

# Rôle de sources ou de réflexions multiples dans la perception de l'environnement

**Noël Chateau,**  
France Telecom,  
Centre National des Études des Télécommunications,  
Laboratoire d'Évaluation et acceptabilité de la Qualité des Services, F-Lannion,  
e-mail : noel.chateau@cnet.francetelecom.fr

**Georges Canévet,**  
Centre National de la Recherche Scientifique,  
Laboratoire de Mécanique et d'Acoustique, F-Marseille,  
e-mail : canevet@alphalma.cnrs-mrs.fr

## Introduction

La localisation auditive est une fonction sensorielle essentielle pour l'homme, qui lui permet d'attribuer une direction aux sons dans l'espace, notamment à ceux dont les sources peuvent se situer hors de portée de son champ de vision. Toutefois, bien que cette fonction intervienne dans des situations extrêmement complexes, incluant la présence de nombreuses sources de réflexions, de stimulations via d'autres modalités sensorielles, elle a principalement été étudiée dans des cas très académiques. C'est ainsi qu'un corpus de données considérable a été établi pendant près d'un siècle sur la localisation d'une source unique. Il est apparu que pour une grande part, les mécanismes de la localisation d'une seule source peuvent être expliqués à partir des informations contenues dans les signaux acoustiques parvenant aux tympans d'un auditeur. Ce constat donne la part belle aux ingénieurs qui se sont aussitôt lancés dans la modélisation de ces mécanismes, et la simulation de l'espace sonore par l'intermédiaire de techniques plus ou moins raffinées de traitement du signal. Or, ce constat est un peu l'arbre qui cache la forêt : plus on avance, et plus on s'aperçoit que de nombreux points restent inexplicables et résistent à la modélisation et à la simulation. Notamment, dès que l'on introduit une deuxième source, il apparaît des distorsions que l'on peut qualifier de non linéaires : le percept associé à la présence de deux sources A et B est quasiment toujours différent de la "somme" des percepts associés aux sources A et B lorsqu'elles sont seules.

La localisation auditive lorsque plusieurs sources sont présentes a surtout été étudiée ces cinquante dernières années. C'est principalement la perception résultant de la présence de deux sources placées dans le plan horizontal qui a été examinée. Des effets d'attraction, de fusion, de répulsion, de changement de taille apparente des sources ont alors été observés. Pour certains, ces phénomènes ont pu de nouveau être expliqués par l'examen des relations physiques entre les signaux présents aux deux tympans.

Plus précisément, des calculs de corrélation, de différences interaurales de temps, de phase, de niveau entre les signaux permettent de comprendre ou de prédire les interactions spatiales entre les sources. Ils nous enseignent également que le système auditif tire grand avantage d'avoir deux entrées (les deux oreilles) lors de la localisation multi-sources.

Au début de l'étude relatée ici, il n'existait pas de travaux traitant de l'examen approfondi des percepts associés à la présence de plusieurs sources hors du plan horizontal.

Or, l'intérêt de la compréhension des mécanismes de la localisation de plusieurs sources dont certaines peuvent être situées hors du plan horizontal est évident, car dans la réalité, les sources de signaux acoustiques ne sont que très rarement situées exactement dans ce plan. Par exemple, les réflexions multiples du signal primaire sur les parois d'une salle ou sur le sol peuvent être vues comme autant de sources secondaires provenant de sites divers.

On voit donc qu'un aspect important de la localisation de sources sonores restait inexploré. C'est pourquoi une investigation a été entreprise, afin d'apporter de nouvelles données contribuant à la compréhension des mécanismes de la localisation de sources sonores multiples selon le site. Pour obtenir ces données, trois types de stimuli ont été employés : des bruits à large bande, qui ont pour particularité d'exciter une grande partie de la membrane basilaire et de fournir de nombreuses informations spectrales au système auditif, des bruits passe-bande, passe-haut, passe-bas et à encoche, qui permettent de distribuer l'énergie dans des régions fréquentielles particulières et enfin des sons purs. Ceux-ci sont particulièrement bien adaptés pour étudier la dépendance fréquentielle des phénomènes perceptifs rencontrés, inhérente à l'organisation tonotopique du système auditif.

L'étude a été concentrée sur les percepts résultant de la présence simultanée ou non de deux sources sonores.

