

CFA '18 LE HAVRE ■ 23-27 avril 2018
14^{ème} Congrès Français d'Acoustique



Une mesure de l'apport du son 3D dans un jeu vidéo sur mobile

J. Moreira^a, L. Gros^a, R. Nicol^a, I. Viaud-Delmon^b, S. Natkin^c et C. Le Prado^c

^aOrange Lab, 2 Avenue Pierre Marzin, 22307 Lannion, France

^bIrcam, 1 place Igor Stravinsky, 75004 Paris, France

^cCnam, 2 rue Conté, 75003 Paris, France

julian.moreira.fr@gmail.com

Nous présentons un protocole expérimental visant à mesurer l'apport du son binaural dans un jeu vidéo sur mobile. Ce protocole répond à deux problématiques distinctes : 1) identifier les types d'apports possibles du binaural, et 2) mettre au point une méthode d'évaluation tenant compte des spécificités du contexte mobile. Nous introduisons trois hypothèses sur la façon dont le binaural peut enrichir l'expérience utilisateur : amélioration du sentiment d'immersion, meilleure mémorisation de la scène virtuelle, et amélioration de la performance (type score de jeu vidéo). Concernant la méthode d'évaluation, nous évoquons d'abord la singularité des contextes mobiles, multiples et imprévisibles. Puis nous proposons une méthode basée sur l'« Experience Sampling Method », qui replace l'expérience dans le contexte d'utilisation quotidien du sujet. Pour mesurer chacun des trois apports, nous proposons de récolter différentes données, séparées entre les données objectives, issues de l'interaction entre l'utilisateur et l'application, et les données subjectives, issues de questionnaires soumis à l'utilisateur. La méthode que nous proposons a également comme ambition d'atténuer l'influence de ces questions subjectives sur l'appréciation que le sujet a de sa propre expérience.

1 Introduction

Les smartphones offrent globalement des performances de plus en plus élevées (CPU, résolution d'écran, appareil photo, etc.). Cela rehausse la qualité vidéo des services média en mobilité, en particulier les applications interactives telles que le jeu vidéo. Cette évolution concernant l'image n'est pas ou peu suivie par l'intégration de systèmes de restitution audio de haute qualité dans ce type de terminal. Or, parallèlement à ces évolutions concernant l'image, des solutions de son spatialisé sur casque, à travers notamment la technique de restitution binaurale basée sur l'utilisation de filtres HRTF (Head Related Transfer Functions) voient le jour [3]. Le binaural constitue alors un candidat idéal à la restitution sonore spatialisée sur mobile : non seulement elle permettrait une immersion accrue, mais aurait également un fort pouvoir isolant des contextes environnants, de par sa restitution sur casque.

Cependant, retransmettre une scène auditive en binaural, plutôt qu'en stéréo ou en mono, entraîne un changement de perception de la scène en question, dont il convient d'étudier les apports potentiels. D'une manière générale, le son spatialisé est souvent évalué en rapport avec la notion de qualité sonore, dont les attributs dépendent essentiellement d'un avis subjectif formulé par l'utilisateur [18, 22]. Vraisemblablement, dans le cadre d'un jeu vidéo, les critères selon lesquels une personne procédera à une évaluation positive sont multiples, qu'ils soient conscients ou non. Ils dépendront non seulement du jeu, de son type, mais aussi du profil du sujet lui-même.

Par ailleurs, une des particularités des terminaux mobiles réside dans l'extrême multiplicité de leurs contextes d'utilisation. Avec une fréquence moyenne de 221 utilisations par jour, mesurée en 2014¹, ils nous accompagnent désormais partout, tout le temps. Une des problématique associée à cette omniprésence est alors la différence d'évaluation qu'un même utilisateur fera de son application selon le contexte, et par conséquent la nécessité de tenir compte des composantes de ce dernier (lieu, heure, humeur du joueur, niveau de mobilité, etc.)

