



Emoacoustics: a study of the physical and psychological dimensions of emotional sound design

*Erkin Asutay*¹, *Daniel Västfjäll*^{1,2}, *Ana Tajadura-Jimenez*³, *Anders Genell*⁴, *Penny Bergman*¹,
*Mendel Kleiner*¹

¹ Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden

² Department of Psychology, University of Gothenburg, Gothenburg, Sweden

³ Department of Psychology, Royal Holloway, University of London, London, England

⁴ Environment and Traffic Analysis, VTI, Gothenburg, Sweden

erkin.asutay@chalmers.se

Abstract

Psychoacoustical research provided indispensable knowledge on how human audition works, which is necessary for successful sound design applications. It may, however, be argued that traditional psychoacoustics, in its narrow focus on signal characteristics, has neglected listener and contextual characteristics. Thus, to demonstrate the influence of meaning the listener attaches to a sound we used an FFT processing to reduce the identifiability of 18 environmental sounds, since source identification is central to meaning attribution. In a listening experiment, 20 subjects listened to and rated all the processed stimuli first and then original stimuli, without being aware of the existence of two sets. Another 20 subjects rated only the processed stimuli, which were primed by their original counterparts. This manipulation was used in order to see the difference when the subject could tell what the sound source is. In both tests subjects rated their emotional experience for each stimulus on the orthogonal dimensions of valence and arousal, as well as perceived annoyance and perceived loudness for each stimulus. They were also asked to identify the sound source. It was found that processing caused correct identification to reduce substantially, while priming recovered most of the identification. While original stimuli induced a wide range of emotional experience, reactions to processed stimuli were affectively neutral. Priming manipulation reversed the effects of processing to some extent. Moreover, even though Zwicker's loudness value of most of the stimuli was reduced after processing, perceived loudness was only decreased for affectively negative stimuli.

Index Terms: auditory-induced emotion, sound design, emoacoustics

1. Introduction

Psychoacoustical models provide methods to link the physical properties of sounds and the resulting sensations [1]. They are widely used in product sound quality and sound design applications, where designer strives to convey auditory information to users. In order to achieve successful auditory displays designers naturally must possess an understanding of how people hear. Traditional psychoacoustical research has offered invaluable knowledge on how human audition works. However, the listener, being in an arbitrary affective state, possessing idiosyncratic knowledge, experience and expectations is a central problem in auditory display design. Traditional psychoacoustics does not resolve these issues. We argue that physical characteristics are not able to fully capture psychological processes in human audition, and focusing

solely on form (i.e. signal) characteristics would be to oversimplify [2].

In our daily life we are subjected to sounds constantly; and sound often induces emotional reactions in its perceiver [3, 4]. Emotions play a central role in many perceptual mechanisms [5-8]. Therefore, understanding how sounds evoke emotions could improve our knowledge on human audition and contribute to the creation of more successful sound design applications (reference to emotional design, [9]). Hence, we name this approach emotional sound design or emoacoustics.

A substantial body of research has been trying to link physical characteristics of sound to certain emotional responses, including annoyance (for a review see [10]). However, we argue that apart from physical properties, the listener and the context are also important factors in evoking emotional responses [11]. In other words, one particular sound could evoke a certain emotional response for one listener-context combination, while it induces completely different emotions for another. This is due to meaning the listener attributes to sounds during the categorization process, to which identification of the sound source is central [12]. In everyday listening, people tend to first identify the sound source [13]. Therefore, in order to investigate the importance of context and meaning on auditory-induced emotion, we focus on the source identification of everyday sounds in the present study. We attempted to reduce the identifiability of a sound source while keeping its physical form as similar to the original as possible by the use of a new FFT processing algorithm. Our aim in the present study is to experimentally manipulate the emotional experience induced by environmental sounds.

2. Experiment

2.1. Methods

2.1.1. Stimuli

Eighteen stimuli were selected among the International Affective Digitized Sounds (IADS) database [14]. Criteria for selection was to make sure that chosen stimuli would cover a wide range of emotional responses, and to avoid stimuli that contain music, erotica and those that were judged to be of poor quality. So, first all stimuli in the IADS database were divided into three groups according to their normative mean valence ratings, i.e. pleasant, neutral and unpleasant. Then, six sounds were selected from each group. All stimuli were sampled at 44.1 kHz and were six-seconds long.