Ces percepts dépendent des relations physiques entre les stimuli que sont les relations temporelles, de niveaux, de compositions spectrales, de degrés de cohérence et bien entendu de positions relatives et absolues dans l'espace. Ainsi, de nombreuses expériences psychoacoustiques ont été conduites dans le but d'étudier l'influence de ces différentes relations. Seules trois de celles-ci, portant sur les relations spatiales, fréquentielles et temporelles, seront relatées par la suite. Ces relations ayant été identifiées comme prépondérantes, elles permettront de proposer une conclusion venant porter un autre regard sur la perception auditive spatiale.

## Influence des relations spatiales et fréquentielles

### L'expérience

L'objet de cette expérience était d'étudier l'influence de la présence d'un bruit à large bande sur la localisation de sons purs dans le plan vertical médian (PVM). Pour cela, deux tests ont été conduits dans une salle anéchoïque (dim. : 5 x 5 x 5 m) : l'un de localisation de sons purs seuls (référence), et le second de localisation de sons purs en présence de bruit à large bande pouvant provenir de cinq directions différentes dans le PVM ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$  et  $180^\circ$  en coordonnées cylindriques, voir Fig. 1). Quatre sons purs ont été sélectionnés pour leurs fréquences représentatives de perceptions bien distinctes dans le PVM (Roffler & Butler [1], Blauert [2]). Ces sons purs avaient les fréquences de 500 Hz, 1, 4 et 8 kHz. Pour le premier test, ils étaient tous émis en face des auditeurs à un niveau de 60 phons. Leur durée était de 300 ms, incluant des durées d'établissement et d'extinction de 20 ms (cosinus carré). Pour chaque présentation, ces sons purs étaient produits 3 fois avec des intervalles de silence de 700 ms.

Pour le second test, ils étaient accompagnés de bruit blanc (100-16 000 Hz) ayant un niveau de 60 dB SPL mesuré au point représentant le centre de la tête. Le bruit blanc avait une durée de 500 ms, incluant des durées d'établissement de 20 ms. Pour chaque présentation, il était produit 3 fois avec des intervalles de silence de 500 ms. Les sons purs étaient centrés temporellement sur le bruit. Par ailleurs, leur niveau était adapté de manière à ce que le rapport signal sur bruit soit de 10 dB dans la bande critique centrée

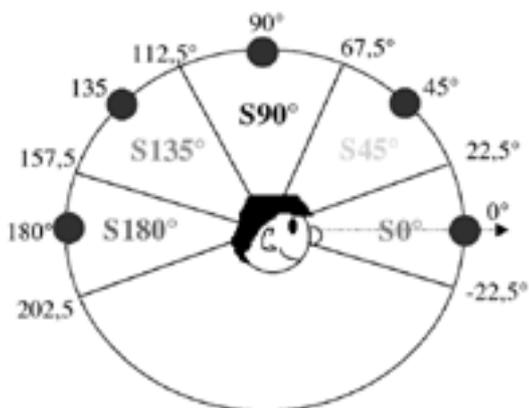


Fig. 1 : Position des sources sonores et composition des secteurs de catégorisation des réponses dans le plan vertical médian

sur leur fréquence [3]. Un tel ajustement permettait de compenser le masquage inégal des sons purs par le bruit blanc en fonction de la fréquence, du fait de l'augmentation de la largeur des bandes critiques avec la fréquence. Les haut-parleurs étaient situés à 2 m du point du centre de la tête des sujets.

Vingt-cinq auditeurs (neuf femmes et seize hommes âgés de 23 à 30 ans) ont pris part à cette expérience. Ils effectuaient tous au préalable un audiogramme attestant qu'aucun n'avait de pertes auditives entre 500 et 8 000 Hz supérieures à 10 dB par rapport à la norme ISO R226. Lors des deux tests, leur tâche était d'indiquer la direction apparente des sons purs par l'intermédiaire d'un micro terminal, en s'aidant éventuellement du dessin figuré ci-dessous (Fig. 2) pour se repérer dans le PVM. Il leur était demandé qu'après chaque réponse, ils ne bougent plus en fixant le haut-parleur leur faisant face et ils ferment les yeux pour éviter l'influence de la vision de celui-ci [4].

