

Interactions non-visuelles pour faciliter le rendu d'une séquence d'informations

Damien Appert

Université de Toulouse & CNRS
IRIT
Toulouse, France
damien.appert@irit.fr

Emmanuel Dubois

Université de Toulouse & CNRS
IRIT
Toulouse, France
Emmanuel.Dubois@irit.fr

Christophe Jouffrais

Université de Toulouse & CNRS
IRIT
Toulouse, France
Christophe.Jouffrais@irit.fr

Résumé

Cet article présente la conception et l'évaluation de techniques d'interaction multimodale non-visuelle pour la perception et la mémorisation d'une séquence d'informations. Cette tâche est représentative de multiples contextes dont notamment la restitution d'itinéraire représenté sous la forme d'une séquence de couple direction-distance. Parmi une vingtaine de techniques que nous avons conçues, quatre ont fait l'objet d'une évaluation utilisateur. Les résultats de cette

étude établissent l'intérêt de deux techniques en termes de perception, de mémorisation et de rappel des informations et une troisième s'avère plus résistante à une tâche d'interférence.

Mots clés choisis par les auteurs

Interactions non-visuelles, multimodalité, mémorisation

Mot clés de la classification ACM

H.5.2 [Information Interfaces and Presentation]: User Interfaces—auditory feedback, evaluation, haptic I/O, voice I/O

Introduction

Dans des situations où la perception visuelle est fortement contrainte ou déficiente, il est nécessaire de rendre perceptible l'information par une autre modalité sensorielle. Dans ces travaux, nous nous intéressons plus particulièrement à la transmission non-visuelle de couples de valeurs de type propriété-mesure, qui peuvent avoir un rôle informationnel, de prévention ou de guidage. Cette étude préliminaire s'inspire du cas particulier de la transmission d'une feuille de route pour piéton non-voyant, générée à partir d'applications facilement disponibles et gratuites mais non dédiées aux déficients visuels (ex. Google itinéraire). La tâche générique consiste alors en la transmission sous forme non-visuelle d'une séquence de couples direction-distance (voir Figure 1).

Dans la majorité des cas, la lecture de la feuille de route s'effectue au moyen d'un lecteur d'écran. Celui-ci permet de parcourir les informations de façon séquentielle et de les restituer à l'utilisateur sous forme non-visuelle : il s'agit le plus fréquemment d'une forme audio, au moyen d'une synthèse vocale, ou d'une forme tactile, à l'aide

- 1 ➡ Tourner à droite
90 m
- 2 ↶ Tourner à gauche
50 m
- 3 ➡ Tourner à droite
6 m

Figure 1: Exemple d'itinéraire composé de trois couples direction-distance : droite-90 ; gauche-50 ; droite-6. Exemple généré à partir d'applications comme Mappy® ou Google Maps®.

© ACM, 2015. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Actes de la 27ème conférence francophone sur l'Interaction Homme-Machine, 2015.
<http://dx.doi.org/10.1145/2820619.2825013>

Modalité direction	Modalité distance
Audio Parole Sonification Tactile Tacton Mouvement saccadé Angle	Audio Parole Durée d'un son Tactile Angle Longueur
Présentation de l'information	
Repère Égocentré (G,D,F) Allocentré (O,N,E)	Plage de valeur Petite (<10) Grande (1à100)
Multimodalité : Fusion ; Séquentiel 1^{ère} information (séquentiel) : Direction-distance ; Distance-direction	

Tableau 1: Synthèse des options envisagées lors des pré-tests. Une modalité de direction + une modalité de distance + choix de la présentation des informations = une technique

d'une plage braille. Cependant, la mise en œuvre de ces solutions alternatives demeure confrontée aux limites de l'utilisateur en termes de mémorisation : un parcours est en effet constitué d'une séquence de couples de valeurs de plus de sept plus ou moins deux mnèmes, limite bien identifiée dans la littérature [10]. Or, les travaux de Dubois [2] ont établi l'intérêt de la multimodalité sur la mémoire de travail dans le cadre de la perception de données de différentes natures. Nous proposons donc d'explorer l'apport des techniques d'interaction multimodale non-visuelle pour optimiser la perception et la mémorisation d'une séquence de couples de données de type propriété-mesure.