2.1.2. Manipulation

The selected stimuli were manipulated using a FFT processing algorithm, which is based on executing FFT, spectral broadening and carrying out inverse FFT [15]. The original sound is divided into time windows which are then transformed into frequency domain. The amplitudes are averaged within a frequency band and the manipulated signal is transformed back into time domain without changing its phase. Our purpose is to distort spectral and temporal information to reduce the possibility to identify it, while keeping physical characteristics as close to the original as possible. Using the proposed algorithm, one can control changes in temporal and spectral domains by carefully selecting the time window and frequency band.

In the present study, spectral averaging was done in 1/3 octave bands; and 75% overlapping, hanning-windows with different lengths are used (4096, 8192 or 16384 samples). The authors decided the processing level informally for each sound. The goal was to determine lowest level of manipulation for reducing identifiability of sound sources. After the initial selection, a group of other individuals, unaware of the purpose of the study, were asked to identify the sound source of the processed versions of the stimuli. The final level of processing for each sound was determined after these evaluations.

2.1.3. Participants

40 subjects (15 females) took part in the test in two different settings. Their ages varied between 22 and 44 (25.1 on average). They were compensated after the test.

2.1.4. Measures

In the test, participants were asked to rate how they felt during each stimulus using the 9-point Self Assessment Manikin (SAM) scales of valence and arousal [16]. Using the valence scale subjects rated the pleasantness of their current feelings, and they rated the arousal level they feel using the arousal scale.

Moreover, they were asked to rate how annoying each stimulus was, using a 9-point unipolar scale from 1 (not at all annoying) to 9 (very much annoying). Also, in a similar scale they marked the perceived loudness level for each sound (1-not loud at all, 9-very much loud). Finally, they were asked to identify the sound source with their own words.

2.1.5. Procedure

The experiment was carried out in a classroom. Sounds were reproduced through headphones (Sennheiser HD-414); and participants rated the sounds on scales using paper and pencil.

Two different conditions were used. In the first setting, 20 participants (8 females, mean age: 24.5) rated all the stimuli that were processed, and then rated all the original stimuli in different orders (for the sake of simplicity from now on these two stimulus sets will be referred to as *processed* and *original* sounds, respectively), without being aware of the existence of two sets of stimuli. They were told that they would listen to and rate 36 everyday sounds.

In the second setting, 20 other participants (7 females, mean age: 25.9) rated only processed sounds, which were primed by their original counterparts. Priming was done by presenting each original sound before their processed versions. They were asked to rate the sounds in the same way as in the first setting (from now on this stimulus set will be referred to as *primed* sounds). We expected that after processing stimuli

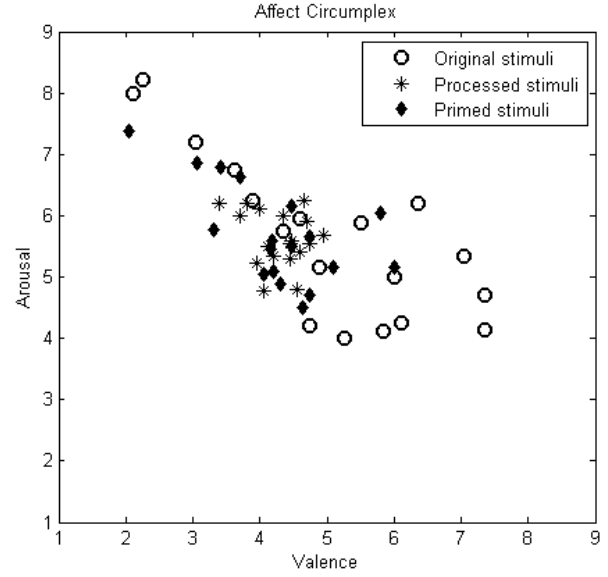


Figure 1: Effect of implemented FFT processing and priming manipulation on the affective responses.

would be unidentifiable. Hence, priming manipulation was introduced in order to see the impact of attaching a correct meaning to a stimulus that is otherwise unidentifiable.