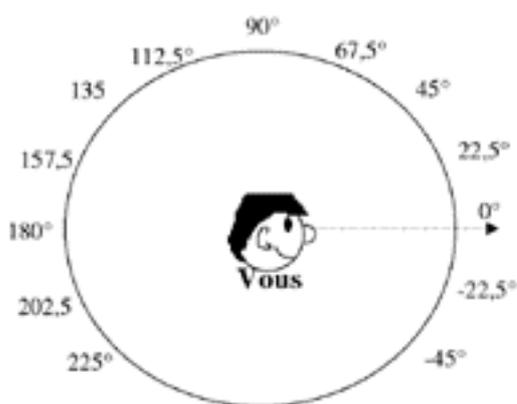


Fig. 2 : Schéma de représentation du PVM fourni aux sujets lors des tests de localisation

Lors du premier test, 32 stimuli étaient présentés dans un ordre aléatoire, soit 8 présentations par fréquences. Dans le deuxième test, 80 stimuli étaient présentés dans un ordre aléatoire, ce qui correspondait à 4 fréquences x 5 directions pour le bruit blanc x 4 présentations pour chaque condition. Un test supplémentaire de localisation du bruit blanc présenté seul dans les 5 directions d'incidences du PVM était effectué, afin de vérifier les aptitudes des sujets à localiser correctement dans ce plan. Les sujets n'ayant pas manifesté de difficultés particulières pour localiser le bruit (localisation généralement correcte), nous ne détaillerons pas les résultats de ce test de moindre intérêt ici (voir [5] pour les détails).

### Résultats

Pour l'analyse, les jugements des sujets ont été regroupés dans 5 secteurs du PVM (voir Fig. 1), ce qui a permis d'établir des pourcentages de réponses dans ces secteurs en fonction des différentes conditions expérimentales.

La figure 3 représente les pourcentages de réponses dans les 5 secteurs du PVM en fonction de la condition de stimulation (son pur seul ou bruit provenant des cinq directions

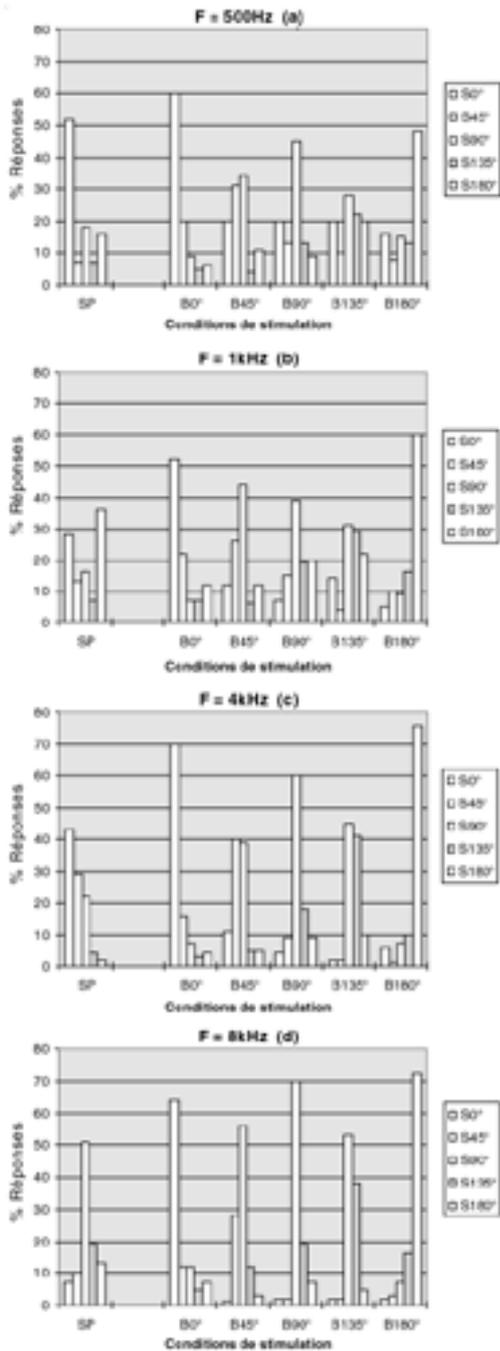


Fig. 3 : Pourcentages de réponses dans les secteurs S0°, S45°, S90°, S135° et S180° du PVM lors de l'expérience de localisation de sons purs en présence ou non de bruit à large bande, en fonction du site de la source de bruit. Le groupe de colonnes de gauche correspond aux pourcentages obtenus pour les sons purs présentés seuls. Chaque graphe représente les résultats obtenus pour une fréquence du son pur ; (a) : 500 Hz, (b) : 1 kHz, (c) : 4 kHz et (d) : 8 kHz. 25 sujets, 4 présentations par sujet