A ces fins, nous avons conçu, dans cette étude préliminaire, différentes techniques d'interaction non-visuelle et multimodale avec l'objectif de favoriser la mémorisation de ce type d'information en répartissant la quantité d'information sur deux modalités utilisées de manière complémentaire. Cet article présente les techniques conçues et synthétise les résultats préliminaires obtenus suite à l'évaluation de ces différentes techniques.

État de l'art

Compenser un déficit de perception visuelle conduit à utiliser les modalités sensorielles auditive ou tactile pour transmettre les informations. Par exemple, il est possible pour un utilisateur d'exprimer une mesure avec l'angle de son coude [3] ou de percevoir une propriété comme une direction au travers d'une ceinture vibrotactile [7]. Il est aussi possible de concevoir des techniques d'interaction basées sur la sonification d'information [1,6] pour transmettre des informations de distance et de direction [15]. L'intensité ou la fréquence d'un son peuvent également être utilisées pour encoder et transmettre une mesure [9,16]. D'autres attributs du son ont également été exploités dans la littérature comme véhicules d'une information à transmettre en particulier à l'aide d'un *panning* (balance des haut-parleurs gauche et droit), d'une réverbération ou de la spatialisation du son [4,8,11]. Certaines études ont aussi cherché à combiner ces différentes techniques pour transmettre un couple de valeurs [12], notamment un

couple direction-distance dans des contextes informationnels où l'information n'a pas besoin d'être mémorisée car elle doit être utilisée immédiatement, comme c'est le cas pour du guidage [14] ou une alerte [5].

Ces travaux issus de la littérature s'inscrivent donc dans un usage immédiat des informations perçues. Notre situation d'étude dépasse ce contexte : le piéton doit se souvenir de l'itinéraire à parcourir pour pouvoir, par exemple, choisir le plus optimal, limiter son stress face à un itinéraire inconnu ou transmettre l'itinéraire à une autre personne (elle aussi déficiente visuelle). Nos travaux visent donc à explorer les apports et les limitations de la multimodalité pour la perception et la mémorisation à moyen terme d'une séquence d'informations de type propriété-mesure.

Pré-études et choix de design

Conception des techniques

Dans le cadre d'une étude comparative, nous avons tout d'abord conçu une technique contrôle, monomodale, basée uniquement sur la synthèse vocale. Celle-ci correspond au rendu des informations grâce au lecteur d'écran et à la synthèse vocale. A partir d'une séance de brainstorming, impliquant des spécialistes en IHM, cognition spatiale et technologie de suppléances pour déficient visuel (DV), puis à la suite de plusieurs séances de réflexion aux solutions techniques, nous avons aussi conçu des techniques multimodales basées sur les modalités tactiles et sonores. Au cours de nombreuses itérations, nous avons testé informellement et avec trois sujets maximum, plusieurs techniques d'interaction multimodale, différant notamment par le rendu de l'information et les associations de modalité. Le tableau 1 synthétise les différentes options explorées et testées.

Les pré-tests nous ont amené à préciser les valeurs à transmettre : trois directions dans un repère égocentré (gauche, droite et en face) et la distance sur une plage de valeurs allant de 1 à 100. A l'issue de ces pré-tests, outre la technique contrôle monomodale, trois techniques multimodales ont finalement été retenues pour transmettre une séquence de couples de valeurs

