THE TEMPORAL WEIGHTING OF ANNOYANCE

Kerstin Dittrich[1,2] and Daniel Oberfeld[1]
[1]Department of Psychology, Johannes Gutenberg-Universitaet Mainz, 55099 Mainz, Germany
[2]Department of Psychology, Albert-Ludwigs-Universitaet Freiburg, 79085 Freiburg, Germany
dittrich@psychologie.uni-freiburg.de

Abstract

The influence of individual temporal portions of a level-fluctuating noise on global annoyance judgments was measured using perceptual weight analysis (cf. Berg, 1989). For loudness judgments it has been found that listeners attach greater weight to the beginning and the ending than to the middle of a stimulus (e.g. Oberfeld & Plank, 2005). Similar weights were expected for annoyance. Annoyance and loudness judgments were obtained from 12 listeners for the same stimuli in a two-interval forced-choice task. The results demonstrated a primacy effect for the temporal weighting of both annoyance and loudness. A recency effect was observed only for annoyance, although the temporal weights for loudness and annoyance were only marginally significantly different. A control experiment showed that the listeners were capable of independently judging the stimuli according to either their loudness or their annoyance: Noises with the same energy-equivalent level but different modulation depths were judged to differ in annoyance but not in loudness.

Several different measures for assessing the annoyance of longer sounds fluctuating in level have been proposed (e.g. N_5, L_{eq} or L_{A}; cf. Zwicker & Fastl, 1999). These measures take into account parameters such as sound pressure level and frequency spectrum. However, the temporal aspect has not received much consideration in these calculations until now. Conventional measures assume that listeners weight the information provided by each temporal segment of a noise uniformly. The present study examined whether this approach is compatible with the perception of annoyance or whether temporal aspects should be considered in the estimation of annoyance.

For loudness it has already been found that if listeners evaluate the overall loudness of a level fluctuating noise, the initial and final portions of the stimulus receive greater weight than its temporal center (e.g. Ellermeier & Schrödl; 2000, Oberfeld & Plank, 2005; Dittrich, Hachgenei & Oberfeld, 2006). Similar weights were expected for annoyance due to the correspondence between loudness and annoyance found in many experiments (cf. Kalivoda & Steiner, 1998). Listeners evaluated the annoyance and the loudness of a level-fluctuating noise in a two-interval forced-choice (2I, 2AFC) task, and the influence of the individual temporal segments on annoyance and loudness judgments was estimated using perceptual weight analysis (cf. Berg, 1989).

In order to make sure that the stimuli used here could be independently judged according to either their loudness or their annoyance, a control experiment was conducted. In this 2I, 2AFC task, noise with the same energy-equivalent level but different modulation depth was presented to the subjects. It was expected that noises with the same energy-equivalent level but different modulation depths were judged to differ in annoyance but not in loudness (e.g. Widmann, 1994).

Method

12 subjects (8 women, 5 men, age 22-51 years) participated in the experiment voluntarily. Most of the listeners were psychology students and participated for course credit. All subjects reported normal hearing; detection thresholds in the right ear measured by a 2I, 2AFC, adaptive procedure with a 3-down, 1-up rule (Levitt, 1971) were better than 15 dB HL at all octave frequencies between 250 and 4000 Hz. The stimuli were generated digitally, played back via two channels of an RME ADI/S digital-to-analog converter (f_S = 44.1 kHz, 24-bit resolution), attenuated (two TDT PA5s), buffered (TDT HB7), and presented diotically via Sennheiser HDA 200 headphones calibrated according to IEC 318 (1970). The experiment was conducted in a single-walled sound-insulated chamber. Listeners were tested individually.

Each level-fluctuating stimulus was a Gaussian wide-band noise consisting of nine temporal segments. The duration of each segment was 100 ms. The trials contained two intervals, with an inter-stimulus-interval of 500 ms. In the main experiment, the sound pressure levels of the nine temporal segments were drawn independently from a normal distribution with mean μ_S = 64.5 dB SPL and standard deviation (SD) σ = 2.5 dB for the interval containing the softer noise. For the interval containing the louder noise, the mean was μ_L = 65.5 dB SPL. The louder noise was presented in interval 1 or interval 2 with identical a priori probability. In the control experiment, the noises with the small and the large modulation depth were generated by independently drawing the sound pressure levels of the nine temporal segments from a normal distribution with mean μ = 65 dB SPL and standard deviation (SD) σ = 2 dB or σ = 4 dB, respectively. The level of each segment of the noise with the large modulation depth was displaced by an identical amount so that L_{eq} was identical to L_{eq} for the noise with small modulation depth. 15 trials with the large standard deviation in the first interval and the small standard deviation in the second interval, and 15 trials with the reverse order of modulation depths were generated and stored before the experiment started. Thus, in the control experiment the subjects evaluated exactly the same set of stimuli according to both their loudness and their annoyance. The same set of 30 trials was used for all subjects.

In both the main and the control experiment, the stimuli were presented in a 2I, 2AFC procedure. Depending on the task, the listeners select the interval containing the louder sound or the more annoying sound. No feedback was provided. The next trial followed the response after an inter-trial interval of 2 s. The subjects were randomly assigned to two experimental groups. Group 1 made only annoyance judgments in the first part of the experiment, and only loudness judgments in the second part. For Group 2, the order of tasks was reversed. Both parts of the experiment consisted of practice blocks, followed by the control experiment, and then the main experiment. In the first part of the experiment, Group 1 received each of the 30 stored trials for the control experiment 10 times in random order and decided which interval had contained the more annoying sound. Subsequently, the listeners received 1000 trials of the main experiment in the annoyance task. In the second part of the experiment, Group 1 again received 300 trials from the control experiment followed by 1000 trials from the main experiment, but this time they made loudness judgments. The stimuli were presented in blocks of 50 trials. For Group 2 that started with the loudness judgments the procedure was analogous. The experiment lasted 6.5 hours divided into 6 sessions.

