International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:11, No:1, 2017

Sound Selection for Gesture Sonification and Manipulation of Virtual Objects

Benjamin Bressolette, Sébastien Denjean, Vincent Roussarie, Mitsuko Aramaki, Sølvi Ystad,
Richard Kronland-Martinet

Abstract**—**New sensors and technologies – such as microphones, touchscreens or infrared sensors – are currently making their appearance in the automotive sector, introducing new kinds of Human-Machine Interfaces (HMIs). The interactions with such tools might be cognitively expensive, thus unsuitable for driving tasks. It could for instance be dangerous to use touchscreens with a visual feedback while driving, as it distracts the driver's visual attention away from the road. Furthermore, new technologies in car cockpits modify the interactions of the users with the central system. In particular, touchscreens are preferred to arrays of buttons for space improvement and design purposes. However, the buttons' tactile feedback is no more available to the driver, which makes such interfaces more difficult to manipulate while driving. Gestures combined with an auditory feedback might therefore constitute an interesting alternative to interact with the HMI. Indeed, gestures can be performed without vision, which means that the driver's visual attention can be totally dedicated to the driving task. In fact, the auditory feedback can both inform the driver with respect to the task performed on the interface and on the performed gesture, which might constitute a possible solution to the lack of tactile information. As audition is a relatively unused sense in automotive contexts, gesture sonification can contribute to reducing the cognitive load thanks to the proposed multisensory exploitation. Our approach consists in using a virtual object (VO) to sonify the consequences of the gesture rather than the gesture itself. This approach is motivated by an ecological point of view: Gestures do not make sound, but their consequences do. In this experiment, the aim was to identify efficient sound strategies, to transmit dynamic information of VOs to users through sound. The swipe gesture was chosen for this purpose, as it is commonly used in current and new interfaces. We chose two VO parameters to sonify, the hand-VO distance and the VO velocity. Two kinds of sound parameters can be chosen to sonify the VO behavior: Spectral or temporal parameters. Pitch and brightness were tested as spectral parameters, and amplitude modulation as a temporal parameter. Performances showed a positive effect of sound compared to a no-sound situation, revealing the usefulness of sounds to accomplish the task.

Keywords**—**Auditory feedback, gesture, sonification, sound perception, virtual object.

Benjamin Bressolette, Mitsuko Aramaki, Sølvi Ystad and Richard Kronland-Martinet are with the LMA, CNRS, UPR 7051 - Aix Marseille Université, Centrale Marseille, 4 impasse Nikola Tesla, CS 40006, F-13453, Marseille cedex 13, France (e-mail: bressolette@lma.cnrs-mrs.fr, aramaki@lma.cnrs-mrs.fr, ystad@lma.cnrs-mrs.fr, kronland@lma.cnrs-mrs.fr).

Sébastien Denjean and Vincent Roussarie are with PSA Group (e-mail: sebastien.denjean@ext.mpsa.com, vincent.roussarie@mpsa.com).

I. INTRODUCTION

THE introduction of new sensors and new technologies in the automotive industry is transforming driver interactions with the central system via new human–machine interfaces (HMIs). For instance, arrays of buttons are replaced by touchscreens with benefits in terms of design and space, or microphones that are implemented for telephony and speech recognition. Consequently, the solicitation of the driver's senses is evolving. For instance, a button could be manipulated with only a haptic feedback where touchscreens require the vision of the driver. This might be a problem in terms of security, since it potentially extends the reaction time of the driver in a dangerous situation.

Gestures can be an alternative to control the central system, with a particular gesture associated to a precise function. Gestures can be performed without turning the eyes away from the road and perceived via proprioception, which is an advantage in our in-car context.

Sonification, which is "the use of non-speech audio to convey information or perceptual data" [1] can be a good feedback to inform the driver on the function currently manipulated. It is already used in several domains and can for instance be a precious help for blind people to cross the street, or even to be guided along more complex paths [2]. Acoustic parking aids have been used by the automotive industry to give the driver feedback on the distance between an obstacle and the car. Sonification for guidance is also used in medicine, in surgery [3], or for rehabilitation [4]-[7]. The use of sounds in a guidance task has been studied in [8] where several sound parameters have been tested. Additionally, auditory interfaces can be comparable to visual interfaces for everyday manipulations, and can be beneficial for driving performances and perceived workload [9].

