

Processus et paramètres de conception de la sonification

Sylvain Daudé

CLIPS-IMAG Equipe IIHM
BP 53 38041 Grenoble cedex 9, France
sylvain.daudé@imag.fr

Laurence Nigay¹

Computing Science Dept, University of Glasgow
17 Lilibank Gardens, Glasgow G128RZ Scotland
laurence@dcs.gla.ac.uk

RESUME

Cet article décrit un cadre d'étude pour la sonification des interfaces en sortie. Les étapes du processus de sonification et leurs paramètres sont présentés en soulignant les transformations successives des données à rendre perceptibles de façon sonore. L'utilité du cadre d'étude à la fois pour la classification de systèmes existants et pour la conception d'interfaces sonores est enfin soulignée.

MOTS CLES : Sonification, métaphore sonore, taxonomie, processus de conception.

ABSTRACT

In this paper we present a unified framework for the design of audible interfaces. We describe the steps of the sonification process and their parameters. The process is modeled as a sequence of transformation functions from the data to present to the produced sounds. The usefulness of the framework for classifying existing audible interfaces and for designing new ones is then discussed.

KEYWORDS : Sonification, Sound Metaphor, Taxonomy, Design Process.

INTRODUCTION

Lors de la conception d'une interface sonore en sortie, une première étape vise à définir les données qui seront rendues de façon sonore. La deuxième étape consiste alors à la sonification ou traduction de ces données dans le domaine sonore [1]. Dans cet article, nous jetons un regard nouveau sur le processus de sonification encore mal cerné, en nous basant sur le processus de visualisation de données de Ed Chi [5]. Par l'identification d'étapes, caractérisées par des paramètres de conception pour la sonification, notre objectif est double : classer les interfaces sonores existantes et fournir une aide au concepteur d'interfaces sonores.

Dans cet article, nous présentons le processus de sonification et ses paramètres en soulignant les transformations successives des données. Nous mettons ensuite en exergue les deux points contributifs de ces travaux : classification et conception d'interfaces sonores. Des perspectives à ces travaux concluent l'article.

PROCESSUS DE SONIFICATION

Comme le montre la Figure 1, quatre états et trois étapes de transformation décrivent le processus de visualisation proposé par Ed Chi [5]. Nous argumentons dans cet article que ce processus en quatre états s'applique à la sonification. La Figure 2 présente notre version du processus de visualisation adaptée à la sonification.

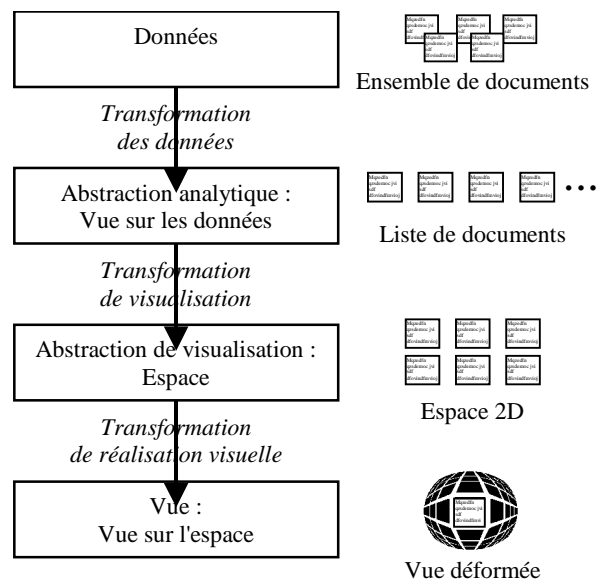


Figure 1 : Processus de visualisation d'Ed Chi [5].
Illustration issue de [15].

Des Données à la Vue sur les Données : Transformation des Données (F1)

Le processus de sonification de la Figure 2 et celui de visualisation de la Figure 1 partagent les deux premiers états et la transformation F1 pour passer de l'un à l'autre.