d'incidence). Pour chaque condition, on obtient un groupe de cinq colonnes donnant les pourcentages de réponses dans les 5 secteurs du PVM (S0°, S45°,...). En ce qui concerne les résultats obtenus pour les sons purs présentés seuls (condition SP), on note qu'ils sont localisés dans un secteur préférentiel en fonction de leur fréquence.

Ces secteurs sont respectivement S0° à 0,5 et 4 kHz, S180° à 1 kHz et S90° à 8 kHz. Ces résultats sont tout à fait en accord avec ceux de la littérature [8, 1] et peuvent s'expliquer de la manière suivante : en l'absence de suffisamment d'informations spectrales fournies au système auditif, les stimuli sont localisés dans la direction correspondant à une amplification maximale pour leur fréquence.

Par exemple, les oreilles externes des auditeurs ont la particularité d'amplifier la région autour de 8 kHz pour les sons provenant du dessus. Ainsi, un son pur de 8 kHz sera assimilé à un stimulus dont l'énergie autour de 8 kHz a été amplifiée (c'est la seule qu'il contienne !), ce qui correspond à un stimulus perçu au-dessus, même si dans notre cas il provient physiquement de face.

Lorsqu'un bruit est présent, on constate figure 3 que la localisation des sons purs est totalement bouleversée. Prenons par exemple le cas du son pur de 4 kHz. Lorsqu'il est présenté seul (condition SP), il est localisé à 42 % dans le secteur S0°, puis à 30 % dans le secteur S45° et à 22 % dans le secteur S90°, les pourcentages dans les autres secteurs étant marginaux. En présence de bruit frontal (condition B0°), il est localisé à près de 70 % dans le secteur S0°, et beaucoup moins dans les 4 autres secteurs. Par conséquent, le bruit va renforcer la localisation de ce son pur dans le secteur frontal. En revanche, lorsque la direction d'incidence du bruit est différente, on observe que le pourcentage de localisation du son pur dans le secteur S0° devient très faible.

En fait, on s'aperçoit que le son pur est quasiment toujours majoritairement localisé dans le secteur d'où provient le bruit. Cet effet est encore plus marqué pour le son pur de 8 kHz, alors qu'il l'est un peu moins pour les sons purs de plus basses fréquences (0,5 et 1 kHz). Ainsi, il existe une réelle attraction des sons purs dans la direction d'incidence du bruit. Cette attraction est d'autant plus forte que la fréquence du son pur est plus élevée. On voit par ailleurs que ce phénomène d'attraction a un caractère stabilisant pour les sons purs, puisque la répartition des jugements dans les différents secteurs est beaucoup moins "hasardeuse" en présence de bruit qu'en son absence. De plus, on peut remarquer que cet effet stabilisant est surtout marqué dans les secteurs S0°, S90° et S180°, quelles que soient les directions d'incidence de bruit et la fréquence des sons purs.

### Influence des relations temporelles

L'effet d'attraction des sons purs par le bruit à large bande ayant clairement été mis en évidence, de nombreuses expériences traitant notamment des relations spatiales et fréquentielles entre son cible à localiser et son perturbateur (masque) ont été conduites. Celles-ci ont toutes été conduites avec les mêmes relations temporelles entre cible et masque que celles décrites précédemment. L'importance de la relation temporelle pouvant être préjugée des nombreux travaux portant sur le masquage [6] et l'effet d'antériorité [7], celle-ci a été étudiée précisément pour un seul son pur, celui de 8 kHz, qui est le plus sensible à l'effet d'attraction.