2.2. Results

2.2.1. Identifiability

Result of the identification task showed that the processing caused identifiability of the source to decrease dramatically ($t=19.9$, $df=17$, $p<.0001$), from a mean of 91% ($std.dev=15\%$) correct identification for original sounds to a mean of 11% ($std.dev=15\%$) for processed stimuli. Moreover, priming the processed stimuli caused identifiability to increase significantly ($t=18.8$, $df=17$, $p<.0001$) to a mean of 78% ($std.dev=12\%$) correct identification compared to the processed condition. Even though the effects of processing and priming on identifiability are not uniform across all sounds because of differences in temporal and spectral properties, changes in each individual sound were statistically significant (at $p<.01$ level).

2.2.2. Affective ratings

Mean valence and arousal values for each stimulus were plotted in the two dimensional affect circumplex [17] to see the overall changes in average emotional reactions (see Fig.1). Replicating the findings of Bradley and Lang [3] the 18 sound scatter across all quadrants of the affect circumplex, with a strong tendency for sounds to spread out from low arousal, neutral valence to high arousal negative or positive valence (a V-shaped function). Most importantly, the emotional reactions to processed stimuli were centered in the middle of the circumplex, i.e. become neutral. For the primed stimuli, mean responses induced by the stimuli were not as neutral; they were more spread on the circumplex than the reactions to processed stimuli. However, still they were not as spread as the reactions to the original sounds.

For both affective dimensions, a three-factor ANOVA, employing processing (*original* vs. *processed*), pleasantness (*pleasant*, *neutral* or *unpleasant*) and sound as within-subject factors, was run. For valence dimension, significant main

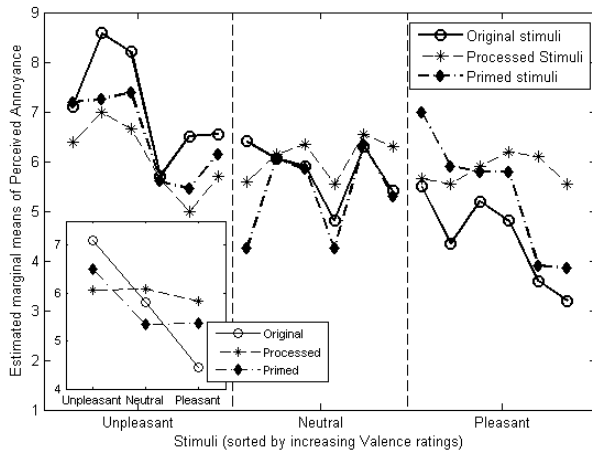


Figure 2: Effect of processing and priming on Perceived Annoyance. Stimuli are sorted according to the valence ratings for the original stimuli. **Inset:** Processing by pleasantness interaction.

effects found for all three factors. Apart from the expected pleasantness ($F(2,24)=52.608$, $p<.0001$) and sound ($F(5,60)=6.383$, $p<.001$) main effects, results showed that reactions along the valence dimension were significantly higher for original than for processed stimuli ($F(1,12)=32.422$, $p<.001$). Also, a significant interaction was found between processing and pleasantness ($F(2,24)=78.046$, $p<.0001$), showing that the effect of processing on valence differed significantly according to the emotional content of original stimuli. When the effect of processing was investigated for separate groups according to pleasantness of original stimuli, it was found that valence ratings of unpleasant sounds increased as a result of processing ($F(1,18)=20.546$, $p<.001$), whereas for pleasant and neutral sounds valence ratings decreased due to processing ($F(1,15)=7.410$, $p<.03$; and $F(1,14)=196.909$, $p<.0001$ for neutral and pleasant stimuli, respectively).

For the arousal dimension, the interaction of processing by pleasantness was also statistically significant ($F(2,20)=30.007$, $p<.0001$). Further analysis showed that while arousal ratings of unpleasant sounds significantly decreased ($F(1,18)=35.532$, $p<.0001$) due to processing, they increased for neutral and pleasant stimuli ($F(1,14)=6.567$, $p<.03$; and $F(1,13)=6.844$, $p<.03$ for neutral and pleasant stimuli, respectively).

