

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



Guidage de personnes déficientes visuelles par audio spatialisé

S. Ferrand, M. Aussal et F. Alouges
Ecole Polytechnique, Route de Saclay, 91120 Palaiseau, France
sylvain.ferrand@polytechnique.edu

Les personnes aveugles sont capables de localiser et de suivre une source sonore avec une efficacité remarquable. C'est une faculté qu'elles utilisent dans la vie de tous les jours par exemple pour suivre les pas de la personne qui les précèdent, afin de se déplacer dans la rue ou dans un bâtiment. Certaines personnes aveugles utilisent également cette faculté dans la pratique sportive, pour suivre un guide valide lors de la pratique de la course à pied ou du roller. Sur le même principe, il est possible de guider des personnes avec des sons spatialisés.

Pour le montrer, nous avons développé un dispositif embarqué autonome capable de guider des personnes en utilisant des techniques de synthèse binaurale. Dans ce dispositif, nous utilisons à la fois, les informations fournies par un système de localisation temps réel (RTLS) précis et à faible latence, et les données d'orientation de la tête fournies par une centrale inertiels à bas cout (IMU).

Ces données de positionnement sont utilisées pour produire des sons spatialisés par filtrage binaural afin de guider en temps réel une personne le long d'un parcours prédéfini. La personne déficiente visuelle va donc se déplacer vers le son, de la même façon qu'elle suivrait les pas d'un guide. Le système a été conçu afin de minimiser la latence globale, à la fois pour le système de localisation et le moteur de rendu audio binaural.

Nous avons ainsi pu montrer qu'il était possible de guider des personnes en leur faisant suivre des sons spatialisés, de façon précise, naturelle et avec un entrainement minimum. Ce système permet déjà de guider des personnes aveugles pour pratiquer le roller dans un gymnase et ainsi leurs permettre de faire du sport de façon plus autonome.

1 Introduction

Notre capacité à localiser une source sonore est liée à la déformation du signal reçu par chaque oreille qui est interprétée par le cerveau. Il est possible de mesurer ou de simuler les fonctions transfert (appelées *Head Related Transfert Fonction* ou *HRTF*) qui caractérisent la façon dont les oreilles reçoivent un son d'un point dans l'espace. Elles peuvent ensuite être utilisées pour synthétiser un son statialisé par convolution. Écouté au casque, l'auditeur aura alors l'illusion que le son provient d'une direction spécifique. Dans la vie quotidienne, cette capacité de localisation, permet un aveugle de suivre d'autres personnes, simplement en entendant de bruit de leurs pas. De la même façon, nous cherchons à guider quelqu'un avec un son spatialisé [1] généré en temps réel et rendu avec un simple casque.

Une application pour la pratique sportive des personnes déficientes visuelles

Nous cherchons à guider des personnes aveugles dans le contexte de la pratique sportive, que ce soit pour le sport de loisir ou de compétition, dans un premier temps pour la course à pied ou le roller.

Aujourd'hui de nombreuses personnes non-voyantes pratiquent ces sports sur piste, avec un guide voyant, mais cette activité se heurte à différents obstacles pratiques et logistiques. En effet il est souvent difficile pour les clubs sportifs de mettre à disposition un guide valide pour chaque sportif déficient visuel. Par ailleurs, dans le contexte du sport de compétition, il est difficile de trouver des guides avec des capacités sportives suffisantes.

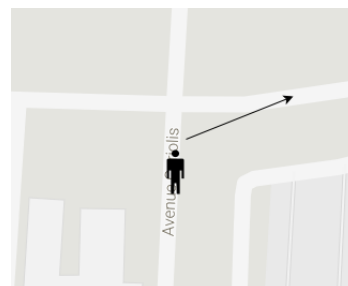
Habituellement, athlètes déficients visuels et guides accompagnateurs sont reliés ensemble par un lien physique comme un bâton de relais ou un élastique, mais certains sportifs non voyants pratiquent également le roller ou la course à pied avec un guidage sonore. L'athlète suit alors son guide grâce aux bruits qu'il produit.

