Evaluation de l'ergonomie d'un dispositif audio 3D pour l'assistance à la navigation dans un environnement virtuel complexe

Gonot Antoine [1,2], Château Noël[3], Emerit Marc[4]

[1] *France Télécom R&D, 22300 Lannion, France, courriel : antoine.gonot@rd.francetelecom.com*
[2] *CNAM, Laboratoire CEDRIC 75003 Paris, France*
[3] *France Télécom R&D, 22300 Lannion, France, courriel : noel.chateau@francetelecom.com*
[4] *France Télécom R&D, 22300 Lannion, France, courriel : marc.emerit@francetelecom.com*

Résumé

Cette étude, porte, dans ses grandes lignes, sur l'utilisation de la spatialisation sonore dans une interface Humain Machine. En particulier, il s'agit d'étudier dans quelle mesure une telle technologie peut enrichir l'expérience d'un l'utilisateur dans le cadre des services proposés par France Télécom. Le contexte retenu, ici, est l'assistance à la navigation dans environnement de jeu, sur ordinateur individuel. L'expérience présentée prend donc la forme d'un jeu de navigation dans une ville virtuelle simplifiée où l'utilisateur doit s'orienter en mettant à profit les informations sonores qui lui sont fournies. Les facteurs expérimentaux concernent, d'une part, le rendu sonore (Stéréo vs. Binaural) et, d'autre part, la représentation spatiale de l'information de position d'une cible ({Direction = coordonnées polaire de la cible} vs. {Direction = Direction du chemin le plus court vers la cible}). L'hypothèse générale qui sous-tend cette expérience est que, plus les informations sonores seront précises, riches et cohérentes avec l'environnement exploré, plus l'ergonomie d'utilisation du système sera bonne, quantifiée, par exemple, par des temps d'arrêt aux intersections plus courts, ou encore par une localisation plus efficace des sources sonores. Bien qu'étonnamment, le trajet parcouru ne soit pas significativement plus court d'une condition à l'autre, ni d'un essai à l'autre, l'analyse des autres résultats tend tout de même à confirmer cette hypothèse.

Introduction

L'objet de cette étude est de déterminer dans quelle mesure l'apport de la spatialisation dans un dispositif de représentation sonore d'information (Auditory Display) peut améliorer l'ergonomie de celui-ci. Or, cette problématique ne nécessite pas uniquement l'étude des mécanismes de la perception auditive. En effet, bien qu'il soit indispensable que les sons utilisés soient perceptibles, il importe avant tout que les informations diffusées par le biais de la modalité auditive permettent d'assister l'utilisateur pour l'accomplissement d'une tâche dans les meilleures conditions. Cela implique, par exemple, de minimiser le temps nécessaire à la prise de décision ou encore de faciliter l'action qui s'ensuit par un feedback adéquat.

De nombreuses études [1] [2] tendent à montrer que le système auditif humain n'accorde pas la priorité absolue à l'indice spatial. En effet, les différences spatiales jouent plutôt un rôle de facilitation, amplifiant la ségrégation basée sur d'autres facteurs, comme l'asynchronisme, les différences de fréquence ou de timbre. D'ailleurs, dans une interface, la spatialisation sonore est principalement utilisée pour aumgenter la sensation d'immersion et de présence (e.g. environnement virtuel [3]), ou encore pour améliorer l'intelligibité lorque plusieurs sources sont présentées de façon concurrente [4] (effet „Cocktail Party"). Il existe, malgré tout, des tâches pour lesquelles la spatialisation sonore présente un intérêt par elle-même, i.e. lorsque l'information est véhiculée par les indices spatiaux seulement. La navigation au moyen d'un dispositif sonore est un bon exemple, puisque les données intiales présentent déjà une dimension spatiale.

G. Kramer [5] a introduit le terme Beacons („balise sonore"), pour décrire des sons servant de référence pour l'exploration et l'analyse d'un ensemble de données sonifiées. Ces balises sont, en quelque sorte, l'équivalent sonore des axes d'un graphe. Les données considérées n'impliquent donc pas nécessairement de dimension spatiale. La navigation auditive, quant à elle, a bien été l'objet de cetaines études (voir [6] pour un récapitualitif de ces travaux), mais elles ont principalement examiné l'effet du type de balise sur la navigation (e.g. parole, impulsion de sonar, bruit large bande, son pur, etc.). Elles ne soulèvent pas explicitement le problème de la représentation d'information par des indices spatiaux et se focalisent plutôt sur des aspects perceptifs que sémantiques. De ce point de vue, l'étude de Walker et Lindsay [7] sur le rayon de capture (i.e. distance à partir de laquelle une cible intermédiaire est estimée atteinte par le système) est plus pertinente, mais elle se limite à l'information de distance, qui plus est, exprimée par une variation de tempo (compteur Geiger).

