Assessing audio clips on affective and semantic level to improve general applicability

Robert Schleicher, Shiva Sundaram, Julia Seebode

Deutsche Telekom Laboratories, TU Berlin, 10587 Berlin, E-Mail: firstname.lastname@tu-berlin.de

Introduction

The relation of emotion and sound has been the topic of research for a while. The relevance of such studies is immediately evident in the case of music, which most people frequently listen to to modulate their current mood. The interested reader is referred to the works of [1] to get an overview about the current state of the art in that area. Compared to music, environmental sounds tend to be more effective in stimulating mental imagery [2]. We contend that studies of emotional impact of general environmental sounds can also be beneficial as it can facilitate to convey implicit information as auditory icons.

Utilizing the affective impact of sound

Emotional stimuli are attended faster and longer than non-emotional ones [3], thus making them a good candidate to initially attract attention of a user.

The most prevalent approach to efficiently assess the subjective emotional impact of a stimulus in psychology is presumably a rating on the dimensions of valence (*unpleasant – pleasant*), arousal (*calm – aroused*), and dominance (*being controlled – being in control*) using a cartoon-like rating scale, the *Self-Assessment-Manikin (SAM)* [4]. It has been successfully used for auditory stimuli by the same authors [5] and adopted for various applied settings like evaluating the annoyance of operating noise [6] or the effectiveness of warnings in vehicles [7, 8].

[8] point out that the urgency of an auditory feedback can be characterized by a combination of negative valence and high arousal, whereas for semantic interpretability perceptual qualities like realism are desired. This latter aspect might need further examination if one attempts to extend the application to other non-warning contexts, for example, an auditory icon for confirmation of an information or notification of an ongoing process.

Adding meaning to affective impact

While the SAM is an established measure to rate all kinds of stimuli, the representation on the axes of valence and arousal can have the disadvantage that material that causes distinct emotions may be grouped quite close in that coordinate system an thus become indistinguishable (e.g. anger and fear which both go along with high arousal and unpleasantness [9, 10]). This may limit the later use of auditory clips rated this way, which is why additional information may be necessary to describe a potential auditory feedback. Based on the work of [11, 12], [13] utilize free text description of users to assess the attributed meaning of an auditory icon, which is then analyzed with an entropy-like measure.

Another way to analyze this kind of data is to exploit the latent information in text using a popular technique known as *latent semantic analysis (LSA)*. It inherently relies on the weights based on the occurrence of terms or words (with semantic value) in a document and across documents to represent the inherent but unobservable semantic structure in them. The representation subsequently derived through this analysis can be used to classify text descriptions. This approach of indexing using term-document frequency has also been successfully extended to example-based classification of audio material [14].

Both approaches have in common that they try to go beyond pure signal characteristics analysis and try to consider the specific meaning a certain sound has for the listener which might be inevitable for a complete emotional evaluation, as emotions are always *about something* and many audio clips in the *International Affective Digitized Sound (IADS)* database [15] depict a short scenery whose interpretation then leads to the emotional reaction.

Directions of research

The previous paragraphs already frame the direction of our current research: considering both basic affective information as well as content description to achieve a comprehensive understanding of impact of auditory scenes or background noise, which then improves the design of complex auditory displays and icons. So far applied research that considers affective ratings for auditory stimuli appears to be characterized by two properties:

- a focus on negative emotional valence, i.e. urgency of alarms [7, 8] or annoyance of noise [6]
- evaluation of a set of sounds that were designed or pre-selected for a specific context

The first aspect can also be noted in the development of *human-computer interaction (HCI) research* in general, which was for a long time mostly focussing on eliminating potential usability flaws or annoyances before turning towards factors that actually cause positive reactions in the user [16]. The second aspect might be caused by the constraints of product developing processes. We try to go the other way by letting subjects rate a standard database of sound effects without imposing a pre-defined context. Exploratory analysis of such an extensive database may help to find groups of auditory stimuli that cover all combinations in the affective space including *calm & pleasant* or *moderately arousing without a high affective valence*. We believe that there is also a need for these less persistent feedbacks in the design of auditory displays to inform the user that a certain process is going on (like charging) without distracting him or her from the main task. To find appropriate sounds for such non-interactive state of the system requires a certain variety of sounds with different emotional saliencies while still maintaining clear semantic meaning. Real-life sounds appear to have a higher intuitiveness and memorability for these purposes [17]. Ideally, several sounds of the same class (i.e. the sound of filling water in a vessel for the aforementioned example of charging), but with different urgency, could allow the user to choose the level of conspicuousness he or she considers appropriate for a certain feedback instead of deactivating it altogether [18].