Nous nous sommes inspirés de cette pratique pour développer notre système de guidage sonore. Muni du dispositif, l'athlète suit une source sonore virtuelle spatialisée par synthèse binaurale. Il va donc suivre le son, un peu de la même façon qu'il a l'habitude de suivre son guide.

2 Les méthodes de guidage par audio spatialisé

L'idée du guidage par audio spatialisé n'est pas nouvelle et plusieurs méthodes ont été proposées, généralement pour des applications de navigation urbaine pour les piétons.

- **Balise sonore** : Dans cette technique l'utilisateur doit se diriger vers une source sonore fixe dans l'espace, en choisissant lui même le chemin à suivre.



- **Guidage GPS** : Cette technique est une extension de la précédente. La source sonore change à chaque intersection (i.e. quand une balise a été atteinte) pour indiquer la prochaine direction à emprunter, exactement à la façon d'un GPS routier. Cette méthode a notamment été utilisée dans les projets de recherche GpsTunes et Ontrack [6-10]



- **Guidage adaptatif binaural temps réel (ou guidage "carotte")** : consiste à guider les personnes avec une source sonore virtuelle qui précède l'utilisateur à une distance constante, comme un *carotte sur un bâton*. Mais, la méthode présente plusieurs difficultés. En effet, si on veut obtenir un guidage précis, la source

sonore doit précéder de peu l'utilisateur (par exemple un mètre). Dans ce cas, il est nécessaire de localiser une personne de façon très précise, car un déplacement même très léger doit être perceptible dans le système audio. Par ailleurs le moteur audio doit également avoir une faible latence. Cette méthode a notamment été utilisée dans le cadre du projet Argus [11], avec un système GPS. Toutefois nous pensons que la précision et la latence introduite par un système GPS pénalise le guidage. Nous expliquerons en partie 4 de cet article comment nous mettons en oeuvre cette méthode pour un guidage naturel et précis des personnes déficientes visuelles.

3 Architecture du dispositif

Les expériences décrites dans cet article sont basées sur un même dispositif. Le dispositif est constitué d'un système informatique embarqué connecté à des capteurs de positionnement 2D précis et un head tracker. La pile logicielle développée inclut un moteur de rendu audio binaural ainsi que nos algorithmes de localisation et de suivi de trajectoire. Les capteurs de positionnement et le head tracker vont alors permettre de simuler une source sonore virtuelle indépendamment du mouvement de l'utilisateur. Le dispositif tient compte à la fois des rotations de la tête et des déplacements de l'utilisateur. Le son est finalement reproduit par un casque.

3.1 Localisation

Nous pensons qu'une application de guidage sonore doit pouvoir donner un feedback immédiat à l'utilisateur en mouvement, c'est pourquoi un système de localisation rapide et précis est nécessaire. Les technologies de positionnement par satellite (GNSS) comme le GPS permettent généralement de se localiser avec une précision de 2 à 10 mètres en horizontal avec une résolution temporelle de quelques Herz, ce qui est très insuffisant. Les GNSS différentiels, de précision décimétrique, sont quant à eux coûteux et restent inutilisables en intérieur. Dans nos expériences nous avons

Technologie	Précision	Rate	Utilisation	Coût
GNSS (GPS)	2 - 5 m	1 - 10 Hz	ext	faible
GPS + IMU	2 - 5 m	100 Hz	ext	faible
UWB	10 cm	10 - 15 Hz	int / ext	moyen
D-GNSS	10 cm	1 - 10 Hz	ext	élevé

TABLEAU 1 – Comparaison des systèmes de positionnement

utilisé un système de localisation utilisant la technologie Ultra Wide Band [12] (UWB) associé à un algorithme de multilatéralisation.