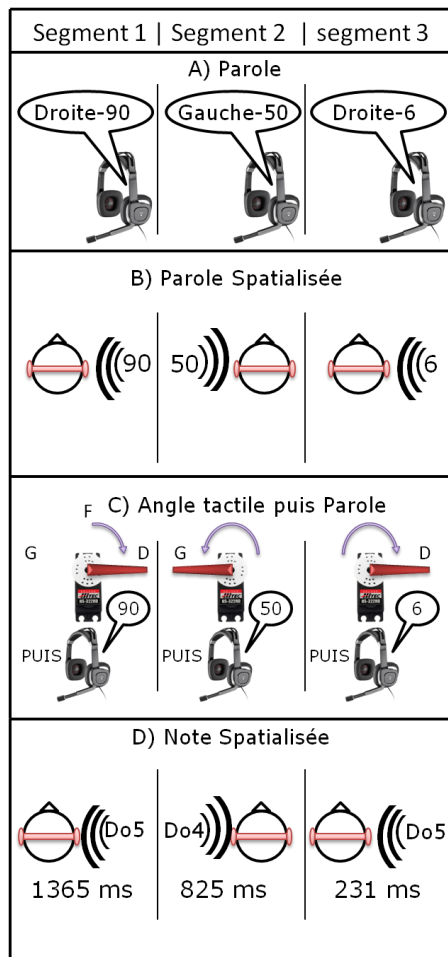


Figure 2 : Illustration d'un même itinéraire de 3 segments, exprimé au travers des quatre techniques d'interaction ; A : TextToSpeech (TTS) ; B : TextToSpeech Latéralisée (TTS-Lat) ; C : PalmGauge-TextToSpeech (PG-TTS) ; D : Mélodie.

direction-distance. Nous décrivons ces quatre techniques ci-après.

Description des techniques retenues

La technique contrôle est la technique TextToSpeech (TTS). Elle repose sur une double synthèse de parole indiquant successivement la direction (parole) puis la distance (parole) pour chaque segment d'un parcours (Figure 2A).

La technique TextToSpeech Latéralisé (TTS-Lat) repose sur une modalité synthèse de parole et une modalité spatialisation. Seule la distance est vocalisée mais son rendu est spatialisé pour indiquer la direction : la source sonore se trouve virtuellement à gauche, en face ou à droite. TTS-Lat est une technique multimodale avec fusion d'informations : la direction et la distance sont deux informations fusionnées en une parole spatialisée (Figure 2B).

La technique PalmGauge-TextToSpeech (PG-TTS) repose sur la rotation d'un curseur placé sous la main qui se positionne à gauche (0°), en face (90°) ou à droite (180°). Puis, la distance est vocalisée (Figure 2C). PG-TTS est une technique multimodale séquentielle.

La technique Mélodie repose sur une note dont la spatialisation, la hauteur et le timbre correspondent de manière redondante à la direction et, dont la durée correspond à la distance (Figure 2D). L'enchaînement des notes correspond alors à une mélodie qui traduit une séquence de couples direction-distance. Mélodie est une technique multimodale avec fusion de données: la direction et la distance sont deux informations fusionnées en une note spatialisée. Elle a été conçue afin de permettre aux utilisateurs de « chanter la mélodie » pour s'en rappeler.

Hypothèses de travail

Nous avons émis trois ensembles d'hypothèses basés sur les raisonnements suivants : TTS (condition contrôle) surcharge l'utilisateur d'une succession d'informations verbales et devrait être la moins utilisable. TTS-Lat, qui est la forme optimisée de TTS, devrait être plus utilisable. PG-TTS et Mélodie sont des techniques

reposant sur deux principes différents permettant de mémoriser plus facilement les couples de valeurs. PG-TTS sépare les informations sur les modalités tactiles et vocales. Mélodie est conçue pour faciliter la mémorisation et le rappel en chantant. Ces deux techniques devraient donc être les plus utilisables dans ce contexte. Afin d'évaluer ces techniques selon ces ensembles d'hypothèses, nous avons conduit une expérimentation.

Protocole expérimental

Les sujets étaient assis sur une chaise, les yeux bandés. Pour chaque sujet, l'expérimentation consistait en 4 blocs, chacun correspondant à une technique, dont l'ordre a été contrebalancé.

Un bloc commençait par quatre essais sur des itinéraires de familiarisation. L'évaluation comportait ensuite deux séries de quatre itinéraires. Les itinéraires de la première série étaient composés de cinq segments (soit 5 couples de direction et distance choisis de façon pseudo-aléatoire) ; les itinéraires de la seconde en contenaient sept. Après chaque itinéraire, le participant devait restituer oralement l'itinéraire perçu (l'expérimentateur le consignait par écrit). Cette tâche de restitution comportait deux rappels (R1 et R2) séparés par une tâche d'interférence [13] (dix décrétements 3 par 3 à partir d'un nombre aléatoire) visant à établir l'apport de la technique en termes de mémorisation.

