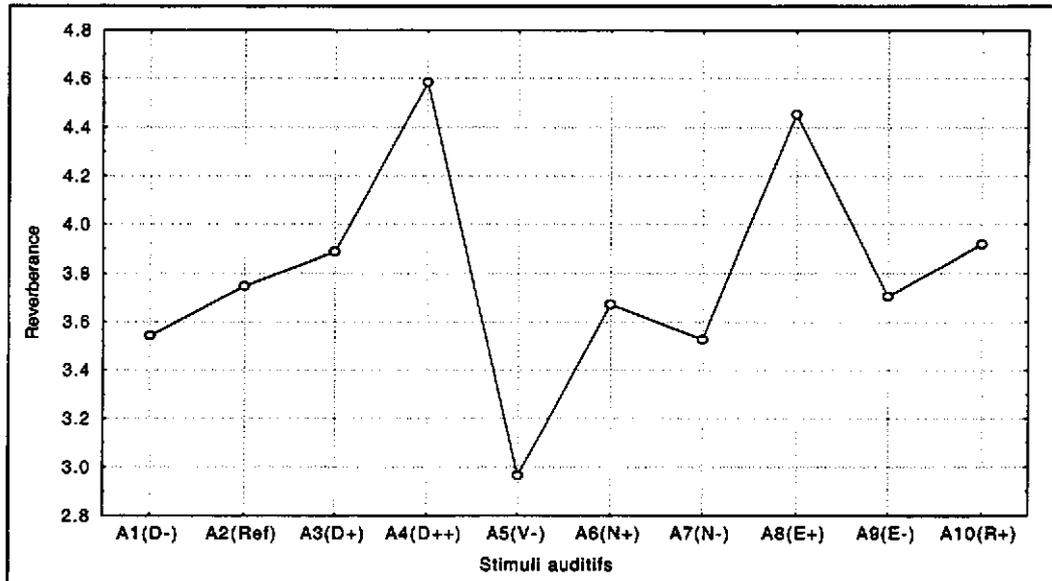
A propos des stimuli sonores, leur répartition sur les deux plans factoriels est cohérente avec les directions des vecteurs des facteurs, c'est à dire que par exemple, les stimuli qui varient sur la distance se trouvent plus ou moins ordonnés le long du vecteur de distance, celui qui est plus loin se trouvant plus loin dans la direction positive de ce vecteur, et vice versa.



1=A1=D, 2=A2=Ref, 3=A3= D<sup>+</sup>, 4=A4= D<sup>++</sup>, 5=A5= V, 6=A6= N<sup>+</sup>, 7=A7= N, 8=A8= E<sup>+</sup>, 9=A9= E, 10=A10= R<sup>+</sup>

**Figure 5.4 :** Représentation du premier et du deuxième plans factoriels pour la partie unimodale de l'expérience. Les variables sont représentées par des vecteurs et les stimuli sonores A1-A10 par des points-chiffres correspondants (1-10) (voir aussi figure 5.1).

La seule exception est pour la réverbérance qui ne semble pas avoir été jugée sur la base des deux stimuli variant sur cet attribut (A2 et A10). En effet, on remarque à l'aide de la **figure 5.5** que ce sont les stimuli A4 (D<sup>++</sup>) et A8 (E<sup>+</sup>) qui sont perçus comme les plus réverbérants et que le stimulus A5 (V<sup>-</sup>) est perçu comme le moins réverbérant.



**Figure 5.5 :** Moyennes des réponses sur la réverbérance pour les 10 stimuli auditifs de la partie unimodale de l'expérience.

Ceci peut s'expliquer comme suit :

Bien que nous ayons pris soin autant que possible de créer des stimuli ne variant les uns par rapport aux autres que sur un aspect perceptif, nous avons une exception notable à cette démarche<sup>1</sup> : le stimulus A4. On observe à l'aide du tableau C.6 de l'annexe C, que le stimulus A4, bien qu'ayant le même Rt (TR) que les autres, de 1.3 s, a un Edt (Early Decay Time) de 1.30 s, tandis que les autres ont un Edt de 1.0 s. L'indice Edt est lié à la réverbérance courante (running reverberance), c'est à dire, à la possibilité de percevoir le « traînage » de la salle dans le discours courant, sans attendre une fin de phrase musicale ou un transitoire marqué (son percussif) ; il est souvent utilisé comme l'indice objectif lié à la réverbérance plutôt que le TR. Dans cette étude, les deux ont été distingués : d'une part la réverbérance tardive liée au TR et associée à la perception de traînage du son, surtout sensible lorsque le message musical s'arrête, et d'autre part, la réverbérance courante liée à l'Edt.

La configuration A4 présente l'Edt le plus fort de la session, il n'est donc pas étonnant de le trouver en position maximale sur la sensation de réverbérance (or, quand on regarde la configuration A10 qui a été construite pour donner une plus grande réverbérance (Rt = 1.7 s), on s'aperçoit que l'Edt n'est que de 1.0 s, c'est à dire comme la plupart des autres configurations). Nous supposons donc que les sujets ont spontanément 'interprété' la notion de réverbérance dans le sens de la réverbérance courante.

<sup>1</sup> La raison pour laquelle il est difficile de créer des situations où l'Edt ne varie pas, est que l'ensemble des relations entre facteurs perceptifs et indices acoustiques fait parfois intervenir des contraintes entre les facteurs (certains deviennent esclaves, et il n'est pas possible d'avoir une totale liberté de réglage). Ainsi, le fait de baisser Es s'accompagne d'une diminution de l'Edt (cela se voit clairement dans le tableau C.6 (annexe C), pour les stimuli A1, A2, A3 et A4).

En ce qui concerne le stimulus A5, il a été construit pour créer une vivacité moindre (ce qui est le cas, puisqu'il se projette le plus bas sur le vecteur de « vivacité »). Or, il est souvent noté que la vivacité – qui correspond à la perception de la durée de réverbération dans les hautes fréquences – a un effet d'entraînement sur la perception globale de la réverbérance : pour deux configurations avec des TR moyens identiques par exemple, la configuration possédant un TR plus élevé aux hautes fréquences aura tendance simultanément, à créer un effet de vivacité, mais également à augmenter la perception globale de la réverbérance<sup>1</sup>.

En ce qui concerne le stimulus A8, nous supposons que sa position élevée dans les jugements de la réverbérance, traduit une certaine confusion dans l'interprétation sémantique des termes. L'hypothèse est que les configurations qui mettent en valeur une sensation liée à la salle (A4 avec un Edt supérieur aux autres, A8 avec un enveloppement supérieur aux autres), ont été « étiquetées » de « plus réverbérantes » ; la réverbérance rassemble toutes les sensations liées à l'effet de salle. A l'inverse, la configuration A5 qui se distinguait par un effet de salle très peu vivant, a été jugée moins réverbérante.

Par ailleurs, des Analyses de Variance effectuées séparément pour chacun des facteurs du questionnaire – avec dans tous les cas les stimuli sonores comme facteur indépendant (10 niveaux) – ont révélé des effets significatifs (à l'exception de la réverbérance). Ce qui indique que des différences perceptives notables sont observées entre les stimuli. De plus, ces analyses ont permis de vérifier les observations obtenues sur les plans factoriels par rapport à la répartition des stimuli sonores : la perception des stimuli se montre en général cohérente avec l'objectif de leur conception, c'est à dire que par exemple, D<sup>+</sup> est perçu plus 'proche' que D<sup>++</sup> ; E<sup>+</sup> plus 'enveloppant' que E<sup>-</sup>, V est perçu moins 'vivant' que tous les autres stimuli, etc. Les corrélations entre ces facteurs ont été calculées, elles sont présentées au **tableau 5.3**.

	force	précision	réverbérance	vivacité	distance	envelopp.	impr.gén.
force	1						
précision	0.70	1					
réverbérance	-0.57	-0.90	1				
vivacité	0.08	-0.40	0.69	1			
distance	-0.88	-0.89	0.84	0.32	1		
envelopp.	0.29	-0.32	0.50	0.80	0.04	1	
impr. gén.	0.84	0.77	-0.52	0.13	-0.83	0.19	1

**Tableau 5.3** : Tableau des corrélations entre variables pour la partie unimodale de l'expérience; résultats obtenus sur 32 sujets.

### 5.2.3.2 La partie bimodale – La perception des stimuli sonores

Dans un premier temps, nous nous intéresserons à la perception des sons indépendamment des conditions visuelles, l'effet de la perception visuelle sur les réponses sera étudié à la suite. C'est pour

<sup>1</sup> En ce sens, c'est d'ailleurs un signe de limite de l'indépendance des facteurs perceptifs.

cette raison que la question de la compatibilité des stimuli hétéromodaux entre eux – question liée à la perception visuelle – n’est pas incluse dans les analyses de cette section.

Les résultats de l’ACP pour la partie bimodale de l’expérience sont présentés au **tableau 5.4** : il présente la proportion de la variance expliquée par les quatre premières composantes principales, la proportion cumulative de cette variance, ainsi que la qualité de représentation des facteurs-variables sur les axes synthétiques.

	Composante 1	Composante 2	Composante 3	Composante 4
Proportion de Variance	0.57	0.27	0.05	0.05
Variance Cumulative	0.57	0.85	0.90	0.95

Variables	Composante 1	Composante 2	Composante 3	Composante 4
force	0.46	0.13		-0.29
précision	0.47	-0.10		0.13
réverbérance	-0.38	0.38	0.56	-0.20
vivacité		0.65	-0.42	-0.52
distance	-0.48		0.23	
enveloppement		0.64		0.74
impression générale	0.44		0.68	-0.19

**Tableau 5.4** : i) Proportion de variance et proportion cumulative expliquées par les quatre premières composantes principales pour la partie bimodale de l’expérience principale ( 10 stimuli sonores présentés x 4 vues = 40 stimuli bimodaux). ii) Qualité de représentation des variables sur ces quatre composantes.

La **figure 5.6** présente les deux premiers plans factoriels (1:2 et 2:3 entre les composantes correspondantes), le **tableau 5.5** donne les notations des stimuli bimodaux projetés sur ces plans.

	A1(D)	A2(Ref)	A3(D <sup>+</sup> )	A4(D <sup>++</sup> )	A5(V)	A6(N <sup>-</sup> )	A7(N <sup>-</sup> )	A8(E <sup>-</sup> )	A9(E)	A10(R <sup>-</sup> )
V1	1	5	9	13	17	21	25	29	33	37
V2	2	6	10	14	18	22	26	30	34	38
V3	3	7	11	15	19	23	27	31	35	39
V4	4	8	12	16	20	24	28	32	36	40

**Tableau 5.5** : Notations des stimuli représentés sur les plans factoriels de la partie bimodale de l’expérience (**figure 5.6**). 1, 2, 3 et 4, par exemple, correspondent au même son A1 présenté sous les quatre conditions visuelles V1, V2, V3 et V4; également 5, 6, 7 et 8 correspondent au son A2, etc.

Nous observons que les résultats sont très similaires à ceux obtenus pour la partie unimodale de l’expérience. Sur la **figure 5.6** (et à l’aide du **tableau 5.5**), les groupes que forment de mêmes stimuli auditifs sous différentes conditions visuelles (groupes A1, A2, etc.), sont distincts, se recouvrent peu et

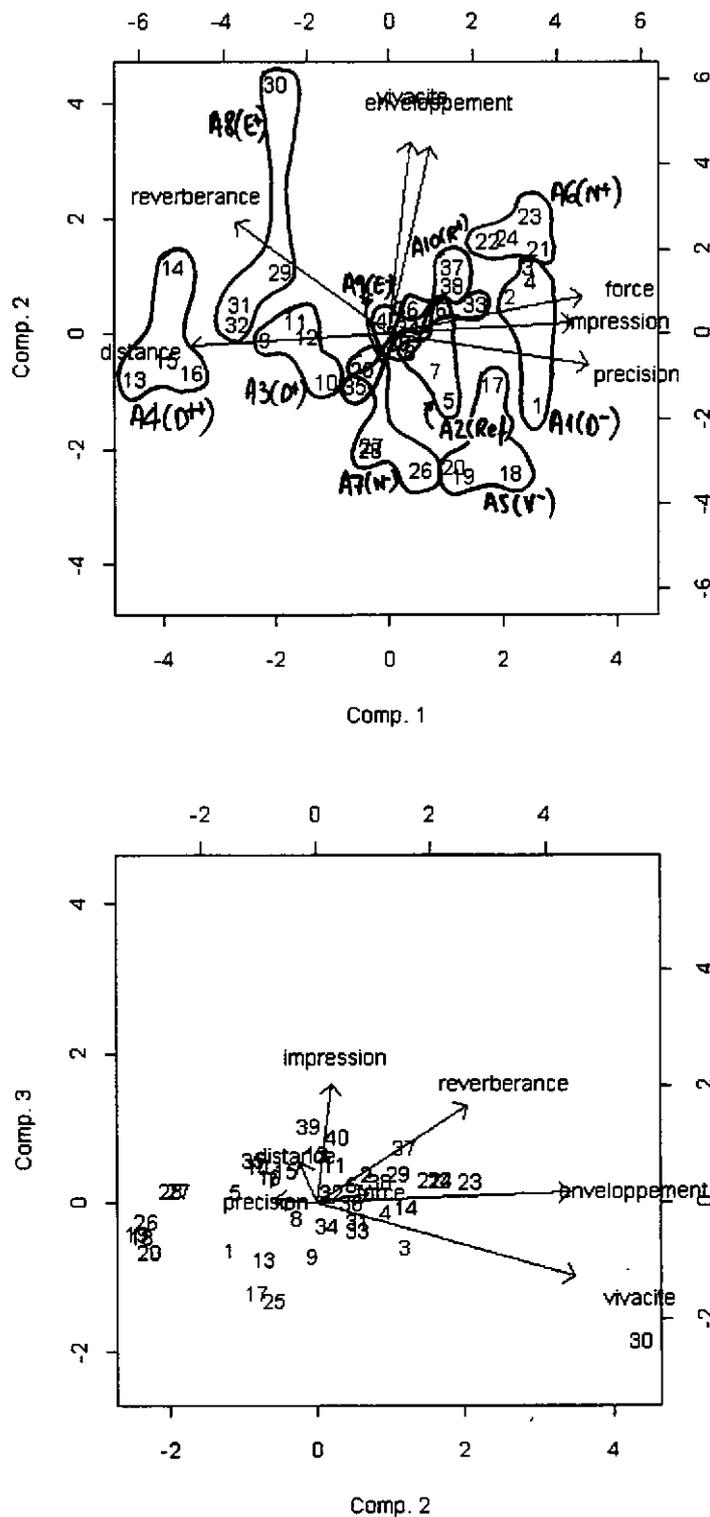


Figure 5.6 : Représentation du premier et du deuxième plans factoriels pour la partie bimodale de l'expérience. Les variables sont représentées par des vecteurs et les stimuli sonores A1-A40 par des points- chiffres correspondants (1-40) (voir notations sur le tableau 5.5).

se trouvent globalement dans les mêmes positions que les stimuli de la partie unimodale. Ce qui montre que l'expérience bimodale ne modifie pas notablement la structure perceptive. Les deux premiers plans factoriels (composantes 1, 2 et 3) expliquent 90 % de la variance totale des réponses ; il s'agit d'une variance un peu plus répartie sur les axes factoriels que celle de l'expérience unimodale (97 % de la variance pour les deux premiers plans factoriels), mais ce sont toujours les mêmes groupes de variables qui se trouvent représentés par les composantes principales.

La première composante regroupe ainsi les facteurs précision, force sonore et distance (inversement) ; elle est aussi inversement liée à la réverbérance.

La deuxième composante regroupe l'enveloppement et la vivacité ; elle est légèrement liée à la réverbérance. Comme pour la partie uniquement auditive de l'expérience, elle semble regrouper les trois sensations liées à l'effet de salle.

Les corrélations entre les facteurs sont données sur le **tableau 5.6**, nous permettant de vérifier les observations obtenues à l'aide des plans factoriels. On observe, entre autres, que l'impression générale est liée surtout à la force sonore, à la précision et inversement à la distance : il s'agit donc d'un son précis, proche et de niveau plutôt élevé, qui semble être le plus apprécié.

On observe également que l'impression générale et la force sonore subissent une petite rotation par rapport à la partie unimodale : dans cette dernière, les gens préfèrent être proches mais avec quand même une sensation de salle ; dans la partie bimodale, ils semblent être exclusivement intéressés par la proximité auditive (au moins pour les vues V2, V3 et V4, cf. figure 5.9).

	force	précision	réverbérance	vivacité	distance	envelopp.	impr.gén.
force	1						
précision	0.81	1					
réverbérance	-0.52	-0.77	1				
vivacité	0.25	-0.04	0.33	1			
distance	-0.90	-0.86	0.70	-0.14	1		
envelopp.	0.27	0.09	0.27	0.67	-0.24	1	
impr.gén.	0.79	0.81	-0.51	0.09	-0.76	0.19	1

**Tableaux 5.6 :** Matrice de corrélation pour la partie bimodale de l'expérience ; résultats obtenus sur 32 sujets.

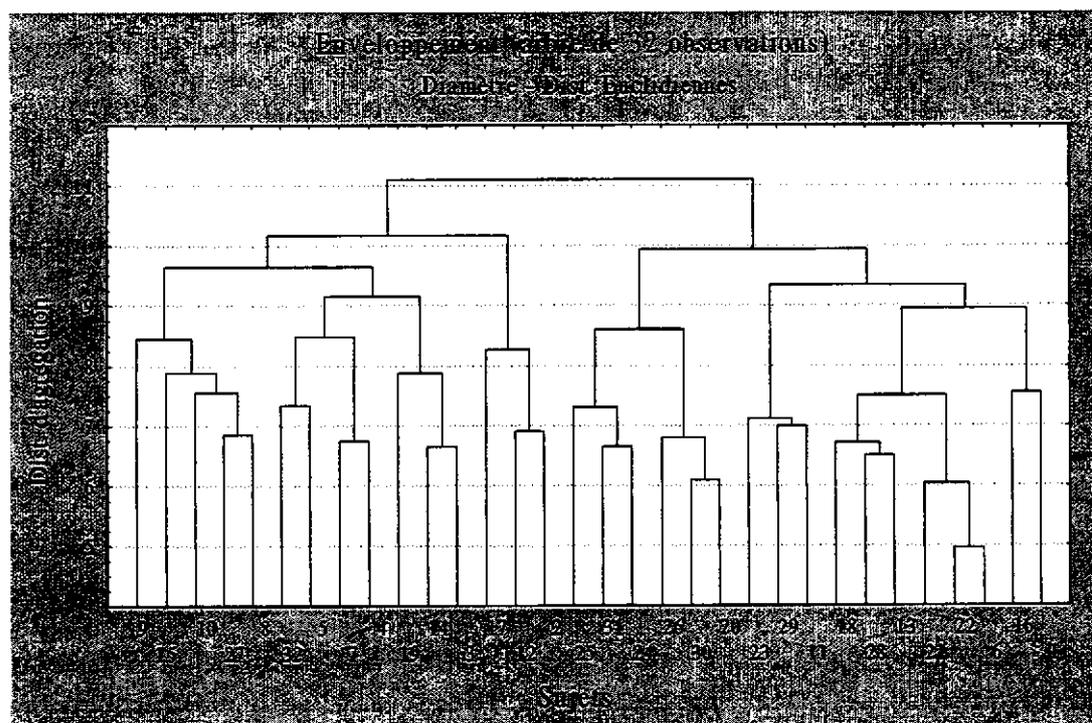
A propos des stimuli sonores, leur répartition sur les deux plans factoriels est cohérente avec les directions de vecteurs représentant les facteurs, c'est à dire que les projections des stimuli variant par exemple sur la distance, se trouvent réparties le long du vecteur de la distance :  $D^{++}$  (stimuli 13, 14, 15 et 16), le plus lointain, se trouve plus loin dans la direction positive de ce vecteur,  $D^+$  (stimuli 9, 10, 11 et 12) et Ref (stimuli 5, 6, 7 et 8) se projettent conformément à l'impression de distance qu'ils représentent, tandis que  $D^-$  (stimuli 1, 2, 3 et 4), le plus proche, se trouve plus loin dans la direction inverse de ce vecteur. Les seules exceptions sont la réverbérance (Ref et  $R^+$ ) et l'enveloppement ( $E^+$ , Ref et  $E^-$ ) ; tous les deux étant jugés, comme nous le verrons plus loin, sur la base d'autres aspects perceptifs.

En ce qui concerne la réverbérance, nous observons que, comme pour la partie unimodale de l'expérience, ce sont les stimuli A4 ( $D^{++}$ ) et A8 ( $E^+$ ) qui sont perçus comme les plus réverbérants et non plus le stimulus A10 ( $R^+$ ). Pour les mêmes raisons présentées dans la partie unimodale, ces résultats ne sont pas surprenants : la réverbérance a été jugée dans le sens de la réverbérance courante,

dont l'indice acoustique est l'EDT (A4 est le stimulus maximisant l'EDT), plutôt que sur l'indice classique de réverbérance qu'est le Tr (Tr maximisé par A10) ; la position de A8 traduit une confusion sémantique entre réverbérance et effet de salle (A8 est le stimulus maximisant la sensation d'enveloppement). La représentation graphique de la réverbérance pour la partie bimodale de l'expérience est très similaire à celle de la figure 5.5, elle n'est pas présentée ici.

La manière dont l'enveloppement a été jugé, nécessite une étude un peu plus approfondie. Dans la partie unimodale de l'expérience, les stimuli A8 (E<sup>+</sup>), A2 (Ref) et A9 (E<sup>-</sup>) se projettent régulièrement sur l'axe 'Enveloppement', en revanche, les trois groupes de stimuli correspondants (A8 (E<sup>+</sup>), A2 (Ref) et A9 (E<sup>-</sup>)), voir tableau 5.6, sont projetés au même endroit (sauf pour le stimulus 30).

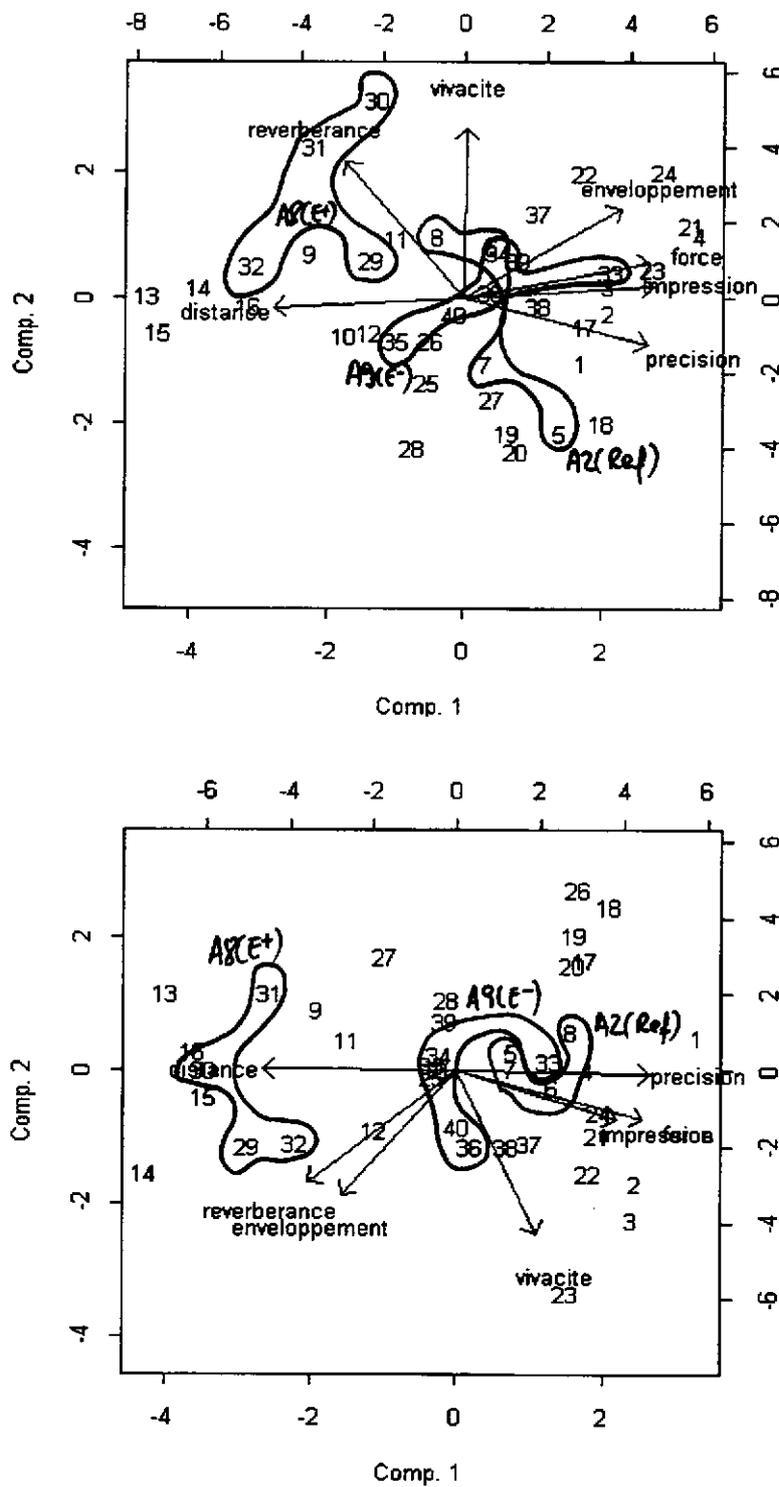
Des analyses intermédiaires effectuées sur des sous-groupes de sujets, ont suggéré que ceux-ci jugeaient l'enveloppement sur la base d'attributs perceptifs différents. Une analyse de classification<sup>1</sup>, réalisée sur les réponses normalisées des sujets, a permis d'étudier ce phénomène; l'arbre de classification obtenu est présenté à la figure 5.7.