An important property presented in this study is the utilization of a virtual object (VO), as an intermediate between the user's gesture and the manipulation of the central system function. This VO is also used to sonify the consequences of the gesture rather than the gesture itself, and can in hereby be considered as the sound source. As vision is dedicated to the driving task, the preferable way to perceive the VO is by its sound. The sonification strategy will consequently be determinant, and will have to transmit the maximum of information about the dynamics of the VO.

Virtual object sonification has been studied by Rath [10], who compared the performances between ecological versus abstract sonification. Some interesting results showed an evolution of performances over time, reflecting different learning effects. This example shows that the choice of sonification can lead to different results.

In this study, we want to know if sounds can help to perform a swipe gesture. We consequently tested several sonification strategies. The experiment is described in Section II including results and a discussion, before a conclusion in Section III.

II. EXPERIMENT

In this experiment, the VO is a spherical object linked to the hand by a spring, with a force $\vec{f} = -k.\vec{x}^{\alpha}$ where k is the stiffness of the spring, α is arbitrarily set by the author, and x is the distance between the VO and the hand.

The aim of the experiment is to send the VO in one of the three delimited zones with a swipe gesture. To succeed, a swipe has to send the VO to the requested zone without exceeding it. The VO is moving along one dimension only, and the three zones don't overlap. The zone delimitations are calculated with the help of Fitts' work and the Shannon reformulation [11] of the index of difficulty [12]:

$$Id = \log_2\left(1 + \frac{D}{W}\right)$$

with Id the index of difficulty, D the distance to the target W the target width.

To keep the Id constant across the three zones, we have:

$$\frac{D_{zone1}}{W_{zone1}} = \frac{D_{zone2}}{W_{zone2}} = \frac{D_{zone3}}{W_{zone3}}$$

As the zones do not overlap, we obtained zones as depicted in Fig. 1. The zone on the right is thin because it is close to the subject. On the contrary, the zone on the left is bigger because of the large distance to the subject.

Fig. 1 Illustration of the system: The hand is on the right, and the three zones to reach on the left. The smallest zone is the one that is closest to the subject, the biggest is the farthest. The swipe gesture is oriented from the right to the left, as expressed by the axis on the bottom of the figure

In this stage, we make the hypothesis that this formulation of the Fitts' difficulty index is valuable in our case.

A. Method

1) Participants: 32 right-handed participants, 14 women and 18 men, from 22 to 53 (mean age 32.0, standard deviation 8.1) were volunteered for this experiment. They were all working for PSA Group and reported no hearing problems.

2) Experimental Variables: Both temporal and spectral parameters can be used to sonify the VO behavior. We here chose one temporal strategy based on amplitude modulation (AM), and two spectral strategies based on pitch and brightness variations.

The AM parameter is constructed with the positive part of a sawtooth signal – to obtain a sharper amplitude variation compared to a sine wave for instance. The frequency of the modulation varies between 1 and 10 Hz, and the modulation depth is set to 80% of the signal – 0% representing no modulations and 100% total modulations. The pitch strategy consists in a pink noise filtered by a sharp band-pass filter – with a decay rate of 25 – with a center frequency varying between 300 and 900 Hz. These limits are chosen to obtain a clear low/high frequency effect. For the brightness parameter, the pink noise is first low-pass filtered with a cutoff frequency of 150 Hz to eliminate the too low frequencies. The signal is then filtered by a high-pass filter with a cutoff frequency between 300 and 900 Hz as for the pitch. This leads to a variation of the spectral centroid parameter between 310 and 630 Hz.

Two VO parameters seem interesting to sonify: The hand – VO distance and the VO velocity. The distance can be a precious help for this task as the success of a throw is judged on the final position of the object compared to the three zones. Then, distance could be a natural parameter to sonify, to help the subjects do the requested task. The VO velocity was also chosen because it seems related to the sound of rolling objects [13], [14] or friction sounds [15].