À l'issue d'un bloc, les sujets répondaient au questionnaire NASA-TLX qui permet d'évaluer la charge cognitive associée à la technique utilisée. Les deux séries étaient séparées par une courte pause (30 s). Les blocs étaient séparés par une pause plus longue (180 s).

Une fois tous les blocs passés, les sujets devaient ordonner les quatre techniques par ordre de préférence. La durée moyenne de l'expérimentation par sujet était de deux heures.

Cette expérimentation suit donc une conception intra-sujet 4x2x2 avec les techniques d'interaction, la taille (5 ou 7) et l'ordre du rappel comme facteurs. Compte tenu des ensembles d'hypothèses énoncés, nous avons analysé la satisfaction des utilisateurs vis-à-vis des

Préférence : moyennes et écarts-types

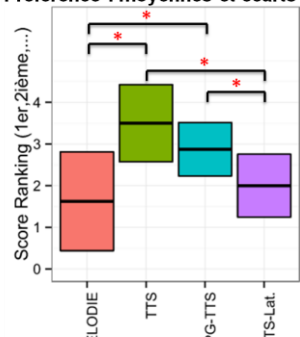


Figure 3 : Ordre moyen de préférence des techniques des sujets Non-DV; Mélodie et TTS-Lat. sont préférées à TTS et PG-TTS ; * : $p < 0,05$ (test post-hoc : Wilcoxon signed rank test)

Robustesse : moyennes et écarts-types

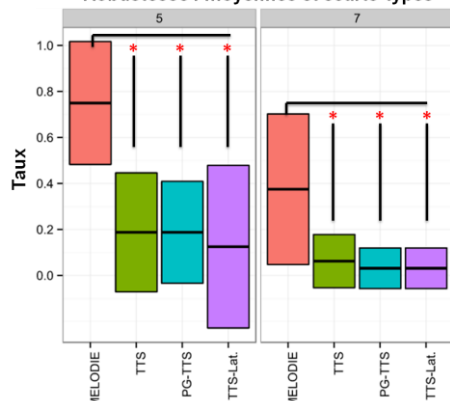


Figure 4 : Robustesse des techniques chez les non-DV (axe-X), selon deux tailles, avec le taux d'itinéraires identiques entre les deux restitutions (axe-Y). * : $p < 0,05$; test post-hoc : Wilcoxon signed rank test

techniques ainsi que la *robustesse* de la mémorisation et l'*efficacité* de chacune des techniques.

La *robustesse* a été mesurée selon le nombre d'itinéraires restitués de la même façon lors d'un premier rappel puis lors d'un second rappel différé. La robustesse était évaluée pour les directions seules, les distances seules puis pour les couples.

L'*efficacité* a été mesurée selon le nombre d'itinéraires restitués correctement lors du premier rappel. L'efficacité a été évaluée pour les directions seules, les distances seules puis pour les couples.

Onze participants ont pris part à l'évaluation : 8 sujets Non Déficiants Visuels (Non-DV, âge : $26,7 \pm 1,7$) et 3 sujets Déficiants Visuels (DV, âge : $35,6 \pm 11,5$). Nous relatons dans un premier temps les résultats relatifs à la population des Non-DV, puis conduisons une analyse succincte des résultats pour les DV.

Analyse des résultats des sujets Non-DV

Lors des comparaisons, nous avons utilisé les tests non-paramétriques de Friedman ou Wilcoxon avec un seuil de significativité à 0,05. Les graphiques représentent des moyennes (fig. 3 à 7) et des écarts-types (fig. 3 à 5).

Il ressort tout d'abord que lors de la première phase de restitution (R1) le nombre d'éléments corrects par itinéraire est en moyenne de 7,68 pour un nombre d'informations égal à 10 pour la taille 5 et à 14 pour la taille 7. La moyenne du nombre de réponses correctes du premier rappel pour chacune des conditions techniques x tailles, est de 6,88 pour la plus faible et de 8,53 pour la plus haute. C'est un résultat cohérent avec la littérature [10] (empan mnésique = 7 ± 2).

