AFFECTIVE MULTIMODAL DISPLAYS: ACOUSTIC SPECTRA MODULATES PERCEPTION OF AUDITORY-TACTILE SIGNALS

Ana Tajadura-Jiménez[1], Aleksander Väljamäe[2], Norimichi Kitagawa[3], & Daniel Västfjäll[1,4]

[1]Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden
[2]Laboratory for Synthetic Perceptive, Emotive and Cognitive Systems (SPECS), @ IUA (Institute of Audiovisual Studies), Universitat Pompeu Fabra, Barcelona, Spain
[3]NTT Communication Science Laboratories, NTT Corporation, Japan
[4]Department of Psychology, Göteborg University, Gothenburg, Sweden

ana.tajadura@chalmers.se; aleksander.valjamae@iua.upf.edu; kitagawa@avg.brl.ntt.co.jp; psydave@psy.gu.se

ABSTRACT

Emotional events may interrupt ongoing cognitive processes and automatically grab attention, modulating the subsequent perceptual processes. Hence, emotional eliciting stimuli might effectively be used in warning applications, where a fast and accurate response from users is required. In addition, conveying information through an optimum multisensory combination can lead to a further enhancement of user responses. In the present study we investigated the emotional response to sounds differing in their acoustic spectra, and their influence on speeded detection of auditory-somatosensory stimuli. Higher sound frequencies resulted in an increase in emotional arousal. We suggest that emotional processes might be responsible for the different auditory-somatosensory integration patterns observed for low and high frequency sounds. The presented results might have important implications for the design of auditory and multisensory warning interfaces.

Keywords: emotions, attention, warning, auditory, somatosensory, multisensory

1. INTRODUCTION

Salient events have the capability of evoking emotional responses. These emotional responses to external stimulation often occur in early stages of stimulus processing, automatically and prior to awareness, modulating the subsequent attentional and perception processes (e.g., [1][2][3]). Emotional events might interrupt the current cognitive focus by eliciting an attentional or behavioral switch towards these events. For instance, in visual dot-probe tasks facilitation in reaction time is observed when the target (a dot-probe) appears after a short-time interval at the same location than emotional stimuli (e.g., an angry face in [4]). Emotional arousal may be transferred also across sensory modalities. For instance, threatening visual stimuli presented close to hands have been shown to attract tactile spatial attention and fasten responses to that modality [5]. Hence, incorporating emotional eliciting stimuli which are able to trigger intuitive responses in users might be beneficial for the design of warning interfaces and/or applications such as those incorporated in vehicles (e.g., aircrafts or cars), emergency systems (e.g., in hospitals) or working environments.

Nowadays, sounds are frequently used in many of these scenarios, for instance, in the form of earcones (abstract musical sounds) or auditory icons (a sound caricature of the intended action the user is supposed to take or has taken) [6]. However, in most of the cases the association between the warning system and the actual event has to be learnt. It is still not completely understood which sounds are more effective in conveying different forms of alerts and warnings in order to attract users' attention and obtain a fast and appropriate response from them. Sounds which carry affective information (e.g., danger) might have this capability of generating an automatic response in users which switch their current focus of attention to this new event.

Human perception has evolved to become a multisensory process. Most of the events taking place in our lives provide information simultaneously to several sensory modalities (e.g., when looking and listening to another person speaking) which our perceptual system integrates to form unified multisensory perceptual events [7][8][9] (for a recent review in multisensory integration see [10]). In general, people tend to respond faster to information simultaneously available at various sensory modalities as compared to unisensory stimuli. This is often referred to as Redundant Signals Effect (RSE), and this response is especially fast if the multisensory information is integrated at a neural level (e.g., [11]). Therefore, multisensory interface systems which integrate information from various sensory modalities might enhance the efficiency in producing a response from users with respect to the unisensory interfaces. For example, in [12] a multisensory interface which combined auditory, visual and vibrotactile spatial warning cues was investigated in search of the most effective combination in directing car drivers' visual spatial attention towards target events; they found that vibrotactile cues were particularly effective for this application. For situations of high information load in one sensory modality, using a different sensory channel, or a combination of different channels, to convey warning information can enhance subsequent responses. Therefore, understanding how the different sensory modalities interact with each other, and how emotion and the perception of saliency are transferred between sensory modalities, might help to enhance the efficiency of multisensory interface systems, since some particular combinations can lead to faster responses. Unfortunately, most of the research investigating the multisensory integration of auditory information with other modalities has been mainly reduced to the use of white noise as auditory stimuli. However, if we attend to the emotional theories mentioned above, it can be hypothesized that multisensory integration might be facilitated for more salient stimuli.