Another set of ANOVAs were carried out so as to investigate the influence of priming on processed stimuli. Therefore, priming was introduced as a between-subjects factor, while pleasantness and sound were kept as within-subject factors. As a result, pleasantness by priming interaction was found to be statistically significant ($F(2,54)=7.809$, $p<.003$). More specifically, it was found that for originally unpleasant sounds priming the processed stimuli caused valence ratings to decrease significantly ($F(1,35)=7.047$, $p<.03$), whereas valence ratings increased when initially pleasant stimuli were primed ($F(1,30)=9.911$, $p<.005$). For the stimuli that were emotionally neutral, the priming manipulation did not cause a significant change in the valence dimension for processed stimuli.

The interaction between priming and pleasantness was also significant ($F(2,48)=3.810$, $p<.05$) for the arousal dimension. Further tests indicated that priming the processed stimuli caused induced arousal to increase only for originally unpleasant stimuli ($F(1,35)=7.996$, $p<.01$). The effect of priming was not significant for the remaining comparisons.

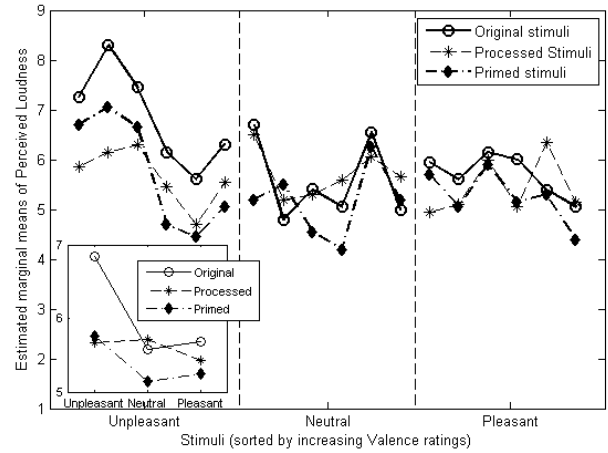


Figure 3: Effect of processing and priming on Perceived Loudness. Stimuli are sorted according to the valence ratings for the original stimuli. **Inset:** Processing by pleasantness interaction.

2.2.3. Annoyance

A three-factor ANOVA was run employing processing (original vs. processed), pleasantness (unpleasant, neutral or pleasant) and sound as within-subject factors to investigate the impact of processing on perceived annoyance (for mean levels see Fig.2). Apart from the significant main effects of pleasantness ($F(2,36)=35.736$, $p<.0001$) and sound ($F(5,90)=7.185$, $p<.0001$), both processing by sound ($F(5,90)=3.110$, $p<.03$), and processing by pleasantness ($F(2,36)=46.855$, $p<.0001$) interactions were found to be highly significant. When the effects were investigated among the separate groups according to the pleasantness of original stimuli, it was found that for both unpleasant and pleasant stimuli there were significant effects of processing ($F(1,18)=31.139$, $p<.0001$; and $F(1,19)=37.387$, $p<.0001$ for unpleasant and pleasant stimuli, respectively) in the opposite direction. In other words, due to FFT processing, auditory-induced annoyance decreased for unpleasant stimuli, while it increased for pleasant stimuli. However, for neutral stimuli there was no significant effect of processing.

Possible differences between processed and primed groups were also investigated using priming as a between subjects factor in the analysis of variance. There was no significant main effect of priming, but the priming by pleasantness interaction was significant ($F(2,76)=7.725$, $p<.003$), indicating that induced annoyance increased for initially unpleasant stimuli and decreased for the rest (see inset in Fig.2).

2.2.4. Loudness

Finally, changes in perceived loudness as a result of processing were also studied (Fig.3). In a three-factor ANOVA, there was a significant main effect of processing (original vs. processed), indicating that perceived loudness for original sounds were higher on the average than processed sounds ($F(1,19)=6.470$, $p<.05$). Furthermore, there were significant interactions of sound by processing ($F(5,95)=4.058$, $p<.05$) and pleasantness by processing ($F(2,38)=28.293$, $p<.0001$), showing that the effect of processing was not uniform on all stimuli. When the differences in loudness was investigated within separate groups according to the pleasantness of the original stimuli, highly significant effect of processing was found for

unpleasant sounds ($F(1,19)=39.107, p<.0001$), indicating that perceived loudness for original stimuli was higher on the average than processed stimuli. However, there were no significant effects of processing for neither pleasant nor neutral stimuli. This implies that FFT processing caused perceived loudness to decrease for unpleasant stimuli only.