¹ En année sabbatique, de l'Université de Grenoble 1, laboratoire CLIPS-IMAG, équipe IIHM

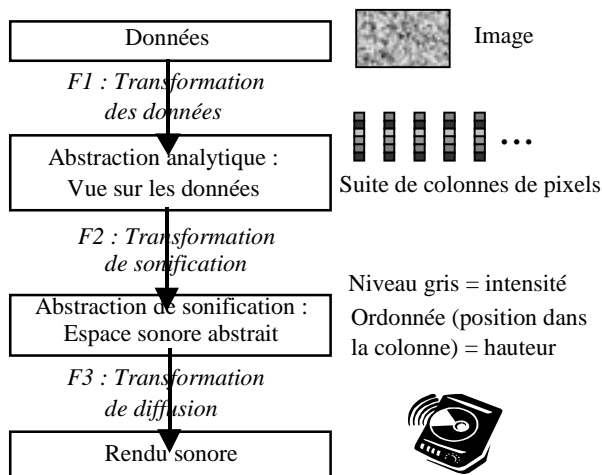


Figure 2 : Notre processus de sonification. Illustration inspirée de [11].

Nous réutilisons donc des caractéristiques connues de visualisation comme le fait que les données sont (i) nominales, ordonnées et quantitatives, ou encore (ii) spatiales (1D, 2D, 3D), temporelles, hiérarchiques, relationnelles et non-structurées (1D, 2D, 3D, nD) [15].

La transformation F1 des données est dépendante de la sémantique des données manipulées et de la tâche de l'utilisateur final. Par exemple, le système AROMA [12] brouille le signal sonore provenant de discussions d'utilisateurs distants (sémantique de la donnée) tout en conservant suffisamment d'informations pour identifier le locuteur (tâche de l'utilisateur), comme le ferait une vitre translucide dans le domaine visuel. La vitre translucide est un exemple d'analogie de la transformation des données F1 avec un phénomène du monde réel.

De la Vue sur des Données à l'Espace Sonore Abstrait : Transformation de Sonification (F2)

Dans le processus de visualisation de la Figure 1, la vue sur les données est transformée pour obtenir un espace euclidien qui sera ensuite représenté à l'écran par un ensemble d'éléments selon des traits graphiques comme la couleur, la taille et la forme. Lors du processus de sonification présenté à la Figure 2, la transformation F2 définit, à partir de la vue sur les données, un espace euclidien ayant pour origine la position de l'utilisateur. Pour définir la présentation d'une donnée, F2 lui associe une durée qui définit la fenêtre temporelle pendant laquelle des traits acoustiques comme le timbre, la hauteur, l'intensité et la position peuvent varier. Les paramètres de F2 sont donc la durée et des traits acoustiques variant dans le temps, ce dernier étant délimité par la durée. Comme les traits graphiques, certains traits acoustiques sont catégoriques comme des schémas de timbre (piano, violon) ou de rythme (code morse), tandis que d'autres sont ordonnés comme la hauteur et l'intensité [1]. Par exemple, une fonction F2 de correspondance univoque, présentée à la Figure 2, consiste à présenter une image

parcourue colonne par colonne [11] en traduisant le niveau de gris et l'ordonnée de chaque pixel (deux traits graphiques ordonnés) respectivement en intensité et en hauteur sonore (deux traits acoustiques ordonnés).

Une Donnée à Représenter

La fonction de sonification F2 met en correspondance la valeur de la donnée avec un ou plusieurs traits acoustiques. Cette correspondance se caractérise selon qu'elle est arbitraire ou non et analogique ou non avec le monde réel [2]. Une correspondance non arbitraire repose sur l'existence d'un système sémantique appris ; ainsi, selon Bernsen [2] les correspondances linguistiques sont non arbitraires, comme les *earcons* [3] où une donnée est traduite par une séquence de sons organisés selon une grammaire. Toutefois l'utilisation d'une grammaire inconnue a priori pose des problèmes de compréhension au-delà de six sons symboliques [1]. A l'inverse, une correspondance directe univoque (valeur, trait) comme pour l'analyse exploratoire de données [1] (exemple de la Figure 2) est une correspondance arbitraire. Orthogonalement, F2 peut générer des traits acoustiques analogiques à un son du monde réel, comme les sons « écologiques ». Ces derniers imitent des sources sonores naturelles, comme des vagues ou une pluie et peuvent s'insérer naturellement dans l'environnement auditif de l'utilisateur [6]. Ce cas correspond aux *Auditory Icons* utilisés dans le *Sonic Finder* [9] où des sons de frottement ou de percussion sont utilisés. Lorsque la fonction F2 est fondée sur une analogie, elle repose fréquemment sur un modèle physique de source sonore [6] ; la donnée à représenter est alors un stimulus qui excite la source sonore. Il faut insister sur le fait que le concepteur peut utiliser une fonction F2 analogique mais adopter une correspondance arbitraire avec la donnée à représenter. Par exemple, il peut décider de représenter l'état de transfert d'un fichier par le grincement d'une porte.