**Figure 5.7 :** Classification des 32 sujets par rapport à leurs réponses sur l'enveloppement. Deux 'classes' se distinguent, une première à 15 sujets (groupe 1) et la deuxième à 17 sujets (groupe 2).

Nous observons que les sujets se distinguent en deux 'groupes' par rapport à leurs jugements sur l'enveloppement : le premier avec 15 sujets et le deuxième avec 17 sujets. Des analyses en composantes principales ont été réalisées séparément pour chacun de ces deux groupes ; les plans factoriels et les matrices de corrélation sont présentés à la figure 5.8 et aux tableaux 5.7 et 5.8.

<sup>1</sup> Les distances entre les différents 'objets' (ici les sujets) sont des distances euclidiennes ; les distances entre classes sont déterminées par la plus grande distance existant entre deux objets de classes différentes (méthode d'agrégation 'diamètre').



**Figure 5.8 :** Représentation des premiers plans factoriels pour les deux groupes de sujets (plans 1:2) : groupe 1 (en haut) et groupe 2 (en bas). Les variables sont représentées par des vecteurs et les stimuli sonores par des points 1-40 (voir notations sur le tableau 5.5). Les plans sont malheureusement inversés par le mode de présentation du logiciel d'analyse.

	Force	Précision	Réverbérance	Vivacité	Distance	Envelopp.	Impr.gén.
Force	1						
Précision	0.70	1					
Réverbérance	-0.36	-0.71	1				
Vivacité	0.15	-0.14	0.51	1			
Distance	-0.89	-0.74	0.48	-0.04	1		
Envelopp.	<b>0.68</b>	0.56	<b>-0.07</b>	0.33	-0.67	1	
Impr. gén.	<b>0.81</b>	0.76	<b>-0.46</b>	0.04	-0.80	<b>0.70</b>	1

	Force	Précision	Réverbérance	Vivacité	Distance	Envelopp.	Impr.gén.
Force	1						
Précision	0.77	1					
Réverbérance	-0.50	-0.61	1				
Vivacité	0.54	0.33	0.22	1			
Distance	-0.85	-0.81	0.72	-0.38	1		
Envelopp.	<b>-0.30</b>	-0.48	<b>0.63</b>	0.25	0.42	1	
Impr. gén.	<b>0.63</b>	0.72	<b>-0.32</b>	0.43	-0.58	<b>-0.30</b>	1

**Tableau 5.7 et Tableau 5.8 :** Matrices de corrélation entre variables présentées séparément pour le premier et le deuxième groupe de sujets (groupe 1 en haut et groupe 2 en bas). La comparaison des valeurs en gras met en évidence les différences perceptives sur l'enveloppement.

Pour le premier groupe de sujets (groupe 1), l'enveloppement est surtout lié à la force sonore et à la précision ( $r = 0.68$  et  $r = 0.56$  respectivement), il n'y a pas de corrélation avec la réverbérance. Au contraire, pour le deuxième groupe (groupe 2), l'enveloppement est lié à la réverbérance ( $r = 0.63$ ) et inversement lié à la force sonore et à la précision ( $r = -0.30$  et  $r = -0.48$  respectivement).

Nous observons d'ailleurs sur la figure 5.8, que pour le groupe 1 – qui associe l'enveloppement à la force sonore – Ref(5, 6, 7 et 8), E<sup>+</sup>(29, 30, 31, et 32) et E(33, 34, 35 et 36) sont projetés sur une même zone du vecteur 'Enveloppement', tandis que pour le groupe 2 – qui associe l'enveloppement à une sensation liée plutôt à l'effet de salle – ces mêmes stimuli se répartissent davantage le long du vecteur 'Enveloppement'.

L'hypothèse que l'on pourrait formuler vis-à-vis de ces résultats, est que le groupe 1 serait constitué de personnes pour qui l'enveloppement signifie « être dans le son » – au sens d'une écoute à fort volume sonore – tandis que le groupe 2 associerait l'enveloppement au fait de « se sentir dans la salle », et ferait la différence entre le son frontal (sources proches) et un son enveloppant (énergie latérale).

Pour les deux groupes de sujets, l'impression générale est liée principalement à la force sonore, à la précision et à la proximité de la source : l'enveloppement est donc considéré comme une qualité par le premier groupe ( $r = 0,70$ ), tandis que le deuxième groupe le juge plutôt neutre ( $r = -0,30$ ).

### 5.2.3.3 La partie bimodale – L'influence des images

Etudier l'influence des images à l'aide des plans factoriels de la figure 5.6 (notations au tableau 5.5) ne s'avère pas suffisamment informatif. On observe quelques variations dans l'espace perceptif – variations qui sembleraient être liées aux conditions visuelles (comme par exemple celles sur la position des stimuli 1, 2, 3 et 4 (son A1 (D<sup>-</sup>) sous les conditions visuelles V1, V2, V3 et V4 respectivement) ou sur la position des stimuli 29, 30, 31 et 32 (son A8 (E<sup>+</sup>) sous les conditions visuelles V1, V2, V3 et V4), etc. – mais ces variations ne sont pas toujours importantes, par ailleurs elles ne se manifestent pas souvent vers la même direction. De plus, pour d'autres stimuli (comme pour les stimuli 21, 22, 23 et 24 (son A6 (N<sup>+</sup>)), ces variations sur les deux premiers plans semblent minimales, ce qui indique que les quatre stimuli sont perçus comme étant identiques.

Ces observations n'étant pas très concluantes, d'autres approches doivent être considérées pour étudier ces mêmes résultats et mettre en évidence l'influence des conditions visuelles sur la perception sonore. Il s'agit :

- de la comparaison des plans factoriels issus des quatre analyses prenant en compte les réponses obtenues sous une seule condition visuelle à la fois (plans 1:2 par exemple, séparément pour les conditions visuelles V1, V2, V3 et V4) ;
- de l'introduction, dans l'analyse de l'espace auditif, d'un facteur représentant le 'visuel' et créé à l'aide des estimations objectives de distance (la mesure de la taille du haut-parleur sur l'écran) ;
- des analyses de variance réalisées séparément pour chacun des facteurs perceptifs du questionnaire, avec comme variables indépendantes, les conditions visuelles et les stimuli sonores.

Ces trois approches sont présentées à la suite.

#### *i) Comparaison des plans factoriels obtenus séparément pour les images V1, V2, V3 et V4*

Le premier plan factoriel (plan 1:2) pour chacune des conditions visuelles V1, V2, V3 et V4 est présenté à la figure 5.9.

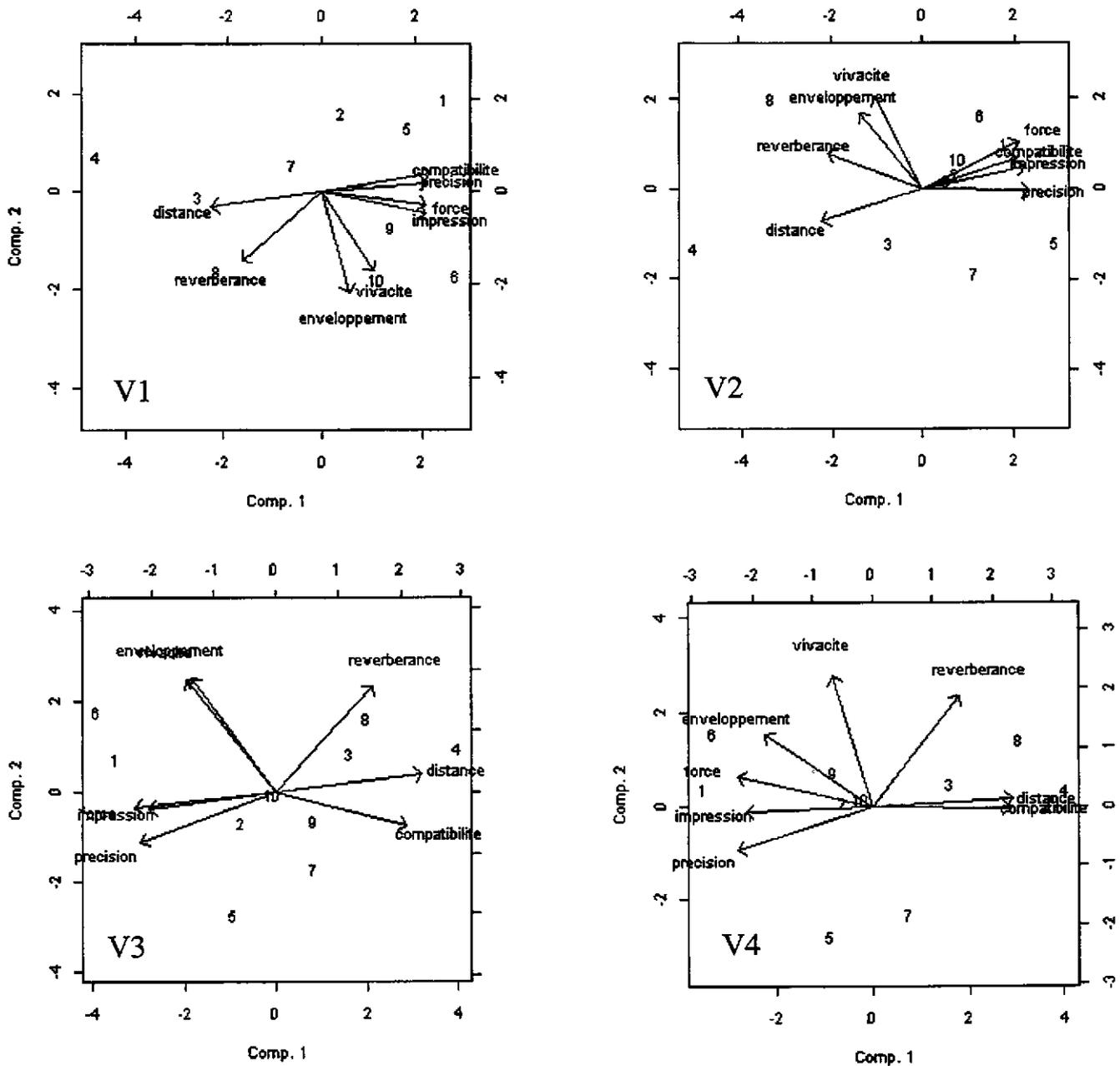
L'orientation des plans de la figure 5.9 n'a pas toujours été respectée par le logiciel d'analyse, ainsi les plans V3 et V4 ont la même orientation mais les plans V1 et V2 sont inversés. Afin de les comparer il faut considérer un vecteur de référence et les imaginer superposés sur ce vecteur. Ce vecteur doit être stable dans les quatre plans, il doit donc représenter un facteur qui risque le moins possible d'être influencé par les images. C'est la force sonore qui semble s'y prêter le plus<sup>1</sup>.

Le premier effet évident que l'on observe est le comportement du vecteur de la compatibilité de stimuli hétéromodaux. Son orientation s'inverse : d'un côté du plan – où il est fortement lié à la force

---

<sup>1</sup> Les analyses suivantes montreront qu'elle n'a pas été influencée par les vues.

sonore (images V1 et V2) – il se trouve en une position presque opposée, liée à la force sonore et très liée à la distance (images V3 et V4). Il s'agit d'un effet attendu : quand les sons sont forts ils sont considérés compatibles avec les images proches (V1 et V2) et peu compatibles avec les images lointaines (V3 et V4).



1=A1=D', 2=A2=Ref, 3=A3= D<sup>+</sup>, 4=A4= D<sup>++</sup>, 5=A5= V, 6=A6= N<sup>+</sup>, 7=A7= N', 8=A8= E<sup>+</sup>, 9=A9= E', 10=A10= R<sup>+</sup>

**Figure 5.9 :** Le plan factoriel (plan 1:2) présenté pour chacune des conditions visuelles V1, V2, V3 et V4 (les plans V3 et V4 ont la même orientation mais les plans V1 et V2 sont inversés).

En ce qui concerne les autres observations qui peuvent être faites sur ces plans, on note en règle générale, une rotation du vecteur de l'enveloppement avec la distance visuelle, distance qui augmente

progressivement avec les vues V1, V2, V3 et V4. Ainsi, l'angle<sup>1</sup> que l'enveloppement forme par exemple avec la distance auditive, augmente avec la distance visuelle (sauf pour la vue V2). Sous la condition visuelle proche (V1), la distance et l'enveloppement ne sont pas corrélés, tandis que sous les conditions visuelles lointaines (V3 et V4), ils sont inversement corrélés ; c'est à dire que lorsqu'un son est jugé lointain, il est rarement jugé enveloppant, et vice versa.

Les matrices de corrélation obtenues pour chacun des quatre espaces perceptifs, vérifient ces observations : le **tableau 5.9** présente quelques-uns des effets observés.

Corrélations	V1	V2	V3	V4
enveloppement - force sonore	0.31	-0.25	0.39	0.74
enveloppement - réverbérance	0.36	0.65	0.44	0.77
enveloppement - distance	-0.12	0.33	-0.45	-0.65
enveloppement - impression générale	0.38	-0.34	0.44	0.77

**Tableau 5.9 :** Corrélations de l'enveloppement avec la force sonore, la réverbérance, la distance et l'impression générale, séparément pour les quatre conditions visuelles.

On observe – à l'exception toujours de la condition visuelle V2 – qu'il semble y avoir une certaine suite dans ces variations (rotations dans les plans factoriels). Les corrélations, par exemple entre l'enveloppement et la distance auditive, baissent avec la distance visuelle (l'angle entre les vecteurs augmente) ; elles augmentent en revanche entre l'enveloppement et la force sonore, la réverbérance et l'impression générale (l'angle entre les vecteurs diminue). Il est intéressant de noter que dans tous les cas, ces corrélations semblent 's'affirmer' pour la condition visuelle V4, comme si, sous des conditions visuelles lointaines et excentrées, lorsque les sujets peuvent apercevoir les côtés de la salle, ils comprenaient mieux la notion de l'enveloppement et appréciaient davantage son importance.

Dans cette partie bimodale de l'expérience – et contrairement à la partie unimodale – l'enveloppement semble donc ne pas être piloté par les stimuli conçus à ce propos (A8 et A9).

Ces effets ne peuvent pourtant être considérées que comme indicatifs de l'influence de la perception visuelle, ils ne permettent pas d'établir l'existence d'une tendance quelconque. Ceci, partiellement parce qu'ils sont complexes (comportement quelque peu différent sous de différentes conditions visuelles), mais aussi à cause de l'ambiguïté avec laquelle le facteur d'enveloppement a été jugé (existence de deux groupes de sujets).

Il faut noter ici que les corrélations entre les autres facteurs sont pratiquement stables, sauf pour la vivacité pour laquelle les effets sont complexes, mais de toute façon, de grandes variations de jugements ont été observées indépendamment des vues pour ce facteur.

<sup>1</sup> Il s'agit en réalité de la projection de cet angle sur le premier plan factoriel.

## ii) Introduction de variable représentant la 'distance visuelle'

Une estimation des distances absolues entre les sujets et la scène – telles qu'elles étaient simulées par les images – étant difficiles à obtenir, les distances relatives ont été calculées en faisant intervenir la taille du haut-parleur qui apparaît dans les images (il était placé au milieu de la scène). Cette taille est inversement liée à la distance à laquelle un observateur voit le haut-parleur (loi de perspective) ; elle a donc été mesurée sur l'écran pour chacune des conditions visuelles (hauteur du haut-parleur : 24,2 cm, 7,5 cm, 6 cm et 4,7 cm respectivement pour les conditions V1, V2, V3 et V4). Les rapports de distances auxquelles le haut-parleur – et par conséquent la scène – sont vus par les sujets, sont ainsi calculés comme étant  $V1:V2:V3:V4 = 1,65:5,33:6,66:8,50$ .

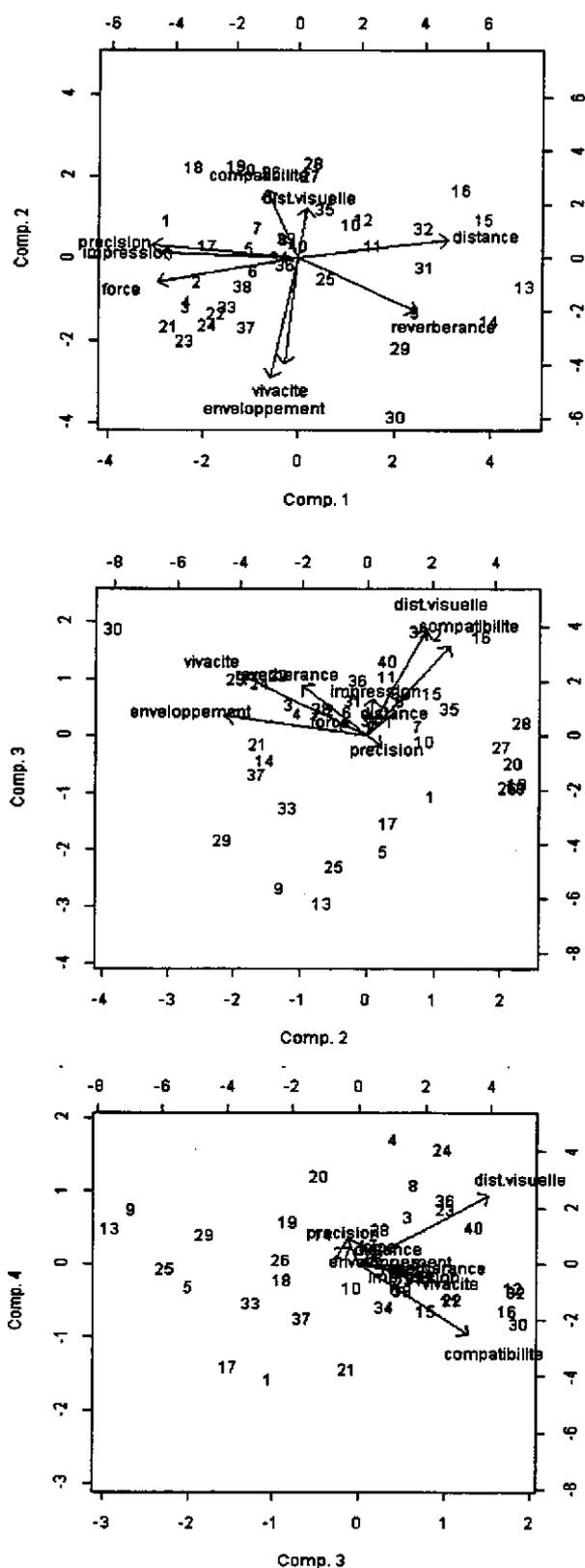
Ce vecteur des rapports a été introduit dans l'analyse comme une nouvelle variable. Afin de lui attribuer un nom explicite, elle est appelée 'distance visuelle' – avec donc une connotation subjective – malgré le fait qu'elle représente une distance objective<sup>1</sup>. La proportion de variance expliquée par les quatre premières composantes et la qualité de représentation des variables sur ces quatre composantes, sont présentées au tableau 5.10 ; la figure 5.10 présente les trois premiers plans factoriels.

	Composante 1	Composante 2	Composante 3	Composante 4
Proportion de Variance	0.45	0.23	0.16	0.06
Variance Cumulative	0.45	0.68	0.84	0.90

Variables	Composante 1	Composante 2	Composante 3	Composante 4
force	-0.45	-0.13		0.14
précision	-0.47			0.25
réverbérance	0.37	-0.28	0.30	
vivacité		-0.55	0.36	-0.13
distance	0.47		0.12	0.12
enveloppement		-0.63	0.12	
impression générale	-0.44		0.22	-0.19
Compatibilité de stimuli	-0.10	0.36	0.54	-0.67
'distance visuelle'		0.26	0.64	0.66

**Tableau 5.10 :** Partie bimodale de l'expérience principale : résultats d'une analyse (ACP) prenant en compte une variable représentant la 'distance visuelle'. i) Proportion de variance et proportion cumulative de variance expliquées par les quatre premières composantes. ii) Qualité de représentation des variables sur les quatre composantes.

<sup>1</sup> Par opposition à 'distance subjective' pour laquelle la réalisation de tests perceptifs visuels sont nécessaires, réalisés par exemple à l'expérience préliminaire visuelle.



**Figure 5.10:** Représentation des premier, deuxième et troisième plans factoriels pour la partie bimodale de l'expérience après l'introduction de la variable représentant la 'distance visuelle'. Les variables sont représentées par des vecteurs et les stimuli sonores A1-A40 par des points-chiffres correspondants (voir notations sur le tableau 5.5).

Nous observons que le vecteur représentant la distance visuelle<sup>1</sup> est assez mal représenté sur le premier plan factoriel, tandis qu'il est bien plus présent aux deuxième et troisième plans. En effet, il est observé, à l'aide du **tableau 5.10**, que la qualité de sa représentation sur les deux premières composantes est très mauvaise, tandis que celle sur les composantes trois et quatre est bien plus élevée. Le même comportement, un peu moins prononcé, est observé pour la compatibilité des stimuli hétéromodaux.

Nous observons ainsi, qu'en règle générale, les aspects auditifs des stimuli sont liés aux deux premières composantes principales, et semblent globalement indépendants des aspects liés à la perception visuelle représentée par les composantes trois et quatre.

	force	précision	réverbér.	vivacité	distance	envelopp.	impr.gén.	compat.	'dist.v'
force	1								
précision	0.81	1							
réverbéranc	-0.52	-0.77	1						
vivacité	0.24	-0.04	0.32	1					
distance	-0.90	-0.85	0.69	-0.14	1				
envelopp.	0.27	0.08	0.27	0.66	-0.23	1			
impr.gén.	0.78	0.80	-0.51	0.09	-0.75	0.19	1		
compat	0.09	0.11	-0.12	-0.05	-0.08	-0.30	0.39	1	
'dist.vis.'	0.00	0.00	0.14	0.01	0.25	-0.20	0.12	0.45	1

**Tableau 5.11** : Matrice de corrélation pour l'espace perceptif comprenant la variable représentant la 'distance visuelle' (32 sujets).

En ce qui concerne la 'distance visuelle' (**tableau 5.11**), bien que les corrélations entre distance visuelle et distance auditive et entre distance visuelle et enveloppement soient supérieures aux autres, elles restent beaucoup trop faibles pour indiquer une tendance quelconque.

### iii) Analyses de variance

Des analyses de variance à mesures répétées ont été réalisées séparément pour chacun des huit facteurs perceptifs du questionnaire. Il s'agit dans tous les cas d'analyses à deux facteurs avec interaction ; ces facteurs explicatifs de la variance sont les conditions visuelles (4 images) et les stimuli sonores (10 sons). Les résultats de ces analyses sont présentés au **tableau 5.12**.

<sup>1</sup> La même analyse a été effectuée en utilisant comme vecteur 'visuel', la distance telle qu'elle était perçue par un groupe de 20 sujets lors du test préliminaire visuel (distance subjective). Les résultats sont très similaires.

Analyse de variance						
	Conditions visuelles		Stimuli auditifs		Interaction	
	F <sub>V</sub> (3, 93)	niveau p	F <sub>A</sub> (9, 279)	niveau p	F <sub>AV</sub> (27, 998)	niveau p
<i>force sonore</i>	0.22	0.88	<b>85.34</b>	<b>*0.00</b>	0.75	0.81
<i>précision</i>	0.14	0.93	<b>11.70</b>	<b>*0.00</b>	0.99	0.46
<i>réverbérance</i>	<b>3.48</b>	<b>*0.02</b>	<b>16.13</b>	<b>*0.00</b>	0.59	0.95
<i>vivacité</i>	0.50	0.68	<b>2,94</b>	<b>*0.02</b>	1.21	0.20
<i>distance auditive</i>	<b>17.60</b>	<b>*0.00</b>	<b>76.68</b>	<b>*0.00</b>	<b>1.75</b>	<b>*0.01</b>
<i>enveloppement</i>	<b>3.74</b>	<b>*0.01</b>	<b>3.00</b>	<b>*0.00</b>	<b>2.11</b>	<b>*0.00</b>
<i>impression générale</i>	<b>4.18</b>	<b>*0.00</b>	<b>2.14</b>	<b>*0.02</b>	<b>4.27</b>	<b>*0.00</b>
<i>compatibilité</i>	<b>11.73</b>	<b>*0.00</b>	<b>20.63</b>	<b>*0.00</b>	<b>8.03</b>	<b>*0.00</b>

**Tableau 5.12 :** Résultats des analyses de variance effectuées pour chacun des huit facteurs du questionnaire (les sept premiers facteurs concernent uniquement l'appréciation auditive, la compatibilité concerne aussi les conditions visuelles). Les effets testés sont les effets principaux des conditions visuelles (4 niveaux) et des stimuli sonores (10 niveaux) ainsi que leur interaction. Les effets significatifs sont en gras et signalés par un astérisque (\*).

Nous constatons que les mêmes observations effectuées auparavant à l'aide de la matrice de corrélations, peuvent être faites ici à l'aide de l'analyse de la variance. La réverbérance, la distance auditive, l'enveloppement et l'impression générale sont influencés par les images, tandis que la force sonore, la précision et la vivacité ne le sont pas.

L'analyse de variance nous donnant la possibilité d'étudier les interactions statistiques entre facteurs expérimentaux, nous constatons que la plupart des effets observés sont complexes, l'influence dépend du 'niveau' de chaque facteur, c'est-à-dire de l'image ou du son présenté.

Afin de pouvoir observer les effets principaux des conditions visuelles, les graphiques pour chacun des quatre facteurs perceptifs influencés sont présentés à la figure 5.11. Le tableau 5.13 présente les moyennes des jugements sur ces quatre facteurs pour chacune des conditions visuelles V1, V2, V3 et V4 ; il présente également des comparaisons Post-Hoc (Tukey HSD) entre les quatre images afin de tester la validité des différences observées.