Technologie Ultra Wide Band (UWB)

La technologie UWB permet d'effectuer des mesures de distance rapides et précises en utilisant le temps de vol (ToF) d'un signal radio. Nous mesurons alors le temps d'aller-retour d'une trame de données envoyée par un transceiver maître (qui est mobile) et les réponses renvoyées par des esclaves (fixes). Le protocole TWR (Two Way Ranging),

mesure le temps écoulé entre l'envoi d'un message par le maître et la réception de la réponse des esclaves. Elle permet de déduire les distances qui les séparent. Comparée à la technologie basée sur la force du signal, l'UWB offre plusieurs avantages : elle est robuste dans les environnements à trajets multiples (en étant relativement insensible à l'ombrage des capteurs et aux interférences RF). Par ailleurs, c'est une technologie à faible coût et à faible consommation.

Dans une zone avec plusieurs esclaves UWB fixes (appelés Ancres) avec des emplacements connus, un algorithme de trilatération est utilisé pour déterminer la position du maître mobile. Dans notre dispositif, le filtrage et la trilatération sont réalisés par un filtre de Kalman, qui contrairement aux techniques basées sur les moindres carrés (LSE) permet d'intégrer un modèle de mouvement et d'estimer la position de l'utilisateur entre les mesures.

Considérant le vecteur de mesure constitué des distances mesurées par les capteurs $z = [dist_{a_1} \ dist_{a_2} \ \dots \ dist_{a_n}]^T$, la fonction de mesure de $h(x_k)$ du filtre de Kalman pour un système avec n ancres est donnée par la distance euclidienne entre chaque ancre et le maître mobile :

$$h(x_k) = \begin{bmatrix} \sqrt{(x - a_{x,1})^2 + (y - a_{y,1})^2} \\ \sqrt{(x - a_{x,2})^2 + (y - a_{y,2})^2} \\ \vdots \\ \sqrt{(x - a_{x,n})^2 + (y - a_{y,n})^2} \end{bmatrix}. \quad (1)$$

Le bruit de mesure peut être considéré comme gaussien, mais la fonction de mesure étant non linéaire, nous avons implémenté ce filtre sous forme d'un Unscented Kalman Filter [13]. Nous utilisons un modèle Position-Vitesse où le vecteur d'état x_k inclut la position et la vitesse (pour une comparaison des modèles utilisés en localisation, voir Zhu, Y. et Al. [14]).

En condition Line-of-sight (LOS - c'est à dire en vue directe), la précision RMSE (Root Mean Square Error) mesurée dans le cas stationnaire est typiquement de 10cm et la fréquence d'acquisition est d'environ 10Hz.

3.2 Head-tracking

Nous devons également connaître l'orientation de la tête de la personne afin de maintenir la source sonore dans une position fixe indépendamment des mouvements de la tête, comme dans une scène sonore réelle. A cet effet, nous utilisons un Head Tracker, c'est à dire un IMU, fixé sur la tête de l'utilisateur et donnant son orientation (angles d'Euler). Dans nos expériences, nous utilisons le Razor IMU de Sparkfun ou le capteur Bosch BNO-055.

3.3 Moteur audio 3D embarqué à faible latence

Nous avons implémenté un moteur sonore binaural (initialement développé en langage Matlab par M. Aussal et F. Alouges [2-4]) basé sur la convolution du son avec des HRTF. Ce logiciel permet une sélection de filtres en temps réel en fonction de la position de l'utilisateur. Pour plus de confort, le logiciel a été conçu pour minimiser la latence et permet une commutation des filtres rapide et imperceptible (technique de fenêtrage + overlap add). Nous utilisons un jeu de HRTF dérivé du sujet 1040 de la base de données Listen

de l'IRCAM [5]. Pour donner des indications de directions précises et limiter les artefacts liés aux changements de filtre, les mesures HRTF ont été interpolées tous les 2° sur une grille régulière. Le rendu audio a été principalement implémenté en Python et utilise Cython pour les tâches intensives.

3.4 Architecture logicielle

Les différentes tâches logicielles à réaliser (Filtrage HRTF, multilatération, calcul de trajectoire) sont assurées par des processus différents qui échangent des messages sur le principe des architectures multi-tiers (API Rest).