3) Sound Strategies: The combination of the three sound parameters and the two VO parameters (i.e. hand-VO distance and VO velocity) resulted in 6 different sound strategies. We also created more "complex" strategies associating both dynamic parameters with a different sound parameter. Finally, as a control condition, the 13^{th} association did not contain any sound. The 13 different strategies are summarized in Table I. In this table, the VO-hand distance and the VO velocity are associated to sound parameters. For instance, strategy 1 corresponds to the sonification of hand-VO distance by the pitch strategy and no sonification of the VO velocity. The "simple" strategies – sonifying either the hand-VO distance or the VO velocity – are strategies 1 to 6, and "complex" strategies – sonifying both the hand-VO distance and the VO velocity with different sonification strategies – are strategies 7 to 12.

TABLE I
CODING OF THE SOUND STRATEGIES

Sound strategy	*Distance*	*Velocity*
1	Pitch	-
2	AM	-
3	Brightness	-
4	-	Pitch
5	-	AM
6	-	Brightness
7	Pitch	AM
8	Pitch	Brightness
9	AM	Pitch
10	AM	Brightness
11	Brightness	Pitch
12	Brightness	AM
13	-	-

4) Experimental Setup: A Leapmotion controller [16], which is based on infrared technology was used to capture the gesture. Subjects were performing the test in a quiet room, with the Leapmotion controller in front of them. Three Fostex PM0.4d loudspeakers were distributed in front of the subject, from the center to the left. The setup is depicted in Fig. 2 with a view of the spatializator [17]. The SDK developed by Leapmotion for the 3D software Unity [18] was used to rebuild the hand on the Unity scene, as in Fig. 3.

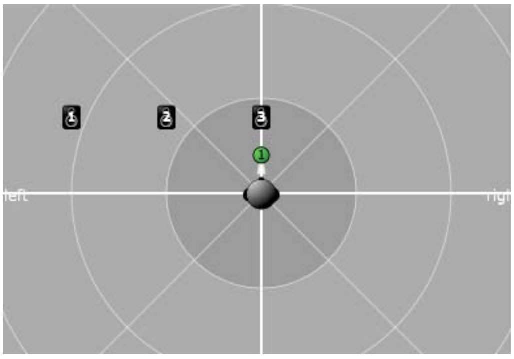

Fig. 2 Top view of the experimental setup given by the spatializator interface. The three loudspeakers are represented, as well as the sound source (the VO) in green

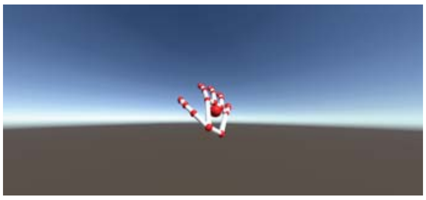

Fig. 3 Rebuilt hand on the right on the Unity scene, the VO is on the left

The sound strategies were lateralized, giving additional information compared to the no-sound condition.

5) Procedure: This test was divided in two phases. For each sonification strategy phase 1 was performed, followed by phase 2 with the same sound strategy. Each strategy was tested, resulting in 13 sessions of 60 swipes. The first phase consisted in a grouped learning, and the second in a distributed learning, closer to the future in-car utilization.

During phase 1, participants were asked to do 45 right-handed swipes. For each zone, 15 successive swipes were requested, with a visual feedback on the performance. The feedback in this phase indicates if the requested zone was reached and if not, whether the VO was thrown too close or too far. The aim of this phase was to help subjects localize the three zones, and to learn the correct gestures that send the VO to the requested zone. The presentation order of the zone was randomized across sonification strategies and subjects.

The second phase consisted in 15 right-handed swipes distributed randomly over the three zones, with 5 swipes for each zone. This task was harder because the zone to reach changed randomly after each trial, and because the feedback only told whether or not the right zone was reached. To prevent a learning gesture phenomenon over the sessions, the stiffness of the spring was modified randomly for each session. 3 stiffness parameters were experimentally chosen representing a soft, normal and hard spring. The 13 sound strategies were tested for each stiffness value, resulting in 13 strategies · 3 stiffness values = 39 conditions.