## L'expérience

Cette expérience employait le même montage que celui décrit précédemment. Le signal cible à localiser était un son pur de 8 kHz, de 300 ms de durée incluant des durées d'établissement et d'extinction de 20 ms (cosinus carré). Son niveau était de 60 dB SPL, mesuré au point du centre de la tête des auditeurs. Le masque était un bruit blanc (100 - 16 000 Hz) pouvant prendre des durées de 200, 300, 500, 1 000 et 5 000 ms, incluant chaque fois des durées d'établissement et d'extinction de 20 ms (cosinus carré). Nous ne décrivons ici que l'expérience portant sur un masque de 500 ms. Son niveau était également de 60 dB SPL, mesuré dans les mêmes conditions que pour le son pur. Pour l'ensemble des conditions, les stimuli étaient émis en face des sujets. Le tableau 1 présente l'ensemble des vingt-deux conditions temporelles testées pour le masque de 500 ms. On voit dans ce tableau que le son pur pouvait débiter de 400 ms avant le début du bruit jusqu'à 600 ms après le début du bruit (soit 100 ms après la fin du bruit). Chaque condition était présentée deux fois, le tout dans un ordre aléatoire, les sujets passant deux fois le test (avec des ordres différents), ce qui permettait d'obtenir quatre données par sujet pour chaque condition.

Une condition supplémentaire était proposée systématiquement à la fin des tests, pour ne pas troubler les sujets. Il s'agissait d'une condition introduite dans le but d'étudier l'influence éventuelle de la présence de capteurs précédant la présentation du son pur cible. L'hypothèse motivant cette condition était qu'il était éventuellement possible de réduire l'effet d'attraction de la cible par le masque en "capturant" la cible dans un flux auditif spatialement indépendant [8]. Pour cela, cible et masque étaient présentés simultanément, six capteurs identiques au son pur à localiser étant présentés avant ce dernier (7e du flux de sons purs). Ces sons purs étaient présentés avec des intervalles de silence de 50 ms.

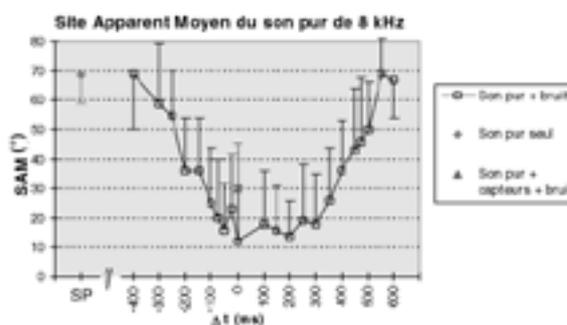
Cinq sujets ont pris part à cette expérience. Ils avaient également participé aux expériences décrites plus haut, ce qui, notamment, permettaient d'avoir en référence leurs données portant sur la localisation du son pur seul dans le PVM. Leur tâche était d'indiquer par l'intermédiaire d'un micro terminal la direction apparente du son pur, en s'aidant éventuellement du graphique déjà présenté figure 3. Il leur était demandé qu'après chaque réponse, ils ne bougent plus en fixant le haut-parleur leur faisant face et ils ferment les yeux pour éviter l'influence de la vision de celui-ci [4].

Retard du son pur sur le bruit (ms)							
-400	-250	-100	-25	150	350	475	600
-300	-200	-75	0	200	400	500	
-200	-150	-50	100	300	450	550	

**Tabl. 1 : Décalage temporel  $\Delta t$  entre les onsets des stimuli. Une valeur négative correspond à une avance du son pur sur le bruit, une valeur positive à une avance du bruit sur le son pur.**

## Résultats

Sur l'ensemble des conditions, 98 % des sons purs ont été localisés avec un site inférieur à  $90^\circ$ , c'est-à-dire qu'extrêmement peu de sons purs ont été perçus derrière les sujets, ce qui était prévisible de par la facilité du son pur de 8 kHz à être attiré par le bruit à large bande. Aussi, dans les très rares cas où les sons purs étaient localisés derrière, nous n'avons pris en compte que le site du signal par rapport au plan horizontal (ex. : localisation à un site de  $160^\circ$  converti en site de  $20^\circ$ ). A partir de ces données, nous avons calculé le site apparent moyen (SAM), avec 20 points de mesure par condition.