In the following section we describe a study where the effect of presenting redundant auditory-somatosensory information on speeded detection was investigated by using auditory stimuli with different acoustic spectra. Section 3 explores emotional responses to sounds differing in acoustic spectra. In section 4, the influence of the perceived saliency of auditory and auditory-somatosensory events in directing spatial attention is discussed. The outcome of this article might have implications for the design of multimodal warnings.

2. ACOUSTIC SPECTRA INFLUENCES AUDITORY-SOMATOSENSORY MULTISENSORY INTEGRATION

We conducted a study on the spatial modulation of auditory-somatosensory multisensory interactions in the region close to head. Spatial proximity of stimuli from two different modalities has been shown to be one of the factors which may facilitate the multisensory integration of audiovisual (e.g., [13]) and visuotactile (e.g., [14]) information. However, auditory-somatosensory studies on this topic have found contradicting results and it continuous being an open research question whether spatial proximity can facilitate auditory-somatosensory integration. In two experiments using white noise as auditory stimuli [15] it was found that a facilitation in multisensory integration occurs for auditory and somatosensory stimuli presented from the same versus different sides of participants' head provided that the auditory stimuli was presented at close distance from participants' head and that somatosensory stimuli was delivered to the head (but not to the hands). Therefore, the spatial modulation of auditory-somatosensory multisensory interactions was dependant on the distance to auditory stimuli and the stimulated body surface. In the following study, we investigated whether this modulation is also dependant on the acoustic spectra.

Low frequency (100-920 Hz) and high frequency (14-17 kHz) noise bursts were used as auditory stimuli. Sound was presented close to participants' head (20 cm) and somatosensory stimuli (electrocutaneous stimulation) was delivered to either of the participants' earlobes. Participants performed a speeded simple detection task to single auditory, somatosensory or double simultaneous auditory-somatosensory stimuli presented from the same versus different sides from the participants head. The experimental setup is shown in Figure 1 (for a more thorough description of this study see [15]

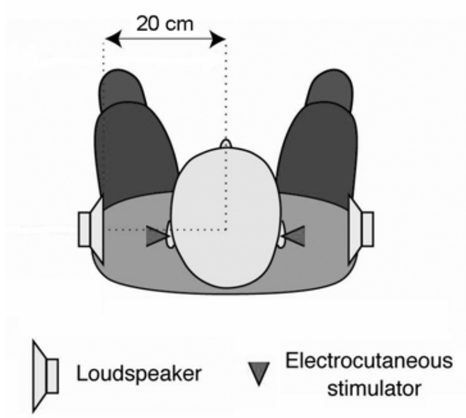

Figure 1. *Experimental setup. Auditory stimulation was delivered from either left or right loudspeakers, and electrocutaneous stimulation from either left or right earlobes.*

The results of this study (see Figure 2) suggest that acoustic spectra might influence the way multisensory information is integrated. Although bimodal information led to faster reaction times (RTs) compared to the unisensory conditions for both types of auditory stimuli (low and high frequency), a further facilitation of presenting auditory and somatosensory stimuli from the same versus different sides of the head was only observed for the high frequency condition. In other words, auditory-somatosensory integration was spatially modulated only for the high frequency sound conditions.