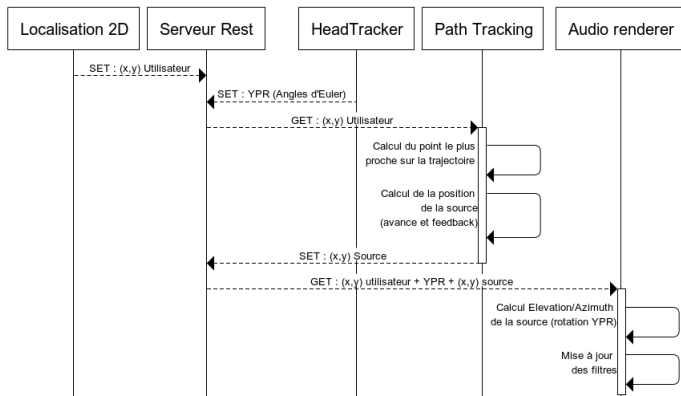


FIGURE 1 – Architecture et messages

La figure 1 montre les échanges de messages et l'organisation des différents processus autour du serveur Rest. En séparant chaque tâche dans des processus différents, le logiciel est capable d'utiliser efficacement les processeurs embarqués multi-cœurs.

L'ensemble de la pile logicielle fonctionne aussi bien sur PC/Mac et sur les systèmes embarqués. Il a notamment été testé sur processeur ARM 8 (Raspberry Pi 3), ARM *Heterogeneous Multi-Processing - Big-Little* (Odroid XU4 et Jetson TX1) et Intel Edison.

Globalement la latence peut se décomposer ainsi :

- Orientation de la tête (Head tracker) : 20 ms
- Localisation 2D : 100 ms mesuré, mais l'interpolation via les prévisions de mouvement du filtre de Kalman permet un rafraîchissement plus rapide.
- Moteur binaural : 23ms

4 Guidage adaptatif

Dans l'expérience suivante, nous faisons en sorte que la source sonore virtuelle précède la personne à guider. Les informations de position et d'orientation issues des capteurs sont utilisées pour déterminer la position de la source et sélectionner le filtre binaural correspondant afin de générer le son adapté. Quand la personne avance, la source sonore virtuelle se déplace en conséquence afin de guider l'utilisateur le long du parcours qui a été prédéfini.

Le filtrage HRTF permet un rendu sonore se rapprochant de la réalité quotidienne des entraînements sportifs où l'athlète non-voyant suit son coach en localisant le bruit

de ses rollers. Cette technique permet donc une prise en main immédiate, sans réel apprentissage avec un rendu relativement naturel. Grâce au filtrage binaural, l'utilisateur peut immédiatement identifier la direction à suivre même s'il perçoit la source derrière lui ce qui ne serait pas possible avec d'autres approches (type balance de volume gauche/droite), cet avantage est particulièrement intéressant pour les trajets très sinueux.

Suivre un parcours inconnu

Dans une première expérience (figure 2), une personne voyante est aveuglée, elle doit suivre un trajet sinueux non connu, avec des virages doux (typiquement une somme de sinusoides dont on fait varier phases et fréquences à chaque essai). Le stimuli est un son alternant sweep et impulsions choisi pour être agréable, riche et régulier. Le casque audio est un modèle Sennheizer HD280 pro. La source sonore est placée à 1m devant la personne à guider. L'utilisateur évolue dans un espace de 8x12m. Dans ces conditions, sans entraînement préalable, la personne est capable de suivre le trajet, mais est limitée à une faible vitesse (marche très lente, environ 3-4km/h).