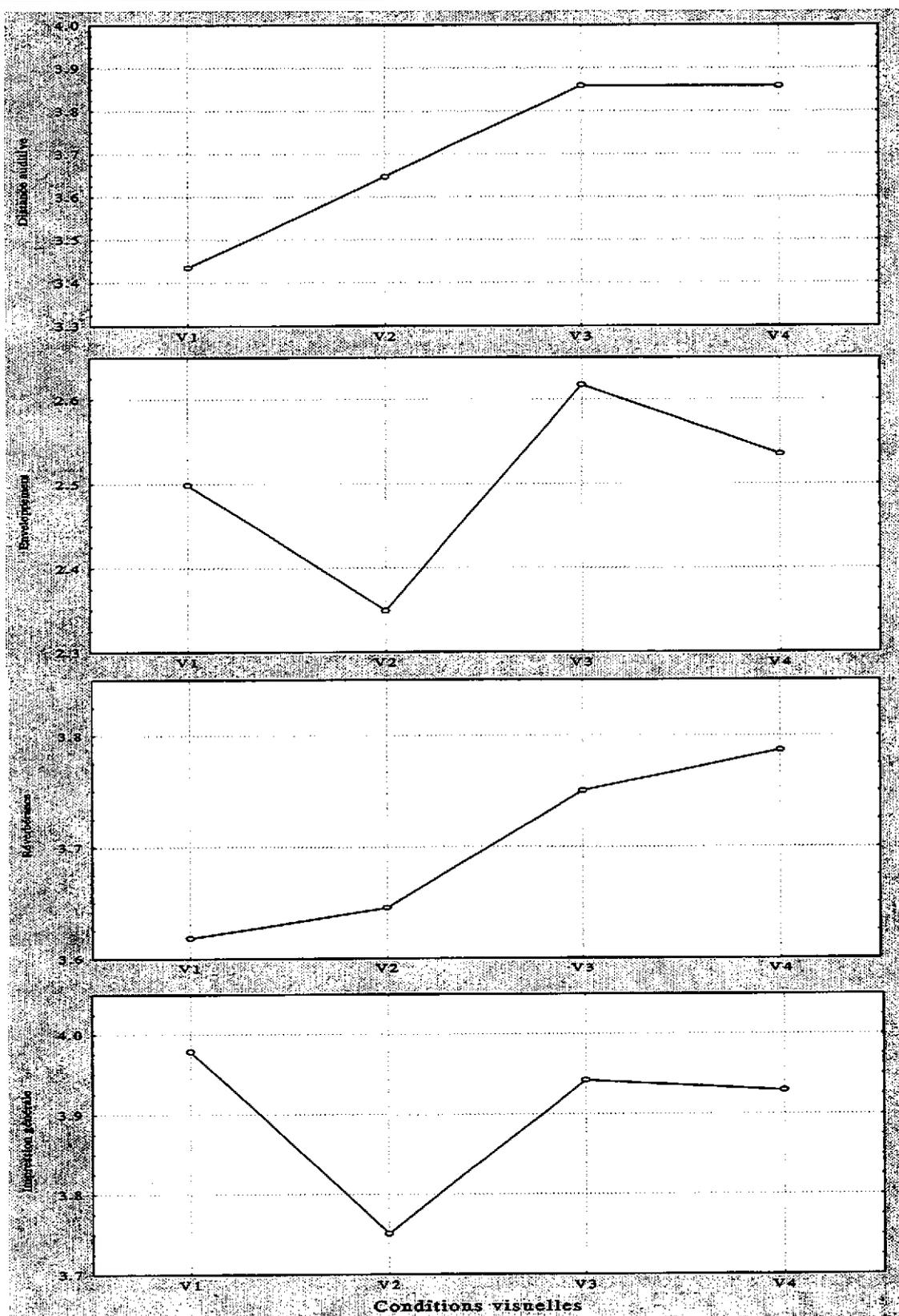


Figure 5.11 : Distance auditive, enveloppement, réverbérance et impression générale en fonction des conditions visuelles. Les valeurs sont moyennées sur les dix stimuli auditifs. L'effet des conditions visuelles est significatif pour les quatre cas (cf. tableau 5.12).



Conditions Visuelles : Post-Hoc (HSD Tukey)				
a. Distance	V1	V2	V3	V4
Moyennes	3.43	3.65	3.86	3.86
V1		*0.01	*0.00	*0.00
V2	*0.01		*0.01	*0.01
V3	*0.00	*0.01		1.00
V4	*0.00	*0.01	1.00	

Conditions Visuelles : Post-Hoc (HSD Tukey)				
b. Enveloppement	V1	V2	V3	V4
Moyennes	2.50	2.35	2.62	2.54
V1		0.27	0.47	0.97
V2	0.27		*0.01	0.11
V3	0.47	*0.01		0.75
V4	0.97	0.11	0.75	

Conditions Visuelles : Post-Hoc (HSD Tukey)				
c. Réverbérance	V1	V2	V3	V4
Moyennes	3.62	3.65	3.75	3.79
V1		0.97	0.15	*0.04
V2	0.97		0.33	0.11
V3	0.15	0.33		0.93
V4	*0.04	0.11	0.93	

Conditions Visuelles : Post-Hoc (HSD Tukey)				
d. Impres. générale	V1	V2	V3	V4
Moyennes	3.98	3.75	3.94	3.93
V1		*0.01	0.95	0.90
V2	*0.01		*0.04	0.06
V3	0.95	*0.04		1.00
V4	0.90	0.06	1.00	

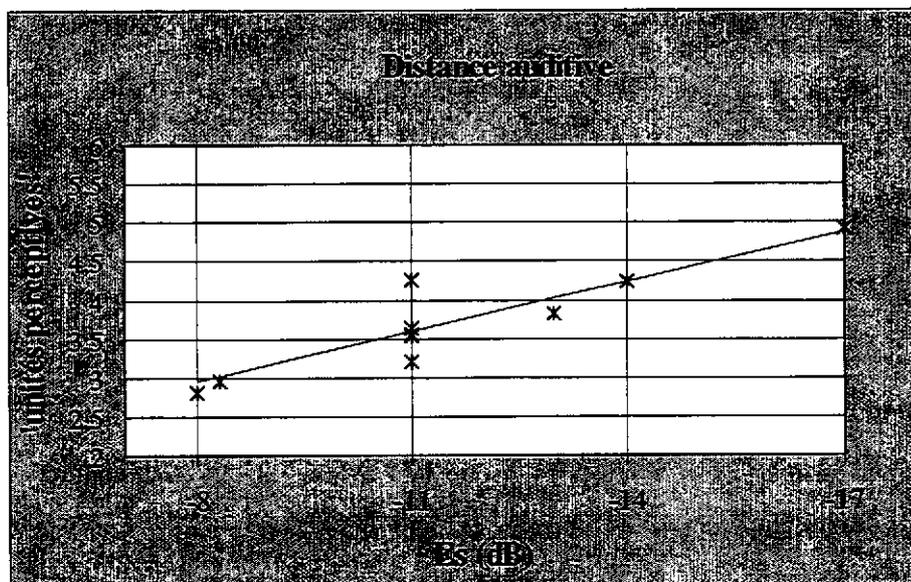
**Tableau 5.13 :** Moyennes et comparaisons Post-Hoc (Tukey HSD) entre les quatre niveaux de conditions visuelles (V1, V2, V3 et V4) séparément, pour a : la distance auditive, b : l'enveloppement, c : la réverbérance et d : l'impression générale. Les différences significatives sont signalées par un astérisque (\*).

La figure 5.11 incite à tenter de trouver une relation quantitative entre l'éloignement visuel et les différentes sensations auditives. Plus particulièrement, il nous intéresse d'estimer à quel changement sur l'indice objectif utilisé pour piloter chacune de ces sensations, correspond l'influence des images – ou pour être plus précis – le rapport de distances des images qui a provoqué ce changement.

Cette tentative est faite pour les attributs de la distance visuelle et de la réverbérance. D'une part, parce que indépendamment des vues, ils présentent une relation presque linéaire entre les indices qui pilotent les sensations correspondantes (Es pour la distance auditive et Edt pour la réverbérance) et les sensations provoqués (cf. figures 5.12 et 5.13). D'autre part, parce que l'éloignement croissant des vues semble avoir une influence régulière sur leur perception (cf. figure 5.11).

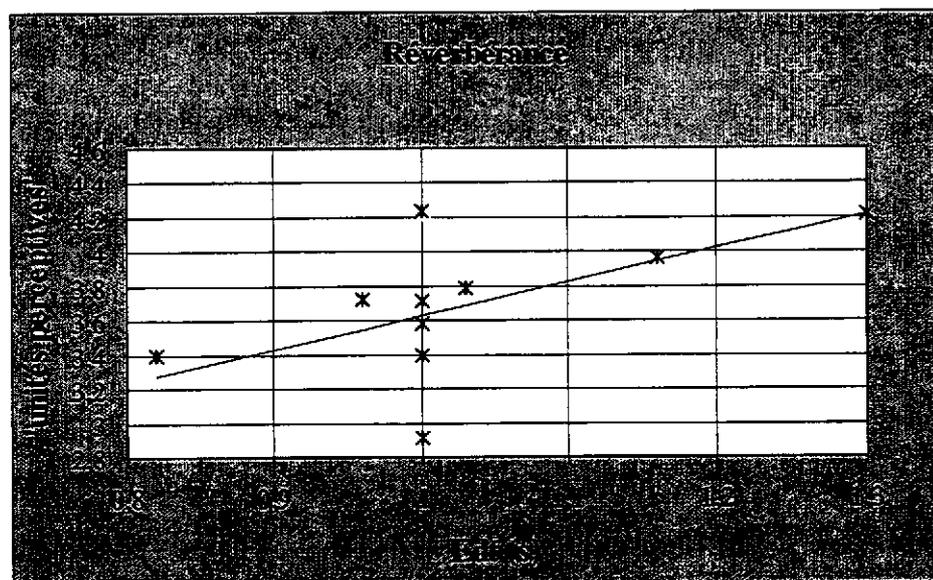
Ainsi, pour la distance auditive, on observe à l'aide de la figure 5.12 qu'un changement de 1 dB sur l'indice Es correspond à une différence d'environ 0.2 'unités perceptives (u.p.)' (pente de la tendance linéaire : 0.21 'u.p./ dB).

L'influence sur la perception de la distance apportée par les vues V2 et V1 est de l'ordre de 0.2 'u.p.' (tableau de la figure 5.13 ;  $3.65 - 3.43 = 0.22$ ), et l'influence apportée par les vues V3 et V2 est également de l'ordre de 0.2 'u.p.' (tableau 5.13 a ;  $3.86 - 3.65 = 0.21$ ). Or 0.2 'u.p.' correspond à une différence de 1dB sur l'indice Es (ce qui est la plus petite différence perceptible sur cet indice), ce qui semble correspondre également aux résultats obtenus précédemment avec les comparaisons Post-Hoc (Tukey HSD).



Stimuli	A1	A6	Ai	A7	A3	A4
Es (dB)	-8	-8.33	-11	-13	-14	-17
Distance auditive	2.81	2.96	3.63	3.83	4.24	4.92

Figure 5.12 : Distance auditive en fonction de l'indice Es (dB) des stimuli (3.63 est la distance moyenne pour les configurations avec Es = -11 dB ; A2 = 3.53, A5 = 3.2, A8 = 4.26, A9 = 3.56, A10 = 3.64) et tendance linéaire. Les valeurs sont moyennées sur toutes les conditions visuelles.



Stimuli	A1	A6	Ai	A10	A3	A4
Edt(s)	0.82	0.96	1	1.03	1.16	1.3
Réverbérance	3.40	3.72	3.60	3.79	3.97	4.22

Figure 5.13 : Réverbérance en fonction de l'indice Edt(s) des stimuli (3.60 est la réverbérance moyenne pour les configurations avec Edt = 1 s ; A2 = 3.58, A5 = 2.91, A7 = 3.4, A8 = 4.24, A9 = 3.71) et tendance linéaire. Les valeurs sont moyennées sur toutes les conditions visuelles.

Par ailleurs, les rapports de distance entre les images (cf. § 3.3.2) sont  $V2/V1 = 3.23$  (provoquant une différence de 0.2 'u.p'),  $V3/V2 = 1.24$  (pour 0.2 'u.p' également), et  $V4/V3 = 1.27$  (pour 0 'u.p'). On observe alors que la distance visuelle - malgré le fait qu'elle influence incontestablement la perception de la distance auditive - n'est pas un bon estimateur de cette influence, puisque les rapports  $V2/V1$  et  $V3/V2$  sont différents alors qu'ils provoquent la même différence auditive, tandis que les rapports  $V3/V2$  et  $V4/V3$  qui sont similaires<sup>1</sup>, ne provoquent pas la même différence.

Enfin, nous remarquons que l'influence observée entre les conditions visuelles  $V3$  et  $V1$  (rapport de distances visuelles  $V3/V1 = 4$ ) est de l'ordre de 0.4 'u.p.', ce qui correspond à une différence de 2 dB sur l'indice  $E_s$ . Cet effet est le double de celui trouvé par l'expérience 5 sur la distance (les vues  $V1$  et  $V3$ , ainsi que quatre des stimuli auditifs sont les mêmes). Or, l'expérience présentée ici adopte une procédure où les images changent, tandis que l'expérience 5 sur la distance est une expérience où les vues sont statiques. Ce fait souligne encore une fois l'importance des effets dynamiques dans la perception (observés initialement à l'occasion des expériences 5 et 6 sur la distance).

En ce qui concerne la réverbérance, on observe à l'aide de la figure 5.13 et le tableau correspondant, qu'un changement de 0.1 s sur l'indice  $E_{dt}$  - le seuil de discrimination auditive en terme de réverbérance (0.1 s  $\equiv$  10 %  $E_{dt}$ ) - correspond à une différence d'environ 0.19 'unité perceptive' (pente de la tendance linéaire : 0.19 'u.p.' / 0.1 s).

L'influence sur la réverbérance apportée par les vues  $V2$  et  $V3$  est de l'ordre de 0.1 'u.p.' (tableau 5.13 c ;  $3.75 - 3.65 = 0.1$  'u.p.'). Il ne s'agit donc pas d'une différence perceptible, ni d'ailleurs d'une influence révélée significative par les comparaisons Post-Hoc. Par contre, si l'on considère la différence apportée à la perception de la réverbérance par les vues  $V1$  et  $V4$ , on observe qu'elle est de l'ordre de 0.17 'u.p.' (tableau 5.13 c ;  $3.79 - 3.62 = 0.17$  'u.p.'), ce qui est juste perceptible et aussi révélé significatif.

En ce qui concerne l'enveloppement, la différence sur les réponses est significative entre les conditions visuelles  $V2$  et  $V3$ , mais comme il est perçu de manière non régulière, aucune tentative pour évaluer objectivement l'effet visuel n'a été faite.

## 5.2.4 Discussion

### 5.2.4.1 La perception des stimuli sonores

La manière dont les stimuli sonores sont perçus est globalement similaire dans les deux parties de l'expérience : la partie unimodale et la partie bimodale ; en effet, l'analyse en composantes principales regroupe dans ces deux cas les mêmes facteurs perceptifs.

Une seule différence semble émerger : la variance des réponses est plus répartie pour la partie bimodale que pour la partie unimodale (90 % de variance représentée par les trois premières composantes dans la partie bimodale, contre 97 % dans la partie unimodale). Cet effet est révélateur d'une faible influence des images sur les réponses ; d'autant plus faible que l'augmentation des stimuli (40 stimuli dans la partie bimodale, contre 10 dans la partie unimodale) peut aussi être liée à une répartition de la variance, ce qui signifierait qu'il s'agirait d'un artefact de l'analyse.

---

<sup>1</sup> A propos de la condition visuelle obtenue grâce à la vue  $V4$ , il faut pourtant dire que, comme elle n'est pas prise sur le même axe de la salle que les autres, ni de la même hauteur, plusieurs paramètres visuels ne sont pas contrôlés, et sa comparaison directe avec les autres est assez délicate. Ce fait pourrait d'ailleurs expliquer l'absence d'effet sur la perception de la distance entre les conditions  $V3$  et  $V4$ .

Deux dimensions indépendantes de perception semblent être révélées par l'analyse (expliquant à elles deux 92 % et 85 % de la variance respectivement pour la partie unimodale et pour la partie bimodale) : l'une des composantes est liée principalement à la précision et à la force sonore (donc à la source sonore), et l'autre composante regroupe surtout les trois sensations liées à la salle : l'enveloppement, la réverbérance et la vivacité. Pour la partie unimodale du test, la force sonore est également représentée par la deuxième composante principale – son indépendance avec l'enveloppement et la vivacité n'est donc pas assurée – tandis qu'elle est indépendante pour la partie bimodale de l'expérience.

Les corrélations de la force sonore ne sont pas surprenantes. En effet, plus on est proche, plus le son est fort et plus il est précis. De plus, la force sonore semble contribuer souvent à la sensation d'enveloppement [Barron, 1988].

La distance auditive est principalement représentée par la première composante, elle semble être une mesure inverse de la force sonore et de la précision.

La réverbérance est liée aux deux premières composantes ; elle est aussi liée à la troisième composante, mais l'effet ne semble pas être stable (ce n'est pas le cas pour la partie unimodale du test par exemple). De plus, la réverbérance n'a pas été perçue sur la base du critère spécialement modifié à ce propos : le temps de réverbération  $Rt(s)$ , mais plutôt sur l'Edt(s), l'indice décrivant la réverbération courante.

La vivacité est plutôt bien perçue dans les deux parties de l'expérience : en effet, le stimulus A5 (V) est jugé dans les deux cas comme le moins vivant. Elle est très corrélée à l'enveloppement, mais aussi dans une moindre mesure, à la réverbérance et à la force sonore (dans la partie bimodale de l'expérience, le son jugé le plus vivant était le stimulus A6 (N<sup>+</sup>). En revanche, les différences perceptives de la vivacité entre les stimuli sonores n'étant pas toujours significatives, l'existence de différences entre les parties unimodale et bimodale de l'expérience, mais aussi la dépendance de la vivacité vis-à-vis de l'enveloppement (jugé, comme nous l'avons vu, de manière ambiguë), ne nous permettent pas d'avoir une grande confiance quant à la stabilité de ces observations. De plus, il est possible que le signal sonore d'origine étant très riche en hautes fréquences, l'aspect de la vivacité ne manquait pas, et que par conséquent elle e soit pas apparue comme une dimension indépendante.

Les résultats sur l'enveloppement sont plus intéressants : ils montrent que les sujets jugent l'enveloppement sur la base d'attributs perceptifs différents. Deux sous-groupes de sujets se distinguent (de 15 et 17 personnes) : le premier groupe considère l'enveloppement comme un attribut lié à la force sonore, tandis que le deuxième comme un attribut lié à la réverbérance et donc l'effet de salle (coefficients de corrélation  $r = 0.68$  et  $r = 0.63$  respectivement). Les corrélations pour les autres facteurs vont dans le même sens, c'est-à-dire que pour le premier groupe de sujets l'enveloppement est très corrélé à la précision et inversement corrélé à la distance, tandis que pour le deuxième groupe, il est inversement corrélé à la force sonore et à la précision et positivement corrélé à la distance auditive (voir tableaux 5.7 et 5.8).

Ces résultats semblent cohérents avec ceux obtenus par Barron sur l'enveloppement : il a également trouvé deux groupes de sujets, le premier a associé l'enveloppement à la réverbérance, et le second à l'intimité ; il a suggéré la force sonore comme étant un aspect commun entre l'enveloppement et l'intimité [Barron, 1988].

L'ambiguïté avec laquelle est jugé l'enveloppement, nous mène à considérer le problème lié à la sémantique (cf. § 2.2.6), c'est à dire la signification du mot 'enveloppement' n'étant pas la même pour tous les sujets, le même mot finit par décrire au moins deux impressions différentes de leur part : l'une est plutôt liée à la force sonore et à la précision, et donc à la source sonore – pour ces personnes, l'enveloppement signifie être 'dans le son' – et l'autre est plutôt liée à la réverbérance et à la vivacité, et donc à la salle, pour ces personnes l'enveloppement signifie être 'dans la salle'.

En ce qui concerne l'impression générale, nous observons que celle-ci est dominée par la force sonore, un phénomène observé aussi par d'autres études optant pour des expériences en laboratoire (travaux de Göttingen, dans [Cremer et Müller, 1982]). La conséquence pour l'enveloppement est qu'il est considéré comme une qualité par le groupe qui l'associe à la source, et comme neutre par le groupe qui l'associe à la salle.

Pourquoi l'espace perceptif obtenu ici n'a-t-il que deux dimensions, tandis qu'en règle générale au moins trois dimensions sont observables ? Nous pouvons faire les considérations suivantes :

Nos expériences ont porté sur la qualité acoustique d'une seule salle (le Théâtre des Champs Elysées simulé auditivement et visuellement) tandis que les expériences de la plupart des études du domaine (comme celles de Hawkes et Douglas [1971], Wilkens [Cremer et Müller, 1982] et Barron [1988], cf. § 2.2.2) ont porté sur un ensemble de salles d'écoute. Ces salles ayant des qualités acoustiques différentes, elles ont permis d'observer davantage de dimensions perceptives que celles qui sont habituellement présentes dans une même salle.

Par ailleurs, la troisième dimension est en général liée à la couleur du son : or, cette dernière est très homogène dans une même salle, alors qu'elle varie énormément d'une salle à l'autre<sup>1</sup>.

Et puis, il est évident qu'il n'y avait pas toujours une vraie dépendance des facteurs perceptifs dans la conception des stimuli. Ainsi, si par exemple la précision est souvent corrélée à la distance (les deux attributs étant basés sur le rapport son direct / son tardif), il est possible qu'elle soit indépendante de la force sonore. En revanche, le corpus de dix sons présenté dans nos expériences ne variait pas spécifiquement sur la précision, ce qui n'était pas le cas pour le niveau sonore global, cela a pu entraîner la très forte dépendance entre les deux attributs et l'apparition d'une seule dimension perceptive au lieu de deux.

De plus, le niveau d'expertise des sujets peut influencer le nombre des dimensions observées : plus nous sommes habitués à une sensation (à une situation, un produit, etc.), plus riches sont nos références ou repères perceptifs, et plus nombreuses deviennent les affinités que nous pouvons y discerner. Barron, par exemple, a choisi pour ses expériences des sujets experts dans le but d'obtenir des réponses indépendantes pour chacune des différentes échelles de son questionnaire [Barron, 1988]. Les sujets qui ont participé aux expériences présentées ici étaient tous de jeunes étudiants sans expérience acquise dans de vraies salles d'écoute : la plupart d'entre eux n'avait jamais assisté à une représentation musicale de ce type. Il est donc fort possible que le nombre peu élevé des dimensions perceptives observées soit lié à ce manque d'expérience.

Par ailleurs, le manque d'expérience n'étant pas nécessairement lié à un manque d'opportunités, il peut être tout simplement lié à un manque d'intérêt. Le problème culturel se pose alors, le type de représentation musicale utilisé dans ces tests (opéra classique) n'étant pas le type d'écoute que les sujets aiment en général, il ne sont pas forcément sensibilisés à ses différentes qualités acoustiques. Et puis, ils peuvent très bien avoir de différentes modalités d'écoute (du type 'écoute du son' ou 'écoute de la salle').

Enfin, un problème complémentaire lié à l'expérience d'écoute, mais aussi à la sémantique, et qui a pu influencer le nombre de dimensions perceptives, est que trois des termes du questionnaire : la vivacité, l'enveloppement et la réverbérance, soit n'avaient aucune signification, soit avaient une signification assez vague pour les sujets. Ils semblent être en effet tous regroupés dans une impression générale liée à l'effet de salle, mais ne se distinguent pas vraiment entre eux. Il est évident que cette ambiguïté a pu entraîner des dépendances entre les différents attributs (la courte session de démonstration des sons qui

---

<sup>1</sup> Barron, communication personnelle.

a précédé les expériences avait justement pour but de faire comprendre aux sujets ce que ces trois termes signifiaient ; mais il ne s'agissait que d'un apprentissage court et forcé, son efficacité est donc discutable).

#### 5.2.4.2 L'influence des images

La méthode a priori la plus immédiate pour étudier l'influence visuelle sur la perception auditive serait de comparer les résultats entre la partie unimodale et la partie bimodale de l'expérience. Si une comparaison générale des plans factoriels peut s'avérer utile pour vérifier que la structure de l'espace perceptif ne change pas fondamentalement sous les deux conditions expérimentales (unimodale et bimodale), la comparaison directe entre stimuli unimodaux et stimuli bimodaux (par exemple entre les réponses pour A1 et les réponses pour A1V1, A1V2, etc.) n'a pas vraiment de sens. Ceci parce que ces deux parties de l'expérience ont pu créer des impressions qualitativement différentes chez les sujets. Ces derniers peuvent leur attribuer des significations différentes et – vu que la manière dont une information est représentée par la mémoire influence son traitement [Glass et Holyoak, 1986] – ces significations peuvent influencer la manière dont les sujets répondent. Il serait donc plus prudent de comparer entre elles des impressions auditives liées à des conditions audiovisuelles similaires dont le 'contenu' cognitif est dans la mesure du possible contrôlé (par exemple les réponses pour A1V1 avec les réponses pour A1V2, A1V3, etc.).

Aussi, la partie unimodale de l'expérience a été présentée au début de chaque séance de tests et a donc précédé la partie bimodale pour tous les sujets. Sachant que, surtout au début des tests, les réponses peuvent évoluer avec le temps, la comparaison des résultats entre les parties unimodale et bimodale pourrait révéler des 'effets' dus justement à cette évolution et non pas aux conditions visuelles. La partie unimodale de l'expérience consiste donc en une étape d'entraînement nécessaire pour réduire la variation des réponses au début des tests, mais elle n'est pas utilisée pour étudier l'influence visuelle.