L'étude présentée ici, s'est donc donnée les objectifs suivants:

- Utiliser les indices naturels de la localisation auditive, non une représentation abstraite, tel que le tempo pour la distance. L'objectif est bien d'étudier particulièrement l'utilisation de la spatialisation sonore pour la représentation d'information, non l'utilisation du son pour la représentation d'informations spatiales.

- Adresser le problème de la relation sémantique entre le concept représenté (données relatives à un emplacement dans une environnement, e.g. une rue dans une ville) et le signe auditif représentant (les indices de la localisation, e.g. distance et azimuth d'une source sonore).
- Evaluer un sytème d'assistance à la navigation dans un environnement virtuel, avec une approche plus globale de la tâche, c'est-à-dire :
  - étudier le cas d'un environnement complexe, présentant une strucuture identifiable, non pas une salle vide.
  - étudier le cas d'une tâche complexe, à savoir, trouver son chemin vers la cible, et pas simplement se déplacer directement d'un point à un autre (prise en compte des contraintes sur le déplacement, imposées par l'environnement).

Dans ce qui suit, sera présenté le design expérimental, tentant de satisfaire ces objectifs, et, plus particulièrement, les variables manipulées Les observables seront ensuite présentés et les résultats discutés, pour conclure, enfin, de façon plus générale sur l'apport du son 3D et du mode de représentation.

L'environnement de test

L'expérience prend la forme d'un jeu de navigation dans un espace visuel simplifié où l'utilisateur doit s'orienter en mettant à profit les informations sonores qui lui sont fournies. Le but du jeu est de trouver le plus rapidement possible neuf sources sonores écologiques ("fanfare", "église", "travaux", etc.), cachées dans certaines rues d'une ville, en navigant (à l'aide des flèches du clavier) avec une vue à la première personne. Ces sources, réparties dans trois zones disjointes de couleur différente (rouge, bleue et verte), sont audibles à tout moment, mais recherchées les unes après les autres. Le point de départ pour la première source se trouve au centre de l'environnement et le sujet part toujours de la position de la source précédente pour trouver la suivante.

Les facteurs expérimentaux

Les variables indépendantes, manipulées concernent, d'une part, le rendu sonore, i.e. la projection de l'espace sonore sur le dispositif physique et, d'autre part, la relation sémantique entre le concept représenté (emplacement dans une ville) et le signe auditif représentant (azimut et distance d'une source sonore).

Le rendu sonore

Il est nécessaire de déterminer une référence pour évaluer la contribution de la spatialisation dans l'interface. La stéréo est choisie tout naturellement, car c'est une technique très courante qui offre les indices essentiels pour la localisation auditive : les différences interaurales de niveau et de phase (IID, ITD). Le mode de rendu stéréo est obtenu grâce à la modélisation d'un couple AB-ORTF, c'est-à-dire, deux microphones cardioïdes espacés de 17 cm, avec un angle de 110° entre les capsules.

Dans la plupart des cas, l'étude de la navigation auditive se fait en situation de mobilité et, sauf quelques rares exceptions, le dispositif d'écoute est un casque. La méthode retenue pour la spatialisation sonore est donc la technique binaurale qui est censée reproduire avec précision le champ sonore induit au niveau des oreilles de l'auditeur, grâce à l'emploi de fonctions de transfert relatives à la tête (HRTF). Celles-ci ne sont pas individualisées, mais ont été validées par des tests subjectifs.

Relation entre les données et les indices sonores

Les informations présentées à l'utilisateur pour l'assister dans sa tâche sont la distance et la direction d'un emplacement dans l'environnement. La direction est simplement représentée par l'azimut d'une source sonore et la distance par son niveau, selon une loi exponentielle classique. Or, d'après H. Hu et D-L Lee [8], l'information de distance et de direction peut avoir différentes significations (sémantique), selon l'application. Concernant la distance, par exemple, certains adopteraient la distance euclidienne, d'autre la longueur du chemin dans un réseau de rue, ou encore l'énergie consommée durant le trajet. Ainsi, au regard de la tâche à accomplir ici, deux relations sémantiques s'opposent, l'une prenant en considération les contraintes du contexte environnemental (configuration de la ville), l'autre pas. Cela définit deux modes de représentation de l'information, l'un dit "contextualisé" et l'autre "décontextualisé".