However, when changes in Zwickers-loudness (N) were studied it was found that N was decreased for almost every stimulus due to processing. Therefore, in order to investigate the relationship between the changes in N and perceived loudness due to processing, a regression analysis was run. All results were pooled over participants and stimuli, and difference in perceived loudness (PL) due to processing was used as dependent variable. As a result, difference in Zwickers-loudness (N) accounted for only 10% of the variance in the difference in PL. Moreover, the same correlation was investigated among separate groups according to the pleasantness of the original stimuli. The difference in N accounted for 3% and 12% of the variance in the difference in PL, for unpleasant and pleasant stimuli, respectively. However, there was no significant correlation for neutral stimuli.

Since previous research has claimed a strong relationship between auditory induced annoyance and loudness for meaningless sounds [18], we studied the dimensional correlation between annoyance and loudness judgments after all results were pooled over participants and observations. Induced annoyance and perceived loudness judgments were positively correlated for all the three stimulus-groups, even though the correlation was weaker for processed stimuli. Pearson Correlation coefficients were .53, .39 and .51 ($p<.05$ for all three) for *original*, *processed* and *primed* stimuli, respectively.

Dimensional correlation between perceived annoyance and Zwickers-loudness was also studied. Auditory-induced annoyance was positively correlated with overall Zwickers-loudness for original stimuli (Pearson $r=.36, p<.05$). However, neither for neutralized nor for primed stimuli, Pearson Correlations were significant (.09 *ns*, and .05 *ns*. for *neutralized* and *primed* stimuli, respectively).

3. Discussion and Conclusions

The present study aimed to show that physical characteristics of acoustic stimuli are not able to fully capture auditory-induced emotion. Auditory emotions are often idiosyncratic and depend on the meaning an individual in a certain circumstance ascribe the sounds [11]. To investigate the context and meaning that is attached to sounds; we reduced the identifiability of a number of environmental sounds in order to distort the meaning attribution processes. The results from the identification task showed that proposed FFT processing worked as intended, i.e. correct identification was reduced significantly for every stimulus. Also, we introduced a priming manipulation, to experimentally vary the context. Correct identification was significantly increased for every processed stimulus in priming condition.

Although the change in correct identification due to processing was not the same for every stimulus, the auditory-induced emotion varied systematically depending on the initial emotional content. The distribution of original stimuli on the affect circumplex was as expected. Processed stimuli, on the other hand, appeared in a smaller area in the middle of the circumplex, i.e. where the emotional experience is neutral. Therefore, one can conclude that influence of the implemented processing algorithm was in the opposite direction for pleasant

and unpleasant stimuli. Similar results were seen on perceived annoyance: processing affected perceived annoyance only when the original stimulus was emotionally charged. Auditory-induced annoyance increased for pleasant stimuli, whereas it decreased for unpleasant stimuli. Furthermore, the priming manipulation reversed the effects of processing on both emotion ratings and perceived annoyance to some extent. These changes on emotional experience due to processing and priming manipulation support the hypothesis that sounds become more emotionally neutral when the stimulus has no context or meaning. However, we claim that even with the processed sounds subjects still heard auditory events rather than frequencies and intensities. Responses to the identification task support this argument, since subjects identified the source of processed sounds or had associations even if they were incorrect. Importantly, however these responses varied much more between individuals than what was the case for the original sounds. Thus, we tentatively propose that the reduction in identifiability caused the sounds to trigger emotionally neutral associations. At the same time, the priming manipulation caused processed, emotionally neutral, stimuli to trigger emotionally loaded associations. As a result, the emotional experience changed significantly for those sounds that were initially emotional, i.e. pleasant or unpleasant.