Dans ces conditions, mettre au point une méthode expérimentale pour mesurer l'apport du binaural dans un jeu vidéo se révèle complexe. Nous proposons ici d'étudier les deux aspects qui nous semblent cruciaux pour une telle tâche : premièrement, identifier les apports possibles du binaural dans un jeu vidéo, et deuxièmement, proposer une méthode *in-situ* tenant compte du contexte d'utilisation. Nous prenons l'exemple d'un type de jeu bien précis,

*Infinite Runner*². Fortement associé au format mobile, il est un bon exemple de type de jeu à la prise en main facile, aux mécaniques de jeu pensées pour le tactile, et aux parties rapides, pouvant se jouer n'importe où (voir Figure 1).



FIGURE 1 – *Temple Run*, un *Infinite Runner* typique : un personnage qui avance automatiquement dans un couloir, des obstacles à éviter, des bonus à collecter. Le joueur ne contrôle que la position latérale du personnage, les sauts et les glissades.

Dans une première section, nous revenons sur les apports possibles du binaural. Dans une deuxième section, nous nous intéressons à la notion de contexte, et aux particularités du contexte mobile. Dans une troisième section enfin, nous proposons une méthode d'évaluation répondant aux problématiques que nous avons introduites.

2. voir par exemple, sur le Google Play Store, *Temple Run*, *Jetpack Joyride*, *Canabalt*, *Subway Surfers*, etc.

1. <http://www.tecmark.co.uk/smartphone-usage-data-uk-2014/>

2 L'apport du binaural

La définition générale de la qualité sonore s'attache particulièrement au ressenti de l'utilisateur [18, 22]. Ce ressenti se bâtit sur les propriétés physiques du son [2], desquelles l'expérimentateur s'éloignera plus ou moins en fonction du niveau de complexité de son critère d'évaluation (simple, comme le ressenti d'un temps d'attaque, ou complexe, comme l'évaluation de la qualité immersive du son).

Concernant le son spatialisé, les critères d'évaluation fréquemment utilisés sont la facilité à localiser les éléments d'une scène sonore [9, 23], et la fidélité de la scène sonore par rapport à une scène de référence, ou sa crédibilité, en l'absence de référence (voir par exemple [19]).

Dans notre cas, il ne s'agit pas tant d'évaluer la qualité sonore du binaural en soi, que sa contribution au ressenti général de la scène virtuelle audiovisuelle. De ce fait, replacée dans le contexte du jeu vidéo, une meilleure efficacité à localiser les sources se traduirait par une hausse de performance et une rapidité accrue pour accomplir une tâche. [17] mentionne déjà cet effet du binaural utilisé en réalité virtuelle. À notre connaissance, aucune étude ne s'est intéressée au cas du mobile. Nous émettons l'hypothèse que cette amélioration de performance est aussi valable. Nous déclinons ici la définition de la performance de deux façons : la performance globale, qui est le score obtenu par le sujet à l'issue d'une partie (distance parcourue dans le cas de l'*Infinite Runner*), et la performance « locale », c'est-à-dire la rapidité du sujet à réagir à une sollicitation soudaine et éphémère (un bonus qui apparaît temporairement à côté du personnage par exemple).

Par ailleurs, la notion d'immersion est aussi largement répandue et étudiée dans le domaine du jeu vidéo [5, 7, 17], bien qu'encore assez peu examinée sur mobile (voir toutefois [4, 27]). L'hypothèse est également qu'un environnement au son spatialisé favorisera l'immersion du joueur.

Enfin [7, 17] argumentent que la sonorisation d'une scène virtuelle en faciliterait sa mémorisation, à travers la représentation mentale qu'on se fait d'une scène virtuelle, et se rapproche en cela de l'antique discipline d'art de la mémoire [31]. Celle-ci consiste à concevoir mentalement un lieu géographique (un palais de mémoire) dans lequel on dissémine différents objets, et à parcourir ce lieu pour se remémorer les objets en question. En s'appuyant sur les travaux déjà cités, nous émettons l'hypothèse que la spatialisation d'une scène virtuelle par le son en permettrait une meilleure représentation mentale, et par conséquent une meilleure mémorisation.

