Whole-Body Vibration Influences Sound Localization in the Median Plane

Ana Tajadura-Jiménez[1,2], Aleksander Väljamäe[1,2], Norimichi Kitagawa[2] and Hsin-Ni Ho[2]
[1]Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden
[2]NTT Communication Science Laboratories, NTT Corporation, Japan
{ana.tajadura@chalmers.se, aleksander.valjamae@chalmers.se, kitagawa@avg.brl.ntt.co.jp, hsinni@avg.brl.ntt.co.jp}

Abstract

The perceived location of events occurring in a mediated environment modulates the users' understanding and involvement in these events. Previous research has shown that when spatially discrepant information is available at various sensory channels, the perceived location of unisensory events might be altered. Tactile "capture" of audition has been reported for lateral sounds. The present study investigates whether auditory localization on the median plane could be altered by concurrent whole-body vibration. Sounds were presented at the front or the back of participants, in isolation or together with vibrations. Subjects made a three alternative forced choice regarding their perceived location of sound ("front", "back" or "center"). Results indicate that vibrations synchronous with sound affected subjects' sound localization, significantly reducing the accuracy on front sound localization in favor of "back" and "center" responses. This research might have implications for the design of multimodal environments, especially for those aiming at creating a sense of presence or inducing affective experiences in users.

Keywords--- **Multimodal environments, spatial sound, multisensory perception, audio-tactile interaction, localization.**

1. Introduction

In media experiences, such as those delivered by virtual reality simulators, cinema or television, it is often desired to create in the user a perceptual illusion of non-mediation [1] or a sense of *'being there'* [2], an illusion which is often referred to as *'presence'* [3]. Our body is intrinsically related to the experience of presence in mediated environments [4]. Every event occurring in our lives can be evaluated from the perspective of our body, a reference frame which we use to establish our position in space and time in relation with other objects [5]. In line with this view, recent embodiment theories (for instance, see [6] and references therein) emphasize the idea that the body is strongly connected to information processing. Finally, our body constitutes also the physical embodiment of self [4]. Therefore, it can be argued that stimuli that are closely related to our body have a stronger potential to be integrated in a body-image, and this might favor immersion in a particular environment. In this light, physical distance between one's body and events occurring in a mediated environment may modulate one's involvement in that experience. Close is arousing, intimate, engaging [7]. Therefore, when designing a media experience, it is important to take into account that the perceived distance to stimuli will affect users' engagement in it. For instance, in [7] was shown that larger and close displays create a greater sense of presence.

Most mediated environments are multimodal, i.e. they deliver information simultaneously to various sensory modalities. However, research on immersion and presence has been mainly focused on the effect of visual cues, paying less attention to the contributions of other sensory information such as sound or touch. This seems surprising, since most people experience the world not only through their eyes [8]. Fortunately, during the last years there has been a considerable increase of studies exploring the role of non-visual cues in inducing a sensation of presence (for instance, see [8] for a review on sound and presence). In addition to studying the contribution of the different sensory modalities in isolation, it is important to understand how the different sensory cues are perceptually integrated. Our perception has evolved to be multisensory, and research has shown that information available to one sensory modality influences the percepts from other sensory modalities (for recent reviews see [9]).

The concurrent presentation of information at various sensory channels may change the perceived location of events, as compared to the purely unisensory case. A classic example of this is the *'ventriloquist effect'*, where a spatially discrepant sound can be perceived as originating from the same position as the associated visual object ([10]; recent reviewed in [11]). In [12], Bertelson and Aschersleben tested this illusion with a flashing light biasing the perceived location of a synchronous sound. In general, vision has been described as the most accurate of the spatial senses [13]. However, this spatial localization dominance is not as clear when comparing auditory and tactile modalities. Regarding this question, in [14] it was investigated how vibrotactile stimulation can affect sound localization, with a methodology similar to that described in [12]. In particular, in this study, subjects were required to do left-right discriminations regarding their perceived location of sound, when only sound was presented and when sound was presented together with vibrotactile stimulation to subjects' fingertips. Results of this study showed that when sound was presented in synchrony with vibrations, it was "captured" towards the vibration location. The effect disappeared when sound and vibration were presented asynchronously. In addition, the authors noted that this '*tactile capture of audition'* was only observed when the sound spatial position was difficult to localize.