B. Index and Statistical Processing

In the aim to analyze the results of this experiment, we could have used the success rate of each sound strategy – noting 1 a good throw and 0 otherwise. However, we wanted to consider a richer information, including the final VO position instead of a binary data. We then build an index, taking into account the success of the throw but also the continuous VO final position. As the task was to reach the zone without any precision, we highly penalized a failure compared to a good throw. We chose to set the index to 0 for a good throw and to 1 for a throw out of the boundaries of the requested zone, plus a penalty related to the VO-zone distance.

The index is represented in Fig. 4 for the three zones.

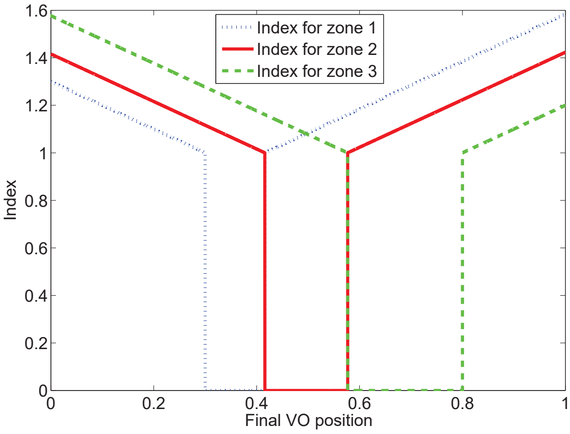

Fig. 4 Index construction for the three zones

Consequently, the closer to 0 the index, the better the strategy.

We then performed an Analysis of Variance (ANOVA) on this index with Strategy as a factor.

C. Results

The results for each phase are depicted in Figs. 5 and 6. The index for each sonification strategy is presented. Strategies are ranked in an increasing order. The ANOVA reported a significant effect of the sound strategies – $F(12, 54892) = 1.979$, $p < 0.05$ for phase 1 and $F(12, 17452) = 2.780$, $p < 0.002$ for phase 2. It allows us to distinguish groups given by the Duncan post hoc analysis. Strategies belonging to the same group are indicated by horizontal lines on the top of the figures. For instance, curves 11, 6, 7 and 8 in Fig. 5 do not have significant differences, and consequently form a group.

Phase 1 informs about the grouped learning phase of each sound strategy. The differences between the strategies are small, as we can see in Fig. 5. However, some strategies reveal significant performance differences: The performance of strategy 11 is significantly higher compared to the control

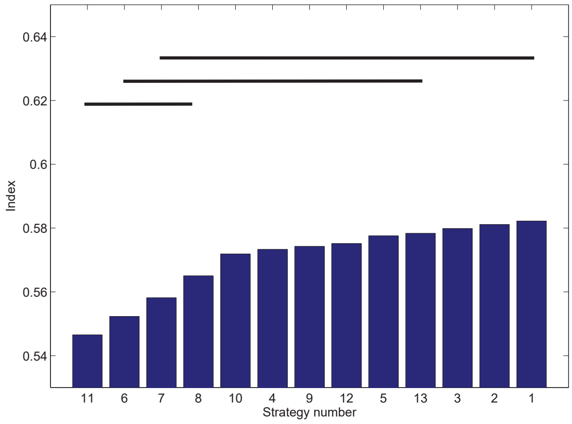

Fig. 5 Index for each strategy for phase 1. Strategies are ranked in an increasing order. Strategies are grouped by the horizontal lines on the top of the figure depending of the Duncan post hoc analysis

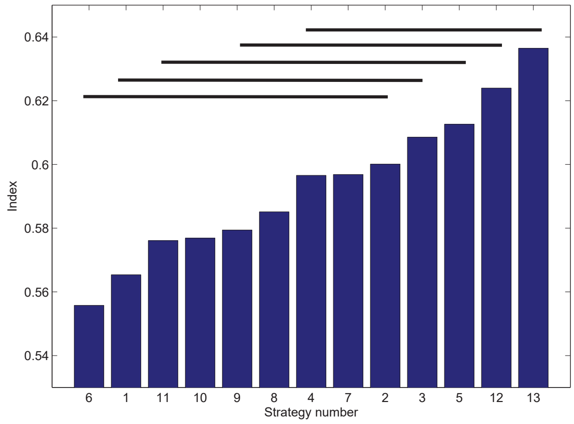

Fig. 6 Index for each strategy for phase 2. Strategies are ranked in an increasing order. Strategies are grouped by the horizontal lines on the top of the figure depending of the Duncan post hoc analysis

strategy for instance. Each success rate is a mean over 32 subjects · 3 stiffness · 3 zones · 15 throws = 4320 values, explaining the significant differences even if the performances seem close.

