Coding Emotions with Sounds

Nadja Schinkel-Bielefeld[*#1], Frederik Nagel,[*#2]
[*]*Fraunhofer Institute for Integrated Circuits IIS, Germany*
[#]*International Audio Laboratories Erlangen, Germany*

[1]nadja.schinkel-bielefeld@iis.fraunhofer.de, [2]frederik.nagel@iis.fraunhofer.de

ABSTRACT

Background

Emotions play a fundamental role in human communication. They are important in the arts, and particularly music and films are capable of eliciting emotions. However, also less artistic sounds encountered in different situations in every-day life are able to elicit strong emotions.

Apart from the stimulus itself there are several additional factors influencing the elicited emotion, for example the mood and personal or cultural biases of the person that is presented with this stimulus (Gabrielsson 2001, Scherer 1991). Thus it is not clear, how consistently the emotions an artist intended to communicate are perceived by different persons or even by just one person at different times when he/she might be in different moods.

Emotions can be mapped onto a three dimensional emotion-space spanned by the dimensions valence/pleasure, arousal, and dominance. To assess the emotions elicited by an image, a sound or a video, one usually relies on the self-report of subjects who rate the emotions associated with the stimulus in these three dimensions. Because emotions unfold and vary over time, they can only be fully captured by continuously tracking them rather than just looking at a single rating after the stimulus presentation.

Aims

The present study assesses the potential of emotion eliciting sounds for communicating emotions. We wanted to find out a) how consistently and reliably one subject associates the sounds with a certain valence and arousal emotion over repeated presentation and b) how similarly the sounds are perceived by different subjects. This includes the question how consistently emotions unfold and vary over time. Furthermore we analyzed the role of context for the perception of these sounds.

Method

For our study we used sounds from the IADS2 (International Affective Digitized Sounds, 2nd Edition) database (Bradley & Lang 2007). These complex natural sounds span a wide range of semantic content from music over mechanic and animal noises to human screams or sobs. Sounds resembling narratives about changing situations which accordingly elicited clearly changing emotions over time were discarded (i.e. a bike bell followed by screeching tires and a scream).

The sounds, presented in isolation, have been rated in dominance, valence, and arousal by more than 100 subjects. Here, we neglected dominance as for music and sounds it is highly correlated with arousal. We were particularly interested in the context dependence of the sound perception. Thus we created sequences of sounds that continuously changed every 5 seconds. For some parts of the sequence the valence and arousal of sounds according to Bradley and Lang varied only gradually, for others it changed randomly and also abrupt changes were possible. Gradually and randomly changing sequences did not share the same sounds. Half of the subjects listened to the gradually changing sequence first. The other half first heard the abruptly changing sequence. Both sequences were also played backwards, meaning we reversed the order of the sounds in the sequence in order to check if the elicited emotions depend on previously played sounds.

For measuring the triggered valence and arousal in real time, we used EMuJoy (Nagel et al 2007), a software for the assessment of emotional self-report to music, pictures or videos. Subjects were instructed to move the mouse in the emotion plane to indicate the emotion they associate with the played sounds. Listeners were trained for 10 minutes with stimuli which also included the sounds most extreme in valence and arousal. This ensured that subjects were not only familiarized with the paradigm but also with the range of elicited emotions. Afterwards they were presented with a 10 minutes stimulus that contained only sounds which were not used during training. In order to assess intra-rater reliability subjects rated the forward sequences twice on the first day and once on the following day. Backward sequences were only presented on the first day.

Results and Discussion

We performed experiments with six male and six female listeners between the age of 20 and 50. Though Perez (1998) showed that emotions can be elicited rather quickly and 250 ms are sufficient to distinguish happy from sad music, in our study subjects reported that especially in the beginning it took effort and concentration to keep up with the speed of the changing sounds. We used Spearman correlations as a measure how consistent elicited emotions were for different subjects. The median inter-rater reliability of all pairs of listeners and all sequence presentations was 0.60 (IQR: 0.23) for valence and 0.52 (IQR: 0.27) for arousal (Fig. 1, left). Especially for arousal there was a large increase of correlation during training. After familiarization with the paradigm further training of a new set of stimuli on that day had no significant effect. However, for rapidly varying sequences valence and arousal increased significantly from the first day to the second day ($p < 0.005$). From visual inspection of histograms we found that the data was not distributed symmetrically. Thus we used bootstrapping (Efron 1979) and permutation testing for paired samples for significance tests. Given that we did not present any new sounds on the second day, it is hard to say whether the increase in correlation is due to training of the paradigm or due to better familiarization to the sounds.