- *Représentation « décontextualisée »* : la distance et l'azimut sont les coordonnées polaires de l'emplacement.
- *Représentation « contextualisée »* : la distance est donnée par la longueur du chemin jusqu'à l'emplacement et l'azimut est donné par le premier noeud (virage ou intersection) de ce chemin.

Procédure

Quarante sujets ont été recrutés pour l'expérience. Ils ont été répartis en quatre groupes de dix, chacun testant l'une des quatre conditions expérimentales issues des facteurs expérimentaux précédents : "BinCont" (binaural contextualisée), "BinDecont" (binaural décontextualsiée), "SteCont" (stéréo contextualsiée) et "SteCont" (stéréo décontextualsiée).

L'ordre dans lequel les neuf sources sont recherchées est différent pour chacun des dix sujets d'une condition. Afin, de minimiser l'influence de l'ordonnancement lors de l'analyse des résultats, une séquence est construite à partir d'un modèle qui attribue à une zone un nombre différent d'entités recherchées successivement : une source dans la première zone, deux dans la deuxième et trois dans la troisième. La liste exhaustive des transitions entre zone respectant cette règle est présentée dans le tableau 1.

Sujet	Séquence de zones	Sujet	Séquence de zones
1	**3** 1 2 1 2 1	6	**1** 2 3 1 2 1
2	**2** 1 2 1 3 1	7	**1** 2 1 2 1 3
3	**2** 1 3 1 2 1	8	**1** 2 1 2 3 1
4	**1** 3 1 2 1 2	9	**1** 2 1 3 1 2
5	**1** 3 2 1 2 1	10	**1** 2 1 3 2 1

Tableau 1: Ordre d'exploration des zones

Après avoir lu les consignes, le sujet suit une phase d'entraînement dans un environnement simplifié, ne contenant que quelques rues, sans zone, où une unique source sonore est présente. Une fois que l'expérimentateur s'est assuré de la bonne compréhension des consignes (il est rappelé que c'est un jeu de rapidité) et de la bonne prise en main de l'interface, le sujet commence le jeu. Lorsque la dernière source sonore a été trouvée, la phase d'évaluation débute immédiatement. Le sujet doit, tout d'abord, replacer les sources recherchées sur une carte de l'environnement, puis évaluer sa propre charge de travail durant la tâche de navigation et, enfin, répondre à douze questions de ressenti. L'intégralité de l'expérience (navigation et évaluation) est réalisée 3 fois. Dans ce qui suit, seuls seront présentés les résultats concernant l'interaction.

Résultats

Une analyse de variance a été réalisée sur les variables dépendantes objectives : la distance parcourue normalisée par la distance optimale, le temps passé à une intersection et la distribution des angles d'écoute. Chaque groupe présente, pour un essai, quatre-vingt-dix réalisations de chaque variable (9 sources x 10 sujets). Les données objectives correspondent aux observables issues des historiques des interactions. Il s'agit des données de position, d'orientation et de temps, dont ont été extraits les indices utiles à l'analyse, présentés ici.

Distance normalisée

Le premier indice est la distance parcourue pour trouver la source, normalisée par la distance optimale.

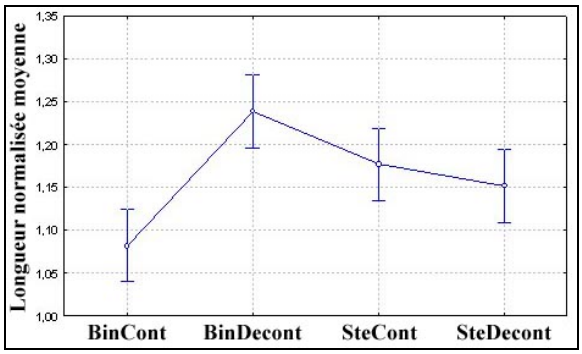

Figure 1**: Effet de la condition sur la longueur normalisée du chemin parcouru**

L'ANOVA révèle un effet hautement significatif de la condition $F(3,356) = 3,0817$, $p < 0,001$. La figure 2 indique de meilleures performances pour la condition "BinCont" (binaural contextualisé). Cependant, la longueur normalisée est proche de 1 quelle que soit la condition. Il est donc délicat de parler d'effet de la condition. Quant à l'apprentissage, l'effet n'est pas significatif, $F(2,72) = 1,9106$, $p = 0,14875$.