Satisfaction. Les techniques préférées des Non-DV sont Mélodie et TTS-Lat (voir Figure 3 ; Friedman rank sum test $X^2 = 10,35$, $df = 3$, $p < 0,02$). Aucune différence significative n'a été établie concernant le score obtenu avec le NASA-TLX (Friedman rank sum test : $X^2 = 4,95$, $df = 3$, $p > 0,1$).

Robustesse. La robustesse (taux) de la Mélodie était de 0,90 pour la direction seule et de 0,70 pour la distance

seule quelle que soit la taille de l'itinéraire. Pour les couples, la robustesse de la Mélodie mesurée en taille 5 (resp. taille 7) atteint 0,75 (resp. 0,37) et était significativement supérieure à celle mesurée pour les autres techniques. En effet, les autres techniques ne dépassaient pas 0,18 en taille 5 et 0,07 en taille 7 (Taille 5 : Friedman rank sum test : $X^2 = 11,7$, $df = 3$, $p < 0,01$; Taille 7 : $X^2 = 12,9$, $df = 3$, $p < 0,01$; voir Figure 4 pour le post-hoc).

Efficacité. Seuls les itinéraires de taille 5 font apparaître des différences notables en termes d'efficacité (taux). Ainsi, en termes d'efficacité pour les couples, la TTS-Lat. (0,44) est significativement supérieure à celle des techniques Mélodie (0,03) et TTS (0,06). De même, l'efficacité pour les couples de la technique PG-TTS (0,31) est significativement supérieure à celle mesurée pour Mélodie (0,03) et TTS (0,06) (Taille 5 : Friedman rank sum test : $X^2 = 12,7$, $df = 3$, $p < 0,01$; voir Figure 5 pour le post-hoc).

Concernant l'efficacité mesurée pour la direction seule, elle est excellente avec une moyenne pour toutes les techniques de 0,71, et en particulier 0,81 pour la

Efficacité : moyennes et écarts-types

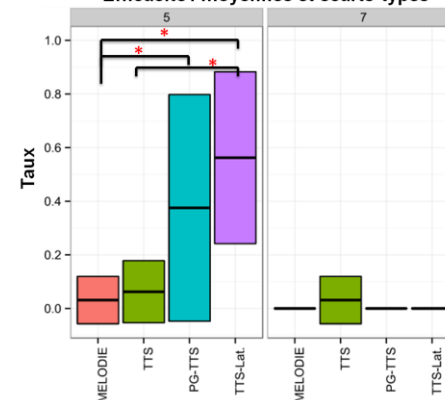


Figure 5 : Efficacité des techniques chez les non-DV (axe-X), selon deux tailles, avec le taux d'itinéraires corrects en première restitution (axes-Y) ; * : $p < 0,05$; test post-hoc : Wilcoxon signed rank test.

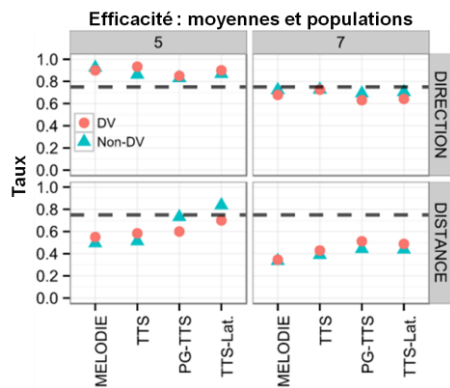


Figure 6 : Efficacité des techniques (axe-X), selon deux tailles pour les deux populations, avec le taux de direction (partie haute) et de distance (partie basse) correctes en première restitution (axes-Y). Les lignes en pointillées indiquent le seuil de 0,75 d'efficacité.