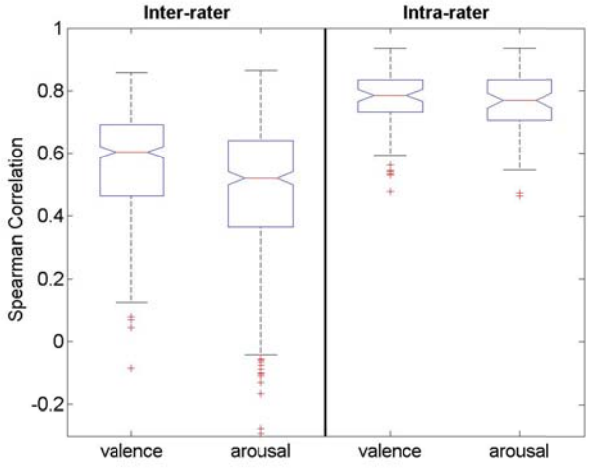

Figure 1: Spearman correlations for valence and arousal between different subjects (left) and between different trials of the same subjects (right).

Median intra-rater correlations of all subjects and all presented sequences was 0.79 (IQR: 0.13) for valence and 0.77 (IQR: 0.10) for arousal (Fig. 1, right). This seems to be in line with the findings of Schubert (2012) who found 80% test-retest reliability for musical stimuli. In our study correlations for valence and arousal increased significantly with repeated presentation ($p < 0.008$). Intra-rater reliability for arousal was stronger for smoothly varying sound sequences than for rapidly varying ones. However, as stimuli were chosen randomly for both sequences, it is not clear to what extent this effect might be caused by the particular sounds used in both sequences.

No significant difference ($p > 0.05$) was found between intra-rater correlation and the correlation between the sequence and the same sequence in reversed order, when adjusting for reaction time and reversing the recorded valence and arousal values for one of the two sequences. Thus, there seems to be no strong dependence of the elicited emotion on previously played sounds in our stimuli.

Generally, sounds with extreme values of valence or arousal are judged more consistently between different subjects. This can be seen from significant correlations (valence: 0.63, arousal 0.39) between the standard deviation over ratings of all subjects at one point in time and the negative absolute value of the mean rating at this time.

Conclusions

Subjects' ratings were fairly consistent once familiarized with the paradigm and sounds. Intra-rater reliability was found to be high with 80%. Inter-rater reliability was lower with still 60% for valence. This suggests that sounds can easily be used to communicate emotions for a single subject. Nevertheless, sounds of extreme values of valence are also useful to communicate emotions amongst subjects.

The presented sounds were chosen because they were already rated by a large population of listeners. However, there may be more intuitive sounds which need less interpretation and cognitive processing and thus are more suitable for fast and reliable communication of emotions. For the used stimuli we observed no effect of the order in which sounds were presented.

Keywords

emotional self report, coding emotions, robust sound perception, communication of emotion, sonification

REFERENCES

Bradley, M. M. & Lang, P. J. (2007). The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. Technical report B-3. University of Florida, Gainesville, Fl.

Efron, B. (1979), "Bootstrap methods: Another look at the jackknife," Ann. Statist., vol. 7, no. 1, pp. 1–26.

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In P. Juslin & J. A. Sloboda (Eds.), Music and Emotion: Theory and Research (pp. 223–248). Oxford, UK: Oxford University Press.

Nagel, F., Kopiez, R., Grewe, O. & Altenmüller, E. (2007). 'EMuJoy' - Software for continuous measurement of perceived emotions in music: basic aspects of data recording and interface features. Behavior Research Methods, 39(2), 283-290.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. Cognition, 68, 111-141.

Scherer, K. R. (1991). Emotion expression in speech and music. In Music, language, speech and brain (pp. 146–156). London: Macmillian.

Schubert, E. (2012). Reliability issues regarding the beginning, middle and end of continuous emotion ratings to music, Psychology of Music, 1-22