To the best of our knowledge, there have not been studies which report a similar effect on sound localization when sounds are presented in the median plane. In our previous study [15], sounds located close or far from the listener were used in combination with synchronous vibrations presented under the seat. The results suggested that vibration might transform the far sound condition into a close stimulation condition, thus representing a case of 'tactile capture of audition'. Post-experimental verbal probing of participants and experimenters own observations also indicated that vibrations affected the localization in depth. However, this experiment was not designed to specifically test the vibration bias in sound localization. Hence, the question of whether the vibrations contributed to a change in the listener's focus from the outer space towards the self still remains to be answered. Besides, the meaning attributed to the vibrotactile stimulation could also play a potential role in this interaction. In [15] vibrotactile stimuli imitated a heartbeat and participants might have integrated it in a self-representation body image. Therefore, it might be that using other stimuli would result in a different audio-tactile interaction.

In the present study, we aimed at exploring the vibration effect in sound localization for sounds presented in the median plane. Having this knowledge would facilitate the development of more involving multimodal environments, giving the possibility of manipulating the perceived distance and location of events, thus modulating users' engagement and presence in the mediated experience. Participants were specifically tested for their localization of sound presented at front or back in the median plane. Sounds were presented in isolation or together with synchronous or asynchronous vibrations.

In particular, the following hypotheses were tested:

- If vibrations affect subjects' sound localization, one would expect a significant difference in subjects' responses between the unimodal and the synchronous bimodal trials.
- If this effect cannot be accounted simply as a distracting effect of vibrations in the sound localization task, there will not be a significant difference between the unimodal and the asynchronous bimodal trials.
- Moreover, if vibrations are able to "capture audition" [14] one would expect more "center" responses in the synchronous condition than in the other two conditions.

2. Materials and methods

2.1. Participants

Twelve subjects (four female), naïve as to the purpose of the experiment, participated in this study. The mean age of participants was 24 (SD = 4). Participants were paid for their participation and gave their informed consent prior to the beginning of the experiment. The experiment was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and approved by the ethical committee of the NTT Corporation.

2.2. Apparatus and materials

The experiment was conducted in a dark laboratory room where participants seated on a chair, and were blindfolded. The chair was placed on a platform (referred to as "vibrating platform" in the following text) which had two mechanical bass shakers (AURA systems) mounted on it. Two loudspeakers were placed 120 cm away, at the front or back, of the participant, at a height of 125 cm. An extra pair of loudspeakers, which were not active during the experiment (referred to as "fake loudspeakers" in the following text), was placed to the right and left of the experimental chair. Finally, a three-button mouse was used for the discrimination task. The experimental setup is shown in Figure 1.

In condition one, only sound, originated either in the front or in the back loudspeaker, was presented (only-sound trials); in condition two, vibrations were presented in synchrony with sound (synchronous trials); in condition three, vibrations, asynchronous with sound, were presented (asynchronous trials). In the asynchronous trials, the 250-msec vibrotactile and auditory pulses were temporally separated by an interstimulus interval (ISI) chosen randomly among four possible values (-300, -200, +200 or +300 msec), thus resembling the inter-stimuli timing described in [14].

The factorial design contained 3 (sound/vibration conditions: sound-only, vibration synchronous or vibration asynchronous) x 2 (sound position: front or back) stimuli types.

All files (mono files with 48 kHz sampling rate) consisted of three 250-msec, 60-Hz sine-wave pulses, separated by 850 msec of silence (see Method section in [14]). An onset/offset Hann half-window ramp of 10 ms was applied to each pulse to avoid possible clicks. For the auditory stimuli, the sound level was set at 37 dBA. Vibrations were delivered at suprathreshold level. Presentation® software [16] (Version 9.90) was used for auditory-vibrotactile stimuli delivery.