Le genre de l'*Infinite Runner* nous apparaît alors comme un support particulièrement pertinent pour tester cette hypothèse, car il s'apparente finalement à une longue promenade sur un chemin sans fin, sur lequel il suffirait de disséminer des objets de mémoire, dont on solliciterait le rappel au joueur en fin de partie.

En résumé, performance, immersion et mémorisation sont les trois composantes identifiées comme étant candidates à améliorer l'expérience par le binaural dans un jeu vidéo mobile. Néanmoins, mesurer ces apports en laboratoire pose le problème de la validité des résultats. En effet, une salle de test en laboratoire n'est qu'un contexte parmi tant d'autres et ne représente pas forcément l'utilisation quotidienne d'un smartphone [13, 15]. Il faut

donc définir la notion de contexte mobile, afin de déterminer les facteurs qui influenceront probablement le plus la qualité d'expérience, c'est-à-dire pour nous l'immersion, la performance et la mémorisation du joueur.

3 Le contexte mobile

[21] définit le contexte de la manière suivante : « *factors that embrace any situational property to describe the user's environment* ». Par ailleurs, selon [6], le contexte est : « *Any information that can be used to characterize the situation of entities (i.e., whether a person, place, or object) that are considered relevant to the interaction between a user and an application, including the user and the application themselves.* » D'après ces définitions, le contexte serait donc constitué de tous les éléments relatifs à l'environnement de l'utilisateur et du service. Dans la seconde définition, qui nous semble plus complète et que nous adoptons, il est intéressant de noter que ces éléments ne peuvent être considérés comme contextuels que dans la mesure où ils jouent un rôle dans le rapport d'interaction entre utilisateur et service. Le contexte est donc par définition dynamique, puisque il évolue en fonction de l'interaction et de la manière dont elle est perçue par l'utilisateur : il évolue d'un utilisateur à un autre, d'un service à un autre, d'une utilisation à une autre, et au sein d'une même utilisation d'un même service, par un même utilisateur.

Utiliser une définition ou une formalisation générique du contexte pour la conception d'une application semble être cependant une tâche ardue [1, 25, 26] : trop de paramètres à gérer, trop de différences entre les paramètres, mais aussi beaucoup de paramètres qui deviennent inutiles d'une situation à l'autre. En pratique, il est de coutume pour les concepteurs d'un service d'envisager les cas d'utilisations qu'on retrouve habituellement avec l'appareil qui le propose (sur un ordinateur dans un espace de travail aux heures de bureau, ou sur une télévision seul chez soi le soir, dans la voiture dans le cas d'un GPS, etc.) Dans notre étude, nous souhaitons nous focaliser sur les terminaux mobiles, pour lesquels le contexte est particulièrement changeant et difficilement prévisible par les concepteurs.

Plusieurs études ont cherché à définir les contextes d'utilisations d'un mobile, qu'on appellera contextes mobiles par la suite. [14], en particulier, décrit le contexte mobile en fonction de différentes composantes :

- composante physique : localisation, propriétés acoustiques du lieu, niveau de lumière ambiante, météo, objets physiques environnants, etc.
- composante temporelle : moment de la journée, de la semaine, du mois, etc. Elle inclut aussi le temps disponible pour effectuer la tâche (important pour le mobile).
- composante de tâche : interruption de tâche, multitâche, interaction de second plan (parler pendant qu'on utilise une application par exemple), charge cognitive requise par l'application.
- composante sociale : autres personnes présentes, leurs rôles, leurs interactions, etc.

- composante technique : reliée aux aspects techniques de l'application, le hardware utilisé, les capteurs utilisés, etc.

Chaque contexte mobile serait donc descriptible à l'aide de ces composantes. Si cette nomenclature nous semble intéressante, chaque composante n'en reste pas moins très riche, contenant beaucoup de facteurs d'influence possibles. Dans notre cas, nous sélectionnerons ceux qui peuvent potentiellement interagir avec l'effet de la présence ou non du son spatialisé.