Lorsque les sons sont spatialisés, leur répartition peut être analogique avec un phénomène spatial du monde réel. Celui-ci peut être sonore, comme l'environnement acoustique d'une cathédrale. Pour illustrer le cas d'une analogie avec un phénomène spatial non sonore, nous considérons le système *Diary in the Sky* [16]. Celui-ci diffuse une séquence de rendez-vous sous forme de sons spatialisés ; l'annonce d'un rendez-vous est associée à une direction, devant l'utilisateur pour midi, à droite pour trois heures etc. Ainsi la répartition spatiale des données est calquée sur celle d'une horloge entourant l'utilisateur.

Lorsque la vue sur les données dispose d'un référentiel, comme des bornes minimale et maximale ou une valeur moyenne, il peut être utile que F2 traduise, outre chaque donnée, ce référentiel. Nous avons identifié ce besoin pour l'exploration de données [8]. Par exemple dans [7] nous trouvons une barre de progression sonifiée par un son d'orgue dont la hauteur se rapproche de celle d'un son de guitare « référence » joué régulièrement.

Ensemble des Données à Représenter

Du point de vue temporel, les données de l'ensemble sont présentées simultanément ou alternativement. Dans le deuxième cas, la concaténation des sons rend plus aisée la perception du fil conducteur des données, tandis que leur séparation temporelle (silence entre deux données) permet une meilleure identification de chaque donnée. Lorsque F2 inclut de plus une spatialisation du son (exposée au paragraphe précédent), les données peuvent être diffusées à partir d'un même endroit de l'espace ou d'endroits distincts. La répartition des données dans l'espace permet une meilleure différenciation et une meilleure mémorisation des données [16] ainsi qu'un sentiment d'immersion, mais peut créer une rupture d'unité sémantique de l'ensemble des données. La Figure 3 présente quatre systèmes se distinguant par la façon dont ils présentent l'ensemble des données.

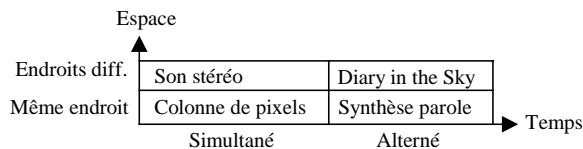


Figure 3 : Classification de quatre systèmes selon la présentation spatiale et temporelle des données.

Dans l'exemple de la Figure 2 [11], tous les sons résultants de la traduction des pixels d'une colonne sont joués en même temps, à partir d'une même direction de l'espace. A l'opposé, un son stéréo répartit deux moitiés d'un son à partir d'endroits différents de l'espace. Le système *Diary in the Sky* diffuse quant à lui les rendez-vous les uns après les autres, à partir de positions différentes de l'espace. Enfin, la synthèse de parole, mais aussi les *auditory icons* du *SonicFinder* [9] ou les *earcons* [3], diffusent les données séquentiellement à partir d'une même direction de l'espace.

De l'Espace Sonore Abstrait au Rendu Sonore : Transformation de Diffusion

La diffusion vise à exprimer l'espace sonore et les traits acoustiques sous forme de signal sonore diffusé par un dispositif physique. Nous caractérisons le dispositif par six paramètres : (1) la précision sonore par rapport aux données à diffuser, (2) le degré de spatialité (mono, stéréo, multicanaux 2D ou multicanaux 3D), (3) la répartition géométrique des haut-parleurs, (4) le degré de partage, c'est-à-dire le nombre d'utilisateurs pouvant percevoir l'information en même temps, (5) la liberté de déplacement qu'il permet à l'utilisateur et (6) le degré d'immersion, traduisant si le dispositif isole l'auditeur du monde réel. Ainsi, le *Soundbeam Neckset* [14] est caractérisé par une bonne précision pour la voix (1) et une diffusion stéréo (2) à travers deux haut-parleurs directionnels situés sur les épaules de l'utilisateur (3). Il est utilisé par un seul utilisateur à la fois (4), permet une liberté de déplacement totale car il est portable (5) ; enfin son degré d'immersion est faible (6).