**Fig. 4 : Site Apparent Moyen (SAM) du son pur de 8 kHz en fonction du retard de l'onset du son pur sur l'onset du bruit. A gauche figure le SAM du son pur en l'absence de bruit (losange). Le triangle donne le SAM obtenu en présence de capteurs. 5 sujets, 4 présentations par sujet.**

La figure 4 illustre le SAM en fonction du retard de la cible sur le masque ( $\Delta t$ ) obtenu pour l'ensemble des conditions temporelles, ainsi pour que pour le son pur présenté seul et le son pur précédé de capteurs. On peut observer une remarquable courbe en forme de U dont le minimum est obtenu pour une présentation simultanée de la cible et du masque ( $\Delta t = 0$ ) et dont le maximum correspond aux valeurs minimales et maximales de  $\Delta t$ . L'effet du retard sur le SAM est extrêmement fort (analyse de variance,  $F(21, 231) = 39,08$ ,  $p < 0,001$ ), le SAM passant de près de  $70^\circ$  pour les fortes valeurs de  $|\Delta t|$  à  $12,5^\circ$  pour  $\Delta t = 0$ .

Pour la condition avec capteurs (triangle), le SAM de la cible n'augmente que peu ( $30,5^\circ$ ) par rapport à celui obtenu lorsque les capteurs sont absents ( $12,5^\circ$ ). Cette augmentation atteint toutefois le seuil de significativité pour un t-test sur variables dépendantes ( $t = 1,82$ ,  $p = 0,049$ ).

Des t-tests ont été conduits entre la condition pour laquelle le son pur est présenté seul et les conditions  $\Delta t$ . Il ressort que l'intervalle de  $\Delta t$  pour lequel le seuil de confiance à  $p = 0,01$  est atteint est  $[-150, 250 \text{ ms}]$ . Ainsi, on peut considérer que si  $\Delta t$  appartient à cet intervalle, la présence du bruit réduit significativement le site du son pur.

## Discussion et conclusion

Les deux expériences présentées ici ont montré qu'il existe un important effet d'attraction de sons purs cibles dans le plan vertical médian dans la direction d'incidence de bruits à large bande. D'autres expériences menées dans le cadre de cette investigation ont montré que cet effet existait également, dans une certaine mesure, lorsque la cible était un son pur et le masque n'était pas un bruit à large bande, mais aussi lorsque cible et masque étaient des bruits à large bande.

En ce qui concerne les sons purs, l'effet d'attraction semble d'autant plus fort que la fréquence du son pur est plus élevée. Lorsque la fréquence est de 8 kHz, l'effet est particulièrement fort lorsqu'il y a coïncidence des onsets de la cible et du masque. Lorsqu'un décalage temporel entre ces onsets est introduit, l'effet diminue jusqu'à complète disparition lorsque cible et masque ne se chevauchent plus dans le temps. Enfin, la présence de capteurs réduit modestement mais significativement l'attraction de la cible par le masque.

Les limites temporelles de validité de l'effet d'attraction (-150, 250 ms) permettent d'éliminer l'interprétation de ce phénomène par l'effet d'antériorité, qui n'agit plus au-delà d'un décalage temporel de 70 ms entre les signaux [7]. De même, le domaine temporel d'action du masquage temporel est très différent de celui de l'effet d'attraction rencontré ici. En effet, celui-ci est important surtout entre -20 ms (masquage rétroactif) et 100 ms (masquage proactif [6]). En fait, les valeurs de validité temporelle de l'effet d'attraction font davantage penser à des processus d'ordre cognitif, ou tout au moins à des traitements opérés par les centres supérieurs du système auditif. Il paraît alors légitime d'inscrire cet effet dans le domaine de la perception des flux auditifs [9].

Au regard des résultats apportés dans cet étude, on peut proposer l'hypothèse suivante : la détermination de la direction apparente du couple de stimuli (cible, masque) se fait par une analyse globale de la mixture sonore qu'ils forment. Lorsque la cible est un son pur, celui-ci n'apporte qu'une raie dans la mixture, tout se passe comme s'il était intégré au masque (large bande). Cette intégration est de même ordre que la capture d'un son dans un flux auditif par la règle du regroupement fréquentiel. Dès lors que cible et masque ne forment qu'un seul flux, ils sont localisés dans la même direction, celle-ci étant imposée par le masque du fait qu'il fournit plus d'informations au système auditif (informations spectrales du fait de son contenu large bande).