Furthermore, the implemented processing algorithm caused overall Zwickers-loudness (N) to decrease for almost every stimulus. However, neither emotional experience nor loudness judgments changed accordingly. When loudness ratings were investigated it was found that perceived loudness decreased only for originally unpleasant stimuli; and change in N accounted for 10% of the variance in change in perceived loudness. Hence, we propose that even loudness judgments might be influenced by the emotional context. Loudness of unpleasant stimuli, which are more salient than neutral or negative stimuli [19], might be emphasized. However, this finding has to be investigated further before drawing any firm conclusions.

In conclusion, the present study demonstrated that focusing only the physical properties of auditory stimuli is not sufficient to understand auditory-induced emotion. One also needs to consider the associations made by the listener. Hence, a sound designer needs to be aware of the different capacities of physical and psychological dimension of an auditory display in causing an emotional reaction in its perceiver, in order to succeed in designing effective auditory displays.

4. References

- [1] Zwicker, E., & Fastl H., "Psychoacoustics: Facts and models", Springer-Verlag, Berlin, 1990.
- [2] Jekosch, U., "Meaning in the context of sound quality assessment", *Acta Acustica*, 85: 681-684, 1999.
- [3] Bradley, M. M., & Lang, P. J., "Affective reactions to acoustic stimuli", *Psychophysiology*, 37(2): 204-215, 2000.
- [4] Tajadura-Jimenez, A., & Västfjäll, D., "Auditory-induced emotion: A neglected channel for communication in human-computer interaction", in C. Peter & R. Beale [Eds.], *Affect and emotion in human-computer interaction: From theory to applications*, 63-74, Springer-Verlag, Berlin / Heidelberg, 2008.
- [5] Damasio, A., "Descartes' Error", Putnam, NY, 1994.
- [6] Zeelenberg, R., Wagenmakers, E., & Retteveel, M., "The impact of emotion on perception: Bias or enhanced processing?", *Psychological Science*, 17(4): 287-291, 2006.
- [7] Phelps, E. A., Ling, S., & Carrasco, M., "Emotion facilitates perception and potentiates the perceptual benefits of attention", *Psychological Science*, 17(4): 292-299, 2006.

- [8] Gan, T., Wang, N., Zhang, Z., Li, H., & Lou, Y., "Emotional influences on time perception: Evidence from event-related potentials", *NeuroReport*, 20: 839-843, 2009.
- [9] Norman, D., "Emotional design: Why we love (or hate) everyday things", Basic Books, NY, 2004.
- [10] Västfjäll, D., "Affect as a component of perceived sound and vibration quality in aircraft", PhD thesis, Chalmers TH, 2003.
- [11] Juslin, P. & Västfjäll, D., "Emotional responses to music: The need to consider underlying mechanisms", *Behavioral and Brain Sciences*, 31(5): 559-621, 2008.
- [12] Bergman, P., Sköld, A., Västfjäll, D. & Fransson, N., "Perceptual and emotional categorization of sound", *Journal of Acoustical Soc. of America*, 126(6): 3156-3167, 2009.
- [13] Gaver, W. W., "What do we hear in the world? An ecological approach to auditory event perception", *Ecological Psychology*, 5(1): 1-29, 1993.
- [14] Bradley, M. M., & Lang, P. J., "International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2)", Gainesville, FL: The Center for Research in Psychophysiology, University of Florida, 1999.
- [15] Genell, A., "Perception of sound and vibration in heavy trucks", PhD thesis, Chalmers TH, 2008.
- [16] Lang, P. J., "Behavioral treatment and bio-behavioral assessment: Computer applications", in J. B. Sidowski, J. H. Johnson & T. A. Williams [Eds.], *Technology in mental health care delivery systems*, 119-137, Ablex Publishing, Norwood, NJ, 1980.
- [17] Russell, J. A., "A circumplex model of affect", *Journal of Personality and Social Psychology*, 39(6): 1161-1178, 1980.
- [18] Västfjäll, D., "Affective reactions to meaningless sounds", *Cognition and Emotion*, in press.
- [19] Tajadura-Jimenez, A., Väljamäe, A., Asutay, E. & Västfjäll, D., "Embodied auditory perception: The emotional impact of approaching and receding sound sources", *Emotion*, 10(2): 216-229, 2010.