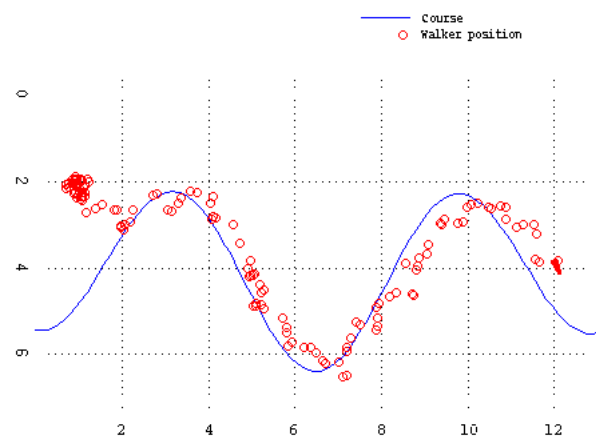


FIGURE 2 – Déplacement le long d'un trajet inconnu. Le parcours à suivre est matérialisé par la sinusoïde et les points représentent la position de l'utilisateur. Départ en haut à gauche.

Parcours élliptique

Dans un deuxième temps, nous avons mené des expériences avec des personnes voyantes et non voyantes sur des parcours de forme elliptique (petit et grand axes d'environ 10 m et 15 m). Dans ces conditions, un effet d'apprentissage intervient rapidement, le guidage devient plus naturel et il est possible de prendre de la vitesse.

Les figures 3 et 4 montrent les parcours effectués par des personnes aveugles respectivement en marche à pied et en roller. Dans les deux cas, ce test a été réalisé avec un casque de conduction osseuse *AfterShokz Sportz titanium* destiné à la pratique sportive. En effet, Bujacz, M. et al. [15] ont montré qu'il était possible d'utiliser des casques osseux pour la localisation sonore, au prix d'une petite perte de précision.

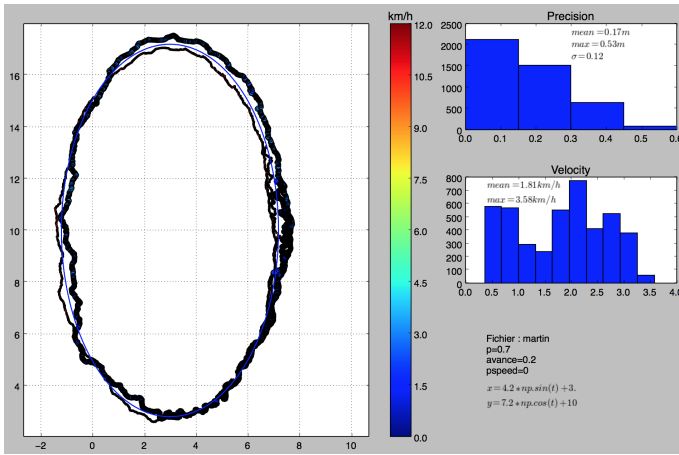


FIGURE 3 – Marche le long d’un chemin connu (Personne aveugle n°1)

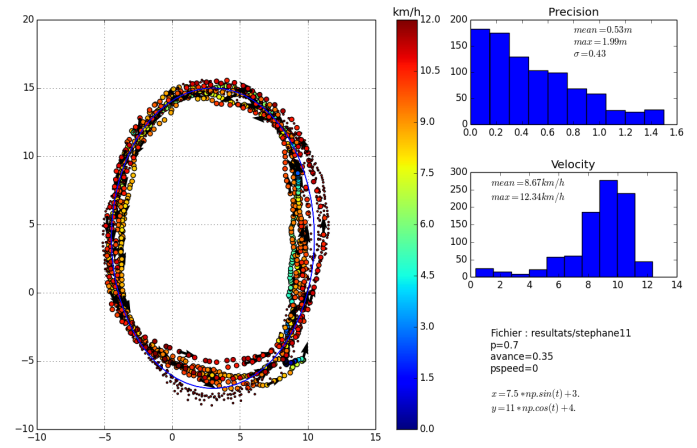


FIGURE 6 – Roler le long d’un chemin connu avec feedback P et $K_p = 0.7$ (Personne aveugle n°2)