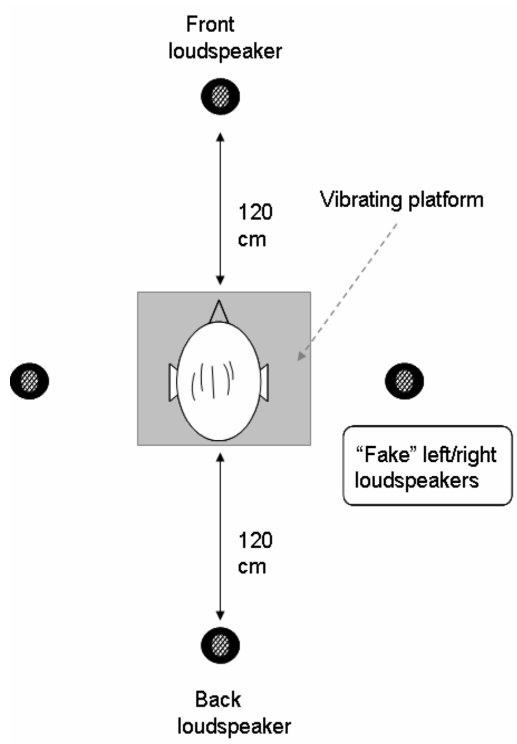

Figure 1 Top-view of the experimental setup

2.3. Procedure

Subjects were required to make a three-alternative forced choice (3AFC) regarding their perceived location of sound ("front", "back" or "center") [17]. The experiment contained 50 repetitions of each of the six stimuli types. Trials were randomized and presented in two blocks of approximately 15 minutes each. Participants did not receive any feedback about their performance.

Participants were instructed verbally on the experimental assignment. They were required to ignore the vibrations and concentrate on performing the three-alternative forced choice based only on the auditory information. Participants were not explicitly told that the sound would be also delivered by the "fake" central loudspeakers, but the fact that they were visible should have insured them that sound potentially could originate on this central position. After being instructed, subjects seated with their feet over the "vibrating platform" and they were blindfolded. A short-training session was carried out (12 trials) and then the actual experiment started. Participants completed the two experimental blocks with a 15-minute break in between. Finally, subjects were debriefed and thanked for their participation.

3. Results

Correct responses for the three-alternative forced choice task were subjected to 3 (sound/vibration condition) x 2 (sound position) repeated measures ANOVA. Alpha level was fixed to 0.05 for all statistical tests. Greenhouse-Geisser correction was used to correct for unequal variances.

Results of the analysis (see Fig. 2) revealed that people were significantly more accurate when sound was presented at their back than at their front ($F(1, 11) = 5.1$, $p < 0.05$). There was also a significant interaction between the sound/vibration condition and sound position ($F(1.7, 19) = 8.3$, $p < 0.005$), which was mainly caused by the fact that in the 'synchronous' trials subjects were significantly less accurate for the 'front', compared to the 'back', sound condition (paired-samples $t(11) = 4.14$, $p < 0.005$). Accuracy for front sound localization was significantly worse in the 'synchronous' than in the 'sound-only' trials ($t(11) = 3.2$, $p < 0.01$); however, there was no significant difference ($p > 0.05$) between the 'sound-only' and 'asynchronous' trials.

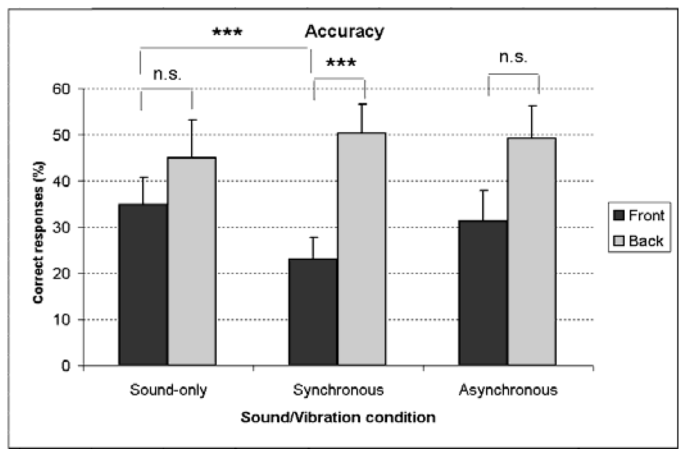

Figure 2 Correct responses (%) for the sound localization task (twelve subjects) consistent with the sound/vibrotactile condition ('sound-only', 'synchronous' vibration and 'asynchronous' vibration) and sound position (front or back) manipulation. The whiskers show the standard errors of the means (* marks significance at p < 0.005 level; n.s. is non-significant).**

A subsequent analysis of the errors in localization was performed (e.g. reporting front or center positions when the sound is presented behind the listener). The comparison of errors shows that the proportion of front-back confusions varied significantly across the three sound/vibration conditions ($F(1,11)=10.4$; $p < 0.01$). The ratio of front-back (FB) to back-front (BF) confusions was highest in the 'synchronous' trials compared to the other two conditions, with means 2.3 ('sound-only'), 3.5 ('synchronous') and 2.7 ('asynchronous'). It is important to note that usually there is a high rate of reversals in the perception of spatial positions of the virtual sources when binaural localization cues are ambiguous (cone of confusion), e.g. front-back confusion when the sound source is presented in the median plane [18]. In general, there is a significant asymmetry between front-back (FB) and back-front (BF) reversals with stronger tendency for FB confusion (roughly 3 times higher than BF) [18].