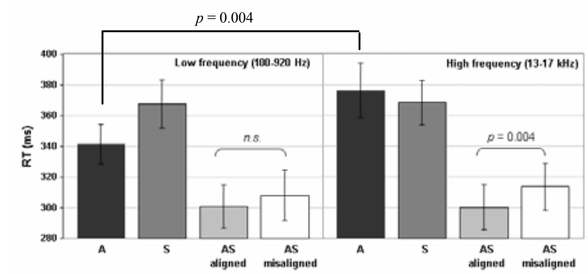

Figure 2. *Mean reaction times (in ms) for the conditions with low frequency sound (left panel) and high frequency sound (right panel). The error bars show the standard error of the means.*

Another disparity was also found in the auditory-alone conditions (see Figure 2). RTs were significantly slower for the high frequency sound than for the low frequency sound (p = 0.004). It might be suggested that this difference is due to the high frequency sound being more difficult to detect than low frequency sound. However, there was no significant difference in percentage of detections between both auditory-only conditions.

The asymmetry in the auditory-somatosensory integration for the different frequency bands of the auditory stimuli might be accounted to different mechanisms. One possible explanation might be found in the emotional or ecological psychology theories. If emotional processing occurs for one of the auditory frequency bands, it might modulate the multisensory integration between auditory and somatosensory information. It may be also suggested that the high frequency sound and the electrocutaneous stimulation form a more salient combination and thus the multisensory integration is favored for spatially coincident stimuli. From auditory research, it is known that sounds with different frequencies may elicit different emotional responses in listeners. In particular, some studies have reported that low frequency sounds are more pleasant than high frequency sounds below 60 dB [16], and that 'sharpness' (i.e., high frequency components) correlates with emotional arousal [17]. Thus, auditory-somatosensory combinations which differ in the type of sound might also elicit different emotional responses which, in turn, influence subsequent perception processes.

Finally, it might be also suggested that the low frequency sounds were more difficult to localize, or that different localization cues are integrated with somatosensory information in a different manner, and that these effects accounted for the lack of spatial modulation of the multisensory integration for the low frequency case. Signals in the frequency range up to approximately 1.5 kHz are lateralized attending to the interaural temporal disparities (ITDs, or differences in the time-of-arrival of sounds at the two ears), while interaural intensity disparities (IIDs) are used at higher frequencies [18].

The following study was designed to explore whether different emotional responses are elicited by sounds differing in their acoustic spectra. Emotional processes might be responsible of the asymmetry in the RTs between high versus low frequency bands. In addition, localization issues were investigated into more detail.

3. AFFECTIVE COLORATION OF ACOUSTIC SPECTRA

In the present study we wanted to investigate the emotional responses to auditory stimuli (band-pass filtered noise bursts) of different frequency bands. We hypothesized that high frequency sounds were more salient in our context, and formed a more natural combination with the electrocutaneous stimulation. If that was the case, presenting somatosensory stimuli at the same location than the highly arousing sounds could have generated a fast response from participants, as part of a defensive behavior [19].

In addition, we explored the capability of participants to lateralize the sounds at different auditory frequencies.

3.1. Methods

3.1.1. Participants

Eight postgraduate students (mean age 26 years; age range from 23 to 29 years; one female) voluntary took part in the study. All of the participants had normal hearing and were naïve as to the purposes of the study. Participants gave their informed consent prior to the beginning of the experiment. The current study was conducted under approval of the local ethics committee.

3.1.2. Apparatus and materials

The experiment was conducted in a laboratory room with the participants seated in a chair. Participants positioned their head on a chin rest when listening to the sounds.

Two identical loudspeakers (GENELEC 1029A – Active motor) were placed 20 cm to the left and right of the center of the participants head, at ear level. An extra loudspeaker, which was not active during the experiment (referred to as "fake loudspeaker" later in this text), was placed at the front of the participants head. The auditory stimuli (mono files with 48 kHz sampling rate) consisted of band-pass filtered noise bursts of 50 ms duration normalized to the same loudness (50 dB(A) as measured from participants' ear position). Four types of auditory stimuli were used which differed in their frequency range: f1 (100-400 Hz; [f1.WAV]), f2 (920-1480 Hz; [f2.WAV]), f3 (2700-4400 Hz; [f3.WAV]) and f4 (9500-23000 Hz; [f4.WAV]). Each of these frequency ranges correspond to three critical bands of hearing, attending to the Bark scale, which is based on equal critical bandwidths. f1, f2, f3 and f4 were specially chosen to be separated an equal distance in the Bark scale, and to be localized by participants by means of different interaural attributes. f1 and f2 are lateralized using ITD cues, while f3 and f4 by means of IID. In addition, the two stimuli with "low" and "high" frequencies (100-920 Hz and 14000-17000 Hz, respectively; [low.WAV] and [high.WAV]) used in the previous experiment (Tajadura et al., submitted) were also included in this study. An onset/offset half-Hanning window ramp of 10 ms was applied to avoid clicks and clipping.