Strategies 11 and 6 turn out to be the best in this phase. In particular, significant differences can be seen between strategy 11 and the control strategy 13 – $p < 0.05$. This difference emphasizes the influence of sound, and that some sound strategies tend to help subjects perform the task. On the contrary, some strategies do not reveal significant differences with the control strategy, as strategy 1 which is the worst strategy in terms of performance.

Results obtained in the distributed learning phase are presented in Fig. 6, revealing significant differences between strategies 6, 1, 11, 10, 9, 8 and the control strategy 13. This suggests that sounds significantly improve the performances for this particular task. Again, the number of tests performed – 32 subjects · 3 stiffness · 3 zones · 5 throws = 1440 – assure significant differences even if the indices are close. The performance differences of each strategy are clearer compared to Fig. 5. This difference with phase 1 can be explained by the higher difficulty of this task, which leads to a clearer discrimination of the sonification strategies.

Results show that strategy 6 seems optimal for this phase. In particular, there is a significant difference between strategies 6 and 13 ($p < 0.05$), confirming the importance of sounds in this case.

Figs. 5 and 6 reveal the effect of the sound on performances. The indexes are relatively high, if we consider that an index of 0 as a success rate of 100% and an index above 1 as performances close to 0%. This can be explained by the difficulty of the task: gestures required to reach the different zones were really close. Sometimes participants were puzzled because they were certain to have performed exactly the same gesture while the obtained results differed.

As the use of future in-car systems will be closer to phase 2 than to phase 1, more importance will be given to the results presented in Fig. 6. The results from the first phase will be discussed afterwards.

The first thing to note is that strategy 6 – velocity linked to brightness – seems to lead to better performances compared to the other sound strategies. Considering only the “simple” strategies, strategy 6 is significantly better than strategies 3 and 5. We can see that the temporal strategies of amplitude modulation seem to be misunderstood by the participants. Concerning the “complex strategies”, strategies 11, 10, 9 and 8 show good results. The good results of strategies 10 and 8 may be due to the velocity - brightness association of strategy 6, but performances may be a bit degraded by the other parameter. It is surprising that strategies 8 and 11 have good results because they are composed of two frequency parameters, which could have resulted in a complex and hard to understand association. In the case of strategy 8, participants might have concentrated on one of the sound parameters, ignoring the other one. For strategies 9 and 11, it is interesting to note that performances of these strategies are better compared to the corresponding “simple” strategies – 2 and 4 for strategy 9, 3 and 4 for strategy 11, even if these differences are not significant. There may have been a “symbiosis effect” in these cases, the two associations of parameters helping the subjects accomplish the task. At this point, this only can be an hypothesis as there are no significant differences. This kind of “symbiosis effect” is also found in the first phase for strategy 11.

We can also see that performances of strategies 5 and 12 are close to the control strategy 13 in Fig. 6. This means that even if there is a sound to help participants accomplish the task, if the sound is not well constructed, the results can be equivalent to an absence of sound. It should be noted that sound strategies 5 and 12 contained spatial clues related to the position of the VO because of the sound spatialization. This should help the subject feel the behavior of the VO, and consequently help to reach the requested zone. We consequently can formulate two hypotheses: either the subjects can not use the spatial information, or strategies 5 and 12 disturbed the subjects sufficiently to impede them to use this information.

Finally, it can be noted that each association of a dynamic VO parameter to a sound parameter leads to different results.

For instance, for the six "simple" strategies neither the distance VO-hand – *for strategies 1, 2 and 3* – nor the VO velocity – *strategies 4, 5 and 6* – seems better compared to the others. The same thing can be seen for the pitch strategy – *strategies 1 and 4*, AM – *strategies 2 and 5* – or brightness strategy – *strategies 3 and 6*. The results do not depend on a particular parameter, but on the association between a VO and a sound parameter.