Figure 3 shows the distribution of responses for the 3AFC (“front”, “back” or “center”) task depending on the sound/vibrotactile condition. It can be seen that the proportion of ‘back’ responses with respect to the ‘front’ and ‘center’ responses dominates in all three sound/vibrotactile conditions ($p < 0.05$ between ‘front’ and ‘back’ responses for all conditions). In addition, the graph shows that when vibrations are presented in synchrony with sound, the proportion of ‘center’ responses is significantly bigger than the proportion of ‘front’ responses ($t(11) = 2.4$, $p < 0.05$). However, this significant difference is not observed in the ‘sound-only’ or ‘asynchronous’ trials.

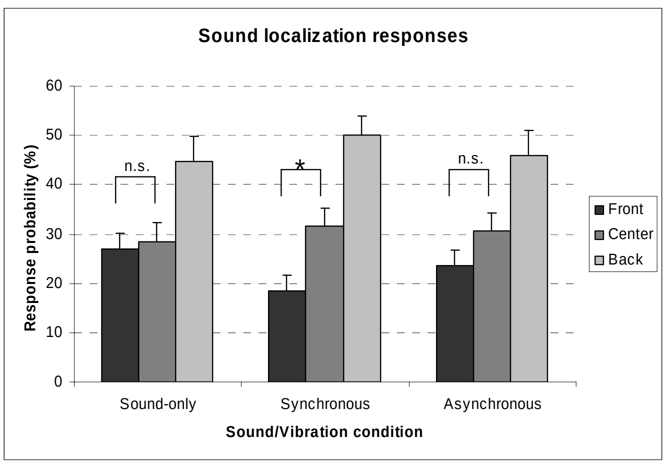

Figure 3 Probability (%) of localizing sound at ‘front’, ‘back’ or ‘center’ position (twelve subjects) consistent with the sound/vibrotactile condition (‘sound-only’, ‘synchronous’ vibration and ‘asynchronous’ vibration) manipulation. The whiskers show the standard errors of the means (* marks significance at p < 0.05 level; n.s. is non-significant). Although it is not indicated in the figure, p < 0.05 between ‘front’ and ‘back’ responses for all conditions.

Therefore, vibrations affected subjects’ sound localization and significantly reduced the accuracy for ‘front’ sound localization, in favor of ‘back’ and ‘center’ sound positions. This effect was only observed when vibrations were presented in synchrony with the sound.

4. Discussion and conclusions

The results of this study are in agreement with the hypotheses proposed in the introduction part. Vibrations affected subjects’ sound localization and significantly reduced the accuracy on ‘front’ sound localization, in favor of ‘back’ and ‘center’ sound localization. This effect was only observed when vibrations were presented in synchrony with the sound, which suggests that it is a case of multisensory integration of sound and vibrotactile information, rather than a distraction effect due to the presence of vibrations during the trial.

These results indicate that the perception of sound location on the median plane is affected by the presence of concurrent vibrotactile information. This effect is consistent with the previous findings for lateral sounds [14].

Nevertheless, further studies are required to test and reveal the nature of the auditory-vibrotactile interaction in the spatial domain. For instance, a different, more refined methodology, could allow approximating the magnitude of the shift in sound location. Another possibility would be to design an experimental paradigm where sound segregation of two auditory streams [19] presented at the same location would only be possible when vibrations are coupled with one of the streams, thus suggesting a tactile spatial capture. In addition, the effect might be different when frequency of vibration or/and sound changes. More interestingly, using stimuli other than pure tones may alter the observed effect. Even though the physical properties of the sound and vibrations undoubtedly play a big role on the reactions induced, other variables related to subjective interpretation and meaning should be considered. Therefore, future research should also include ecologically valid stimuli.