Two circular electrodes (diameter 2 cm; referred to as "fake electrocutaneous stimulators" later in this text), which were not active during the experiment, were attached to the lower tip of the participants' left and right earlobes.

Finally, a small display and keyboard were used for collecting the participants' responses. Presentation® software (Version 11.3) controlled stimuli delivery and recorded responses.

The experimental setup is shown in Figure 3.

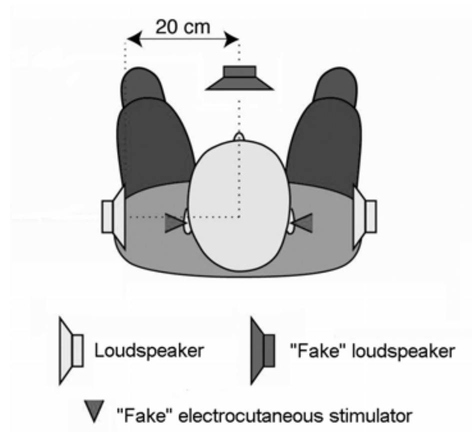

Figure 3. *Experimental setup. Auditory stimulation was delivered from left or right loudspeakers. Electrocutaneous stimulators and the front loudspeaker were not active during the experiment.*

3.1.3. Design

During the experiment, auditory stimuli were presented from one of the two active loudspeakers (left or right). The factorial design contained 6 frequency ranges (f1, f2, f3, f4, 'low' and 'high' frequency ranges) x 2 sound positions (left or right) conditions. Thus, the experimental design involved twelve different types of trials.

3.1.4. Procedure

At the start of the experimental session, participants seated, the "fake electrocutaneous stimulators" were positioned and written and verbal instructions on the experimental assignment were given. Participants were required to listen to the sounds and after each stimulus rate their affective reaction in terms of valence (i.e. pleasure; positive versus negative) and arousal (i.e. activation; excited versus calm) to the stimulus by using the Self-Assessment manikin (SAM) developed by Lang [20]. These two dimensions, valence and arousal, allow to place emotional responses in a two-dimensional affective space, following a dimensional approach to emotions [19][21][22]. Specifically, for the valence dimension, SAM 9-point pictorial scale ranges from a figure showing a wide smile (rated as 9) to a frowning figure (rated as 1). On the other hand, for the arousal dimension, SAM scale varies from a highly excited (rated as 9) to a low excited, sleepy figure (rated as 1). SAM scale was displayed on the screen and participants rated picture valence and arousal by using a keyboard.

In addition, participants were required to make a three-alternative forced choice (3AFC) regarding their perceived location of sound ("left", "front" or "right"). This question was included in order to test if auditory stimuli were localizable. Participants were not explicitly told that the sound would be also delivered by the "fake" front loudspeaker, but the fact that it was visible should have insured them that sound potentially could originate on this central position.

Participants were told that in some of the trials sound would be accompanied by electrocutaneous stimulation to their earlobes. Although the stimulators were not active during this experiment, they were included in the setup in order to keep the same context than in the previous study.

The experiment contained 3 repetitions of each stimulus configuration. The resulting 36 experimental trials were randomized and presented in a block which lasted for an average duration of 15 minutes. After participants carried out a short-training session (2 trials) to familiarize with the paradigm, they completed the experimental block. Finally, they were debriefed and thanked for their participation.