4 Proposition d'un protocole expérimental

Le protocole que nous proposons vise à replacer l'expérience dans l'environnement naturel du sujet. Nous nous appuyons sur une méthode bien connue des psychologues, nommée *Experience Sampling Method* (ESM) [11]. Cette méthode a déjà été largement utilisée pour mesurer certains usages sur mobile [10, 12, 28]. Elle consiste à diviser l'expérience en de courtes sessions, *in situ*, répétées plusieurs fois par jour sur plusieurs semaines. L'application est installée sur un téléphone que le sujet porte avec lui en permanence. Il lui est ensuite notifié à des moments aléatoires de la journée qu'il doit accomplir sa session. Celle-ci ne doit pas durer plus de quelques minutes, pour éviter la lassitude du sujet. Pendant la session, on récupère les données dont on a besoin, données d'interaction, données récoltées par questionnaire, données contextuelles, etc. Ces données sont ensuite transmises à un serveur distant, après chaque session. L'expérimentateur peut avoir un contrôle sur le bon déroulement des opérations en consultant la base de données, éventuellement en contactant le sujet. Il peut aussi coupler cette méthode avec des sessions hebdomadaires d'entretien avec les sujets, pour compléter les données. L'expérience se termine au bout d'un temps donné (en général plusieurs semaines), ou lorsque l'utilisateur a accompli un nombre suffisant de sessions.

La force de cette méthode réside dans la multiplicité des sessions, garantissant de capturer un maximum de contextes différents. Il faut cependant s'assurer de ne pas être trop invasif (pas trop de sessions quotidiennes, pas trop longues). Lorsqu'un questionnaire est soumis en fin de session, il est également intéressant de le diviser en 2 ou 3 blocs, et d'alterner la présentation de ces blocs. On atténue ainsi l'influence du questionnaire sur l'appréciation que le sujet a de sa propre expérience (on peut aussi parfois simplement supprimer le questionnaire).

La méthode ESM est utilisée dans cette expérimentation pour mesurer l'apport du son binaural en termes d'immersion, de mémorisation et de performances, dans le cadre d'un jeu vidéo développé sur le modèle de l'*Infinite Runner*. Ce dernier garantit des parties rapides, et favorise la répétition de sessions sur une journée. Pour la performance, les données d'interaction de la session (pour la performance « globale », la distance totale parcourue, pour la performance « locale », le temps de réaction moyen, ou le ratio d'items bonus collectés sur le nombre total croisé) sont transmises. Pour la mémorisation, les objets de mémoire sont disséminés sur le parcours du joueur. À la fin de la session, un questionnaire demande à l'utilisateur de restituer le nom

des objets qu'il aura croisés, par ordre d'apparition. La liste est alors comparée à celle d'origine, et permet d'évaluer l'empan mnésique du sujet. Pour l'immersion enfin, nous nous appuyons sur des questionnaires déjà existants, relatifs à l'immersion et le sentiment de présence, dans le jeu vidéo [5], en environnement virtuel [29], ou pour évaluer du son spatialisé [20]. À partir de cette méthode, on infère une note d'immersion basée sur le ressenti de l'utilisateur. En cela elle est dite subjective, en opposition aux méthodes dites objectives, qui élaborent une note via un modèle, construit sur des données diverses, physiologiques [8], issues du service [16] ou de l'interaction [30]. Néanmoins, si ces méthodes sont transparentes aux yeux de l'utilisateur, il est difficile de construire un modèle solide face aux contextes multiples et imprévisibles.