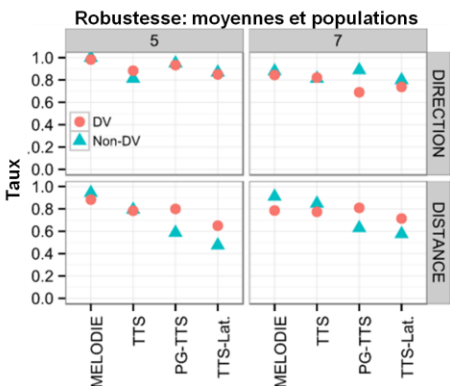


Figure 7 : Robustesse des techniques (axe-X), selon deux tailles pour les deux populations, avec le taux de direction (partie haute) et de distance (partie basse) identique entre les deux restitutions (axes-Y).

mélodie et 0,72 pour la TTS-Lat. Notons en revanche, que l'efficacité mesurée pour la distance seule est très faible pour la Mélodie (<0,04).

Au travers de l'analyse globale de l'efficacité, une tendance singulière semble se dégager : si pour une taille d'itinéraire et une technique données, l'efficacité pour la direction et la distance dépassent le taux de 0,75, alors l'efficacité pour les couples atteindra 0,40 (cf. Figure 6).

Analyse des résultats préliminaires concernant les sujets DV

Nous avons complété cette première analyse par une analyse spécifiquement liée aux participants DV. Compte tenu de l'échantillon très réduit, cette analyse est plus succincte que la précédente et ne peut faire appel à des tests statistiques.

Il ressort ainsi que lors de la première phase de restitution (R1) le nombre d'éléments corrects par itinéraire est en moyenne de 7,65 (min: 7,17 ; max: 8,08). Comme pour la population des non-DV ce résultat est cohérent avec la limite bien connue de 7 ± 2 .

Nous avons également observé le nombre moyen de directions et de distances correctes de la première restitution, pour chaque technique et selon les deux tailles et les deux populations. Il apparaît que les résultats des DV sont très similaires à ceux des Non-DV (points et triangles se chevauchant, voir Figure 6).

Il en va de même pour la robustesse (voir Figure 7), avec toutefois des différences (graphiques) plus marquées avec PG-TTS, notamment pour les directions et les distances.

Certaines formes de multimodalité pourraient donc faire apparaître une différence plus marquée chez les DV que chez les Non-DV.

Discussion & Conclusion

Nous avons conçu et évalué quatre techniques d'interaction non-visuelles dont trois étaient multimodales. Ces techniques ont été conçues afin d'évaluer leur impact sur la mémorisation et le rappel

d'une séquence d'informations représentative, par exemple, d'un itinéraire pour piétons.

Nous avons étudié les résultats d'un groupe de sujets Non-DV et un groupe de sujets DV. Nos résultats ne montrent pas de différences (graphiques) entre les deux populations (Non-DV/DV) à part pour PG-TTS. Il est intéressant de noter que la limite théorique de l'empan mnésique (7 ± 2) aurait dû rendre impossible la mémorisation de tous les couples de valeurs pour des itinéraires courts (10 éléments), mais surtout pour des itinéraires longs (14 éléments). Néanmoins certaines techniques ont obtenue une efficacité correcte, ce qui d'après les résultats précédents de la littérature, semblent particulièrement prometteur. Par conséquent, cette étude établie que les techniques TTS-Lat. et PG-TTS semblent contribuer à optimiser la capacité de perception et de mémorisation d'une séquence de couples de valeurs.

Par ailleurs, la Mélodie a montré en taille 5, la meilleure efficacité pour les directions seules (0,81) et la pire pour les distances seules (<0,04). Néanmoins sa robustesse pour ces mêmes itinéraires et les couples, dépasse le taux de 0,75. Il est possible qu'un mécanisme d'autosuggestion ait été utilisé par les participants pour restituer des distances (fausses) et que ce mécanisme optimise la mémorisation des valeurs. Compte tenu des retours utilisateurs et du fait qu'elle est la technique préférée des participants, nous pensons toutefois que c'est la technique elle-même qui joue ce rôle en permettant de chanter l'information. Mélodie bénéficiait aussi d'un triple encodage de la direction : la redondance dans la multimodalité pourrait donc jouer un rôle dans la perception et la mémorisation des données. Une expérimentation complémentaire est en cours pour étudier ces deux facteurs.

Notre technique de référence, la TTS ne s'est pas particulièrement démarquée. Cette technique était la moins préférée (avec PG-TTS), la moins efficace (avec Mélodie) et n'était pas robuste.