3.2. Results

Self-reported valence and arousal (SAM ratings) were used as dependent variables for a multivariate analyses of variance (MANOVA) with factors 6 frequency ranges (f1, f2, f3, f4, 'low' and 'high' frequency ranges) x 2 sound positions (left or right). Alpha level was fixed at 0.05. Wilks' Lambda was used as the multivariate criterion. The results of this analysis (see Figure 4) showed that there was a significant main effect of frequency range on self-reported emotional experience (Wilks' $\Lambda = 0.57$, $F_{(10,\ 68)} = 2.22$, $p = 0.028$). Independent univariate tests (Greenhouse-Geisser correction) for valence and arousal showed that the effect of frequency range was only significant for the arousal ratings ($F_{(3,\ 21)} = 3.95$, $p = 0.022$). Increasing the frequency range was translated in a significant increase in self-reported arousal. The within-participant polynomial contrast with frequency range as factor revealed a significant trend for the arousal dimension ($F_{(1,\ 7)} = 12.4$, $p = 0.01$).

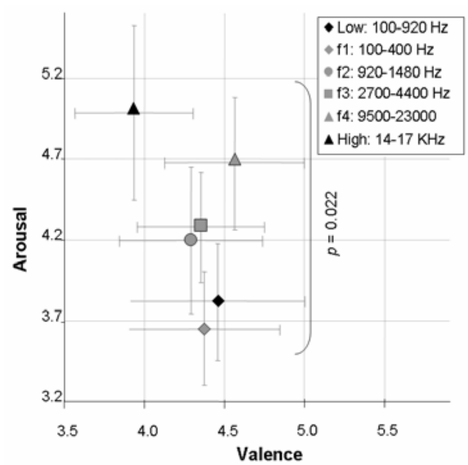

Figure 4. *Effect of frequency band on participants' emotional response to auditory stimuli. Emotional response is represented in a two-dimensional space with valence and arousal as coordinates. Valence and arousal are rated in a 9-point scale. The error bars show the standard errors of the means.*

In addition, we tested the localization performance across conditions. Participants were able to lateralize sounds of all frequency ranges. There were only two mislocalization errors across all participants and conditions (two different participants made one localization error), which were made in the "high" frequency condition, but this did not lead to a significant effect.

4. DISCUSSION AND CONCLUSIONS

The results of this study show a significant linear trend with sounds with a higher frequency range evoking a more intense emotional experience in listeners. Auditory stimuli in the high frequency band might be perceived as more threatening or, in other words, ecologically salient. In addition, high frequency sounds seem to form a more natural combination with the electrocutaneous stimulation; in words of one of the participants in this study "if I felt electrical stimulation was together with the high frequency sounds". Lateralization results allow discarding the hypothesis that the observed multisensory interaction for high but not for low frequency sounds can be rooted in the spatial hearing mechanisms.

These results, together with those on auditory-somatosensory interactions previously reported (Tajadura et al., submitted), suggest that the perceived saliency of auditory and auditory-somatosensory events might be altered by acoustic spectra. They also suggest that more alerting events not always lead to faster responses. In this case, although the high frequency sounds made people felt more arouse, they led to a delay in participants' responses. In some situations salient events may perturb the cognitive processes involved in the performance of the task (e.g., [23]). However, an optimum multisensory combination can fasten the response, as in the case of presenting high frequency sound with electrocutaneous stimulation from the same location.

This research might apply to the design of emergency and warning applications which intend to attract users' attention and generate a fast response from them. A careful design of multimodal environments including auditory and somatosensory information may provide an optimum way of eliciting an automatic response in users. Nevertheless, further studies are required to understand the role of each sensory modality and how the modalities might interact between each other in such displays. A very important factor in such studies would be the contextual information and meaning attributed by users to particular uni- and multimodal stimulation in order to find the most effective sensory combinations.

5. ACKNOWLEDGMENTS

This work was carried out as part of the PRESENCCIA project, a European Union–funded Integrated Project under the IST Program (Project Number 27731). We thank all participants in the study.

6. REFERENCES

[1] A. Öhman, "Fear and anxiety as emotional phenomena: Clinical phenomenology evolutionary perspectives, and information-processing mechanisms", in M. Lewis and J.M. Haviland, *Handbook of emotions*, pp. 511-536. Guilford, New York, 1993.

[2] E.A. Phelps and J.E. LeDoux,, "Contributions of the amygdala to emotion processing: From animal models to human behavior.," *Neuron*, vol. 48, pp. 175-187, 2005.