La diffusion s'accompagne d'un traitement du son dépendant du dispositif physique, par exemple de son degré de spatialité. Ainsi, diffuser un son spatialisé sur un haut-parleur revient à projeter ce son sur une droite (car la notion de distance est conservée). Une autre caractéristique du dispositif exploitée par la transformation de diffusion est la répartition géométrique des haut-parleurs : la connaissance de la position de deux haut-parleurs permet de limiter leurs interférences par une méthode de dé-corrélation et ainsi produire une spatialisation plus précise [10].

Outre des facteurs liés au dispositif physique, la diffusion peut prendre en compte un modèle de l'auditeur ; par exemple la spatialisation par écouteurs utilise souvent un traitement par filtre *Head-Related Transfer Function*, qui vise à compenser le filtrage fréquentiel du conduit auditif. Elle peut aussi utiliser un modèle de l'environnement acoustique de l'utilisateur, comme les filtres *Binaural Room Impulse-Response* qui traitent le signal pour donner à l'utilisateur muni d'écouteurs l'illusion que les sons proviennent de la pièce où il se trouve.

CONTRIBUTION Classification

L'identification des étapes du processus de sonification permet tout d'abord de différencier les systèmes de sonification existants. Ainsi les quatre classes de sonification habituellement distinguées, les fonctions de correspondance univoques non analogiques, les *earcons* [3], les *auditory icons* [9] et la synthèse de parole, se différencient par la fonction F2. Les *auditory icons* et la synthèse de parole génèrent des sons analogiques au monde réel, à l'inverse des *earcons* et des fonctions univoques ; en outre, contrairement aux *auditory icons* et aux fonctions univoques non analogiques, la synthèse de parole et les *earcons* utilisent une grammaire.

Par ailleurs des systèmes qui auparavant étaient classés dans une même catégorie peuvent désormais être distingués. Ainsi les sonneries de téléphones portables, les *earcons* de Blattner [3] et les « *earcons* parallèles » de Brewster [4] étaient classées dans une grande catégorie « *earcons* ». Pour la sonnerie, les données initiales sont une liste de hauteurs et de durées pré-enregistrées et F1 et F2 se réduisent à la fonction identité. Dans les autres exemples, F1 définit et organise les données à sonifier, mais F2 diffère du point de vue temporel : les sons sont présentés alternativement pour les *earcons* de Blattner, simultanément pour les *earcons* parallèles de Brewster.

Conception

L'apport de ce cadre d'étude de la sonification est multiple pour la conception. Tout d'abord, l'indépendance fonctionnelle des transformations aide le concepteur à différencier les niveaux d'abstraction du processus et permet une plus grande flexibilité ; ainsi la vue sur les données peut être simultanément sonifiée et visualisée, ou une même transformation de rendu peut traiter plu-

sieurs espaces sonores abstraits. De plus à chaque étape du processus, nous avons identifié des paramètres qui incitent le concepteur à se poser les bonnes questions au bon niveau d'abstraction. Enfin ces paramètres peuvent servir de fondement pour définir des heuristiques de conception. Prenons l'exemple du suivi d'activités d'arrière-plan ; les sons correspondants doivent si possible être courts et s'intégrer dans l'environnement acoustique de l'utilisateur. La première contrainte peut inciter à sonifier les différents champs de la donnée en même temps, comme dans le cas des *earcons* parallèles [4] tandis que la deuxième peut impliquer le choix de sons écologiques, comme le système SOFA [6]. Si les étapes du processus sont indépendantes fonctionnellement, la configuration d'une étape peut néanmoins induire des choix de conception dans les autres étapes ; par exemple si le dispositif sonore est monophonique, il est inutile d'utiliser une fonction de sonification F2 spatialisée. De même la fiabilité des données initiales a un impact sur les étapes de transformation suivantes. Par exemple, dans un système de localisation d'une balise dans l'océan [1], le canal d'acquisition de la position de la balise étant peu fiable, il est préférable de ne pas utiliser une sonification émettant continûment un signal sonore jusqu'à l'acquisition d'une nouvelle donnée de localisation.