La dernière étape de l'élaboration de ce protocole consiste justement à collecter les données contextuelles. Nous nous appuyons sur les composantes décrites en Section 3 pour récupérer les informations suivantes, par questionnaire : le lieu (à la maison, au travail/à l'école, dans la rue, en transport, autre intérieur, autre extérieur), les personnes environnantes (seul, avec une ou plusieurs personnes connues, entouré d'une ou plusieurs personnes inconnues), le niveau d'occupation (allant d'un esprit complètement libre à un esprit très occupé), et le niveau de mobilité de l'utilisateur (assis/allongé, debout, en train de marcher). D'autres données seront récupérées automatiquement (e.g., l'heure), ou en début d'expérience (e.g., modèle du téléphone, modèle du casque). Précisons que ces données seront exploitées après l'expérience, pour déterminer l'influence possible du contexte sur les résultats, et pas en temps-réel, comme le font les applications *context-aware*. Celles-ci adaptent leur mode d'interaction selon un contexte inféré [1, 6, 24], pour en atténuer l'influence. Cela supposerait une connaissance a priori de l'effet du contexte que nous n'avons pas.

5 Conclusion

Nous avons proposé un protocole basé sur la méthode dite « *Experience Sampling Method* », répondant à la problématique de la mesure en contexte écologique. Ce protocole devrait nous permettre de mesurer trois apports possibles du binaural que nous avons identifiés : l'immersion, la mémorisation et la performance. Outre le déploiement de l'expérience et l'analyse des résultats, il serait intéressant d'examiner plus en profondeur la validité de cette méthode. Une manière de faire serait par exemple de mener une étude comparative des résultats obtenus avec ceux d'une expérience plus traditionnelle, en laboratoire. Une autre possibilité serait d'étendre l'expérience à d'autres types de jeux vidéo, voire à d'autres applications (services de communication ou consommation de contenus multimédia par exemple).

Références

- [1] Gediminas Adomavicius and Alexander Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook*, pages 191–226. Springer, 2015.