[3] E.A. Phelps, S. Ling and M. Carrasco, “Emotion facilitates perception and potentiates the perceptual benefits of attention”, *Psychological Science,* vol. 17, pp. 292-299, 2006.

[4] B.P. Bradley, K. Mogg and N.H. Millar, “Covert and overt orienting of attention to emotional faces in anxiety”, *Cogn. Emotion*, vol. 14, pp. 789-808, 2000.

[5] E. Poliakoff, E. Miles, X. Li and I. Blanchette, “The effect of visual threat on spatial attention to touch”, *Cognition*, vol. 102, pp. 405-14, 2007.

[6] A. Sköld, P. Bergman, D. Västfjäll, A. Tajadura-Jiménez and P. Larsson, “Emotional Reactions to Information and Warning Sounds”, *Acta Acustica* (submitted).

[7] R.B., Welch and D.H. Warren, “Intersensory interactions”, in K.R. Boff, L. Kaufman and J.P. Thomas, *Sensory processes and perception. Handbook of perception and performance*, vol. 1, pp. 25-1 - 25-36. Wiley, New York, 1986.

[8] A.A. Ghazanfar and C.E. Schroeder, “Is neocortex essentially multisensory?”, *Trends in Cognitive Sciences*, vol. 10, pp. 278-285, 2006.

[9] C. Spence, “Audiovisual multisensory integration”. *Acoustical Science and Technology*, vol. 28, pp. 61-70, 2007.

[10] C. Spence and J. Driver, *Crossmodal Space and Cross-modal Attention*. Oxford University Press, Oxford, 2004.

[11] J. Miller, “Discrete versus continuous stage models of human information processing: In search of partial output”, *Journal of Experimental Psychology: Human Perception & Performance*, vol. 8, pp. 273-296, 1982.

[12] C. Ho, C. Spence and H.Z. Tan, “Warning signals go multisensory”, in *Proc. the 11th Annual Conference on Human-Computer Interaction*, Las Vegas, Nevada, July 22-27, 2005.

[13] L.K. Harrington and C.K. Peck, “Spatial disparity affects visual-auditory interactions in human sensorimotor processing”, *Experimental Brain Research*, vol. 122, pp. 247-252, 1988.

[14] A. Diederich, H. Colonius, D. Bockhorst and S. Tabeling, “Visual-tactile spatial interaction in saccade generation”, *Experimental Brain Research,* vol. 148, pp. 328-337, 2003.

[15] A. Tajadura-Jiménez, N. Kitagawa, A. Väljamäe, M. Zampini, M.M. Murray and C. Spence, “Auditory-tactile interactions are spatially modulated by stimulated body surface and acoustic spectra”, *Brain Research* (submitted).

[16] P.C. Vitz, “Preference for tones as a function of frequency and intensity”, *Perception & Psychophysics*, vol. 11, pp. 84-88, 1973.

[17] D. Västfjäll and M. Kleiner, “Emotion in Product Sound Design”, in *Proc. of Journées Design Sonore*, Paris, France, March 20-21, 2002.

[18] J. Blauert, Spatial hearing: The psychophysics of human sound localization. MIT Press, Cambridge, MA, 1997.

[19] P.J. Lang, “The Emotion Probe: Studies of Motivation and Attention”, *American Psychologist*, vol. 50, pp. 372-385, 1995.

[20] P.J. Lang, “Behavioral treatment and bio-behavioral assessment: computer applications”, in J.B. Sidowski, J.H. Johnson and T.A.Williams, *Technology in Mental Health Care Delivery Systems,* pp. 119-137. Ablex Publishing, Norwood, NJ, 1980.

[21] C. Osgood, G. Suci, and P. Tannenbaum, *The Measurement of Meaning*. University of Illinois Press, Urbana, IL, 1957.

[22] J.A. Russell, “The Circumplex Model of Affect”, *Journal of Personality and Social Psychology*, vol. 39, pp. 1161-1178, 1980.

[23] C.F. Zink, G. Pagnoni, J. Chappelow, M. Martin-Skurski and G.S. Berns, “Human striatal activation reflects degree of stimulus saliency”, *Neuroimage,* vol. 29, pp. 977-983, 2006.