Temps passé à une intersection

L'analyse de variance révèle encore un effet hautement significatif de la condition sur le temps passé à une intersection (temps nécessaire à la prise de décision), $F(3,356) = 19,054$, $p<0,001$.

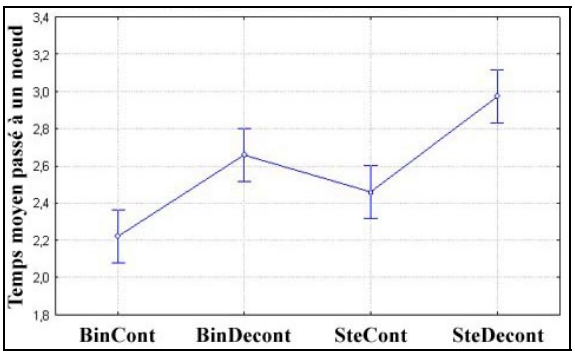

Figure 2**: Effet de la condition sur le temps moyen passé à un nœud (intersection)**

Cette fois-ci, l'effet de la condition semble cohérent avec ce qui était attendu. La figure 3 montre que pour un rendu sonore identique, la contextualisation de la représentation permet une prise de décision plus rapide. De la même façon, pour une représentation de l'information identique, le rendu binaural offre de meilleures performances. En outre, l'effet d'apprentissage est cette fois-ci hautement significatif, $F(2,712) = 279,95$, $p < 0,001$.

Distribution des angles d'écoute

Afin d'analyser le comportement de rotation, dû à une localisation dynamique des sources sonore, le nombre d'arrêts dans une direction donnée est calculé à chaque intersection. La figure 4 illustre les distributions des angles d'écoute (l'angle reporté ici est l'azimut de la source recherchée), pour une partition en seize secteurs angulaires, comparant les conditions contextualisée et décontextualisée.

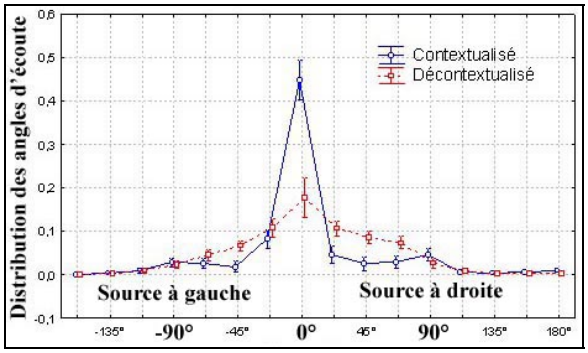

Figure 3**: Effet croisé de l'azimut et de la condition sur la fréquence des angles d'écoutes**

L'ANOVA indique un effet croisé de l'azimut et de la condition hautement significatif, $F(15,5370) = 149,31$, $p < 0,001$. On remarque tout d'abord que l'azimut frontal (0°) est celui le plus fréquemment utilisé pour la localisation, quelle que soit la condition. Cependant, on observe une différence dans la forme de la distribution. Pour la condition contextualisé, il y a une émergence particulière des azimuts latérales (+/- 90°). La distribution de la condition décontextualisé, quant à elle, ressemble plutôt à une gaussienne. Il n'y a pas d'émergence des azimuts latéraux. Une analyse de variances pour chacun des azimuts de l'intervalle [-67,5°;67,5°], indique des différences hautement significatives, ce qui tend à valider ces différences de distribution.

Discussion

Deux tâches principales peuvent être distinguées dans ce test. Une tâche globale de navigation, qui consiste à trouver la source et à se repérer dans l'environnement, et une tâche locale d'orientation, sous-jacente à la précédente, qui consiste à prendre la bonne direction à chaque intersection.

La distance parcourue est un bon indice de l'ergonomie de l'interface pour la tâche globale. Or, il a été observé que les performances étaient relativement bonnes quelle que soit la condition. Il semblerait donc que la navigation ait été relativement aisée. Des différences plus significatives auraient sans doute pu être observées si les erreurs d'orientation avaient été plus pénalisantes (e.g. avec une configuration de la ville plus complexe ou avec des distances plus importantes). Malgré tout, la consigne étant explicitement de trouver la source sonore le plus vite possible, il semble que les sujets aient plutôt cherchés à diminuer le temps de prise de décision qu'à diminuer la distance parcourue. L'absence de progression avec les essais pour ce dernier observable tend à confirmer cette hypothèse.