This research results might have important implications for the telepresence community. The simultaneous presentation of audio-tactile information may lead to sound localization confusions and care should be taken when designing such multimodal environments. On the other hand, knowledge about “tactile capture of audition” might provide possibilities to overcome technical limitations in these environment simulations. For instance, in large virtual environments (e.g. a CAVE), where loudspeakers sound reproduction is used, it is often difficult to create a virtual sound source close to the person. However, by adding synchronous vibrations (e.g. using floor vibrations or a vibrating vest [20]), the perceived sound location might be shifted towards the user. The possibility of presenting to a user with events occurring in their close space allows for creating more intimate and intense situations [7]. Therefore, this knowledge might be of special interest for the design of virtual affective environments. Currently, new auditory-vibrotactile experiments in this line of research are being conducted by our group.

Acknowledgements

This work was carried out as part of the PRESENCCIA project, an EU funded Integrated Project under the IST programme (Project Number 27731).

References

[1] M. Lombard, T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer Mediated Communication, 3* (2), [on-line], 1997.

[2] J. Freeman, S.E. Avons, D.E. Pearson, W.A. IJsselsteijn. Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators and Virtual Environments, 8*(1), 1-13, 1999.

[3] M. Meehan, B. Insko, M. Whitton, F.P. Brooks. Physiological measures of presence in virtual environments. In *Proceedings of 4th International Workshop on Presence. Philadelphia*, USA, 21-23 May, 2001.

[4] M. Slater, M. Usoh. Body Centred Interaction in Immersive Virtual Environments. In Magnenat Thalmann N, Thalmann D, eds. *Artificial Life and Virtual Reality.* John Wiley and Sons, pp. 125-148. 1994.

[5] A. Damasio. The Feeling of What Happens: Body and Emotion in the Making of Consciousness. Harcourt Brace & Company. 1999.

[6] P.M. Niedenthal, L.W. Barsalou, P. Winkielman, S. Krauth-Gruber, F. Ric. Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9*, 184–211. 2005.

[7] B. Reeves, C. Nass. The media equation: How people treat computers, television, and new media like real people and places. New York: Cambridge University Press. 1996.

[8] P. Larsson. Virtually hearing, seeing, and being : room acoustics, presence, and audiovisual environments. Doctoral dissertation thesis at Chalmers University of Technology, ISSN 0346-718X. 2005.

[9] C. Spence, J. Driver (Eds.). *Crossmodal space and cross-modal attention*. Oxford: Oxford University Press. 2004.

[10] H.A. Witkin, S. Wapner, T. Leventhal. Sound localization with conflicting visual and auditory cues. *Journal of Experimental Psychology, 43*, 58–67. 1952.

[11] J. Vroomen, B. De Gelder, B. Perceptual effects of cross-modal stimulation: The cases of ventriloquism and the freezing phenomenon. In G. Calvert, C. Spence & B. E. Stein (Eds.), *Handbook of multisensory processes* (pp. 141-150). Cambridge: MIT Press. 2004.

[12] P. Bertelson, G. Aschersleben. Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review, 5*, 482-489. 1998.

[13] R.B. Welch, D.H. Warren. Intersensory interactions. In K. R.Boff, L. Kaufman, & J. P. Thomas (Eds), *Handbook of perception and human performance: Vol. 1. Sensory processes and perception* (pp. 25-1 to 25-36). New York: Wiley. 1986.

[14] A. Caclin, S. Soto-Faraco, A. Kingstone, C. Spence. Tactile "capture" of audition. *Perception & Psychophysics 2002; 64*:616-630.

[15] A. Tajadura-Jiménez, A. Väljamäe, D. Västfjäll (in press). Self-representation in mediated environments: the experience of emotions modulated by auditory-vibrotactile heartbeat. (*to appear in CyberPsychology and Behavior*).

[16] Presentation URL: http://www.neurobs.com/

[17] M. Usher, N. Donnelly. Visual synchrony affects binding and segmentation in perception. *Nature, 394*: 179-182. 1998.

[18] D.R. Begault, 3D Sound for Virtual Reality and Multimedia. Academic Press, 1994

[19] W.J. Macken, S. Tremblay, R.J. Houghton, A.P. Nicholls, D.M. Jones. Does Auditory Streaming Require Attention? Evidence From Attentional Selectivity in Short-Term Memory. Journal of Experimental Psychology:Human Perception and Performance, 29(1): 43–51. 2003.

[20] J.B.F. Van Erp, H. A. H. C. van Veen. A multi-purpose tactile vest for astronauts in the international space station. In I. Oakley, S. O_Modhrain, & F. Newell (Eds.), Proceedings of Eurohaptics 2003 (pp. 405–408). Dublin, Ireland: Trinity College. 2003.