L'influence des conditions visuelles sur la perception auditive pour la partie bimodale de l'expérience a été étudiée à l'aide de trois approches différentes :

La première était la comparaison du premier plan factoriel des quatre analyses prenant en compte les réponses obtenues séparément pour les conditions visuelles V1, V2, V3 et V4 (cf. § 5.2.3.3). Mis à part une rotation du vecteur 'compatibilité' observée entre ces quatre plans (effet attendu puisque la question sur la 'compatibilité' concerne également la stimulation visuelle), on observe une rotation du vecteur 'enveloppement'. Les corrélations de ce dernier augmentent avec la distance visuelle et semblent s'affirmer pour la condition visuelle V4. Pour tenter d'expliquer cet effet, nous pouvons distinguer les conditions visuelles en deux catégories : une 1<sup>re</sup> catégorie où l'on ne voit pas la salle (vues V1 et V2), et la 2<sup>ème</sup> catégorie où l'on voit la salle (vues V3 et V4). Dans les deux cas, l'enveloppement semble être piloté surtout par la force sonore et inversement par la distance. Le fait que la salle apparaisse visuellement entre les deux catégories, entraîne une augmentation de l'enveloppement. C'est comme si le terme 'enveloppement' prenait plus de sens sous des conditions visuelles lointaines et excentrées, là où l'on aperçoit les cotés de la salle. Ces observations n'ont qu'une valeur indicative, elles ne donnent pas lieu à l'établissement d'une tendance générale quelconque.

Ensuite, un facteur représentant la 'distance visuelle' a été introduit dans l'analyse de l'espace auditif, ce facteur fut créé à l'aide des estimations objectives de la distance visuelle. Les corrélations de ce facteur visuel avec les facteurs auditifs ne sont en aucun cas significatives. Malgré le fait que les corrélations pour la distance auditive et pour l'enveloppement soient un peu plus élevées ( $r = 0.25$  et  $r = -0.20$  respectivement) que pour les autres attributs perceptifs, elles ne sont en aucun cas significatives et leurs valeurs sont uniquement indicatives.

Enfin, des analyses de variance ont été effectuées séparément pour chacun des facteurs perceptifs du questionnaire. Les résultats ont montré des effets significatifs dus aux stimuli sonores pour tous les attributs, mais aussi des effets significatifs dus aux conditions visuelles pour les attributs de la réverbérance, la distance auditive, l'enveloppement, l'impression générale et la compatibilité. Des tests 'post-hoc' ont révélé des différences significatives entre les conditions visuelles, uniquement pour la distance auditive, l'enveloppement, la réverbérance et l'impression générale. Des interactions significatives entre les stimuli auditifs et les conditions visuelles montrent que l'effet visuel en général n'est pas simple : il dépend largement du stimulus sonore impliqué.

L'effet visuel sur l'évaluation de la distance auditive est bien plus marquant que pour les autres attributs. Des tests 'post-hoc' ont révélé des différences significatives entre les quatre conditions visuelles (sauf entre les conditions V3 et V4) et l'interaction entre les conditions visuelles et les stimuli sonores, quoique significative, est faible. L'influence visuelle correspond à une variation de l'ordre de 1 à 2 dB sur l'indice objectif Es utilisé pour piloter l'impression de l'éloignement auditif (pour des rapport de distances visuelles entre 1.24 et 4). En revanche, la distance visuelle n'est pas un bon estimateur de l'influence visuelle sur l'audition, parce qu'elle ne permet pas de prédire la différence induite sur la distance auditive. On peut observer l'influence des conditions visuelles et même la faible interaction (surtout entre les conditions visuelles V3 et V4) sur la figure 5.14.

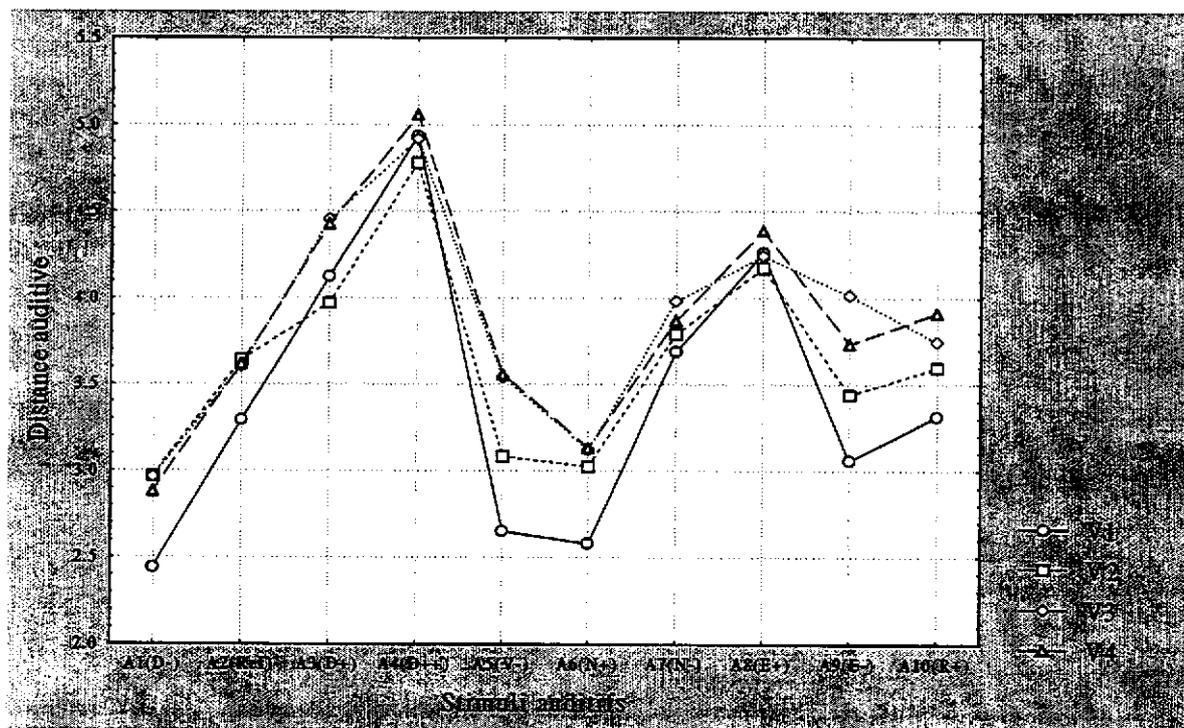


Figure 5.14 : Distance apparente auditive de chacun des stimuli auditifs en fonction des conditions visuelles V1, V2, V3 et V4.

L'influence visuelle sur l'enveloppement est moins prononcée mais significative. Elle se manifeste surtout entre les conditions visuelles V2 et V3 et pour la majorité des stimuli sonores (tous hormis pour les stimuli A7, A8 et, dans une moindre mesure, pour le stimulus A6) (figure 5.15).

Cet effet pourrait être considéré comme une différence entre les deux « catégories » des vues (V1 et V2 où l'on ne voit pas les côtés de la salle, V3 et V4 où on les voit) : nous avons vu que

l'enveloppement augmente entre ces catégories. Puis, dans chacune des catégories, on observe que c'est la condition visuelle la plus proche qui favorise davantage l'enveloppement. Cette différence pour les vues V1 et V2 pourrait s'expliquer par le fait qu'être proche de la scène fait que les sujets se sentent plus impliqués visuellement, et ils traduisent cette sensation en terme d'enveloppement. En ce qui concerne la différence observée entre les conditions V3 et V4, elle doit être due au fait que la vue V4 étant excentrée, elle donne une impression d'enveloppement moindre.

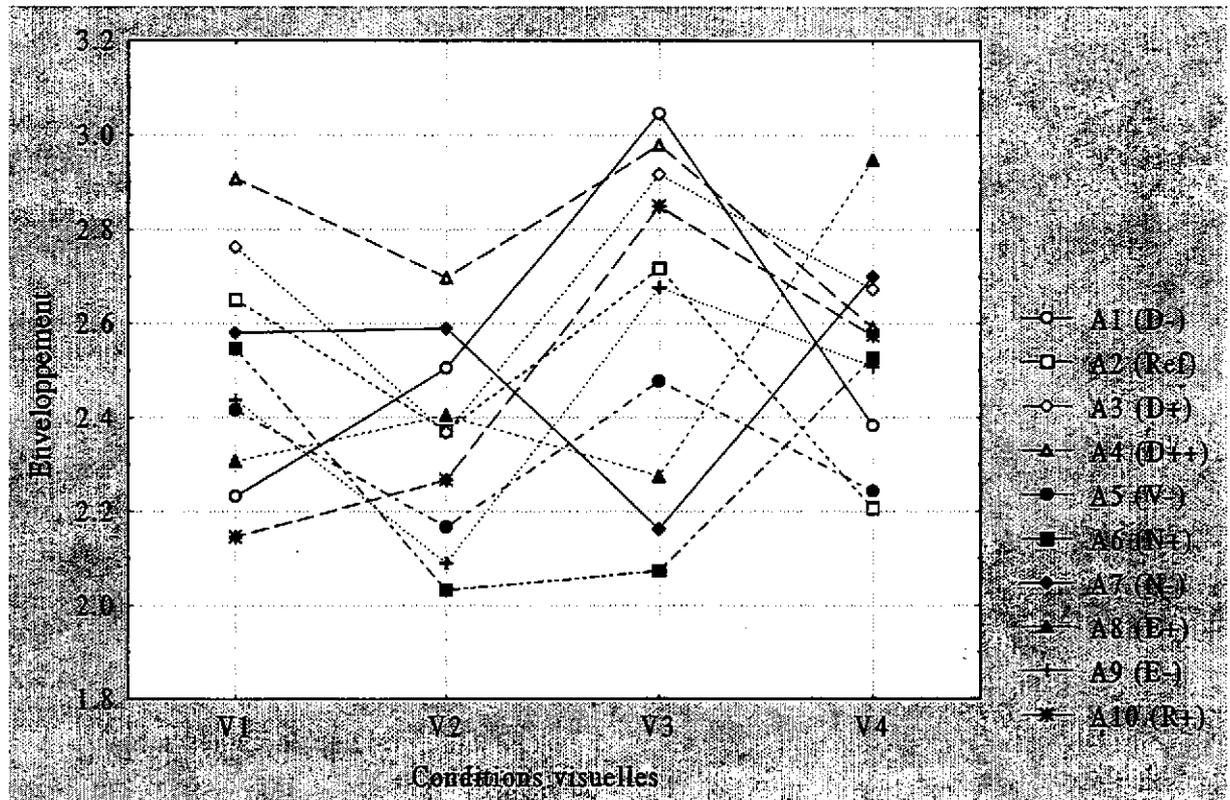


Figure 5.15 : Enveloppement des stimuli auditifs (A1-A10) sous chacune des quatre conditions visuelles (V1, V2, V3 et V4).

Les résultats sur la distance auditive et l'enveloppement vérifient les hypothèses formulées au début de ce travail (cf. § 2.5) sur la nature des facteurs auditifs qui risquent d'être les plus influencés par la perception visuelle, c'est à dire les facteurs décrivant la dimension spatiale du son. De plus, ils justifient le choix de la distance auditive comme l'unique attribut perceptif étudié lors de la première partie de ce travail : on constate à ce propos que la simulation visuelle de la distance est plus facile que la simulation visuelle de l'enveloppement – tant du point de vue de sa conception que du point de vue pratique – ce fait a pu être à l'origine de l'influence visuelle bien plus marquante sur la distance auditive.

L'influence de la perception visuelle sur l'impression générale auditive révèle que sous la condition visuelle V2 les stimuli sonores ne sont pas spécialement appréciés. Les spectateurs-auditeurs semblent alors préférer être, soit proches de la scène, là où ils se sentent impliqués visuellement (même si du fait qu'il n'y avait pas de mouvement dans la simulation visuelle, il n'y avait pas non plus de détail 'changeant' à percevoir), soit relativement loin de la scène, là où ils peuvent percevoir les côtés de la salle et apprécier son esthétique. Il est intéressant également de remarquer que l'influence visuelle sur

l'impression générale suit à peu près le même comportement que celle sur l'enveloppement (cf. figure 5.11).

Pour conclure sur l'influence des images dans cette expérience, on observe que ce n'est que l'influence de la distance visuelle sur la distance auditive qui se manifeste de manière certaine. Pour les autres facteurs du questionnaire, comme l'enveloppement ou l'impression générale, l'effet visuel observé ne peut être considéré qu'indicatif. Des expériences supplémentaires devraient être menées, dont des expériences unimodales préalables aux tests bimodaux, afin de choisir les conditions visuelles et les stimuli auditifs les plus adéquats.

En ce qui concerne la méthode employée, l'approche des questionnaires habituellement utilisée pour étudier la qualité acoustique des salles se prête peu pour révéler les effets dus aux conditions visuelles (au moins dans le cas des expériences en laboratoire où, comme ici, la simulation des conditions visuelles est relativement pauvre). Ni l'étude séparée des plans factoriels obtenus pour chacune des conditions visuelles, ni l'étude des corrélations avec le vecteur visuel, n'ont donné des résultats significatifs (alors que l'analyse de variance en a donné). Il semble donc peu surprenant que les effets des conditions visuelles de salles de spectacle, apparaissent rarement dans les résultats des études classiques adoptant des analyses multidimensionnelles.

Il faut néanmoins reconnaître que, même sous ces conditions expérimentales peu favorables pour l'apparition de l'influence visuelle, l'analyse habituelle (étude des plans factoriels, étude des corrélations) a permis de fournir des indications valables confortées à l'aide de l'analyse de variance (par exemple, pour tous les attributs perceptifs pour lesquels les corrélations avec le vecteur visuel étaient différentes de zéro, l'ANOVA a révélé des effets significatifs des conditions visuelles).

Aussi, dans une tentative d'évaluation quantitative de ces influences, les effets significatifs relevés par les comparaisons Post-Hoc (Tukey HSD) correspondent aux différences qui sont de l'ordre du seuil de la discrimination auditive (1 dB et 0.1 s respectivement pour les indices Es et Edt qui pilotent la distance auditive et la réverbérance).

## 5.2.5 Conclusion

Une méthodologie empruntée aux travaux classiques de la qualité acoustique des salles a été utilisée pour cette deuxième partie des expériences. Il s'agit de la méthode des questionnaires portant sur un ensemble de facteurs perceptifs couramment utilisés pour décrire la qualité acoustique des salles : la *puissance sonore*, la *réverbération*, la *précision d'écoute*, la *vivacité*, l'*enveloppement* et l'*impression générale*.

La méthode a été adaptée de manière à incorporer des indices visuels et d'étudier ainsi l'influence des informations visuelles sur la perception de la qualité acoustique. Une expérience principale a ainsi été organisée, composée d'une partie unimodale (sans stimulation visuelle) et d'une partie bimodale (avec la projection de quatre images représentant la scène du théâtre simulé).

L'étude des plans perceptifs d'un point de vue uniquement auditif, tant pour la partie unimodale que pour la partie bimodale de l'expérience, ne révèle pas de différences marquantes quant à la structure générale de l'espace perceptif. Dans les deux cas, deux dimensions perceptives principales semblent se manifester expliquant entre 92 % et 85 % de la variance : la première est liée principalement à la perception de la source sonore (réunissant les attributs de force sonore et de précision), et la seconde est principalement liée à la perception de la salle (réverbérance, enveloppement et vivacité).

Un résultat intéressant a été révélé à propos de l'enveloppement. Les sujets semblent se partager en deux groupes par rapport à la signification qu'ils donnent à cet attribut auditif : un premier groupe associe l'enveloppement à la source sonore (force sonore) et le second groupe l'associe à l'effet de salle (réverbérance).

La force sonore domine les jugements sur l'impression générale, l'enveloppement est ainsi considéré être une qualité par le premier groupe de sujets, et être plutôt neutre par le deuxième groupe.

En ce qui concerne l'influence des images, des effets existent pour la distance auditive, l'enveloppement, l'impression générale et la réverbérance. Ces effets sont faibles et complexes (interactions entre stimuli auditifs et conditions visuelles significatives). En règle générale, on observe qu'un son est perçu plus lointain sous des conditions visuelles lointaines et plus proche sous des conditions visuelles proches. Une vue de la scène qui permet aux spectateurs de percevoir une bonne partie des côtés de la salle, ou une place de préférence au milieu de la salle, semble favoriser la perception de l'enveloppement sonore (vues V3 et V4). Enfin, l'impression générale auditive est généralement inversement proportionnelle à la distance et proportionnelle à l'enveloppement : si l'on considère les deux catégories des vues (V1 et V2 où l'on ne voit pas la salle, et V3 et V4 où on la voit), on observe qu'au sein de chaque groupe, la vue préférée est la vue la plus proche (V1 par rapport à V2, et V3 par rapport à V4), tandis que sur l'ensemble des catégories, la vue préférée est celle qui favorise le plus l'enveloppement (vue V3).

Ces résultats ont permis de valider les hypothèses avancées par rapport à l'influence visuelle sur les attributs spatiaux du son, ils ont permis également de valider le choix de la distance auditive pour les expériences de la première partie de cette étude. La distance auditive est en fait le seul facteur perceptif pour lequel on peut associer une influence certaine de la perception visuelle. Des résultats sur les autres attributs sont également présents, mais doivent être considérés indicatifs. Des expériences plus poussées, avec un soin particulier sur les choix des stimuli visuels et auditifs, devraient être menées afin d'établir des influences plus pertinentes.

Enfin, nous considérons que la méthode habituellement adoptée par les études sur la qualité acoustique des salles se prête peu pour l'étude de l'influence visuelle. Du moins sous des conditions expérimentales similaires à celles de notre travail (expériences en laboratoire avec simulation visuelle de qualité moyenne). Les conditions visuelles semblent engendrer quelques modifications indicatives sur les plans factoriels, mais ces influences peuvent facilement être attribuées aux incertitudes courantes dans les analyses factorielles et peuvent ainsi être associées uniquement à la perception auditive. Il a fallu en effet recourir à la technique de l'analyse de la variance pour observer les effets visuels significatifs mentionnés ci-dessus.



## **6 DISCUSSION GENERALE**

Une discussion sur les résultats les plus importants de cette étude est présentée dans cette partie. Elle est organisée en fonction des objectifs principaux de notre travail définis à la fin de l'étude bibliographique (cf. § 2.5) et présentés au début de chaque section correspondante.

Une discussion supplémentaire suit, portant sur les améliorations que l'on pourrait apporter aux méthodes employées pour étudier l'influence visuelle sur la qualité acoustique des salles, ainsi qu'une discussion relative aux applications dans d'autres domaines qui pourraient bénéficier de nos résultats sur la perception multimodale.

## 6.1 INFLUENCE DES INFORMATIONS VISUELLES SUR L'APPRECIATION DE LA QUALITE ACOUSTIQUE DES SALLES.

Notre objectif initial est d'étudier l'influence des informations visuelles sur l'appréciation de la qualité acoustique des salles. Cette étude devrait constituer une étape nécessaire vers une amélioration des corrélations des facteurs perceptifs avec les indices objectifs, informations qui sont très utiles pour la conception ou la réhabilitation de salles d'écoute.

Ce travail est composé de deux parties distinctes visant à étudier l'influence des informations visuelles sur la qualité acoustique dans des perspectives quelque peu différentes, et adoptant chacune des méthodes propres aux domaines scientifiques correspondants à ces perspectives.

Ainsi, la première partie utilise la méthodologie classique des études portant sur les interactions sensorielles entre vision et audition pour étudier de manière plus approfondie l'influence de la distance visuelle de la scène d'une salle sur la perception de la distance auditive dans cette même salle.

La deuxième partie adopte la méthode traditionnellement utilisée par les travaux sur la qualité acoustique des salles pour étudier le problème de l'influence visuelle dans un sens plus large, mais aussi dans un contexte expérimental propre à des travaux spécifiques du domaine. L'influence des informations visuelles a ainsi été étudiée sur un ensemble de facteurs perceptifs couramment utilisés pour évaluer la qualité acoustique des salles. Il s'agit de la force sonore, la précision, la réverbérance, la vivacité, la distance auditive, l'enveloppement et l'impression générale.

Les résultats de la première partie des expériences ont montré qu'il existe effectivement une influence de la distance visuelle de la scène sur la perception de la distance auditive dans une salle. Des mêmes stimuli sont perçus auditivement plus lointains sous des conditions visuelles lointaines, et plus proches sous des conditions visuelles proches. Cette influence est faible mais significative et robuste, elle apparaît dans toutes les expériences en projection 3-D, indépendamment du signal et de la méthode de diffusion sonore. Elle peut être de l'ordre d'une variation de 1 dB sur l'indice objectif  $E_s$  pour un rapport de distances visuelles de  $V_3/V_1=4$ .

L'influence visuelle pourrait être expliquée en considérant que les sujets, dans leurs jugements auditifs spatiaux, utilisent – en dehors des repères purement auditifs – l'espace qui les entoure comme un cadre de référence. Or, cet espace est construit de manière multimodale, plusieurs modalités sensorielles y participent dont la modalité visuelle qui est en effet considérée comme la modalité dominante dans cette représentation. Enfin, il a aussi été observé que l'influence visuelle dépend des facteurs attentionnels et cognitifs particulièrement présents dans les cas d'évaluations subjectives multimodales, facteurs qui sont discutés dans la partie suivante (cf. § 6.2).

Les résultats de la deuxième partie de nos expériences ont montré une influence visuelle assez complexe, souvent faible, et en tous cas très dépendante de l'attribut auditif en question. Si l'on prend en compte uniquement les résultats des analyses classiques en acoustique des salles (analyse factorielle, étude des corrélations), les effets visuels ne s'avèrent pas significatifs mais fournissent des indications qui sont validées par les analyses de variance employées. Ainsi, la distance auditive, l'enveloppement, l'impression générale et étonnamment la réverbérance, semblent être influencés par les conditions visuelles. Ce n'est pas le cas pour la force sonore, la précision et la vivacité, ce qui semble valider l'hypothèse selon laquelle ce sont surtout les attributs perceptifs spatiaux du son qui sont susceptibles d'être influencés par la perception visuelle.

L'effet visuel pour la distance auditive est le plus important et le moins complexe, c'est à dire qu'il dépend moins des stimuli auditifs (l'interaction entre stimuli auditifs et conditions visuelles est

significative mais relativement faible, cf. tableau 5.12 et figure 5.12). Cet effet peut être équivalent à une variation de l'ordre de 1 à 2 dB sur l'indice objectif  $E_s$  (les variations de 1 dB sont obtenues pour des rapports de distances visuelles  $V_2/V_1 = 3.23$  et  $V_3/V_2 = 1.24$ , et la variation de 2 dB pour un rapport  $V_4/V_1 = 4$ ). Par conséquent, la distance visuelle ne semble pas un indice pertinent pour estimer l'influence visuelle (puisque de différents rapports de distances visuelles peuvent provoquer les mêmes influences sur l'audition).

Les effets pour les autres attributs influencés peuvent être considérés comme indicatifs, mais ils ne permettent pas d'établir une tendance quelconque. De manière générale, on observe que la perception de l'enveloppement semble être favorisée par une vue de la scène qui permet aux spectateurs de percevoir une bonne partie des côtés de la salle (vues  $V_3$  et  $V_4$ ). L'impression générale semble bénéficier de ces emplacements, qui sont souvent des emplacements proches : vis-à-vis de chaque catégorie d'images,  $V_1$  est préférée à  $V_2$ , et  $V_3$  à  $V_4$ . Enfin, la sensation de réverbérance semble également augmenter avec la distance visuelle<sup>1</sup> ; cette influence peut être équivalente à une variation de l'ordre de 0,1 s sur l'indice objectif  $Edt$  qui pilote cet attribut (il s'agit de la différence juste perceptible sur cet indice  $Edt$ ). Pour tous ces attributs, des expériences supplémentaires devront être menées, dont des expériences préliminaires unimodales, préalables aux tests bimodaux, afin de choisir les stimuli auditifs et surtout les conditions visuelles les plus adéquates.

Enfin, il est important de noter à la fin de cette section, que la qualité de la simulation influence l'ampleur de l'effet visuel. Nous considérons donc que, si même sous des conditions de simulation visuelle relativement pauvres (cf. § 6.3), des effets visuels ont pu être observés, la question de leur pertinence en est a priori renforcée. Leur apparition est ainsi susceptible d'être bien plus accentuée sous des conditions de simulation visuelle améliorées, ou dans des conditions réelles de salles de spectacle où le réalisme des conditions auditives et visuelles est évidemment assuré.

---

<sup>1</sup>  $Edt$  et  $E_s$  sont d'ailleurs corrélés objectivement.

## 6.2 EVALUATION SUBJECTIVE MULTIMODALE

L'un des objectifs de ce travail est de permettre une meilleure compréhension du processus de l'évaluation subjective, une évaluation qui implique le plus souvent des informations issues de plusieurs modalités sensorielles. L'influence des facteurs visuels bien sûr, mais aussi celle des facteurs attentionnels et cognitifs particulièrement importants dans la perception multimodale, sont ainsi discutées.

C'est surtout la première partie de nos expériences sur la distance auditive qui, ayant apporté diverses modifications sur la procédure expérimentale, nous a permis d'observer d'intéressants aspects du processus de l'évaluation subjective.

L'influence de la distance visuelle sur la distance auditive observée globalement dans nos expériences, montre que des indices visuels peuvent être pris en compte de manière involontaire dans l'évaluation subjective de la distance auditive, mais aussi dans celle d'autres facteurs auditifs liés principalement aux aspects spatiaux du son.

Le fait que cette influence est observée dans les expériences en projection 3-D et non pas dans les expériences en projection 2-D, indique qu'il s'agit surtout de la perception de profondeur des images 3-D qui provoque cette influence, et non pas des associations mentales liées à la taille d'objets connus par exemple – qui *suggèrent* la distance – ni à des consignes d'orientation vers la modalité auditive, mal interprétées par les sujets (si les sujets avaient compris qu'ils pouvaient utiliser les vues dans leurs jugements de la distance, ils auraient été également influencés par les images en 2-D puisqu'elles donnent aussi l'impression de l'éloignement). Par conséquent, il s'agit d'une influence qui trouve son origine essentiellement dans la stimulation sensorielle physique et non pas dans des informations de nature cognitive.