Cette hypothèse soulève la question de la non indépendance des traitements relevant de l'identification et de la localisation des sources sonores. En effet, le système auditif, contrairement au système visuel, ne dispose pas de représentation topographique de l'espace. De par son organisation tonotopique, il ne peut accéder à une représentation complète de l'espace que par une analyse spectrale des stimuli (les indices spectraux dans le plan vertical médian). Or, lorsque plusieurs sources sonores sont présentes et forment une mixture, l'identification des sources est elle aussi basée sur une analyse spectrale de la mixture. Cette analyse obéit à des lois de groupements

fréquentiel et temporel, décrites par la Psychologie de la Gestalt [10, et appliquées par Bregman [11] à l'audition. On peut donc penser que la dualité identification/localisation dans l'analyse spectrale peut se traduire par une interaction des sources sonores dans l'espace, comme l'ont déjà proposé Bregman & Steiger[8].

Ainsi, alors que les mécanismes de la localisation d'une source unique dans l'espace peuvent être expliqués par une analyse physique des signaux acoustiques présents aux tympans des auditeurs, ceux impliqués dans la localisation de sources multiples sont beaucoup plus complexes. Ils semblent faire intervenir des processus relevant de la perception de flux auditifs et mettent en évidence les limitations du système auditif à avoir un accès direct à une représentation de l'espace. Actuellement, il peut donc apparaître utopique de vouloir modéliser la fonction de localisation en élévation du système auditif, tant elle semble intimement liée à la fonction d'identification des sources sonores. Les modèles (et les systèmes de simulation de l'environnement dérivés) devront donc se baser sur une approche globale de l'analyse de scènes auditives, ne dissociant pas l'identité et la position des sources sonores dans l'espace.

La psychoacoustique étudie aujourd'hui des problématiques de plus en plus complexes, plus représentatives des environnements sonores réels. La répartition spatiale des sources sonores constitutives de ces environnements a un effet non négligeable sur les dimensions subjectives généralement étudiées (perception des environnements sonores urbains, qualité et confort acoustique dans les automobiles, les télécommunications...) et les démarches mises en place se doivent de les prendre en compte. Comme on le voit, une simple prise en compte de l'espace par des modèles pondérant par exemple telle ou telle source sonore selon sa position sera insuffisante, du fait de la non indépendance des traitements auditifs mis en œuvre. De nombreux efforts devront encore être fournis afin de mieux connaître les liens entre les différentes dimensions subjectives relevant de la perception auditive et de l'audition spatiale.

## Références bibliographiques

- [1] Roffler SK, Butler RA, Localization of tonal stimuli in the vertical plane, *J. Acoust. Soc. Am.*, 43, PP 1260-1266, 1968.
- [2] Blauert J, Sound localization in the median plane, *Acustica*, 22, PP 205-213, 1969-1970.
- [3] Scharf B, Critical bands in Foundations of modern auditory theory, (Vol. 1), Tobias JV Ed., Academic Press, New York, USA, 1970.
- [4] Jack C E, Thurlow W R, Effects of degree visual association and angle displacement on the "ventriloquism" effect, *Perceptual and Motor Skill*, 37, PP 967-979, 1973.
- [5] Château N, Localisation de sources sonores multiples dans l'hémisphère supérieur, Thèse de l'Université d'Aix-Marseille-II, France, 1996.
- [6] Fastl H, Temporal masking effects : I. Broad band noise masker, *Acustica*, 35, PP 287-302, 1977.
- [7] Blauert J, Spatial hearing, The MIT Press, London, England, 1983.
- [8] Bregman AS, Steiger H, Auditory streaming and vertical localization : Interdependence of "what" and "where" decisions in audition, *Perception & Psychophysics*, 28, PP 539-546, 1980.
- [9] Bregman, AS, Auditory scene analysis. The perceptual organization of sound, The MIT Press, Cambridge, Massachusetts, USA, 1990.
- [10] Koffka K, Principles of Gestalt Psychology, Hartcourt, Brace and World, New York, USA, 1935.
- [11] Guillaume P, La psychologie de la Forme, Nouvelle Bibliothèque Scientifique, Flammarion Ed., Paris, France, 1937.

## Remerciements

Les auteurs remercient le ministère de l'environnement pour le financement de cette étude. ■