Pour compléter cette base de trois techniques d'interaction multimodale et non-visuelle pour la

perception et la mémorisation de couples de valeurs, des améliorations sont requises afin d'accroître la robustesse de TTS-Lat et PG-TTS d'une part et d'autre part d'améliorer l'efficacité des trois techniques, notamment celle de la Mélodie. Cette étude a également permis de mettre en évidence un biais potentiel du protocole expérimental : il apparaît en effet nécessaire, d'adapter le mode de restitution en fonction de la technique d'interaction car la mélodie requiert une transformation (mélodie vers texte) susceptible d'affecter la restitution. De même pour une analyse plus réaliste des résultats, la notion de « distance correctement restituée » nécessite d'être affinée. Enfin, cette étude se limite à évaluer des techniques dans un cas d'utilisation stricte qui ne permet pas de généraliser sur l'efficacité des techniques *in situ*.

Nous avons également envisagé d'autres contextes basés sur l'utilisation de couples propriété-mesure, comme par exemple en géographie (région-superficie ou ville-habitant) pour lesquels de telles techniques pourraient s'avérer pertinentes et dans lesquels il conviendra d'étudier la réplicabilité des résultats observés.

Références

- [1] Barrass, S. and Kramer, G. Using sonification. *Multimedia Systems - Special issue on audio and multimedia* 7, 1 (1999), 23–31.
- [2] Dubois, V., Gyselinck, V., Choplin, H. Multimodalité et mémoire de travail. *Environnements informatiques pour l'apprentissage humain*, (2003), 187– 198.
- [3] Enokibori, Y. and Mase, K. Human joint angle estimation with an e-textile sensor. *Proceedings of the 2014 ACM International Symposium on Wearable Computers - ISWC '14*, ACM Press (2014), 129–130.
- [4] Fontana, F. and Rocchesso, D. Auditory distance perception in an acoustic pipe. *ACM Transactions on Applied Perception (TAP)* 5, 3 (2008), 16:1–16:15.
- [5] Fujimoto, E. and Turk, M. Non-Visual Navigation Using Combined Audio Music and Haptic Cues. *ICMI '14*, ACM Press, 411–418.
- [6] Hermann, T., Hunt, A., Neuhoff, J.G., and Vickers, P. Chap 18-Sonification for Process Monitoring. In *The Sonification Handbook*. 2011, 455–491.
- [7] Heuten, W., Henze, N., Boll, S., and Pielot, M. Tactile wayfinder: a non-visual support system for wayfinding. *NordiCHI '08*, ACM Press, 172–181.
- [8] Larsen, C.H., Lauritsen, D.S., Larsen, J.J., Pilgaard, M., and Madsen, J.B. Differences in human audio localization performance between a HRTF- and a non-HRTF audio system. *AM '13*, (2013), 5:1–5:8.
- [9] Mershon, D.H. and King, L.E. Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics* 18, 6 (1975), 409–415.
- [10] Miller, G.A. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* 63, 2 (1956), 81–97.
- [11] Mills, A.W. On the Minimum Audible Angle. *The Journal of the Acoustical Society of America* 30, 4 (1958), 237–246.
- [12] Nigay, L. and Coutaz, J. A design space for multimodal systems: concurrent processing and data fusion. *Proceedings of the INTERACT '93 and CHI '93*, ACM Press (1993), 172–178.
- [13] Peterson, L. and Peterson, M.J. Short-term retention of individual verbal items. *Journal of Experimental Psychology* 58, 3 (1959), 193–198.
- [14] Rümelin, S., Rukzio, E., and Hardy, R. NaviRadar: a novel tactile information display for pedestrian navigation. *UIST '11*, ACM Press, 293–302.
- [15] Yamano, S., Hamajo, T., Takahashi, S., and Higuchi, K. EyeSound: Single-modal Mobile Navigation Using Directionally Annotated Music. *AH'12*, 22:1–22:4.
- [16] Zahorik, P., Brungart, D.S., and Bronkhorst, A.W. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica* 91, 3 (2005), 409–420.