Toujours au sujet de l'influence visuelle, nous pouvons évoquer quelques réflexions sur les raisons de son apparition : dans la psychologie expérimentale, il est généralement admis que la localisation spatiale dépend du système de référence utilisé. Or, nos perceptions visuelles et auditives sont spatialisées et cet espace perçu constitue un cadre de référence construit ainsi à partir de l'intégration d'informations sensorielles en provenance de ces deux modalités (ainsi que d'autres, comme le toucher), mais aussi de nos représentations cognitives du monde. Il s'agit en effet d'un cadre 'amodal' puisque plus d'une modalité sensorielle contribue à sa formation. La modalité visuelle consiste en la modalité dominante dans la construction de cet espace, de la même manière que la modalité auditive est considérée dominante dans l'analyse de la dimension temporelle. Enfin, il est généralement admis que la modalité visuelle aide à organiser l'espace sonore et à maintenir la mémoire spatiale.

Ainsi, dans nos expériences, et dans le cas des jugements spatiaux, les sujets semblent utiliser les sons mais aussi les images pour créer un tel cadre spatial, qui pris comme référence, peut influencer leurs jugements auditifs<sup>1</sup>. Les sujets ont en effet constaté que les images ne les perturbaient pas dans leur tâche auditive (sauf pour le cas des vues changeantes), et qu'au contraire, elles les aidaient à se situer dans l'espace.

---

<sup>1</sup> Cela ne signifie pas l'espace perceptif auditif ait été modifié considérablement par la présence du cadre visuel : il a d'ailleurs été observé qu'il reste pratiquement identique.

Si l'influence visuelle dans nos expériences peut être considérée comme le résultat d'une stimulation sensorielle, son ampleur dépend non seulement de la distance visuelle des images mais aussi d'une multitude de facteurs cognitifs, ce qui est discuté ci-après.

L'information sensorielle est le plus souvent insuffisante pour identifier les messages qui nous proviennent de l'environnement. Des processus cognitifs interviennent de manière nécessaire dans la structuration et l'interprétation des informations. Ceci est encore plus vrai pour le cas de la perception multimodale, où le 'contenu cognitif' lié aux stimulations sensorielles est encore plus difficile à contrôler, puisque les représentations d'une modalité font constamment et automatiquement appel aux représentations des autres modalités. Ces 'appels' dépendent de la connaissance et de la mémoire relatives à l'apparition précédente et simultanée des stimuli provenant des deux modalités. Autrement dit, il y a beaucoup plus de possibilités d'interprétation liées aux événements multimodaux, ce fait peut justement entraîner davantage d'influences de nature cognitive.

Dans nos expériences, nous avons essayé de contrôler le paramètre cognitif au niveau de la conception des stimuli (cf. § 4.10), ce qui ne nous donne pas la possibilité d'observer des influences de ce genre dans nos résultats. Néanmoins, d'autres paramètres de nature cognitive, introduits au niveau de la procédure, et notamment des facteurs attentionnels, semblent influencer fortement l'ampleur de l'influence visuelle.

Il a été observé que l'influence visuelle des expériences à vues statiques était bien moins forte que celle observée dans les expériences à vues changeantes. On considère que cette différence est principalement due au fait que l'attention des sujets est davantage attirée vers les images changeantes que vers les images statiques. Ce phénomène est connu dans la psychologie cognitive : les êtres humains sont davantage interpellés par des situations changeantes (et davantage encore lorsque ce changement va vers une condition inconnue) que par des situations statiques (cf. § 2.4). Comme notre capacité d'attention est limitée, dans le cas des vues changeantes les sujets tendent à porter moins d'attention sur les stimuli auditifs, ce qui entraîne une plus forte incertitude par rapport aux jugements. Les images sont alors davantage utilisées comme référence, et entraînent des influences visuelles plus grandes.

Et puis, il y a un aspect de l'appréciation subjective qui a été étudié à l'occasion de ce travail et qui, sans concerner directement l'influence visuelle, mérite d'être cité : il s'agit du fait que le fonctionnement de notre système perceptif n'est pas immuable au cours du temps. L'apprentissage et l'exercice peuvent en effet modifier le processus de l'évaluation subjective.

Nous avons ainsi observé une augmentation des réponses sur la distance auditive avec le temps (entre la première et la deuxième présentation des stimuli, c'est à dire au bout des 5 premières minutes de tests environ) et une stabilisation de ces réponses ensuite (jusqu'à la septième présentation des stimuli au moins, cf. expérience principale 2, § 4.5). Cette tendance à sous-évaluer les distances est observée aussi par d'autres travaux sur la perception de la distance auditive [Coleman, 1962], [Mershon et Bowers, 1979], les sujets semblent répondre prudemment dans l'expectative de sons à venir, en utilisant probablement leur propre position comme point de référence (cf. § 2.3.3).

Nous avons également observé une 'polarisation' des réponses vers les extrémités de l'échelle proposée, c'est à dire une tendance à donner des notes plus proches de ces valeurs et à négliger en quelque sorte le milieu de l'échelle. Ce phénomène est observé au début des tests, il disparaît à la deuxième présentation des stimuli (cf. expérience 2, § 4.5). Il pourrait être dû, d'une part au doute initial des sujets dans leur jugement, ils se basent au début par rapport aux stimuli de référence (ici le stimulus le plus proche et le stimulus le plus éloigné), d'autre part, il pourrait être lié au fait que souvent, au début d'une expérience perceptive – dans le sens large du terme – on ne distingue que quelques niveaux d'un attribut perceptif (ici un son 'lointain' et un son 'proche'). Avec l'apprentissage, notre capacité d'évaluation s'affine et l'on peut distinguer davantage de niveaux de cet

attribut. Il s'agit en effet de la théorie sur 'l'affinement croissant de la différenciation perceptive produite par des apprentissages naturels et artificiels' adopté par Gibson [Casalis, 1991].

Pour conclure, comme le fonctionnement de notre système perceptif se base sur la perception de changements entre grandeurs physiques, et non pas sur la perception de grandeurs absolues, notre organisme a besoin de références ou de repères pour effectuer ses jugements. Dans le cas général, ces repères sont intramodaux, c'est à dire issus de la modalité sensorielle en question, ici la modalité auditive. Dans le cas d'expositions multimodales en revanche, ces repères peuvent être de provenance intermodale, ici visuelle, ou même cognitive (connaissances, mémoire, attentes). De plus, des paramètres attentionnels et évolutifs liés au changement des réponses avec le temps, peuvent également influencer de manière significative le processus de l'évaluation perceptive.

### 6.3 DISCUSSION SUR LA METHODOLOGIE – DIFFICULTES DE LA SIMULATION BIMODALE

L'un des buts de ce travail est d'étudier et de mettre en place une méthodologie cohérente qui permette d'étudier les interactions sensorielles, tant dans le cadre de la simulation d'environnements réels que dans le cadre de la simulation d'environnements artificiels.

D'une part, la méthode généralement utilisée par la psychologie expérimentale pour tester les interactions sensorielles a été adoptée, afin de tester l'influence visuelle sur un unique attribut perceptif de manière plus approfondie, celui de la distance auditive. Elle a permis d'observer des effets significatifs de la distance visuelle sur la distance auditive, et de constater à quel point la procédure expérimentale peut influencer l'ampleur de ces influences.

D'autre part, la méthode des questionnaires habituellement adoptée par les études sur la qualité acoustique des salles, a été modifiée de manière à introduire des indices visuels présents dans les salles, et fut utilisée pour tester la perception d'un ensemble d'attributs auditifs descriptifs de cette qualité acoustique. Elle a ainsi permis une étude plus large de l'influence visuelle sur la perception auditive. Aussi, elle a permis de valider le choix de la distance auditive pour la première partie des expériences, et de montrer que c'est surtout les aspects spatiaux du son qui sont les plus susceptibles d'être influencés par la vision. Par contre, la plupart de ces résultats sont indicatifs et ne peuvent pas être considérés comme preuves de l'influence visuelle.

On considère par conséquent, que la méthode adoptée par les études sur la qualité acoustique des salles se prête plutôt pour une étude investigatrice de l'espace perceptif, et que d'autres méthodes, du type des méthodes utilisées pour étudier d'habitude les interactions sensorielles, doivent être adoptées afin de valider les effets visuels de manière isolée, c'est à dire sur chacune des composantes de cet espace.

Quelle que soit la méthode adoptée, son efficacité dépend largement de la qualité de la simulation bimodale. Lors d'une simulation, l'objectif idéal est de donner aux êtres humains l'impression du réalisme. Or, une telle impression n'est pratiquement jamais donnée, la raison principale étant la capacité remarquable chez les êtres humains pour distinguer ce qui appartient au monde réel de ce qui est artificiel. Par conséquent, ce que les sujets jugent – consciemment ou inconsciemment – lors d'une simulation, c'est la plausibilité de la situation simulée, c'est à dire la réponse à la question : *'est ce que cela aurait pu être réel ?'*, et non pas vraiment son réalisme.

Lors d'une simulation unimodale, il est déjà difficile de donner aux êtres humains une telle impression. Par contre, d'un point de vue méthodologique la simulation unimodale semble simple : il suffit de faire tout ce qui est possible pour améliorer la qualité de la condition simulée. Or, pour une simulation bimodale, il ne s'agit pas d'améliorer simplement la simulation sonore et la simulation visuelle séparément. L'efficacité d'une simulation bimodale dépend non seulement de la qualité des simulations considérées séparément, mais aussi de l'équilibre entre ces qualités. Par équilibre, on entend tous les paramètres liés à la compatibilité « physique » entre deux situations : compatibilité spatiale, compatibilité temporelle, cohérence du rythme et du mouvement, équilibre des intensités (cf. § 2.4.3). Si par exemple, une simulation visuelle est très bonne, et la simulation auditive qui l'accompagne est relativement pauvre, il se peut que, comme ces deux représentations du même objet ne correspondent pas, elles ne soient pas associées par le cerveau des sujets, et qu'en conséquence, la situation bimodale résultante ne soit pas considérée plausible, crédible, etc. Dans un tel cas, et si l'on ne peut pas améliorer la qualité de la simulation 'inférieure', il vaudra mieux 'abaisser' la qualité de la simulation 'supérieure'.

Egalement, d'un point de vue cognitif, la simulation bimodale est bien plus difficile à réaliser correctement que la simulation unimodale. Son efficacité dépend du contexte d'une multitude de

paramètres cognitifs, tels que la connaissance, la mémoire, l'entraînement, l'orientation d'attention, les informations préliminaires aux tests données aux sujets, etc. (cf. § 2.4.3). Il s'agit en effet de prendre en compte ces paramètres dans la conception des stimuli, et d'essayer de contrôler ou d'équilibrer leur influence. La qualité de la simulation dépend du degré de l'association cognitive qui s'effectue entre stimuli sonores et visuels : on associe facilement une voix féminine à l'image d'une femme qui parle, moins facilement la même voix à l'image d'un jeune enfant et encore moins facilement à l'image d'un homme. Egalement, on associe plus facilement un son rythmique à une image qui change de la même manière, qu'à une image stable par exemple. Ou encore, on associe mieux deux stimuli hétéromodaux dont on nous a dit qu'ils concernent le même événement, plutôt que deux stimuli dont on ne possède pas d'information sur la provenance. Il est évident que plus les stimuli conçus respectent ou prennent en compte ces paramètres, plus la qualité de la simulation est élevée.

Dans nos expériences, le paramètre lié à la connaissance est plutôt respecté, c'est à dire que la condition audiovisuelle est crédible, mais ne renvoie pas pour autant à une association forte : la vue d'un haut-parleur au milieu d'une scène de théâtre peut effectivement créer l'attente d'un son, même s'il s'agit d'un objet immobile – et ceci est vrai pour des personnes qui ont déjà entendu un haut-parleur émettre du son – mais l'absence de mouvement dans le champ visuel résulte en une association plutôt faible des stimuli auditifs et visuels. La vue d'une chanteuse qui interprète la chanson serait évidemment une bien meilleure – mais plus coûteuse – solution.

En ce qui concerne les informations préliminaires aux tests, par rapport à la provenance des stimuli hétéromodaux, aucun effort n'a évidemment été fait pour convaincre les sujets que le son venait de l'image projetée. L'association à ce niveau de connaissance est très basse, les sujets de nos expériences savent que le son et l'image sont deux événements distincts, ce qui pénalise fortement l'ampleur des interactions.

Nous avons vu que d'un point de vue méthodologique et cognitif, une simulation bimodale est beaucoup plus complexe qu'une simulation unimodale. Ceci est également vrai, considéré d'un point de vue attentionnel<sup>1</sup> lorsque les sujets, même s'ils sont orientés vers une seule modalité sensorielle – comme c'est le cas dans nos expériences – sont 'exposés' à des stimulations provenant de plusieurs modalités. Leur attention peut involontairement être orientée d'une modalité vers l'autre et influencer l'association des stimuli. Il faut donc essayer de respecter un certain équilibre attentionnel, ou bien, dans le cas où cet équilibre n'est pas respecté, de signaler et de tester ces influences séparément (comme par exemple par la réalisation de tests en vues statiques et de tests en vues changeantes dans le présent travail).

Pour conclure, nous constatons que les difficultés de la simulation bimodale sont bien plus nombreuses que celles de la simulation unimodale. Une technique de simulation améliorée est nécessaire (pour assurer par exemple la synchronisation et la coïncidence spatiale des stimuli), ainsi que des connaissances psychologiques poussées, relatives aux associations cognitives des informations hétéromodales.

Nous constatons aussi que la qualité de la simulation de nos expériences est plutôt moyenne, en partie parce que la simulation visuelle était relativement pauvre par rapport à la simulation auditive (pas de temps réel et donc pas de mouvement à l'image, absence de présence humaine sur la scène et dans la salle), mais aussi à cause du manque d'équilibre attentionnel, de la connaissance de la présence de deux événements distincts de la part des sujets, etc. Certaines améliorations qui pourraient être envisagées sont, vis-à-vis de la sensation visuelle, l'introduction du mouvement dans le champ visuel, ou même l'utilisation d'une technique d'immersion dans l'image (petits écrans LCD correspondants chacun à un œil, etc.), et vis-à-vis de la sensation auditive, l'utilisation d'un "suiveur de position" permettant le mouvement de la tête des auditeurs tout en respectant la localisation des sources sonores.

---

<sup>1</sup> L'attention est aussi un des objets étudiés par la psychologie cognitive. Elle est mentionnée ici séparément pour la distinguer de la connaissance.

## 6.4 PERSPECTIVES

Les perspectives de ce travail concernent deux directions principales :

- d'une part le domaine de l'acoustique des salles, directement lié à la présente étude,
- d'autre part, une multitude de domaines, scientifiques ou techniques, où la perception multimodale est importante ; c'est à dire presque tous les domaines où l'évaluation subjective d'un produit ou d'un événement est nécessaire.

En ce qui concerne l'acoustique des salles, et plus particulièrement l'étude de l'influence visuelle sur la perception de la distance, il serait fortement souhaitable d'améliorer la méthode de simulation visuelle en y introduisant le mouvement (par exemple la vidéo d'une chanteuse qui interprète la chanson que les sujets écoutent). L'introduction d'un 'suiveur de position' pour réactualiser la position des sources sonores en temps réel pourrait aussi avoir un effet bénéfique, surtout en ce qui concerne la 'crédibilité' de la situation simulée. En tous cas, il faut toujours respecter l'équilibre des simulations hétéromodales pour éviter l'orientation involontaire de l'attention des sujets, ou pire, que ces derniers dissocient complètement les deux situations.

Pour les autres attributs perceptifs, il serait surtout intéressant de tester ceux liés à la dimension spatiale du son, comme l'enveloppement, la largeur de source ou l'impression spatiale. Egalement, il serait intéressant d'étudier l'influence de la perception visuelle sur l'attribut perceptif très important qu'est l'intimité, bien que non directement lié à la dimension spatiale du son, il se révèle par certaines études lié à la distance auditive, l'enveloppement, etc. Enfin, parmi les études supplémentaires nécessaires sur la qualité acoustique, doit figurer aussi une étude sur la dimension hédonique : l'influence de la perception visuelle sur l'impression générale auditive mériterait également d'être étudiée.

Pour tous ces attributs, une méthode similaire à celle employée pour la distance auditive pourrait être utilisée, en prenant un soin particulier dans le choix des stimuli. Des expériences préliminaires unimodales, préalables aux tests bimodaux, doivent ainsi être menées afin de choisir les stimuli auditifs et visuels les plus adéquates. De plus, des recherches sémantiques devraient éventuellement être menées afin de déterminer ce que les sujets signifient dans l'emploi des différents termes (par exemple, deux groupes de sujets ont été observés dans cette étude par rapport à l'interprétation qu'ils donnent à l'enveloppement), mais aussi afin d'accéder aux représentations qu'ils se font par rapport aux conditions expérimentales [Maffiolo, 1999].

La deuxième direction vers laquelle s'ouvre ce travail est bien plus vaste : elle concerne, d'une part l'évaluation subjective unimodale en général, et d'autre part, de nombreuses applications de la perception multimodale.

Les interactions sensorielles peuvent ainsi être considérées comme l'un des multiples problèmes de l'évaluation subjective. Une évaluation sur une qualité unimodale – ici la modalité auditive – effectuée dans des conditions réelles d'exposition, implique dans la plupart des cas, des informations provenant des autres modalités sensorielles. Du fait du traitement des informations par notre système perceptif, de manière non pas parallèle, mais interactive et complexe, ces informations interviennent forcément dans l'appréciation subjective. Presque tous les domaines où l'évaluation subjective d'un produit ou d'un événement est effectuée dans des conditions réelles, sont ainsi concernés : l'appréciation de la qualité sonore d'une chaîne hi-fi, d'un système audiovisuel, d'une voiture ou même d'une ville, peut être influencée par son aspect visuel, même si cet aspect n'est pas directement étudié.

Même dans des conditions d'expériences spécifiquement organisées (en laboratoire ou in situ), souvent caractérisées par le souci d'isoler la modalité sensorielle en question en contrôlant les autres modalités, le problème des interactions persiste. Si, dans une même expérience ou dans une série d'expériences, les conditions visuelles peuvent être maintenues stables, il est souvent délicat de comparer les résultats avec ceux d'études différentes. Il semble ainsi risqué de comparer des résultats, par exemple sur un attribut spatial du son, si ceux-ci sont obtenus d'un côté dans une grande salle d'écoute, et de l'autre dans le local restreint d'une chambre anéchoïque.

Les interactions sensorielles prennent toute leur importance dans le cas des applications liées à la perception multimodale [Hollier et al., 1999] : ainsi, les systèmes de communication et les applications multimédias, comme les téléconférences, la télévision, les ordinateurs, les jeux électroniques, les programmes audiovisuels, sont des applications dont l'efficacité et l'impact reposent en grande partie sur la collaboration des sens.

Les techniques d'adaptation son - image, fondamentales pour ces applications, mériteraient d'être étudiées dans une perspective moderne, permettant d'améliorer leur qualité 'audiovisuelle' – et non pas auditive et visuelle séparément – dans le cadre des études subjectives visant à minimiser les efforts et les coûts, tout en assurant un couplage des sens réussi.

L'interactivité ouvre aussi de nouveaux horizons. Passer de la place de 'spectateur' à celle 'd'acteur', introduit une multitude de nouveaux paramètres dont les interactions avec les sens doivent être comprises et prises en compte, comme la fonctionnalité, l'ergonomie, la lisibilité des scènes virtuelles, etc.

Le son s'y trouve avec l'image dans une position privilégiée ; et le son a d'ailleurs l'avantage de pouvoir donner une impression davantage 'spatialisée' par rapport à la vision, c'est à dire de créer une impression plus 'enveloppante', plus 'immersive' (du fait notamment que les basses fréquences sont plus omnidirectionnelles).

Enfin, une direction très importante, qui tire son intérêt du problème de la simulation multimodale, est celle qui consiste à étudier des méthodes expérimentales qui permettraient de tester la perception multimodale sans décorréler les sens. Malgré le fait que ceci présente actuellement la seule solution pour étudier les interactions sensorielles, il est en effet paradoxal de découpler les sens dans un premier temps, afin d'essayer de les coupler à nouveau dans le but d'étudier leurs relations. Or, le mal est déjà fait : les sens se trouvent souvent décorrélés dans la tête des sujets pendant une bonne partie des expériences !

## 7 CONCLUSION GENERALE

L'objectif principal de ce travail concerne l'influence des informations visuelles sur l'appréciation de la qualité acoustique des salles. Il a été réalisé visant une meilleure compréhension des paramètres qui pilotent cette qualité et une amélioration des corrélations qui relient les sensations auditives et les indices objectifs utilisés pour son appréciation. Deux objectifs supplémentaires sont, d'une part la mise en place d'une méthode pertinente permettant d'étudier les interactions sensorielles entre vision et audition, et d'autre part, l'étude de l'influence des informations intermodales sur l'évaluation subjective en général.

Une série d'expériences a été réalisée afin d'étudier ces aspects perceptifs relatifs aux interactions sensorielles entre vision et audition. Deux différentes approches méthodologiques ont été adoptées, correspondant chacune à l'un des deux domaines scientifiques directement concernés. La première est empruntée au domaine de la psychologie, cette approche traite les interactions sensorielles et permet d'étudier uniquement l'influence de la distance visuelle de la scène d'une salle sur la perception de la distance auditive dans cette même salle. La deuxième approche est inspirée de la méthodologie traditionnellement employée dans les études du domaine de la qualité acoustique des salles, et concerne l'influence visuelle sur un ensemble de facteurs perceptifs descriptifs de cette qualité : la force sonore, la précision, la réverbérance, la vivacité, la distance auditive, l'enveloppement et l'impression générale.

Les résultats ont mis en évidence une influence de la distance visuelle de la scène sur la perception de la distance auditive dans une salle : Les mêmes stimuli auditifs sont perçus plus lointains sous des conditions visuelles lointaines et plus proches sous des conditions visuelles proches. Cette influence est faible mais robuste, elle peut être équivalente perceptivement à un changement de 1 à 2 dB de l'indice objectif  $E_s$  – utilisé pour piloter l'impression de l'éloignement dans nos expériences – pour un quadruplement de la distance visuelle. L'ampleur de cette influence dépend largement de la procédure expérimentale employée (par exemple, de paramètres attentionnels et d'autres paramètres cognitifs introduits par cette procédure) ainsi que de la qualité de la simulation bimodale. Par ailleurs, les présentes conditions de simulation visuelle étant plutôt pauvres, ce phénomène est susceptible d'être renforcé sous des conditions de simulation visuelle améliorée ou dans des conditions réelles d'écoute.

En ce qui concerne les autres attributs perceptifs testés, pour certains d'entre eux, les résultats ont montré une influence visuelle significative mais complexe et faible. Ces attributs sont l'enveloppement, l'impression générale et la réverbérance (l'influence est de l'ordre de 0.1 s sur l'indice  $E_{dt}$  pour un quadruplement de la distance, ce qui correspond à la variation juste perceptible de la sensation de la réverbérance). Les effets peuvent être considérés uniquement indicatifs et ne permettent pas d'établir une tendance quelconque. Ils incitent pourtant à poursuivre les recherches avec des expériences spécifiques sur chaque attribut, telles qu'elles ont été réalisées pour la distance auditive.

Parmi les deux approches méthodologiques employées, celle empruntée à la psychologie expérimentale s'avère mieux adaptée pour tester les interactions sensorielles ; elle permet en effet de révéler des influences même faibles, mais robustes. En revanche, la méthode habituellement utilisée dans le domaine de l'acoustique des salles – adaptée ici afin d'incorporer des indices visuels – est moins adaptée, tout du moins dans le cas des expériences en laboratoire : elle n'apporte aucune modification notable sur l'espace perceptif tel qu'il est observé sous des conditions d'exposition uniquement auditives ; malgré le fait qu'en ce qui concerne la distance auditive, l'influence visuelle est certaine. Il faut pourtant reconnaître que l'utilisation de cette méthode, du fait d'avoir étudié

l'influence visuelle sur plusieurs attributs perceptifs à la fois, a permis leur comparaison directe ainsi que la validation de l'hypothèse avancée au début de ce travail, selon laquelle ce sont surtout les attributs perceptifs spatiaux du son qui sont plus susceptibles d'être influencés par la perception visuelle.

Par ailleurs, les principales conclusions que cette étude nous amène à tirer pour l'évaluation subjective, concernent les différents repères qui sont nécessaires aux processus de comparaisons que notre système est amené à effectuer.

Ainsi, comme le fonctionnement de notre système perceptif se base sur la perception de différences entre grandeurs, et non pas sur la perception de grandeurs absolues, notre organisme a besoin de références ou de repères pour effectuer ses jugements. Dans le cas général, des repères intramodaux<sup>1</sup> sont utilisés, c'est à dire des repères issus de la modalité sensorielle en question (la modalité auditive dans notre cas). Dans le cas d'exposition multimodale, et sous certaines conditions – comme l'absence de repères pertinents de la modalité principale, la présence de repères forts d'une autre modalité, ou sous des conditions expérimentales qui favorisent cette deuxième modalité – des repères intermodaux peuvent alors intervenir dans l'évaluation (dans notre cas des repères visuels). En ce qui concerne plus particulièrement les jugements spatiaux, il semble que le cadre de référence spatial chez l'homme soit créé de manière amodale, c'est à dire que non pas une seule, mais plusieurs modalités sensorielles y participent, dont la modalité visuelle qui est en effet considérée comme la modalité dominante dans la représentation de l'espace. Par conséquent, l'utilisation de ce cadre influence les jugements spatiaux-auditifs.

Enfin, des repères autres que les repères physiques peuvent intervenir dans l'évaluation subjective. Il s'agit des repères cognitifs, particulièrement importants dans le cas de l'étude des interactions sensorielles. Ils interviennent, d'une part au niveau de la conception des stimuli hétéromodaux, en faisant appel par exemple à la connaissance et à la mémoire qui déterminent la compatibilité cognitive des stimuli. Ils interviennent d'autre part au niveau de la procédure expérimentale, soit sous la forme de paramètres attentionnels – par exemple, les effets dynamiques attirent davantage l'attention des sujets vis-à-vis des effets statiques – soit en prenant un caractère évolutif, se traduisant par le changement des réponses avec le temps. Ainsi, tous ces paramètres tels que la connaissance, la mémoire, l'attention, l'intérêt, l'apprentissage, l'habitude, les consignes etc. peuvent consciemment ou non, influencer le jugement sur une grandeur physique, et ils doivent absolument être pris en compte dans les cas de conditions d'exposition multimodale.