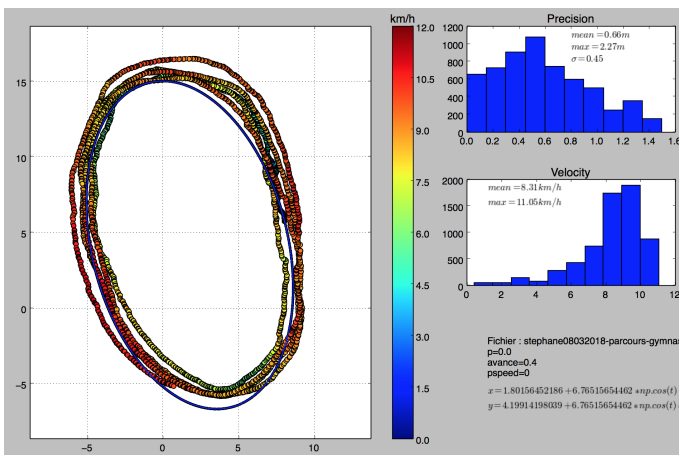


FIGURE 4 – Roler le long d’un chemin connu sans feedback (Personne aveugle n°2)

Dans la mesure ou les personnes déficientes visuelles ne souhaitent pas être isolées de l’extérieur, nous utilisons principalement des écouteurs à conduction osseuses.

Dans nos expériences, nous mesurons à chaque instant la position de l’utilisateur et la distance par rapport à la courbe du trajet à suivre, notée e_t . Pour m mesures, nous pouvons alors mesurer l’erreur moyenne sur un trajet :

$$e_{moyen} = \frac{\sum |e_t|}{m}$$

Dans l’expérience de la figure 3 (marche à pied, utilisateur 1), avec une vitesse moyenne de 1.8km/h, l’écart moyen par rapport au parcours idéal (e_{moyen}) était de 17cm (max 53cm, $\sigma = 0,12$). Dans la même expérience, réalisé en roler (figure 4), l’utilisateur a atteint une vitesse moyenne de 8.3km/h (max 11km/h) et l’erreur moyenne était de 0.66m (max 2.27m, $\sigma = 0.45$).

Nous avons également introduit un feedback proportionnel :

$$F_t = K_p e_t$$

où la source sonore est décalée de F_t par rapport à la position théorique sur le parcours, proportionnellement à l’erreur e_t (voir illustration 5), ainsi qu’un feedback

proportionnel-intégral :

$$F_t = K_p e_t + K_i \sum_j e_j$$

tenant compte de l’accumulation de l’erreur sur une fenêtre de temps.

Nous avons donc effectué des tests de guidage en conditions réelles en mettant en oeuvre le feedback proportionnel, avec pour conséquence une augmentation sensible de la précision. La figure 6, montre le parcours de l’utilisateur 2 avec feedback, dans les même conditions que le test de la figure 4.

Dans cette expérience, l’utilisateur a atteint la vitesse 12 km/h et l’erreur moyenne était d’environ 0,5 m (maximum 1,3 m, $\sigma = 0.43$).

Ces méthodes de guidages sont efficaces. Après avoir effectué quelques tours à faible vitesse, les utilisateurs (y compris non-voyants) sont capable de percevoir la configuration du trajet et peuvent prendre de la vitesse. Le système de guidage permet alors à l’utilisateur d’ajuster sa position en temps réel sans effort supplémentaire.

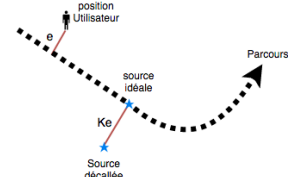


FIGURE 5 – Feedback proportionnel

5 Conclusion

La pratique du sport par des personnes en situation de handicap visuel est un enjeu important. En effet, elle est encore plus importante que pour les personnes valides puisqu’elle permet de lutter contre l’isolement social et d’améliorer la coordination et la motricité.