- [2] Jens Blauert. Concepts Behind Sound Quality : Some Basic Considerations. *Internoise*, 9 :72 – 79, 2003.
- [3] Jens Blauert. *The technology of binaural listening*. Springer, 2013.
- [4] Cheryl Campanella Bracken and Gary Pettey. It is really a smaller (and smaller) world : Presence and small screens. In *Proceedings of the 10th International Workshop on Presence.(Barcelona, Spain)*, pages 283–290, 2007.
- [5] Jeanne H Brockmyer, Christine M Fox, Kathleen A Curtiss, Evan McBroom, Kimberly M Burkhart, and Jacquelyn N Pidruzny. The development of the game engagement questionnaire : A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, 45(4) :624–634, 2009.
- [6] Anind K Dey, Gregory D Abowd, and Daniel Salber. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-computer interaction*, 16(2) :97–166, 2001.
- [7] Huong Q Dinh, Neff Walker, Larry F Hodges, Chang Song, and Akira Kobayashi. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Virtual Reality, 1999. Proceedings., IEEE*, pages 222–228. IEEE, 1999.
- [8] Ulrich Engelke, Daniel P Darcy, Grant H Mulliken, Sebastian Bosse, Maria G Martini, Sebastian Arndt, Jan-Niklas Antons, Kit Yan Chan, Naeem Ramzan, and Kjell Brunnström. Psychophysiology-based qoe assessment : a survey. *IEEE Journal of Selected Topics in Signal Processing*, 11(1) :6–21, 2017.
- [9] Matthias Frank, Franz Zotter, Hagen Wierstorf, and Sascha Spors. Spatial audio rendering. In *Quality of Experience*, pages 247–260. Springer, 2014.
- [10] Jon Froehlich, Mike Y Chen, Sunny Consolvo, Beverly Harrison, and James A Landay. Myexperience : a system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of the 5th international conference on Mobile systems, applications and services*, pages 57–70. ACM, 2007.
- [11] Joel M Hektner, Jennifer A Schmidt, and Mihaly Csikszentmihalyi. *Experience sampling method : Measuring the quality of everyday life*. Sage, 2007.
- [12] Selim Ickin, Katarzyna Wac, Markus Fiedler, Lucjan Janowski, Jin-Hyuk Hong, and Anind K Dey. Factors influencing quality of experience of commonly used mobile applications. *IEEE Communications Magazine*, 50(4), 2012.
- [13] Satu Jumisko-Pyykkö and Miska M Hannuksela. Does context matter in quality evaluation of mobile television? In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, pages 63–72. ACM, 2008.
- [14] Satu Jumisko-Pyykkö and Teija Vainio. Framing the context of use for mobile hci. *Social and Organizational Impacts of Emerging Mobile Devices : Evaluating USE : Evaluating Use*, pages 217–219, 2012.
- [15] Anne Kaikkonen, Aki Kekäläinen, Mihael Cankar, Titti Kallio, and Anu Kankainen. Usability testing of mobile applications : A comparison between laboratory and field testing. *Journal of Usability studies*, 1(1) :4–16, 2005.
- [16] Asiya Khan, Lingfen Sun, and Emmanuel Ifeakor. Qoe prediction model and its application in video quality adaptation over umts networks. *IEEE Transactions on Multimedia*, 14(2) :431–442, 2012.
- [17] Pontus Larsson, Daniel Vastfjäll, and Mendel Kleiner. Better presence and performance in virtual environments by improved binaural sound rendering. In *Audio Engineering Society Conference : 22nd International Conference : Virtual, Synthetic, and Entertainment Audio*. Audio Engineering Society, 2002.
- [18] T Letowski. Sound quality assessment : cardinal concepts. In *87th Convention of the Audio Engineering Society, J. Audio Eng. Soc.(Abstracts)*, volume 37, page 1062, 1989.
- [19] Alexander Lindau and Stefan Weinzierl. Assessing the plausibility of virtual acoustic environments. *Acta Acustica united with Acustica*, 98(5) :804–810, 2012.
- [20] Rozen Nicol, L Gros, C Colomes, E Roncière, and JC Messonier. Etude comparative du rendu de différentes techniques de prise de son spatialisée après binauralisation. *CFA/VISHNO*, 2016.
- [21] Ulrich Reiter, Kjell Brunnström, Katrien De Moor, Mohamed-Chaker Larabi, Manuela Pereira, Antonio Pinheiro, Junyong You, and Andrej Zgank. Factors influencing quality of experience. In *Quality of experience*, pages 55–72. Springer, 2014.
- [22] Francis Rumsey. Spatial Quality Evaluation for Reproduced Sound : Terminology, Meaning, and a Scene-Based Paradigm. *Journal of the Audio Engineering Society*, 50(9) :651–666, 2002.
- [23] Francis Rumsey, Sławomir Zieliński, Rafael Kassier, and Søren Bech. On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *The Journal of the Acoustical Society of America*, 118(2) :968–976, 2005.
- [24] Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In *Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on*, pages 85–90. IEEE, 1994.
- [25] Robert Schleicher, Tilo Westermann, and Ralf Reichmuth. Mobile human–computer interaction. In *Quality of experience*, pages 339–349. Springer, 2014.
- [26] Sakari Tamminen, Antti Oulasvirta, Kalle Toiskallio, and Anu Kankainen. Understanding mobile contexts. *Personal and ubiquitous computing*, 8(2) :135–143, 2004.
- [27] Matt Thompson, A Imran Nordin, and Paul Cairns. Effect of touch-screen size on game immersion. In *Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers*, pages 280–285. British Computer Society, 2012.

- [28] Katarzyna Wac, Selim Ickin, Jin-Hyuk Hong, Lucjan Janowski, Markus Fiedler, and Anind K Dey. Studying the experience of mobile applications used in different contexts of daily life. In *Proceedings of the first ACM SIGCOMM workshop on Measurements up the stack*, pages 7–12. ACM, 2011.
- [29] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments : A presence questionnaire. *Presence*, 7(3) :225–240, 1998.
- [30] Wanmin Wu, Ahsan Arefin, Raoul Rivas, Klara Nahrstedt, Renata Sheppard, and Zhenyu Yang. Quality of experience in distributed interactive multimedia environments : toward a theoretical framework. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 481–490. ACM, 2009.
- [31] Frances Amelia Yates. *The art of memory*, volume 64. Random House, 1992.