Next to auditory feedback design, research on noise disturbance may also benefit from an extensive data set of this kind: instead of trying to minimize the annoyance of a certain sound, e.g. operation noise, one could try to change it into mimicking a less unpleasant natural sound, given that you have knowledge about which sounds are less obtrusive or easily re-appraised as something neutral due to their assumed origin.

On the long run, we plan to extend our approach of combining low-level affective measures and high-level semantic descriptions to other multimedia material. The motivation attributed to deliberate music exposure in the introduction, conscious emotion regulation, surely also holds for accessing databases like youtube, which are frequently visited to browse for and share 'touching' video clips. So far, the material there is mostly labelled based on content description, less with regard to its emotional content. The already available user tags may facilitate the semantic analysis, but which signal features are indicative of the affective quality is still an open question and may even vary for different types of video content.

Literature

[1] Juslin, P.N. & Sloboda, J.A., eds. Music and Emotion. Oxford University Press, New York, NY, 2001

[2] Gomez, P. & Danuser, B.: Affective and physiological responses to environmental noises and music. International Journal of Psychophysiology, (2004), 53(2), 91-103

[3] Lang, P.J., Bradley, M.M. & Cuthbert, B.C.: Motivated Attention: Affect, Activation, and Action, in Attention and orienting: Sensory and motivational processes, P.J. Lang, R.F. Simons, and M.T. Balaban, Editors. Lawrence Erlbaum Associates: Mawah, NJ. pp. 97-135, 1997

[4] Bradley, M.M. & Lang, P.J.: Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. Journal of Behavior Therapy and Experimental Psychiatry, (1994), 25(1), 49-59

[5] Bradley, M.M. & Lang, P.J.: Affective reactions to acoustic stimuli. Psychophysiology, (2000), 37, 204-215

[6] Västfjäll, D., Kleiner , M. & Gärling, T.: Affective Reactions to Interior Aircraft Sounds. Acta Acustica united with Acustica, (2003), 89, 693-701

[7] Larrson, P., Opperud, A., Fredriksson, K. & Västfjäll, D., Emotional and behavioral response to auditory icons and earcons in driver-vehicle interfaces (2009), URL: www-nrd.nhtsa.dot.gov/pdf/esv/esv21/09-0104.pdf

[8] Bergman, P., Skold, A., Vastfjall, D. & Fransson, N.: Perceptual and emotional categorization of sound. The Journal of the Acoustical Society of America, (2009), 126(6), 3156-67

[9] Stevenson, R.A. & James, T.W.: Affective auditory stimuli: characterization of the International Affective Digitized Sounds (IADS) by discrete emotional categories. Behavior Research Methods, (2008), 40(1), 315-21

[10] Schleicher, R. & Galley, L.: Continuous Rating and Psychophysiological Monitoring of Experienced Affect While Watching Emotional Film Clips. Psychophysiology, (2009), 46(1), S51

[11] Gaver, W.W.: Auditory Icons: Using Sound in Computer Interfaces. Human-Computer Interaction, (1986), 2, 167-177

[12] Ballas, J.A.: Common Factors in the Identification of an Assortment of Brief Everyday Sounds. Journal of Experimental Psychology: Human Perception and Performance, (1993), 19(2), 250-267

[13] Fernström, M., Brazil, E. & Bannon, L.: HCI Design and Interactive Sonification for Fingers and Ears. IEEE Multimedia, (2005), 12(2), 36-44

[14] Sundaram, S. & Narayanan, S.: Classification of sound clips by two schemes: using onomatopoeia and semantic labels. IEEE International Conference on Multimedia and Expo, (2008), 1341 - 1344

[15] Bradley, M.M. & Lang, P.J., The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. 2007, University of Florida: Gainesville, FL.

[16] Schleicher, R. & Trösterer, S.: The 'Joy-of-Use'-Button: Recording Pleasant Moments While Using a PC, in Human-Computer Interaction - INTERACT 2009. Springer: Heidelberg, 2009

[17] Garzonis, S., Jones, S., T., J. & O'Neill, E.: Auditory Icon and Earcon Mobile Service Notifications: Intuitiveness, Learnability, Memorability and Preference. Proceedings of the 27th international conference on Human factors in computing systems, (2009), 1513-1522

[18] Block, F.E., Nuutinen, L. & Ballast, B.: Optimization of alarms: a study on alarm limits, alarm sounds, and false alarms, intended to reduce annoyance. Journal of Clinical Monitoring and Computing, (1999), 15, 75-83