PERSPECTIVES

Nous identifions donc de nombreuses perspectives à ces travaux, comme la prise en compte d'autres paramètres au sein du processus et la mise en place d'heuristiques de conception. Nous proposons aussi deux extensions qui concourent à ne pas considérer ce processus isolément.

Vers des Sonifications Sensitives au Contexte

La prise en compte du contexte suppose que le processus de sonification soit paramétré par des données du contexte comme la tâche courante entreprise par l'utilisateur ou l'environnement physique. Ces données peuvent influencer chaque étape du processus ; par exemple si la tâche liée à la sonification devient secondaire, F1 peut être modifiée pour que la vue sur les données ne contienne que les changements des valeurs des données et non leurs contenus, afin de libérer l'attention de l'utilisateur. L'environnement physique peut aussi avoir un impact sur la transformation de sonification F2 uniquement : ainsi Ronkainen [13] cite la possibilité de changer automatiquement la sonnerie d'un téléphone portable afin qu'elle soit en harmonie avec l'environnement sonore courant de l'utilisateur. Enfin la transformation de diffusion F3 peut être modifiée, par exemple dans le cas d'un utilisateur mobile se déplaçant d'une pièce à une autre, chaque pièce étant équipée d'un dispositif différent.

Sonifications Simultanées

Nous avons limité cette étude à la sonification d'un seul ensemble ou source de données ; la prise en compte de plusieurs ensembles soulève des problèmes de conflit entre les sonifications, comme la confusion entre deux

espaces sonores sonifiés simultanément. Cette perspective constitue un de nos prochains axes de recherche.

BIBLIOGRAPHIE

1. Barrass, S. et Kramer, G. *Using sonification*. *Multi-media Systems*, Vol. 7, No 1, 1999, pp. 23-31.
2. Bernsen, O. *Foundations of multimodal representations. A taxonomy of representational modalities*. *Interacting with Computers* 6(4), 1994, pp. 347-371.
3. Blattner, M., Sumikawa, D. et Greenberg R. *Earcons and Icons : Their Structure and Common Design Principles*. In *Proceedings of HCI 1989*, pp. 11-44.
4. Brewster, S.A., Wright, P.C. et Edwards, A.D.N. *Parallel Earcons : Reducing the Length of Audio Message*. *International Journal of Man-Machine Studies*, Vol. 43, No 2, 1995, pp. 153-175.
5. Chi, E. et Riedl, J. *An Operator Interaction Framework for Visualisation Systems*. In *Proceedings of Info Vis 1998*, pp. 63-70.
6. Conversy, S. *Conception d'Icônes Auditives Paramétrées pour les Interfaces Homme-Machine*. Thèse de Doctorat, Université de Paris-Sud, 2000.
7. Crease, M. et Brewster, S.A. *Making progress with sounds - the design and evaluation of an audio progress bar*. In *Proceedings of ICAD 1998*.
8. Daassi, C., Fauvet, M.C. et Nigay, L. *Multiple Visual Representation of Temporal Data*. In *Proceedings of DEXA 2002*, à paraître.
9. Gaver, W. *The SonicFinder : An interface that uses auditory icons*. Actes de la conf. HCI'89, pp. 67-94.
10. Kraemer, A. *Two speakers are better than 5.1*. *IEEE Spectrum*, mai 2001, pp. 70-74.
11. Meijer, P.B.L. *An Experimental System for Auditory Image Representations*. *IEEE Transactions on Biomedical Engineering*, Vol. 39 No 2, 1992, p.134.
12. Pedersen, E. et Sokoler, T. *AROMA: abstract representation of presence supporting mutual awareness*. In *Proceedings of CHI 1997*, pp. 51-58.
13. Ronkainen, S. *Earcons In Motion - Defining Language for an Intelligent Mobile Device*. In *Proceedings of ICAD 2001*, pp. 146-151.
14. Sawhney, N. et Schmandt, C. *Design of Spatialized Audio in Nomadic Environments*. In *Proceedings of ICAD 1997*, pp. 129-135.
15. Vernier F. *La multimodalité en sortie et son application à la visualisation de grandes quantités d'information*. Thèse de doctorat, Université Joseph Fourier, 2001.
16. Walker, A., Brewster, S.A., McGookin D. et Ng A., *Diary in the sky: a spatial audio display for a mobile calendar*. In *Proceedings of CHI 2001*, pp. 531-540.