Enfin, les perspectives qui s'ouvrent devant ce travail sont multiples :

D'une part, et en restant dans le cadre de la présente étude, une méthode similaire à celle utilisée pour tester l'influence visuelle sur la distance auditive, devrait être adoptée pour tester les autres attributs perceptifs, descriptifs de la qualité acoustique des salles. Notamment ceux qui sont liés à la dimension spatiale du son (comme l'enveloppement, l'impression d'espace) ou ceux qui présentent un caractère sémantique plus complexe (comme encore l'enveloppement, l'intimité ou même la réverbérance). Un soin particulier dans la conception des stimuli auditifs et visuels doit être pris, notamment par l'utilisation de techniques de simulation améliorées (par exemple du mouvement dans le champ visuel). Enfin, des études sémantiques doivent éventuellement compléter les études plus classiques,

---

<sup>1</sup> Nous rappelons ici la note présentée au § 2.4 : par *intramodale*, on entend une opération effectuée dans une même modalité ; par *intermodale*, une opération effectuée entre deux modalités (c'est le cas de l'interaction) ; par *multimodale*, on entend une opération où plusieurs modalités sont impliquées ; par *amodale*, une opération qui n'est pas associée à une seule modalité mais à plusieurs (dans ce sens, amodale est similaire à multimodale). Par ailleurs, on parle de stimuli *hétéromodaux* quand il s'agit de stimuli de différentes modalités ; de stimuli *bimodaux* quand il s'agit de stimuli où deux modalités sont impliquées ; et enfin, de stimuli *unimodaux* quand ils ne concernent qu'une seule modalité.

afin d'accéder aux représentations que les sujets se font par rapport aux différentes situations simulées ainsi qu'aux expériences.

D'autre part les problèmes liés à la simulation bimodale sont tels qu'ils nécessitent des études spécifiques. Une technique de simulation améliorée est évidemment avantageuse mais ne nous épargne pas du problème de 'couplage' des sens, si difficile sous des conditions expérimentales avec simulation. La question qui se pose – et qui se posera de plus en plus dans l'avenir – est celle du développement de méthodes qui permettront d'étudier les interactions sensorielles sans obliger un découplage des sens, et engendrant ainsi d'office l'affaiblissement du phénomène que nous voulons justement étudier.

Et puis, la direction qui s'ouvre vers l'évaluation subjective 'multimodale' représente un intérêt croissant et vraiment vaste. Tout d'abord, presque toute évaluation subjective considérée comme 'unimodale', implique des informations provenant des autres modalités sensorielles (ou des représentations cognitives) : ainsi, un jugement sur le bruit d'une voiture, d'un téléphone ou d'un ordinateur, peut être influencé par la couleur, la taille ou la forme de l'objet (ou également par la réputation d'une marque, de son prix, etc.). Ces informations doivent donc être étudiées et prises en compte pour arriver à une évaluation des produits plus complète, et en tous cas, moins 'biaisée'. Enfin, le domaine d'applications des communications ou de 'multimédia' qui visent à la création d'une illusion de la réalité (ou de plausibilité, de vraisemblance), reposent justement sur la perception multimodale. Nombreux sont les domaines qui pourraient être explorés : le couplage des sens bien sûr, mais aussi leur interaction avec les aspects fonctionnels et ergonomiques des interfaces, notions correspondant aux évolutions apportées par le besoin croissant d'interactivité entre l'homme et le monde artificiel.



## 8 ANNEXES

### 8.1 ANNEXE A : MATERIEL DES EXPERIENCES

**Projecteurs :** KODAK EXTAPRO 5000, Slide Projector, # 052032500D et # 052086500D

**Lentilles :** Docteur Optic WetZCAR, DOCTARLUX MC, PC 2.8/25, f= 26.9.

Lentilles grand angle f= 26.9, [25mm, f/28 MC] à distance focale réglable et à faisceaux dirigeables permettant de créer une image très peu déformée

**Ecran :** WEBER altuglass (translucide, diffusant léger)

**Haut-parleurs :** TANNOY D100 –Rosewood (Serial No :019720M et 019721M)

**Amplificateur :** Harman/Kardon HK6500, Intergrated Amplifier (Serial No : S154-02451)

**DAT :** Panasonic Model :SV – 37000EG-H, Professional Digital Audio Tape

**Système de commande :** DATATON AC PAX Adaptor

**Casque audio :** Sennheiser HD 545

## 8.2 ANNEXE B : EXPERIENCES SUR LA DISTANCE AUDITIVE

### Expérience préliminaire

#### i) Séquences sonores

Les quatre séquences de 18 stimuli (3 présentations x 6 sons) utilisées dans les tests sont :

A2 – A6 – A3 – A5 – A4 – A1 –/– A6 – A4 – A3 – A2 – A5 – A1 –/– A3 – A1 – A5 – A6 – A2 – A4

A5 – A3 – A4 – A2 – A1 – A6 –/– A1 – A2 – A3 – A6 – A5 – A4 –/– A6 – A2 – A5 – A1 – A4 – A3

A3 – A5 – A2 – A4 – A6 – A1 –/– A4 – A5 – A3 – A1 – A2 – A6 –/– A4 – A1 – A5 – A6 – A3 – A2

A3 – A5 – A1 – A2 – A6 – A4 –/– A1 – A3 – A5 – A6 – A4 – A2 –/– A1 – A5 – A3 – A2 – A4 – A6

#### ii) Résultats

V	P	A	Stimulus	Moyenne	Ecart-type
1	1	1	A1P1V0	0,29	1,142
1	1	2	A2P1V0	1,52	1,286
1	1	3	A3P1V0	3,31	2,015
1	1	4	A4P1V0	6,67	1,724
1	1	5	A5P1V0	4,75	2,591
1	1	6	A6P1V0	8,09	0,791
1	2	1	A1P2V0	0,33	0,940
1	2	2	A2P2V0	2,23	2,126
1	2	3	A3P2V0	4,52	1,852
1	2	4	A4P2V0	5,76	1,715
1	2	5	A5P2V0	5,02	1,139
1	2	6	A6P2V0	8,08	0,671
1	3	1	A1P3V0	0,13	1,018
1	3	2	A2P3V0	1,52	1,208
1	3	3	A3P3V0	4,13	1,328
1	3	4	A4P3V0	5,27	1,863
1	3	5	A5P3V0	5,69	1,487
1	3	6	A6P3V0	8,35	0,802
2	1	1	A1P1V1	0,20	0,718
2	1	2	A2P1V1	2,14	1,578

2	1	3	A3P1V1	4,16	1,315
2	1	4	A4P1V1	6,30	0,824
2	1	5	A5P1V1	4,93	2,245
2	1	6	A6P1V1	8,34	1,051
2	2	1	A1P2V1	0,27	0,936
2	2	2	A2P2V1	2,31	1,814
2	2	3	A3P2V1	4,69	1,712
2	2	4	A4P2V1	6,48	1,297
2	2	5	A5P2V2	5,91	1,706
2	2	6	A6P2V1	8,33	0,749
2	3	1	A1P3V1	0,13	0,656
2	3	2	A2P3V1	1,87	1,230
2	3	3	A3P3V1	4,69	1,100
2	3	4	A4P3V1	5,25	1,966
2	3	5	A5P3V2	6,39	1,065
2	3	6	A6P3V1	8,42	0,631
3	1	1	A1P1V3	0,38	0,974
3	1	2	A2P1V3	3,02	1,666
3	1	3	A3P1V3	4,67	2,133
3	1	4	A4P1V3	6,65	1,044
3	1	5	A5P1V3	4,82	2,707
3	1	6	A6P1V3	8,30	1,562
3	2	1	A1P2V3	0,84	1,220
3	2	2	A2P2V3	2,73	1,321
3	2	3	A3P2V3	4,87	1,214
3	2	4	A4P2V3	6,32	1,145
3	2	5	A5P2V3	5,83	0,974
3	2	6	A6P2V3	8,34	0,705
3	3	1	A1P3V3	0,70	0,911
3	3	2	A2P3V3	2,30	1,191
3	3	3	A3P3V3	5,87	0,835
3	3	4	A4P3V3	6,00	1,800
3	3	5	A5P3V3	6,21	1,022
3	3	6	A6P3V3	8,70	0,717

Tableaux B.1 : Résultats de l'expérience préliminaire.

## Expérience 1

### i) Description des stimuli auditifs

**Tableau B.2 : Interface haut-niveau**

config.	Source Perception			Room Preception			Late Room Decay		
	source presence Es (dB)	source warmth Desl(dB)	source brilliance Desh(dB)	room presence Rev (dB)	running reverb. Edt (s)	envelop. Rd1(dB)	late reverb. Rt (s)	heaviness Drth (lin)	liveness Drth (lin)
A1	-5	0	0	-26	0,6	-12	1,3	1,2	0,5
A2	-8	0	0	-26	0,83	-12,1	1,3	1,2	0,5
A3 (ref)	-11	0	0	-26	1,04	-12,1	1,3	1,2	0,5
A4	-14	0	0	-26	1,20	-12,1	1,3	1,2	0,5
A5	-17	0	0	-26	1,34	-12,1	1,3	1,2	0,5
A6	-5	0	0	-27	0,55	-18,7	1,3	1,2	0,5
A7	-7,66	0	0	-26	0,82	-15,7	1,3	1,2	0,5
A8	-13,3	0	0	-26	1,16	-10,2	1,3	1,2	0,5
A9	-15,3	0	0	-25	1,28	-8,3	1,3	1,2	0,5
A10	-8	0	0	-23	1,04	-12,1	1,3	1,2	0,5
A11	-14	0	0	-29	1,04	-12,1	1,3	1,2	0,5

## Expérience 1 (suite)

**Tableau B.3 : Interface bas-niveau**

config.	Energy			
	direct OD (dB)	early R1(dB)	cluster R2(dB)	rev R3(dB)
A1	-5,9	-12	-14	-16
A2	-8,9	-15	-18	-17
A3 (ref)	-11	-18	-22	-17
A4	-14	-21	-25	-17
A5	-17	-24	-27	-18
A6	-5,1	-18	-23	-17
A7	-8	-18	-22	-17
A8	-14	-18	-22	-17
A9	-17	-18	-21	-17
A10	-8	-15	-19	-14
A11	-14	-21	-25	-20

**Tableaux B.2 et B.3 :** Description des paramètres acoustiques haut et bas-niveau du *Spatialisateur* pour les 11 stimuli de l'expérience 1. A1, A2, A3, A4 et A5 variaient sur Es, A6, A7, A8 et A9 sur OD et A10, A3 et A11 sur le niveau global G (les valeurs sont données par rapport à un niveau de référence qui correspond au niveau du signal d'entrée).

### ii) Niveaux sonores

Les niveaux sonores ont été mesurés à la position de la tête des auditeurs uniquement pour les stimuli A6, A3 et A5 ; le stimulus A6 représente le son le plus fort, A3 est un des sons de niveau moyen et A5 est le son le plus faible. Des valeurs instantanées sont données ici à titre indicatif : A6 = 73 dB(A), A3 = 68 dB(A) et A5 = 63 dB(A) pour S.Vega, tandis que A6 = 74 dB(A), A3 = 68 dB(A) et A5 = 61 dB(A) pour P.Bryn-Jullson.

### iii) Calcul des erreurs

*Les trois séries de stimuli.*

	6 dB	3 dB	0 dB	-3 dB	-6 dB
Es	A1	A2	A3	A4	A5
OD	A6	A7	A3	A8	A9
G		A10	A3	A11	

**Tableau B.4 :** Stimuli auditifs de l'expérience principale 1. Le stimulus A3 représente la configuration de référence.

## Expérience 1 (suite)

Les erreurs sont calculés par la formule :  $(x_i - x_{reg})^2$ , l'erreur globale par  $\Sigma(x_i - x_{reg})^2 / n - 1$ , où  $x_i$  sont les valeurs moyennes de distance auditive,  $x_{reg}$  les valeurs correspondantes des droites de régression et  $n$  le nombre de stimuli par indice (5 pour Es et OD, 3 pour G).

« Tom's dinner » de S.Vega :

dB	6	3	0	-3	-6	
<b>Es</b>	0,90	2,33	4,94	6,38	8,89	
<b>OD</b>	1,03	2,23	4,94	5,99	8,07	
<b>G</b>		2,72	4,94	6,01		
<b>Es reg</b>	0,68	2,69	4,69	6,69	8,69	
<b>OD reg</b>	0,89	2,67	4,45	6,24	8,02	
<b>G reg</b>	1,27	2,91	4,56	6,20	7,85	
						<b>Erreur globale</b>
<b>Es erreur</b>	0,048	0,125	0,062	0,096	0,038	<b>0,092</b>
<b>OD erreur</b>	0,021	0,190	0,235	0,061	0,003	<b>0,127</b>
<b>G erreur</b>		0,036	0,144	0,036		<b>0,108</b>

« La Somnambule » de Bellini interprété par Phyllis Bryn-Jullson :

dB	6	3	0	-3	-6	Erreur globale
<b>Es</b>	0,92	2,54	4,40	6,70	8,51	
<b>OD</b>	0,57	2,16	4,40	5,78	7,77	
<b>G</b>		3,52	4,40	6,63		
<b>Es reg</b>	0,75	2,68	4,62	6,55	8,48	
<b>OD reg</b>	0,53	2,33	4,14	5,94	7,74	
<b>G reg</b>	1,73	3,29	4,85	6,41	7,97	
<b>Es erreur</b>	0,030	0,020	0,044	0,023	0,001	<b>0,030</b>
<b>OD erreur</b>	0,001	0,029	0,072	0,026	0,001	<b>0,032</b>
<b>G erreur</b>		0,050	0,199	0,050		<b>0,149</b>

**Tableaux B.5 et B.6 :** Valeurs moyennes de distance auditive et droites de régression linéaires correspondantes pour «Tom's dinner» de S.Vega et « La Somnambule » de Bellini interprétée par Phyllis Bryn-Jullson.

## Expérience 2

### i) Description des stimuli auditifs

**Tableau B.7 : Interface haut-niveau**

Tableau B6	Source Perception			Room Preception			Late Room Decay		
	config.	Source presence Es (dB)	source warmth Desl(dB)	source brilliance Desh(dB)	room presence Rev (dB)	running reverb. Edt (s)	envelop. Rd1(dB)	late reverb. Rt (s)	heaviness Drtl (lin)
A1	-2	0	0	-26	0,39	-12	1,3	1,2	0,5
A2	-5	0	0	-26	0,6	-12	1,3	1,2	0,5
A3	-8	0	0	-26	0,83	-12,1	1,3	1,2	0,5
A4	-11	0	0	-26	1,04	-12,1	1,3	1,2	0,5
A5	-14	0	0	-26	1,20	-12,1	1,3	1,2	0,5
A6	-17	0	0	-26	1,34	-12,1	1,3	1,2	0,5

**Tableau B.8 : Interface bas-niveau**

config.	Energy			
	direct OD (dB)	early R1(dB)	cluster R2(dB)	rev R3(dB)
A1	-3	-8,7	-6,7	-16
A2	-5,9	-12	-14	-16
A3	-8,9	-15	-18	-17
A4	-11	-18	-22	-17
A5	-14	-21	-25	-17
A6	-17	-24	-27	-18

**Tableaux B.7 et B.8 :** Description des paramètres acoustiques haut et bas-niveau du *Spatialisateur* pour les 6 stimuli de l'expérience 2 : A1, A2, A3, A4 et A5; les stimuli varient principalement sur Es (les valeurs sont données par rapport à un niveau de référence qui correspond au niveau du signal anéchoïque d'entrée).

## Expérience 2 (suite)

### ii) Séquences

Les deux séquences de 42 stimuli (7 x 6) utilisées dans les tests étaient :

#### Séquence 1 :

A3 - A6 - A5 - A2 - A1 - A4 - /	A2 - A3 - A5 - A1 - A6 - A4 - /	A2 - A6 - A4 - A1 - A3 - A5		
1 <sup>ère</sup> prés.	2 <sup>ème</sup> prés.	3 <sup>ème</sup> prés.		
A4 - A6 - A2 - A1 - A5 - A3 - /	A5 - A6 - A3 - A4 - A1 - A2 - /	A5 - A3 - A2 - A6 - A1 - A4		
4 <sup>ème</sup> prés.	5 <sup>ème</sup> prés.	6 <sup>ème</sup> prés.		
<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center; border: 1px solid black; padding: 5px;">A3 - A4 - A2 - A5 - A6 - A1</td> </tr> <tr> <td style="text-align: center;">7<sup>ème</sup> prés.</td> </tr> </table>			A3 - A4 - A2 - A5 - A6 - A1	7 <sup>ème</sup> prés.
A3 - A4 - A2 - A5 - A6 - A1				
7 <sup>ème</sup> prés.				

#### Séquence 2 :

A5 - A3 - A2 - A6 - A1 - A4 - /	A6 - A4 - A2 - A1 - A5 - A3 - /	A2 - A6 - A4 - A1 - A3 - A5		
1 <sup>ère</sup> prés.	2 <sup>ème</sup> prés.	3 <sup>ème</sup> prés.		
A3 - A6 - A5 - A4 - A1 - A2 - /	A3 - A4 - A2 - A5 - A6 - A1 - /	A3 - A6 - A5 - A2 - A1 - A4		
4 <sup>ème</sup> prés.	5 <sup>ème</sup> prés.	6 <sup>ème</sup> prés.		
<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center; border: 1px solid black; padding: 5px;">A2 - A3 - A5 - A1 - A6 - A4</td> </tr> <tr> <td style="text-align: center;">7<sup>ème</sup> prés.</td> </tr> </table>			A2 - A3 - A5 - A1 - A6 - A4	7 <sup>ème</sup> prés.
A2 - A3 - A5 - A1 - A6 - A4				
7 <sup>ème</sup> prés.				

### iii) Niveaux sonores

Les niveaux sonores ont été mesurés à la position de la tête des auditeurs. Ils ont été enregistrés pour les deux durées totales d'intégration de 5 s et 10 s (ces durées correspondent environ à la durée pendant laquelle les sujets écoutaient les signaux avant de répondre).

Durée d'intégration	A1	A2	A3	A4	A5	A6
5 s	74,5	71	68	64	62	59,5
10 s	78	75	71	67	64,5	61

**Tableau B.9 :** Niveaux sonores équivalents en dB(A), arrondis au demi-décibel le plus proche.



## Expérience 5 (suite).

## iii) Résultats

V	P	A	Stimulus	Moyenne	Ecart-type
1	1	1	A1P1V0	0.53	0.58
1	1	2	A2P1V0	1.65	1.13
1	1	3	A3P1V0	3.15	1.25
1	1	4	A4P1V0	4.78	1.72
1	1	5	A5P1V0	7.12	1.06
1	1	6	A6P1V0	8.35	0.89
1	2	1	A1P2V0	0.83	0.70
1	2	2	A2P2V0	2.12	1.24
1	2	3	A3P2V0	3.92	0.86
1	2	4	A4P2V0	5.71	1.39
1	2	5	A5P2V0	7.40	0.80
1	2	6	A6P2V0	8.62	0.62
2	1	1	A1P1V1	0.65	0.63
2	1	2	A2P1V1	1.60	0.94
2	1	3	A3P1V1	3.15	1.19
2	1	4	A4P1V1	5.15	1.30
2	1	5	A5P1V1	6.98	0.93
2	1	6	A6P1V1	8.33	0.93
2	2	1	A1P2V1	0.55	0.74
2	2	2	A2P2V1	1.75	1.09
2	2	3	A3P2V1	3.65	1.26
2	2	4	A4P2V1	5.46	1.37
2	2	5	A5P2V2	6.95	0.90
2	2	6	A6P2V1	8.35	0.74
3	1	1	A1P1V3	0.68	0.68
3	1	2	A2P1V3	1.67	1.05
3	1	3	A3P1V3	3.70	1.42
3	1	4	A4P1V3	5.26	1.22
3	1	5	A5P1V3	6.86	1.01
3	1	6	A6P1V3	8.24	1.00
3	2	1	A1P2V3	0.82	0.69
3	2	2	A2P2V3	2.12	0.95
3	2	3	A3P2V3	3.72	1.19
3	2	4	A4P2V3	5.94	1.20
3	2	5	A5P2V3	7.67	0.86
3	2	6	A6P2V3	8.56	0.57

Tableau B.10 : Résultats de l'expérience 5 (vues statiques) pour 28 sujets.

## Expérience 6 (vues changeantes).

### i) Description et niveau sonore des stimuli auditifs

Voir expérience principale 2.

### ii) Séquences

Sept séquences de 18 stimuli (3 conditions visuelles x 6 sons) ont été utilisées lors des tests.

Afin de compenser les effets d'ordre de présentation des stimuli –inévitablement présents lors de présentations successives de stimuli sonores – plusieurs séquences bimodales ont été utilisées. La règle générale de leur agencement fut d'éviter la répétition de l'apparition d'un couple de sons dans un certain ordre bien plus de fois que les autres couples. Ainsi, les k séquences d'un test étaient créées ensemble, de telle manière que si le son A1 apparaissait trois fois après le son A5 pendant une séquence, le son A5 devait y apparaître également trois fois après le son A1, autant par exemple que le son A2 après le son A4, etc. (ou à une apparition près quand cela était impossible). Cette règle fut d'ailleurs appliquée également pour les tests en vues statiques; la contrainte supplémentaire ici fut que la même image ne pouvait apparaître deux fois à la suite.

Deux exemples sur les 7 séquences créées en tout, sont donnés ci-après :

#### 1<sup>ère</sup> séquence :

##### 1<sup>ère</sup> présentation :

A3V1 A6V3 A5V2 A2V3 A1V2 A4V3 / A2V1 A3V3 A5V1 A1V3 A6V2  
A4V1 / A3V2 A1V1 A5V3 A4V2 A6V1 A2V2

##### 2<sup>ème</sup> présentation :

A5V3 A3V1 A2V2 A4V1 A1V2 A6V3 / A3V2 A6V1 A5V2 A2V3 A1V1  
A4V2 / A3V3 A6V2 A1V3 A2V1 A4V3 A5V1

#### 2<sup>ème</sup> séquence :

##### 1<sup>ère</sup> présentation :

A4V1 A3V2 A2V1 A5V3 A6V2 A1V1 / A2V2 A4V3 A6V1 A3V3 A5V1  
A1V3 / A4V2 A6V3 A3V1 A5V2 A2V3 A1V2

##### 2<sup>ème</sup> présentation :

A3V1 A4V3 A2V2 A6V3 A5V1 A1V2 / A4V1 A5V2 A3V3 A6V1 A2V3  
A1V1 / A6V2 A2V1 A3V2 A5V3 A4V2 A1V3

## Expérience 6 (suite).

## iii) Résultats

V	P	A	Stimulus	Moyenne	Ecart-type
1	1	1	A1P1V1	0.56	0.64
1	1	2	A2P1V1	1.36	0.87
1	1	3	A3P1V1	2.61	1.17
1	1	4	A4P1V1	4.32	1.60
1	1	5	A5P1V1	7.04	1.65
1	1	6	A6P1V1	8.18	0.99
1	2	1	A1P2V1	0.91	0.71
1	2	2	A2P2V1	2.09	1.09
1	2	3	A3P2V1	3.30	0.88
1	2	4	A4P2V1	5.26	1.30
1	2	5	A5P2V2	7.10	0.97
1	2	6	A6P2V1	8.34	0.57
2	1	1	A1P1V2	0.80	0.93
2	1	2	A2P1V2	1.59	1.67
2	1	3	A3P1V2	3.07	1.03
2	1	4	A4P1V2	5.01	1.66
2	1	5	A5P1V2	6.84	1.07
2	1	6	A6P1V2	8.38	0.81
2	2	1	A1P2V2	1.06	0.82
2	2	2	A2P2V2	2.16	1.07
2	2	3	A3P2V2	4.07	1.12
2	2	4	A4P2V2	5.75	1.21
2	2	5	A5P2V2	7.01	0.92
2	2	6	A6P2V2	8.50	0.74
3	1	1	A1P1V3	0.83	0.97
3	1	2	A2P1V3	1.47	0.85
3	1	3	A3P1V3	3.62	1.20
3	1	4	A4P1V3	6.18	1.27
3	1	5	A5P1V3	7.77	1.29
3	1	6	A6P1V3	8.40	0.58
3	2	1	A1P2V3	1.06	0.97
3	2	2	A2P2V3	2.15	1.25
3	2	3	A3P2V3	3.68	1.03
3	2	4	A4P2V3	5.57	1.08
3	2	5	A5P2V3	7.31	0.88
3	2	6	A6P2V3	8.67	0.80

Tableau B.11 : Résultats de l'expérience 6 (vues changeantes) pour 28 sujets.