D. Discussion

Velocity seems to be a natural parameter to sonify as it has been used in several studies on sonification: to characterize motions evoked by sounds [19], to transmit information of a virtual rolling ball [10] or to recreate a friction sound [15].

Houben studied the sound of rolling balls in several ways, and the perception of size and speed of wooden balls rolling along a plane [13]. He excluded clues of loudness and amplitude modulation to study the fine structures of these rolling sounds. By combining spectral and temporal clues of different rolling sounds of wooden balls, Houben and al. [20] showed that subjects link the perception of speed to spectral cues. More precisely, the "spectral centroid" seems to influence the perception of size and speed of wooden balls [21]. This spectral centroid can be related to our brightness parameter as it seems to correspond to the clues used to judge the speed of the balls. Strategy 6 can then be seen as a "natural sonification", as it may reproduce the sound of natural rolling objects. Houben also showed that the perception of speed can be improved adding an amplitude modulation related to the angular speed. Indeed, irregularities on the perfect circularity of the rolling ball induce variations on the height of the center of mass c between c_1 and c_2, which leads to an AM proportional to the height variation $c_2 - c_1$ [10]. In our case, the amplitude of the modulations may have been too rough to be taken into account by the participants. Furthermore, the frequency of the modulations may not have been related to a natural modulation caused by the non perfect circularity of a rolling ball.

Strategy 1, relying the VO-hand distance to pitch, led to good results on phase 2, which may rely on the fact that the task to perform was based on judgments of the distance. Furthermore, pitch is perceived with good precision, supporting our idea of an "abstract sonification" as seen by Rath [10]. In this study, "natural" and "abstract" sonifications were compared, showing differences in performances before and after the training phase: the natural sonification had good results even before the training and showed little improvement after the training, whereas the abstract sonification had really bad performances before the training but spectacularly improved after the training. In our case, strategy 1 was ranked as the 13^{th} strategy in phase 1 and as the 2^{nd} strategy in phase 2, showing a high improvement between the two phases. On the contrary, strategy 6 showed good performances in both phases. This may be linked to the same effect: a natural sonification is learned very quickly but show no real improvements after a learning phase, whereas an abstract sonification is not understood at first sight but leads to good performances after some time.

III. CONCLUSION

In this study, we investigated some aspects of a virtual object sonification. The VO is used to sonify the consequences of the user gestures, and is then considered as the sound source. In our experiment, we asked the subjects to throw the VO in one of the three available zones. This experiment was tested with 12 different sound feedbacks, and one control situation without sound. The 12 sound situations were created with two VO parameters – hand-VO distance and VO velocity – and three sound parameters – pitch, AM and brightness. For each situation, a learning phase preceded a phase close to the future in-car utilization.

Results obtained showed that performances with sound are significantly better compared to the no-sound situation. Consequently, sounds seem to provide valuable feedback to the users to improve their task performance. The comparison concerning the nature of the sound parameters indicates that spectral parameters seem to be better fitted to transmit information about the VO dynamics compared to temporal parameters. A sound strategy obtained performances comparable to the no-sound situation, suggesting that sounds are not necessary helpful even if they are spatialized. If the sound isn't well constructed, it could even be disturbing for the users. On the contrary, several strategies revealed significantly better performances compared to the no-sound situation, suggesting a real contribution of sound to accomplish the task. In particular, the association of the VO velocity to brightness and the combination of VO-hand distance to pitch seems promising, with different behaviors for the two phases. The VO velocity to brightness association provided good results in both phases, whereas the VO-hand distance to pitch combination obtained low performances in the first phase, but good results in the second phase. These evolutions can be related to the study of Rath [10] where an "abstract" sonification – low performances before a learning phase, but good results after – was compared to a "natural" sonification – with good results before and after the learning phase. The combination of the VO velocity to brightness may be seen as a "natural" sonification – as suggested by Houben's studies [13], [20], [21], and the VO-hand distance to pitch association as an "abstract" sonification, adapted to the task to perform.