Pour développer ce dispositif, nous avons cherché à nous associer autant que possible avec des personnes déficientes visuelles, pour rester au plus proche de leurs besoins. De plus, de nombreuses pistes d’améliorations sont à envisager. L’utilisation d’éléments réalistes d’écholocalisation ou de réverbération est une piste qui reste à explorer. Nous travaillons également sur l’évaluation des stimulis de guidage, en particulier sur l’élaboration de filtre plus

réguliers et optimisés pour la localisation dynamique en azimuth. Nous sommes également en train de mettre au point un protocole d'évaluation subjective plus précis du moteur binaural pour le guidage. Notre système doit également évoluer pour être utilisé en extérieur afin d'être étendu à d'autres sports (karting, ski et vélo).

Nous travaillons en partenariat étroit avec plusieurs associations sportives d'aide aux personnes en situation de handicap. Notre prototype est régulièrement testé avec les sportifs de ces associations. Il permet déjà de guider une personne non-voyante à roller dans un grand gymnase à environ 12km/h grâce au guidage binaural adaptatif. A terme, notre dispositif pourrait permettre à un plus grand nombre de personnes déficientes visuelles de pratiquer du sport, par exemple avec un seul accompagnateur pour tout un groupe. Il pourrait également faciliter la pratique de la compétition.

Références

- [1] Blauert, J. (1983). Spatial hearing.
- [2] Aussal, M. (2014). Méthodes numériques pour la spatialisation sonore, de la simulation à la synthèse binaurale (Doctoral dissertation, Palaiseau, Ecole polytechnique).
- [3] MyBino Binaural Engine, Available from : <http://www.cmap.polytechnique.fr/xaudio/mybino> (date last viewed 19/03/18).
- [4] Salmon, F., Aussal, M., Messonnier, J. C., & Millot, L. Optimization of interactive binaural processing for object-based audio mixing.
- [5] Warusfel, O. (2003). Listen HRTF database. online, IRCAM and AK, Available : <http://recherche.ircam.fr/equipes/salles/listen/index.html> (date last viewed 01/06/16).
- [6] Albrecht, R., Vaananen, R., and Lokki, T. (2016). Guided by music : pedestrian and cyclist navigation with route and beacon guidance. *Personal and Ubiquitous Computing*, 20(1), 121-145.
- [7] Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., & O'Modhrain, S. (2005, September). GpsTunes : controlling navigation via audio feedback. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services* (pp. 275-278). ACM.
- [8] Jones, M., Jones, S., Bradley, G., Warren, N., Bainbridge, D., & Holmes, G. (2008). ONTRACK : Dynamically adapting music playback to support navigation. *Personal and Ubiquitous Computing*, 12(7), 513-525.
- [9] Zwinderman, M., Zavialova, T., Tetteroo, D., & Lehouck, P. (2011, August). Oh music, where art thou ?. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 533-538). ACM.
- [10] Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (1998). Navigation system for the blind : Auditory display modes and guidance. *Presence*, 7(2), 193-203.
- [11] Carrasco, E., Loyo, E., Otaegui, O., Fösleitner, C., Dubielzig, M., Olmedo, R., ... & Spiller, J. (2014, July). ARGUS autonomous navigation system for people with visual impairments. In *International Conference on Computers for Handicapped Persons* (pp. 100-107). Springer International Publishing.
- [12] Sahinoglu, Z., Gezici, S., & Guvenc, I. (2008). *Ultra-wideband positioning systems* (Vol. 2). Cambridge, UK : : Cambridge university press.
- [13] Wan, E. A., & Van Der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000* (pp. 153-158). Ieee.
- [14] Zhu, Y., & Shareef, A. (2006, August). Comparisons of three kalman filter tracking algorithms in sensor network. In *Networking, Architecture, and Storages, 2006. IWNAS'06. International Workshop on* (pp. 2-pp). IEEE.
- [15] Bujacz, M., Kropidlowski, K., Ivanica, G., Moldoveanu, A., Saitis, C., Csapo, A., ... & Rotnicki, M. (2016, July). Sound of Vision-Spatial audio output and sonification approaches. In *International Conference on Computers Helping People with Special Needs* (pp. 202-209). Springer, Cham.