### 8.3 ANNEXE C : EXPERIENCES SUR UN ENSEMBLE DE FACTEURS PERCEPTIFS

#### Expérience préliminaire auditive

*i) Description des stimuli auditifs*

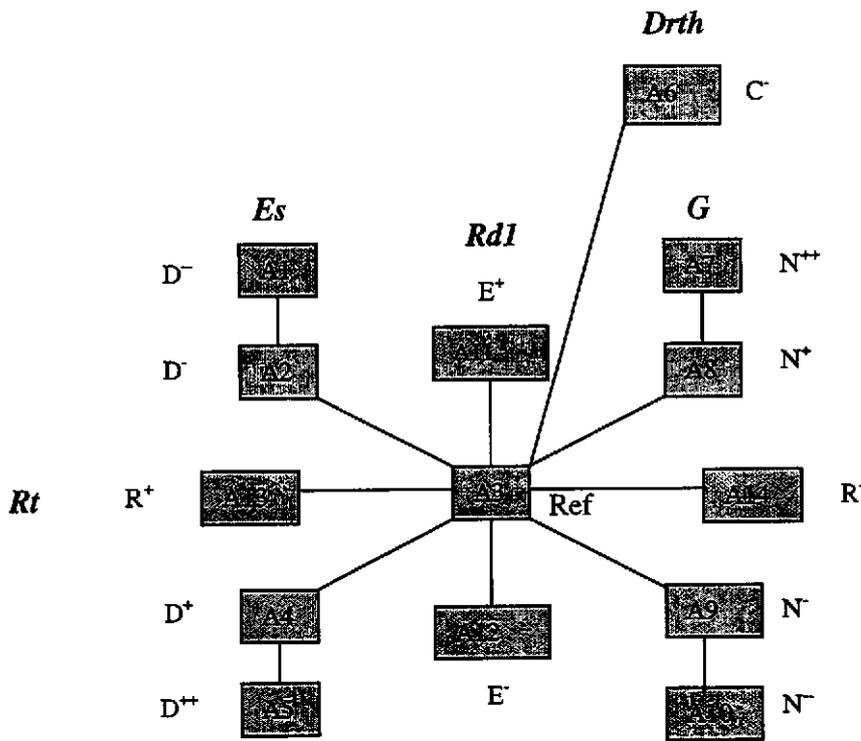


Figure C.1: Présentation schématique des 14 stimuli auditifs du test préliminaire auditif.

## Expérience préliminaire auditive (suite)

Tableau C.1 : Interface haut-niveau

config.	Source Perception			Room Preception			Late Room Decay		
	source presence Es (dB)	source warmth Desl(dB)	source brilliance Desh(dB)	Room presence Rev (dB)	running reverb. Edt (s)	envelop. Rd1(dB)	late reverb. Rt (s)	Heavi-ness Drth (lin)	liveness Drth (lin)
A1 (D <sup>-</sup> )	-5	0	0	-26	0,6	-12	1,3	1,2	0,5
A2 (D <sup>-</sup> )	-8	0	0	-26	0,815	-11,8	1,3	1,2	0,5
<b>A3 (Ref)</b>	<b>-11</b>	<b>0</b>	<b>0</b>	<b>-26</b>	<b>1,00</b>	<b>-12,5</b>	<b>1,3</b>	<b>1,2</b>	<b>0,5</b>
A4 (D <sup>+</sup> )	-14	0	0	-26	1,16	-11,2	1,3	1,2	0,5
A5 (D <sup>++</sup> )	-17	0	0	-26	1,30	-11,2	1,3	1,2	0,5
A6 (C <sup>-</sup> )	-11	0	0	-26	1,00	-12,5	1,3	1,2	0,2
A7 (N <sup>++</sup> )	-7,33	0	0	-23	0,96	-13,0	1,3	1,2	0,5
A8 (N <sup>+</sup> )	-8,33	0	0	-24	0,96	-13,0	1,3	1,2	0,5
A9 (N <sup>-</sup> )	-13	0	0	-28	1,00	-12,5	1,3	1,2	0,5
A10 (N <sup>-</sup> )	-14	0	0	-29	1,00	-12,5	1,3	1,2	0,5
A11 (E <sup>+</sup> )	-11	0	0	-26	1,00	-8,5	1,3	1,2	0,5
A12 (E <sup>-</sup> )	-11	0	0	-26	1,00	-19,5	1,3	1,2	0,5
A13 (R <sup>+</sup> )	-11	0	0	-26	1,03	-12,5	1,7	1,2	0,5
A14 (R <sup>-</sup> )	-11	0	0	-26	0,92	-12,5	1,0	1,2	0,5

Tableau C.2 : Interface bas-niveau

config.	Energy			
	direct OD (dB)	Early R1(dB)	cluster R2(dB)	rev R3(dB)
A1 (D <sup>-</sup> )	-5,9	-12	-14	-16
A2 (D <sup>-</sup> )	-8,9	-15	-16	-17
<b>A3 (Ref)</b>	<b>-11</b>	<b>-19</b>	<b>-17</b>	<b>-17</b>
A4 (D <sup>+</sup> )	-14	-20	-19	-17
A5 (D <sup>++</sup> )	-17	-23	-21	-18
A6 (C <sup>-</sup> )	-11	-19	-17	-17
A7 (N <sup>++</sup> )	-7,9	-16	-14	-14
A8 (N <sup>+</sup> )	-8,9	-17	-15	-15
A9 (N <sup>-</sup> )	-13	-21	-19	-19
A10 (N <sup>-</sup> )	-14	-22	-20	-20
A11 (E <sup>+</sup> )	-13	-14	-17	-17
A12 (E <sup>-</sup> )	-11	-39	-17	-17
A13 (R <sup>+</sup> )	-11	-19	-18	-19
A14 (R <sup>-</sup> )	-11	-19	-15	-15

**Tableaux C.1 et C.2** : Description des paramètres acoustiques haut et bas-niveau du *Spatialisateur* pour les 14 stimuli de l'expérience préliminaire auditive. A1, A2, A3, A4 et A5 variaient sur 'source présence' Es, A3 et A6 sur 'liveness' Drth, A7, A8, A9 et A10 sur le niveau global G, A11, A3 et A12 sur 'enveloppement' Rd1, A13, A3 et A14 sur 'late réverbérance' Rt (les valeurs sont données par rapport à un niveau de référence qui correspond au niveau du signal d'entrée).

## Expérience préliminaire auditive (suite)

## ii) Résultats

	Composante 1	Composante 2	Composante 3	Composante 4
Proportion de Variance	0.67	0.18	0.09	0.03
Variance Cumulative	0.67	0.95	0.94	0.97

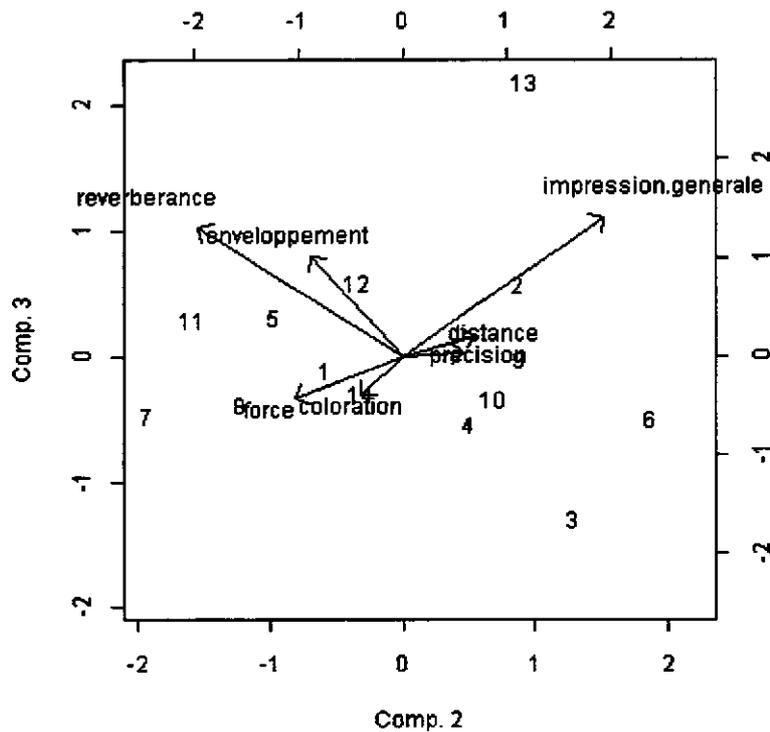
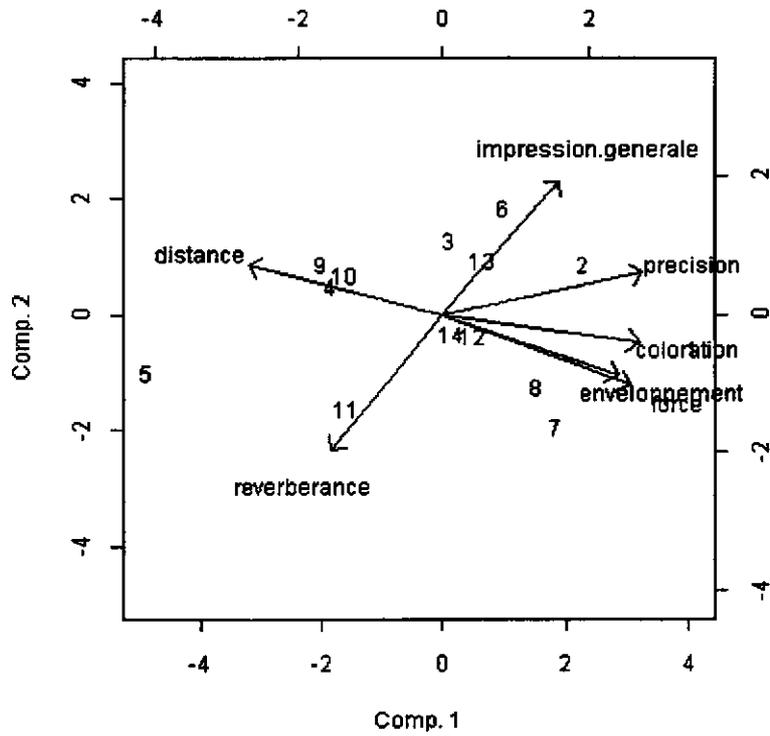
Variables	Composante 1	Composante 2	Composante 3	Composante 4
force	0.41	-0.32	-0.19	-0.34
précision	0.44	0.18		-0.44
réverbérance	-0.25	-0.61	0.58	-0.44
coloration	0.43	-0.12	-0.18	
distance	-0.43	0.22		
enveloppement	0.39	-0.27	0.45	0.68
impression générale	0.25	0.60	0.62	-0.17

**Tableau C.3 :** Expérience préliminaire auditive : résultats d'une analyse (ACP) prenant en compte sept variables : force sonore, précision, réverbérance, coloration, distance auditive; enveloppement, impression générale i) Proportion de variance et proportion cumulative de variance expliquées par les quatre premières composantes. ii) Qualité de représentation des variables sur les quatre composantes.

	force	précision	réverbér.	coloration	distance	envelopp.	impr.gén.
force	1						
précision	0.77	1					
réverbéranc	-0.28	-0.60	1				
coloration	0.86	0.85	-0.45	1			
distance	-0.92	-0.83	0.37	-0.87	1		
envelopp.	0.75	0.68	-0.14	0.76	-0.83	1	
impr. gén.	0.20	0.66	-0.52	0.35	-0.32	0.40	1

**Tableau C.4 :** Matrice de corrélation pour l'espace perceptif comprenant la variable représentant la 'distance visuelle' (32 sujets).

### Expérience préliminaire auditive (suite)



1:A1=D-, 2:A2=D-, 3:A3=Ref, 4:A4=D+, 5:A5=D++, 6:A6=C-, 7:A7=N++, 8:A8=N+, 9:A9=N-, 10:A10=N-,  
 11:A11=E+, 12:A12=E-, 13:A13=R+, 14:A14=R-

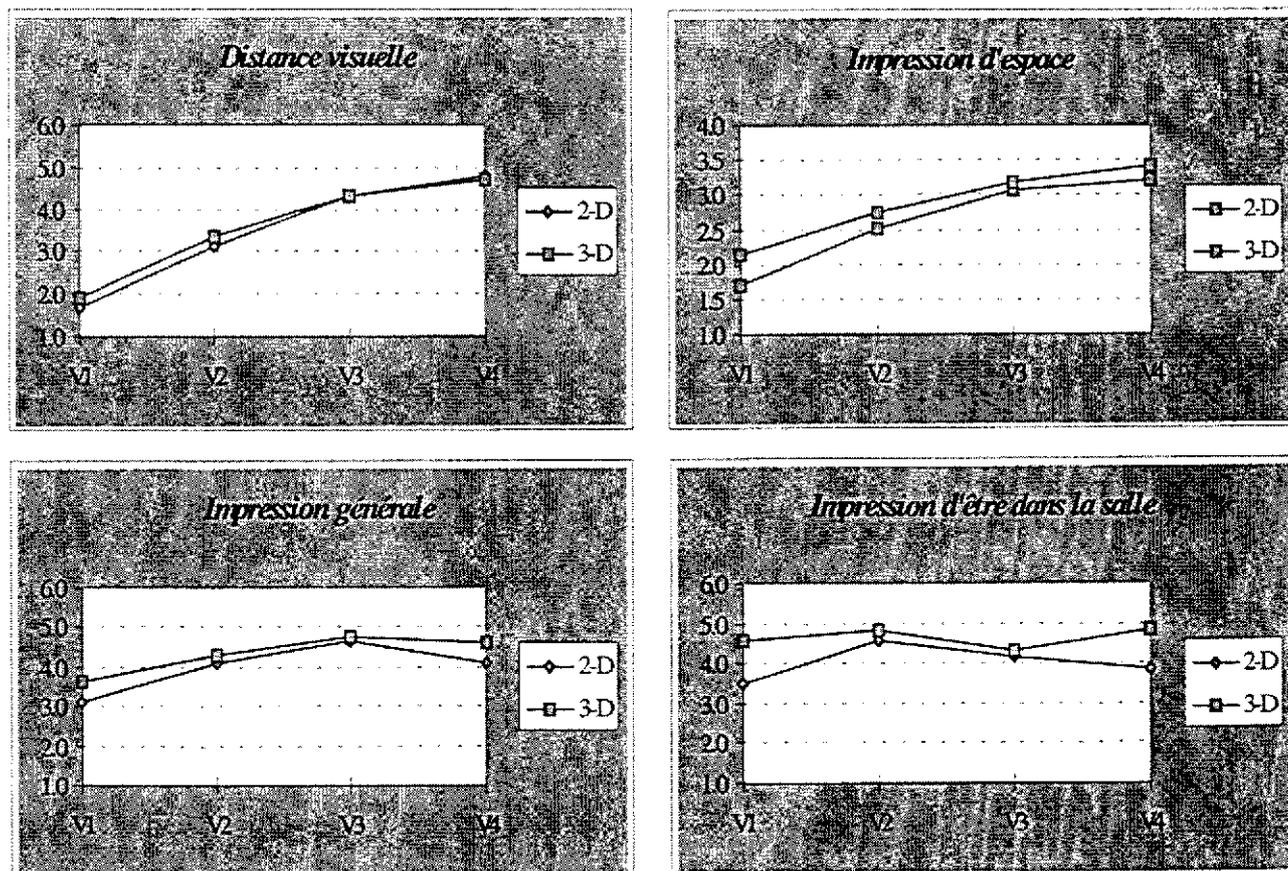
**Figure C.2 :** Premier et deuxième plans factoriels de l'expérience préliminaire auditive avec 14 sons et sept variables : force sonore, précision, réverbérance, coloration, distance auditive; enveloppement, impression générale (valeurs moyennées sur 11 sujets).

## Expérience préliminaire visuelle

Les 4 photos du Théâtre des Champs Elysées utilisées lors des tests principaux, V1, V2, V3 et V4, ont été projetées en 2-D et en 3-D (8 images au total). 20 sujets ont participé à l'expérience; pour la moitié des sujets, les 4 photos 2-D ont été présentées au début des tests et les 4 photos 3-D ensuite. L'ordre a été inversé pour l'autre moitié des sujets. De plus, l'ordre de présentation des images (dans chaque sous-ensemble de quatre images 2-D ou 3-D) a été alterné plusieurs fois pour éviter les effets d'ordre.

Le questionnaire donné aux sujets portait sur 4 aspects 'visuels' des images : la 'distance' de la scène (échelle différentielle de 1 à 6), 'l'impression d'espace' (échelle de 1 à 4), 'l'impression d'être dans la salle' (échelle de 1 à 6), et 'l'impression générale' (échelle de 1 à 6).

Les quatre graphiques qui suivent présentent les principaux résultats séparément pour les quatre aspects visuels des images et les deux méthodes de projection, 2-D et 3-D.



**Figure C.3 :** Moyennes des réponses du test préliminaire visuel sur la 'distance visuelle', 'l'impression d'espace', 'l'impression générale' et 'l'impression d'être dans la salle' pour les quatre vues V1, V2, V3 et V4 utilisées dans l'expérience principale, en fonction des méthodes de projection 2-D et 3-D (20 sujets).

## Expérience préliminaire visuelle (suite)

L'analyse en composantes principales est également effectuée sur les réponses. La répartition de la variance sur les deux premières composantes et le premier plan factoriel sont présentés à la suite.

	Composante	Composante 2
Proportion de Variance	0.75	0.21
Variance Cumulative	0.75	0.97

Variables	Composante	Composante 2
distance visuelle	0.53	-0.37
impr. d'espace	0.56	-0.19
impr. générale	0.55	
impr. d'être dans la salle	0.30	0.90

Tableau C.5 : Répartition de la variance sur les deux premières composantes.

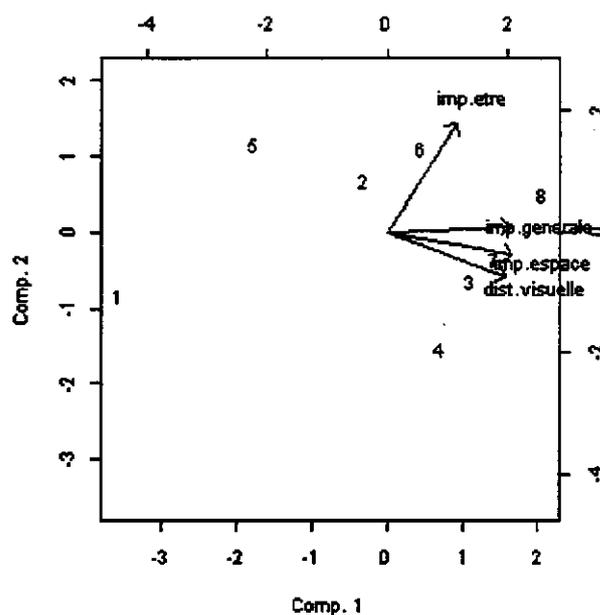


Figure C.4 : Premier plan factoriel du test préliminaire visuel. Les variables sont la 'distance visuelle', 'l'impression d'espace', 'l'impression générale' et 'l'impression d'être dans la salle', les points 1, 2, 3, 4 correspondent respectivement aux images V1, V2, V3 et V4 en projection 2-D, tandis que les points 4, 5, 6 et 7 correspondent aux mêmes images en projection 3-D.

## Expérience principale

### i) Description des stimuli auditifs

Tableau C.6 : Interface haut-niveau

config.	Source Perception			Room Perception			Late Room Decay		
	source presence Es (dB)	source warmth Desl(dB)	source brilliance Desh(dB)	room presence Rev (dB)	running reverb. Edt (s)	envelop. Rd1(dB)	late reverb. Rt (s)	heaviness Drtl (lin)	liveness Drth (lin)
A1 (D <sup>-</sup> )	-8	0	0	-26	0,82	-11,8	1,3	1,2	0,5
<b>A2 (Ref)</b>	<b>-11</b>	<b>0</b>	<b>0</b>	<b>-26</b>	<b>1,00</b>	<b>-12,5</b>	<b>1,3</b>	<b>1,2</b>	<b>0,5</b>
A3 (D <sup>+</sup> )	-14	0	0	-26	1,16	-11,2	1,3	1,2	0,5
A4 (D <sup>++</sup> )	-17	0	0	-26	1,30	-11,2	1,3	1,2	0,5
A5 (V <sup>-</sup> )	-11	0	0	-26	1,00	-12,5	1,3	1,2	0,2
A6 (N <sup>-</sup> )	-8,33	0	0	-24	0,96	-13,0	1,3	1,2	0,5
A7 (N <sup>-</sup> )	-13	0	0	-28	1,00	-12,5	1,3	1,2	0,5
A8 (E <sup>-</sup> )	-11	0	0	-26	1,00	-6,5	1,3	1,2	0,5
A9 (E <sup>-</sup> )	-11	0	0	-26	1,00	-19,5	1,3	1,2	0,5
A10 (R <sup>-</sup> )	-11	0	0	-26	1,03	-12,5	1,7	1,2	0,5

Tableau C.7 : Interface bas-niveau

config.	Energy			
	direct OD (dB)	early R1(dB)	cluster R2(dB)	rev R3(dB)
A1 (D <sup>-</sup> )	-8,9	-15	-16	-17
<b>A2 (Ref)</b>	<b>-11</b>	<b>-19</b>	<b>-17</b>	<b>-17</b>
A3 (D <sup>+</sup> )	-14	-20	-19	-17
A4 (D <sup>++</sup> )	-17	-23	-21	-18
A5 (V <sup>-</sup> )	-11	-19	-17	-17
A6 (N <sup>-</sup> )	-8,9	-17	-15	-15
A7 (N <sup>-</sup> )	-13	-21	-19	-19
A8 (E <sup>-</sup> )	-13	-14	-17	-17
A9 (E <sup>-</sup> )	-11	-39	-17	-17
A10 (R <sup>-</sup> )	-11	-19	-18	-19

Tableaux C.6 et C.7 : Description des paramètres acoustiques haut et bas-niveau du *Spatialisateur* pour les 10 stimuli de l'expérience principale. A1, A2, A3 et A4 variaient sur 'source presence' Es, A2 et A5 sur 'liveness' Drth, A6, A2 et A7 sur le niveau global, A8, A2 et A9 sur 'envelopment' Rd1, A2 et A10 sur 'late réverbérance' Rt (les valeurs sont données par rapport à un niveau de référence qui correspond au niveau du signal d'entrée).

## Expérience principale (suite)

### ii) Séquences

La même philosophie que celle employée lors des tests sur la distance, a été suivie ici pour la conception des ordres de présentation des stimuli (voir Annexe B, Expérience 6, ii. Séquences).

Un exemple de séquence est donné ci-dessous, au total, 8 séquences différentes de ce type (avec 50 stimuli en tout) ont été produites :

*10 stimuli de la partie unimodale du test :*

A5 A1 A4 A9 A7 A3 A2 A10 A8 A6

*40 stimuli de la partie bimodale du tests :*

A1V2 A9V1 A2V3 A8V2 A6V4 A5V1 A3V3 A7V4 A10V3 A4V2  
 A1V1 A7V2 A9V4 A3V2 A4V1 A8V4 A6V3 A2V4 A10V1 A5V3 /  
 A6V2 A10V4 A7V1 A4V4 A1V3 A2V2 A8V1 A4V5 A9V3 A3V1  
 A10V2 A8V3 A1V4 A6V1 A4V3 A2V1 A9V2 A3V4 A7V3 A5V2

### iii) Niveaux sonores

Stimuli sonores	Niveau sonore global
A1	83.5 dB(A)
A2	81 dB(A)
A3	78 dB(A)
A4	75.5 dB(A)
A5	81 dB(A)
A6	83.5 dB(A)
A7	79 dB(A)
A8	79 dB(A)
A9	81.5 dB(A)
A10	80.5 dB(A)

**Tableau C.8 :** Niveaux sonores équivalents sur la durée de diffusion de base (45s) du signal (boucle continue de 45s), arrondis au demi-décibel le plus proche.

## Expérience principale (suite)

### iv) Estimation du niveau sonore moyen dans le Théâtre des Champs Elysées.

Afin d'estimer le niveau sonore moyen dans une salle, il y a deux paramètres qui doivent être pris en compte : la puissance sonore d'une voix chantée du même type que celle de nos expériences, et l'amplification d'une salle similaire à la salle du Théâtre des Champs Elysées.

En ce qui concerne la puissance sonore d'une voix chantée, elle peut être estimée en s'appuyant sur des considérations d'équilibre entre masses orchestrales et ensembles choristes. Les valeurs de 99 dB pour un chanteur soliste (type opéra wagnérien), de 96 dB pour un style classique ou baroque, et enfin de 93 dB pour un choriste ont été estimées [Kahle, 1995]. Pour la chanson de Phyllis Bryn-Jullson ('La Somnambule' de Bellini), un niveau de 96 dB semble préférable et assez réaliste.

En ce qui concerne l'amplification liée à la salle :

L'amplification notée généralement  $G$  est le rapport entre le *niveau total de la réponse impulsionnelle* et le *niveau sonore de la source en champ libre à 10 m*.

Ici, nous utilisons la force sonore  $F$  qui est le rapport entre le *niveau total de la réponse impulsionnelle* et la *puissance de la source* calculée à 1 m.

La relation entre l'amplification  $G$  et la force sonore  $F$  est :  $G = F - (-20 -11)$

(-20 dB à cause de la multiplication de la distance par un facteur de 10, et -11 parce que pour une source omnidirectionnelle de puissance 0 dB le niveau sonore à 1 m est inférieur de -11 dB).

La force sonore  $F$  du Théâtre des Champs Elysées est de l'ordre de -30 dB en moyenne (en référence à la puissance d'une source de 0 dB). La variation à l'intérieur de la salle dépend de sa taille (plus la salle est grande, plus large est la variation). Typiquement, pour la salle des Champs Elysées, ces valeurs couvrent une plage de -22 dB (pour les places les plus proches, c'est à dire à environ 10 m) à -35 dB pour les places les plus lointaines ou situées sur les côtés (sans visibilité).

Compte tenu de ces valeurs, nous avons donc un niveau sonore absolu du Théâtre des Champs Elysées de l'ordre de : *Puissance \* Force Sonore* = 96 dB - 30 dB = 66 dB en moyenne.

Cette valeur doit être corrigée par la directivité de la voix, ce qui amènerait à des valeurs sensiblement supérieures de quelques dB, par exemple, pour une augmentation due à la directivité de 6 dB le niveau sonore moyen deviendrait 72 dB.

La comparaison de cette dernière valeur avec celle du tableau C.8 montre que les niveaux sonores utilisés dans ces expériences étaient relativement élevées par rapport aux niveaux estimés dans le véritable Théâtre des Champs Elysées.