Ainsi, ces résultats indiquent que la représentation contextualisée, comme le rendu binaural, apportent un avantage certain pour la tâche locale d'orientation, mais pas vraiment pour la tâche globale de navigation. En effet, la contextualisation enlève l'ambiguité quant à la direction à prendre, puisqu'une source sonore est toujours dans la direction d'une rue. Cela implique naturellement des temps de prise de décision plus court, par rapport à une situation où la source est derrière un obstacle. De même, le rendu binaural, en augmentant l'efficacité de la localisation auditive, facilite cette décision. Ces résultats sont confirmés par le fait que les rotations nécessaires pour la localisation dynamique des sources sont moins importantes. Une analyse de variance sur la quantité de mouvement (ou "distance angulaire parcourue") à chaque intersection indique une supériorité hautement significative du binaural (1,45 rad.) sur la stéréo (1,77 rad.), $F(3,358) = 18{,}536$, $p < 0{,}001$, et de la contextualisation (1,46 rad.) sur la décontextualisation (1,75 rad.), $F(3,358) = 15{,}970$, $p < 0{,}001$.

L'observation de la distribution des angles d'écoute précise un peu plus les raisons pour lesquelles la localisation est plus efficace pour la représentation contextualisée. En effet, un comportement typique consiste à tendre l'oreille vers la source (azimut +/-90°) en maximisant les différences interaurales, puis à tourner la tête face à la source (azimut 0°), là où l'acuité est la plus grande, pour affiner la localisation. Or, pour la représentation contextualisée, mise à part les azimuts +/- 22.5 ° qui semblent correspondre à des erreurs de localisation, les azimuts frontal et latéral sont bien les plus fréquemment utilisés. Le comportement semble conforme à celui qui a été décrit. En revanche la distribution de la représentation décontextualisée indique que les azimuts les plus proches de la position frontale sont les plus utilisés. Bien qu'il soit logique que les erreurs de localisation soient plus nombreuses pour cette condition, puisque le sujet n'a pas d'*a priori* sur l'azimut potentiel de la source sonore, le comportement de localisation parait tout de même être différent du précédent. Nous faisons l'hypothèse suivante que chercherons à vérifier lors de travaux futurs. Il est possible que l'azimut latéral (+/-90°) soit peu ou pas utilisée et que le comportement de localisation ressemble plutôt à une recherche à taton de l'azimut frontal (0°).

Conclusion

Cette étude a partiellement validé l'hypothèse selon laquelle, plus les informations sonores étaient précises, riches et cohérentes avec l'environnement exploré, plus l'ergonomie d'utilisation du système est bonne. En effet, la contextualisation de la représentation et le rendu binaural, garants tout deux de ces qualités, apportent un réel avantage sur la tâche locale d'orientation, mais en apporte peu sur la tâche globale de navigation. La spatialisation sonore a plutôt joué un rôle de facilitation et c'est sur les aspects perceptifs et non sémantique que la contextualisation s'est montrée la plus pertinente.

Références

[1] Bregman, A.S., (1994), *L'analyse de scène auditives : l'audition dans des environnements complexes*. Penser les sons, Psychologie et sciences de la pensée, Presses Universitaires de France, Chap. 2, pp 11-39.

[2] Deutsch, D., Roll, P.H., (1976), *Separate "What" and "Where" Decision Mechanisms In Processing a Dichotic Tonal Sequence*. Journal of Experimental Psychology: Human Perception and Performance, Vol. 2, No. 1, pp 23-29.

[3] Larsson, P., Vätfjäll, D., Kleiner, M., (2001), *Ecological Acoustics and the multimodal perception of rooms: Real and Unreal experiences of auditory-visual virtual environments.* Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, (July 29 – August 1), pp 245-259.

[4] Arons, B., (1992), *A Review of The Cocktail Party Effect*. Journal of American Voice I/O Society, pp 35-50.

[5] Kramer, G. (1994), *Some Organizing Principle for Representing Data with Sound.* Auditory Display: Sonification , Audification and Auditory Interface, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, USA, pp 202-208.

[6] Walker, B. N., Lindsay, J. (2003), *Effect of Beacon Sounds on Navigation Performance in a Virtual Reality Environment*. Proceedings of the 2003 International Conference on Auditory Display, Boston, MA (July 6-9), pp 204-207.

[7] Walker, B. N., Lindsay, J., (2004), *Auditory navigation Performance is Affected by Waypoint Capture Radius.* Proceedings of the 2004 International Conference on Auditory Display, Sydney, Australia, (6-9 July).

[8] Hu H., Lee D-L, (2004), *Semantic location Modeling for Location Navigation in mobile environment*. Proc. 5th IEEE International Conference on Mobile Data Management, pp 52-61.