In future researches, it could be interesting to test strategies associating sound parameters to the same VO parameter. For instance, it would be interesting to associate the velocity to brightness but also to AM, to recreate the natural modulation of rolling balls as described in [10]. A longer learning phase could also be helpful to observe if the VO-hand distance to pitch performances could surpass the VO velocity to brightness performances. The study of other gestures can be another path to explore, starting for instance with back-handed swipes.

ACKNOWLEDGMENT

We would especially like to thank Lionel Bringoux and Christophe Bourdin (Aix-Marseille Universite, CNRS, ISM UMR 7287, 13288 Marseille Cedex 09, France) for their advices and for the fruitful discussion we had.

This work was partly funded by the French National Research Agency (ANR) under the SoniMove: Inform, Guide and Learn Actions by Sounds project (ANR-14-CE24-0018-01).

REFERENCES

[1] G. Kramer et al., *Sonification report: Status of the field and research agenda*, 2010.
[2] J. M. Loomis, R. G. Golledge, and R. L. Klatzky, *Navigation system for the blind: Auditory display modes and guidance*, Presence: Teleoperators and Virtual Environments, vol. 7, no. 2, pp. 193203, 1998.
[3] K. Wegner, *Surgical navigation system and method using audio feedback*, 1998.
[4] M. Dozza, L. Chiari, and F. B. Horak, *A portable audio-biofeedback system to improve postural control*, in Engineering in Medicine and Biology Society, 2004. IEMBS04. 26th Annual International Conference of the IEEE, 2004, vol. 2, pp. 47994802.
[5] J. Danna et al., *The effect of real-time auditory feedback on learning new characters*, Human Movement Science, vol. 43, pp. 216228, Oct. 2015.
[6] J. Danna et al., *Handwriting movement sonification for the rehabilitation of dysgraphia*, in 10th International Symposium on Computer Music Multidisciplinary Research (CMMR)-Sound, Music and Motion-15-18 oct. 2013-Marseille, France, 2013, pp. 200208.
[7] M. Aramaki, C. Gondre, R. Kronland-Martinet, T. Voinier, and S. Ystad, *Thinking the sounds: an intuitive control of an impact sound synthesizer*, in International Conference on Auditory Display (ICAD09), 2009, pp. 119124.
[8] G. Parseihian, C. Gondre, M. Aramaki, S. Ystad, and R. Kronland-Martinet, *Comparison and evaluation of sonification strategies for guidance tasks*, IEEE Transactions on Multimedia, vol. 18, no. 4, pp. 674686, 2016.
[9] G. Jakus, C. Dicke, and J. Sodnik, *A user study of auditory, head-up and multi-modal displays in vehicles*, Applied Ergonomics, vol. 46, pp. 184192, Jan. 2015.
[10] M. Rath and R. Schleicher, *On the relevance of auditory feedback for quality of control in a balancing task*, Acta Acustica United With Acustica, vol. 94, no. 1, pp. 1220, 2008.
[11] I. S. MacKenzie, *Fitts law as a research and design tool in human-computer interaction*, Human-computer interaction, vol. 7, no. 1, pp. 91139, 1992.
[12] P. M. Fitts, *The information capacity of the human motor system in controlling the amplitude of movement.*, Journal of experimental psychology, vol. 47, no. 6, p. 381, 1954.
[13] M. M. J. Houben, *The sound of rolling objects : perception of size and speed*, PhD Thesis, 2002.
[14] S. Conan, O. Derrien, M. Aramaki, S. Ystad, and R. Kronland-Martinet, *A synthesis model with intuitive control capabilities for rolling sounds*, IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 8, pp. 12601273, 2014.
[15] E. Thoret, M. Aramaki, R. Kronland-Martinet, J.-L. Velay, and S. Ystad, *From sound to shape: auditory perception of drawing movements.*, Journal of Experimental Psychology: Human Perception and Performance, vol. 40, no. 3, p. 983, 2014.
[16] *Leap Motion — Mac & PC Motion Controller for Games, Design, Virtual Reality & More* (online). Available: https://www.leapmotion.com/
[17] *Spat* (online). Available: http://forumnet.ircam.fr/fr/produit/spat/
[18] *Unity - Game Engine* (online). Available: https://unity3d.com/fr/
[19] A. Merer, M. Aramaki, S. Ystad, and R. Kronland-Martinet, *Perceptual characterization of motion evoked by sounds for synthesis control purposes*, ACM Transactions on Applied Perception, vol. 10, no. 1, pp. 124, Feb. 2013.
[20] M. M. Houben, A. Kohlrausch, and D. Hermes, *Auditory cues determining the perception of the size and speed of rolling balls*, 2001.
[21] M. M. J. Houben, A. Kohlrausch, and D. J. Hermes, *Perception of the size and speed of rolling balls by sound*, Speech Communication, vol. 43, no. 4, pp. 331345, Sep. 2004.