## Expérience principale (suite)

## v) Résultats

	sons	stimuli bimodals	Force sonore		Précision		Vivacité		Réverbérance	
			moyenne	écart type	moyenne	écart type	moyenne	écart type	moyenne	écart type
1	D-	A1V1	4.57	0.65	4.52	0.90	2.38	0.85	3.12	0.91
2		A1V2	4.42	0.69	4.37	1.30	2.51	1.03	3.62	1.00
3		A1V3	4.59	1.00	4.57	1.25	2.86	0.98	3.46	0.97
4		A1V4	4.65	0.58	4.53	1.17	2.45	0.86	3.44	0.96
5	Ref	A2V1	3.81	0.72	4.43	0.76	2.15	0.91	3.45	0.77
6		A2V2	3.86	0.80	4.18	0.93	2.51	0.85	3.55	0.92
7		A2V3	3.74	0.66	4.44	0.74	2.20	1.02	3.57	0.96
8		A2V4	4.05	0.69	4.42	0.65	2.51	0.83	3.80	0.79
9	D+	A3V1	3.19	0.87	4.07	0.86	2.31	0.87	3.95	0.94
10		A3V2	3.11	0.85	4.01	0.89	2.25	0.84	3.96	0.80
11		A3V3	3.03	0.79	3.87	0.86	2.42	0.81	4.03	0.90
12		A3V4	3.23	0.74	4.10	0.95	2.50	0.97	3.94	0.85
13	D++	A4V1	2.39	0.71	3.34	0.86	2.22	0.95	4.15	0.89
14		A4V2	2.39	0.83	3.61	1.00	2.41	0.99	4.23	0.93
15		A4V3	2.43	0.77	3.58	0.72	2.30	1.18	4.32	0.90
16		A4V4	2.49	1.07	3.72	0.87	2.42	1.00	4.22	0.92
17	V-	A5V1	3.58	1.06	4.58	1.30	2.26	1.08	2.92	0.98
18		A5V2	3.88	0.83	4.54	0.74	2.05	0.91	2.78	0.95
19		A5V3	4.03	0.53	4.44	0.68	2.01	0.93	2.94	0.93
20		A5V4	3.82	0.89	4.47	0.84	2.02	0.66	3.01	0.87
21	N+	A6V1	4.73	0.49	4.40	0.83	2.45	0.87	3.61	0.84
22		A6V2	4.48	0.63	4.30	0.70	2.85	0.99	3.85	0.75
23		A6V3	4.58	0.79	4.73	0.54	2.91	1.02	3.69	1.14
24		A6V4	4.60	0.67	4.47	0.91	2.69	0.94	3.78	1.10
25	N-	A7V1	3.38	0.75	4.16	0.66	2.54	0.96	3.40	0.86
26		A7V2	3.62	0.63	4.38	0.88	2.16	0.90	3.25	0.85
27		A7V3	3.57	1.06	4.40	0.94	2.16	0.99	3.54	1.04
28		A7V4	3.30	0.69	4.36	0.61	2.05	0.82	3.45	0.75
29	E+	A8V1	3.54	1.05	3.85	0.95	2.34	1.21	4.15	1.10
30		A8V2	3.24	0.73	3.94	1.07	2.69	1.28	4.10	1.09
31		A8V3	3.48	0.82	3.61	0.98	2.39	1.09	4.27	1.08
32		A8V4	3.46	0.86	3.57	1.48	2.61	1.00	4.45	0.98
33	E-	A9V1	4.04	0.87	4.47	0.90	2.52	1.07	3.60	0.87
34		A9V2	4.04	0.48	4.08	0.85	2.49	0.73	3.60	0.84
35		A9V3	3.82	0.81	3.97	0.75	2.22	0.79	3.87	0.84
36		A9V4	3.95	0.67	4.28	1.02	2.63	1.28	3.80	0.85
37	R+	A10V1	3.85	0.75	4.34	0.63	2.62	0.70	3.85	0.92
38		A10V2	3.72	0.70	4.49	0.63	2.45	0.67	3.51	0.64
39		A10V3	3.69	0.78	4.16	1.00	2.21	0.92	3.82	0.96
40		A10V4	3.67	0.71	4.22	0.74	2.33	1.09	3.98	0.80

**Tableau C.9** : Moyennes et écart-types des 40 stimuli bimodaux (10 stimuli sonores x 4 conditions visuelles) pour chacun des 8 facteurs du questionnaire (valeurs normalisées).

## Expérience principale (suite)

	sons	stimuli bimodals	Distance auditive		Enveloppement		Impression générale		Compatibilité	
			moyenne	écart type	moyenne	écart type	moyenne	écart type	moyenne	écart type
1	D-	A1V1	2.44	0.83	2.23	1.06	4.13	0.82	3.15	0.80
2		A1V2	2.97	0.69	2.50	0.97	3.86	1.08	3.02	0.70
3		A1V3	2.96	0.66	3.05	1.08	4.28	0.91	2.95	0.96
4		A1V4	2.88	0.66	2.38	1.00	3.44	0.98	2.72	0.84
5	Ref	A2V1	3.29	0.86	2.65	0.88	4.38	0.76	2.73	0.76
6		A2V2	3.64	0.60	2.37	1.01	4.10	0.96	3.33	0.67
7		A2V3	3.60	0.59	2.72	0.88	4.35	0.69	2.81	0.67
8		A2V4	3.62	0.82	2.21	1.01	3.54	1.14	3.25	0.75
9	D+	A3V1	4.12	0.84	2.76	1.00	4.02	1.13	2.36	0.91
10		A3V2	3.97	0.67	2.37	1.07	3.04	1.03	1.14	0.92
11		A3V3	4.45	0.54	2.92	0.89	4.38	1.07	2.34	0.96
12		A3V4	4.43	0.68	2.67	1.01	3.87	0.91	2.33	1.14
13	D++	A4V1	4.94	0.73	2.91	0.90	4.02	1.06	2.11	0.74
14		A4V2	4.78	0.70	2.70	1.43	3.33	1.01	1.98	0.90
15		A4V3	4.92	0.70	2.98	0.92	4.12	0.98	2.19	1.00
16		A4V4	5.06	0.62	2.59	0.76	3.88	0.84	2.85	0.82
17	V-	A5V1	2.65	0.87	2.42	0.75	3.94	0.91	2.21	0.98
18		A5V2	3.08	0.63	2.17	1.41	3.54	0.98	3.03	0.91
19		A5V3	3.54	0.58	2.48	1.20	3.62	0.69	1.90	0.95
20		A5V4	3.56	0.65	2.24	0.70	3.95	0.88	2.87	0.71
21	N+	A6V1	2.58	0.76	2.55	0.75	4.31	0.85	2.83	0.70
22		A6V2	3.03	0.76	2.03	1.03	3.79	1.07	3.42	0.80
23		A6V3	3.12	0.98	2.07	0.81	4.00	0.77	2.67	0.70
24		A6V4	3.13	0.84	2.53	0.66	3.94	0.89	2.59	0.90
25	N-	A7V1	3.69	0.92	2.58	0.93	3.89	0.79	2.85	0.76
26		A7V2	3.79	0.65	2.59	0.82	3.90	0.93	2.86	1.03
27		A7V3	3.98	0.89	2.16	0.82	3.84	0.82	2.81	0.76
28		A7V4	3.87	0.70	2.70	0.73	4.23	0.91	2.29	1.03
29	E+	A8V1	4.26	0.80	2.31	0.82	3.91	0.95	2.41	1.03
30		A8V2	4.17	1.15	2.40	1.12	4.17	0.90	3.00	0.65
31		A8V3	4.24	0.75	2.27	0.96	3.84	1.03	2.97	0.86
32		A8V4	4.39	0.82	2.95	0.83	4.26	0.82	2.94	0.80
33	E-	A9V1	3.06	0.79	2.44	1.06	3.23	0.88	1.30	0.69
34		A9V2	3.44	0.56	2.09	0.84	3.95	0.85	2.67	0.95
35		A9V3	4.02	0.67	2.68	0.96	3.53	1.11	1.48	0.94
36		A9V4	3.73	0.89	2.51	0.81	4.25	0.71	2.94	0.79
37	R+	A10V1	3.31	0.78	2.15	0.94	3.96	0.88	2.67	0.77
38		A10V2	3.60	0.71	2.27	0.82	3.83	1.01	2.49	0.78
39		A10V3	3.75	0.66	2.85	1.42	3.44	1.02	2.62	1.07
40		A10V4	3.91	0.68	2.57	0.78	3.92	0.87	2.94	0.81

**Tableau C.9 :** Moyennes et écart-types des 40 stimuli bimodaux (10 stimuli sonores x 4 conditions visuelles) pour chacun des 8 facteurs du questionnaire (valeurs normalisées).



## 9 BIBLIOGRAPHIE

- B.S. Atal, M.S. Schroeder et G.M. Sessler. (1965). Subjective reverberation time and its relation to sound decay. *Proc. 5<sup>th</sup> International Congress on Acoustics*, Liege, G 32.
- M. Barron. (1971). The subjective effects of first reflections in concert halls – The need for lateral reflections. *J. Sound and Vibration*, 15(4), pp. 475-494.
- M. Barron. (1978). The Glubenkian Great Hall, Lisbon, II : an acoustic study of a concert hall with variable stage. *J. Sound and Vibration*, 59, p. 481.
- M. Barron et A.H. Marshall. (1981). Spatial impression due to early lateral réflexions in concert halls : the derivation of a physical measure. *J. Sound and Vibration*, 77(2), pp. 221-232.
- M. Barron. (1988). Subjective study of British symphony concert halls. *Acustica*, 66, pp. 1-14.
- M. Barron. (1993). Auditorium Acoustics and Architectural Design. E et FN Spon, London.
- M. Barron. (1996). Loudness in concert halls. *Acustica acta acustica*, 82 (Suppl. 1), p. 21.
- S. Bech, V. Hansen et W. Woszczyk. (1995). Interaction between audio-visual factors in a home theater system : Experimental results. *99th AES Convention, New-York*, October 1995, Preprint n° 4096.
- D. Begault. (1999). Auditory and non-auditory factors that potentially influence virtual acoustic imagery. *Proc. de AES, 16<sup>th</sup> International Conference on Spatial Sound Reproduction*, Finland.
- L.L. Beranek. (1962). Music, Acoustics and Architecture. Wiley, New York.
- L.L. Beranek. (1992). Concert hall acoustics. *J. Acoust. Soc. Am.*, 92(1), pp.1-39.
- L.L. Beranek. (1995). Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as mesures of acoustical quality of concert halls. *J. Acoust. Soc. Am.*, 98(2), pp. 988-1007.
- L.L. Beranek. (1996). Concert and Opera Halls : How They Sound. Acoustical Society of America.
- I.H. Bernstein. (1988). Applied multivariate analysis. Springer-Verlag, New-York.
- J. Blauert, K. Gruber et W. Lindemann. (1984). Untersuchung des höreignismersmerkmals 'Räumlichkeit' mit hilfe eines rechner-modells der binauralen signalverarbeitung. Forschungsbericht für DFG, Ruhr-Universität Bochum, AEA Bochum.
- J. Blauert. (1996). Spatial hearing. The MIT Press, Cambridge, Massachusetts, London, England.
- C. Bonnet. (1986). Manuel pratique de psychophysique. Armand Colin Editeur, Paris.
- J.H. Bray et S.E. Maxwell. (1985). Multivariate analysis of variance. Sage University Paper Series, Quantitative applications in the social sciences n° 07-054.
- J.M. Brown, K.L. Anderson, C.A. Fowel et C. Carello. (1998). Visual influences on auditory distance perception. *J. Acoust. Soc. Am.*, 104, 3, Pt. 2, p. 1798.

- D. Brungart. (1993). Distance simulation in virtual audio displays. Proc. of the IEEE 1993 *National Aerospace and Electronics Conference*, NAECON 1993, 2, pp. 612-617.
- O. Cahen. (1990). L'image en relief : de la photographie stéréoscopique à la vidéo 3D. Masson, Paris.
- G. Canevet. (1994/1995). Eléments de psychoacoustique. Cours de DEA de Mécanique, Laboratoire de Mécanique et d'Acoustique, CNRS, Université Aix-Marseille II, Marseille.
- D. Casalis. (1991) (coordination éditoriale). Grand Dictionnaire de la Psychologie. Larousse, Canada.
- C. Cavé, R. Ragot et M. Fano (June 1992). Perception of sound-image synchrony in cinematographic conditions. *Fourth Workshop on Rhythm Perception et Production* - Bourges, France.
- N. Château. (1998). Les interactions sensorielles entre le son et l'image lors de la perception d'événements audiovisuels. *Musique & Arts Plastiques*, Edité par H.Genevois et Y.Orlarey. Rencontres Musicales Pluridisciplinaires 1998, Lyon.
- P. Cochran, J. Troop et W.E. Simson. (1968). Estimation of distance of a source of sound. *Am. J. Psychol.*, 81, pp. 198-206.
- F.B. Colavita. (1974). Human sensory dominance. *Perception et Psychophysics*, 16, pp. 409-412.
- P. Coleman. (1962). Failure to localize the source distance of an unfamiliar sound. *J. Acoust. Soc. Am.*, 34(3), pp. 345-346.
- P. Coleman. (1963). An analysis of cues to auditory depth perception in free space. *Psychological bulletin*, 60(3), pp. 302-315.
- P. Coleman. (1968). Dual role of frequency spectrum in determination of auditory distance. *J. Acoust. Soc. Am.*, 44(3), pp. 631-632.
- L. Cremer et H.A. Müller. (1982). Principles and Applications of Room Acoustics, Vol 1 et 2 ; traduit par T.J.Shultz. Applied Science Publishers, London.
- P. Damaske et Y. Ando. (1972). Interaural crosscorrelation for multichannel loudspeaker reproduction. *Acustica*, 27, pp. 232-238.
- N.F. Dixon et L. Spitz (1980). The detection of audiovisual desynchrony. *Perception*, 9, pp. 719-721.
- B. Dodd et R. Campbell. (1987). *Hearing by Eye: The Psychology of Lip Reading*. Erlbaum, London.
- R.E. Easton. (1983). The effect of head movements on visual and auditory dominance. *Perception*, 12, pp. 63-70.
- B. Escofier et J. Pages. (1988). *Analyses factorielles simples et multiples*. Dunod, Paris.
- M. Gardner. (1968). Proximity image effect in sound localization. *J. Acoust. Soc. Am.*, 43, p. 163.
- M. Gardner. (1969). Distance estimation of 0° or apparent 0° -Oriented speech signals in anechoic space. *J. Acoust. Soc. Am.*, 45(1), pp. 47-53.
- A.L. Glass et K.J. Holyoak. (1986). *Cognition*. McGraw-Hill International Editions.
- A. Gribenski. (1978). Interactions Sensorielles: Audition et Vision. *Journal de Psychologie*, 2, pp. 147-171.

- D. Griesinger. (1997). The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces. *Acustica-acta acustica*, 83, pp. 721-731.
- P. Guillaume. (1937). La psychologie de la forme, Nouvelle Bibliothèque Scientifique, Flammarion Ed., Paris, France.
- R.J. Hawkes et H. Douglas. (1971). Subjective acoustic experience in concert auditoria. *Acustica*, 24(5), pp. 235-250.
- M.P. Hollier, A.N. Rimell, D.S. Hands et R.M. Voelcker. (1999). Multi-modal perception. *BT Technol. J*, Vol 17, n° 1.
- C.V. Jackson. (1953). Visual factors in auditory localisation. *Quarterly Journal of Experimental Psychology*, 5, pp. 52-65.
- B. Jones (1975). Spatial perception in the blind. *British Journal of Psychology*, 66, pp. 461-472.
- V.L. Jordan. (1980). Acoustical Design of Concert halls and Theaters. Applied Science Publishers, London.
- J.M. Jot, O. Warusfel, E. Kahle et M. Mein. (1993). Binaural concert hall simulation in real time. Addendum to Proc. IEEE Workshop on Applications of Digital Signal Processing to Audio and Acoustics.
- J.M. Jot, V. Larcher et O. Warusfel. (1995). Digital signal processing issues in the context of binaural and transaural stereophony. In Proc. 98<sup>th</sup> Conv. Audio Eng. Soc., Preprint n° 3980.
- J.M. Jot, O. Warusfel. (1995). "Spat~ a spatial processor for musicians and sound engineers". CIARM'95 Conference, Ferrara - Italie. Mai 1995.
- J.M. Jot. (1996). Synthesizing three dimensional sound scenes in audio or multimedia production and interaction with human-computer interfaces. Proc. Of 5<sup>th</sup> Int. Conf. on Interface on real and virtual worlds, pp. 399-415.
- J.M. Jot, J-M. Cerveau et O. Warusfel. (1997). Analysis and synthesis of room reverberation based on a statistical time-frequency model. 103<sup>rd</sup> AES convention, New York, septembre 1997.
- J. Jouhaneau. (1997). Acoustique des salles et sonorisation. Lavoisier Technique et Documentation, Paris.
- J-P. Jullien, E. Kahle, S. Winsberg, et O. Warusfel. (1992). Some results on the objective and perceptual characterization of room acoustical quality in both laboratory and real environments. *Proc. Institute of Acoustics*, Vol.XIX, n° 2.
- J-P. Jullien. (1995). Structured model for the representation and the control of room acoustical quality. In Proc. 15<sup>th</sup> International Conf. on Acoustics, Trondheim - Norvège, 2, pp. 517-520.
- E. Kahle. (1995). Validation d'un modèle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d'opéras. Thèse de Doctorat. Université du Maine. Le Mans.
- W. de V. Keet. (1968). The influence of early lateral reflections on the spatial impression. 6<sup>th</sup> International Congress on Acoustics, Tokyo, August 1968.
- O. Klemm. (1909). Lokalisation von Sinneseindrücken bei disparaten Nebenreizen. *Psychol. Studien*, 5, pp. 73-162.

- V.O. Knudsen. (1931). Acoustics of music rooms. *J. Acoust. Soc. Am*, pp.434-467.
- A. Kohlrausch et S.de Par. (1999). Auditory-Visual interaction: from fundamental research in cognitive psychology to (possible) applications. Proceedings of SPIE Reprint. 25-28 January 1999, San Jose, California.
- H. Kuttruff (1979). Room Acoustics. Applied Science Publishers, London.
- J. de Lagarde. (1983). Initiation à l'analyse des données. Dunod, Paris.
- S. Lakatos (1993). Comparison of the perception of motion on auditory sources in the front and in the rear. Submitted to *Perception* (January, 1993).
- V. Larcher, J-M. Jot et G. Vandernoot. (1998). "Equalization methods in Binaural Technology", *105th convention of the Audio Eng. Soc.* in San Francisco, Preprint n° 4858.
- C. Lavandier. (1989). Validation perceptive d'un modèle objectif de la qualité acoustique des salles. Thèse de Doctorat. Université du Maine. Le Mans.
- H. McGurk et J. McDonald. (1976). Hearing lips and seeing voices. *Nature*, 264, pp. 746-748.
- V. Maffiolo. (1999). De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain. Thèse de Doctorat, Université du Maine, Le Mans.
- L.E. Marks. (1987). On cross-modal similarity: Auditory-Visual Interactions in speeded discrimination. *Journal of Experimental Psychology*, 13, pp. 384-394.
- N. Marmaras, P. Vassilakis et G. Dounias. (1995). Factors affecting accuracy of producing time intervals. *Perceptual and Motor Skills*, 80, pp. 1043-1056.
- A.H. Marshall. (1967). A note on the importance of room cross-section in concert halls. *J. Sound and Vibration*, 5, pp. 100-112.
- J.M. Martel et R. Nadeau. (1988). Statistique en gestion et en économie. Editions Gaëtan Morin, Paris.
- D. Mershon et E. King. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics*, 18(6), pp. 409-415.
- D.H. Mershon et W.C. Gogel. (1975). Failure of familiar size to determine a metric for visually perceived distance. *Perception & Psychophysics*, 1975, 17, pp. 101-106.
- D. Mershon et J. Bowers. (1979). Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, 8, pp.311-322.
- G.A. Milliken et D.E. Johnson. (1984). Analysis of messy data: Vol. I. Designed experiments. Van Nostrand Reinhold, Co. New York.
- H. Møller. (1992). Fundamentals of binaural technology. *Applied Acoustics*, Vol. 36, pp.171-217
- F.R. Moore. (1983). A general model for spatial processing of sounds. *Computer Music Journal*, vol.7, no. 6.
- M. Mzali. (1997). Influence de la directivité des sources sur la perception d'une scène sonore. Rapport de stage DEA ATIAM. Université Paris VI.

- C. Nathanail, C. Lavandier et J.D. Polack (1996). Influence of visual information on auditory perception. Consequences on the objective characterization of room acoustics. Forum Acusticum, Antwerpen, April 1996, Supplément à *Acustica*, 82, S216.
- C. Nathanail, C. Lavandier, J.D. Polack. et O. Warusfel. (1997a). Etude de l'influence de la perception visuelle sur la perception auditive. Conséquences sur la caractérisation perceptive et objective de l'acoustique des salles. Actes du 4ème Congrès Français d'Acoustique, Marseille, France (14-18 Avril 1997), 1, pp. 275-278.
- C. Nathanail, C. Lavandier, J.D. Polack. et O. Warusfel. (1997b). Influence of visual information on auditory perception. Consequences on the objective characterization of room acoustics. Proc. ASVA 97 (International Symposium on Simulation, Visualisation and Auralisation for Acoustic Research and Education), Tokyo, Japan (2-4 April 1997), pp. 285-290.
- C. Nathanail, C. Lavandier, J.D. Polack. et O. Warusfel. (1997c). Influence of sensory interactions between vision and audition on the perceptual characterization of room acoustics. Proc. ICMC 97 (International Computer Music Conference), Thessaloniki - Grèce (25-30 Septembre 1997), pp. 414-417.
- S. Nielsen. (1991). Distance perception in hearing. Ph.D. thesis, Aalborg University Press, Aalborg, Denmark.
- S. Nielsen. (1993). Auditory distance perception in different rooms. *J. Audio Eng. Soc.*, 41(10), pp. 755-770.
- J-D. Polack. (1982). Cours de Perception et Acoustique des Salles. DEA d'Acoustique Appliquée de l'Université du Maine. Le Mans.
- J-D. Polack, X. Meynial, G. Dodd, et A. Marshall. (1992). Midas system for all scale room acoustics measurements. *11<sup>th</sup> International AES Test et Measurement Conference*, 29-31 May, Portland, Oregon, USA.
- J-D. Polack. (1994). General Discussion. *Proc. on the Int. Conf. on Acoustic Quality of Concert Halls*. 24-25 October, Madrid, p. 136.
- R. Ragot, C. Cavé et M. Fano (1988). Reciprocal effects of visual and auditory stimuli in a spatial compatibility situation. *Bul. of the Psychon. Soc.*, 26(4), pp. 350-352.
- M. Richelle, J. Requin, M. Robert. (1994). *Traité de psychologie expérimentale*. Presses Universitaires de France.
- E. de Rocquigny. (1993). Etude de l'empacement optimal d'un nouvel orgue à la Primatiale St-Jean de Lyon. Etude réalisée à la demande du Ministère de la Culture, par le Groupe d'Acoustique Musicale, sous la direction de M.Castellengo. Laboratoire d'Acoustique Musicale, Unité Mixte de Recherche 9945, CNRS, Ministère de la Culture, Université Paris 6.
- W.C. Sabine. (1922). *Collected Papers on Acoustics*. Harvard Univ. Press, Los Altos, California. Reprinted by Peninsula Publishing, Acoustical Society of America., 1992.
- W.C. Sabine. (1932). *Acoustics and Architecture*. McGraw-Hill Book Company, Inc, New York and London, 1932.
- SAS. (1994). General discussion. Proc. of the International Conference in Acoustic Quality of Concert Halls. Spanish Acoustical Society (SAS), Octobre 1994, Madrid.

- M.R. Schroeder, D. Gottlob et K.F. Siebrasse. (1974). Comparative study of european concert halls : correlation of subjective preference with geometric and acoustic parameters. *J. Acoust. Soc. Am.*, 56, p. 1195.
- B.R. Shelton et C.L. Searle (1980). The influence of vision on the absolute identification of sound-source position. *Perception et Psychophysics*, 28(6), pp. 589-596.
- G.A. Soulodre et J.S. Bradley. (1995). Subjective evaluation of new room acoustics measures. *J. Acoust. Soc. Am.*, 98(1), pp. 294-301.
- S.S. Stevens. (1975). *Psychophysics. Introduction to its perceptual, neural and social prospects*. J. Wiley, New York, 1975.
- W.H. Sumby et I. Pollack. (1954). Visual contribution to speech intelligibility. *J. Acoust. Soc. Am.*, 26, pp. 212-215.
- G. Thomas. (1941). Influence of vision on sound localisation. *Journal of Experimental Psychology*, 28, pp. 163-175.
- J-P. Vian. (1983). Critères permettant de caractériser l'acoustique d'une salle. Conférence I.T.B.T.P., CSTB.
- J-P. Vian. (1996). La qualité acoustique des espaces, comment la caractériser ?. *Acoustique et Techniques*, 5, pp. 21-26.
- S. Violon et C. Lavandier. (1999). A sound judgement depending on the visual setting ?. In Proc. Joint Meeting ASA/EAA/DEGA : Forum Acusticum 1999, 15-19 mars, Berlin.
- C. Vogel. (1999). Etude sémiotique et acoustique de l'identification des signaux d'avertissement en contexte urbain. Thèse de doctorat, Université Paris 6, Paris.
- W.M. Wagenaars. (1990). Localization of sound in a room with reflecting walls. *J. Audio Eng.Soc.*, 38(3), pp.99-110.
- D.H. Warren (1970). Intermodal interactions in spatial localization. *Cognitive Psychology*, 1, pp. 114-133.
- D.H. Warren. (1979). Spatial localisation under conflict situations : Is there a single explanation ? *Perception*, 8, pp. 323-337.
- D.H. Warren, R.B. Welch et T.J. McCarthy. (1981). The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, 30(6), pp. 557-564.
- B. J Winer. (1962). *Statistical principles in experimental design*. McGraw-Hill. New York.
- W. Woszczyk, S. Bech, and V. Hansen (1995). Interaction between audio-visual factors in a home theater system: Definition of subjective attributes. *99th AES Convention*, New-York, October 1995, Preprint n° 4133.