Benjamin Bressolette received a MS degree from École Centrale Marseille, Marseille, France in 2014. He is currently preparing a PhD on gesture sonification for human-computer interaction, in collaboration with PSA Group and Laboratoire de Mécanique et d'Acoustique (CNRS: National French Research Center).

Sebastien Denjean received a MS degree from École Centrale Marseille, Marseille, France in 2010 and his Ph.D. degree in acoustics from the Aix-Marseille University, Marseille, France, in 2015 for his work on in-car sonification for electric cars. He is currently working at PSA Group on multi-modal interfaces and sonification.

Vincent Roussarie received the Ph.D. degree from Université du Maine, Le Mans, France in 1999 for his work on the perception of vibrating structures. He is currently working at PSA Group as a research project manager on the team "Cognitive Science and Human Factors" dealing the in-vehicle user perception, with subjects concerning the human-machine interaction, the perception of vibrations or the integration of new technologies.

Mitsuko Aramaki received the Ph.D. degree from the Aix-Marseille University, Marseille, France, in 2003, for her work on analysis and synthesis of impact sounds using physical and perceptual approaches. She is currently a researcher at the National French Research Center (CNRS) in Marseille, France. She has been working at the Institut de Neurosciences Cognitives de la Méditerranée (Marseille, France) from 2006 to 2011 and since 2012, she joined the Laboratoire de Mécanique et d'Acoustique. Her research mainly focused on sound modeling, perceptual and cognitive aspects of timbre, neuroscience methods and multimodal interactions in the context of virtual/augmented reality. She has been involved in industrial contracts with renowned companies such as Orange Labs and PSA Group. She is a member of the CMMR (Computer Music Multidisciplinary Research) steering committee and is regularly involved in the chair panel of the conference. She published 20 articles in international peer reviewed journals, more than 30 articles in international conference proceedings and co-edited 4 books published by Springer in their series LNCS.

Sølvi Ystad received her degree as a civil engineer in electronics from NTH (Norger Tekniske Hgskole), Trondheim, Norway in 1992 and her Ph.D. degree in acoustics from the Aix-Marseille University, Marseille, France, in 1998. She is currently a researcher at the Laboratoire de Mécanique et d'Acoustique (CNRS: National French Research Center) in Marseille, France. Her research activities are related to sound modeling with a special emphasis on the identification of perceptually relevant sound structures to propose intuitive user interfaces for controlling synthesized sounds. She was in charge of the ANR funded research project Towards the sense of sounds, (http://www.sensons.cnrs-mrs.fr) from 2006-2009 and is currently participating in the ANR funded projects SoniMove (http://sonimove.lma.cnrs-mrs.fr).

Richard Kronland-Martinet has a scientific background in theoretical physics, supplemented by a specialization in acoustics. In 1989 he obtained a Doctorate of Science (Habilitation Research) from the Aix-Marseille University, for his work on analysis, synthesis and processing of sounds by time-frequency and time-scale (wavelet) approaches. Since 1998 he is Director of Research at the Laboratoire de Mécanique et d'Acoustique (CNRS: National French Research Center) in Marseille, France. His scientific activity concerns the science of sounds addressed from a multidisciplinary sight. He has been an important actor in the development of time-scale analysis methods and in their use for analysis, processing and synthesis of sound signals and music. The combination of physical concepts and methods of non-stationary signal processing has enabled him to develop many paradigms of synthesis, especially for instrumental sounds (piano, flute) but also for environmental sounds. His interest in perceptual and cognitive aspects associated with sounds have more recently led him to undertake research on the intuitive control of sounds over the process able to reproduce perceptual effects corresponding to high-level attributes.