Results

In the main experiment trial-by-trial correlation analyses were used to determine the relative perceptual weights assigned to the nine temporal segments (see Richards & Zhu, 1994; Lutfi, 1995). For each trial and each segment (i = 1 … 9), the difference between the level of segment *i* in interval 2 and the level of segment *i* in interval 1 was computed. The biserial correlations between these level differences and the binary response (1: "First noise louder / more annoying"; 2: "Second noise louder / more annoying") were taken as the weight estimates. The weights were normalized individually such that the sum of the absolute values was unity (see Kortekaas, Buus & Florentine, 2003), resulting in a set of relative temporal weights for each listener and each task. For annoyance, the mean weights of all subjects reflected a primacy effect [$t(11) = 3.97$, $p = .002$] as well as a recency effect [$t(11) = 2.751$, $p = .019$]. For the mean weights of loudness a primacy effect was found, too [$t(11) = 3.060$, $p = .011$], but no significant recency effect was observed [$t(11) =.419$, $p = .683$]. Figure 1 displays the mean weights of annoyance and loudness for the two experimental groups.

A three-factorial repeated-measures analysis of variance (ANOVA) with Huynh-Feldt correction for the degrees of freedom was conducted. The two within-subjects factors were segment and task (annoyance judgments / loudness judgments). The order of tasks (i.e., experimental group) was included as a between-subjects factor. As expected, there was a significant effect of segment [$F(8, 80) = 7.00$, $\varepsilon = .43$, $p = .001$]. The Segment × Task interaction was marginally significant [$F(8, 80) = 2.04$, $\varepsilon = .847$, $p = .064$], partially confirming the observation that there was a primacy effect for both annoyance and loudness, while only the annoyance weights showed a recency effect. The ANOVA also showed a significant effect of order of tasks [$F(1, 10) = 5.58$, $p = .040$], presumably due to the stronger primacy effect produced by Group 1 which started with the annoyance judgments. Neither the Segment × Order of Tasks interaction [$F(8, 80) = 0.96$], nor the Task × Order of Tasks interaction [$F(1, 10) = 0.08$] interaction was significant. Due to the normalization of the weights, the effect of task was also not significant [$F(1, 10) = 0.47$].

In the control condition the proportion of trials in which the noise with the large standard deviation was chosen as louder or more annoying was analyzed. For loudness, the mean proportion was .52 (SD = .085), compatible with the hypothesis that the two noises should be perceived as equally loud. For annoyance, the mean proportion was .66 (SD = .084), indicating that as expected the noise with the larger modulation depth was perceived as more annoying. A repeated-measures ANOVA with the within-subjects factor task and the between-subjects factor order of tasks showed a significant effect of task [$F(1, 10) = 45.75$, $\varepsilon = 1.0$, $p = 0.001$]. Neither the effect of order of tasks [$F(1, 10) =2.47$], nor the Task × Order of Tasks interaction [$F(1, 10) = 0.25$] interaction was significant. A one-sample *t*-test showed that the mean proportion of trials in which the noise with the large standard deviation was chosen as louder was not significantly different from .5 [$t(11) = .632$, $p = .54$]. For annoyance, however, the mean proportion differed significantly from .5 [$t(11) = 6.791$, $p = .001$].

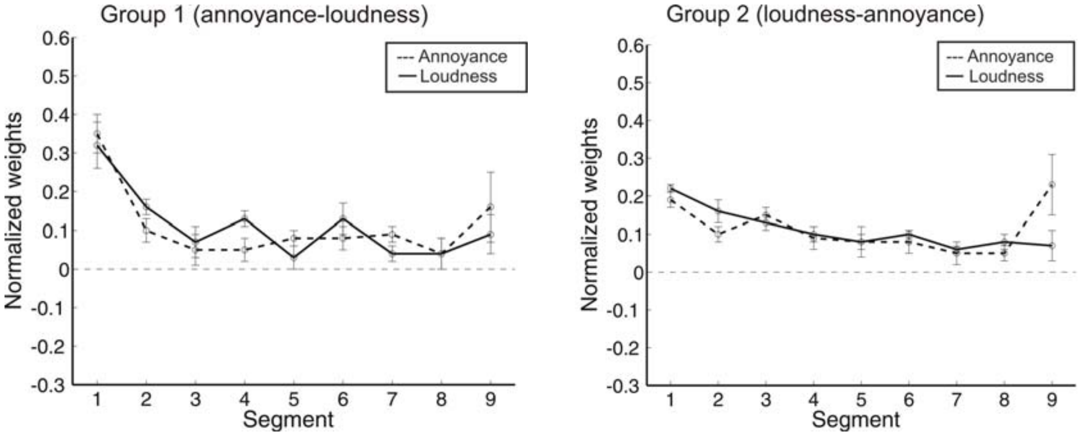

Fig. 1. Main experiment. Relative normalized temporal weights for annoyance and loudness. The two panels show the result from the two experimental groups who received a different order of task. Error bars show ± 1SEM.

Discussion

The results demonstrated a primacy effect for the temporal weighting of both annoyance and loudness. A significant recency effect was observed only for annoyance, but the Segment × Task interaction was only marginally significant. The control experiment showed that the listeners were capable of independently judging the stimuli according to either their loudness or their annoyance: Noises with the same energy-equivalent level but different modulation depths were judged to differ in annoyance, but not in loudness.

One potential explanation for the similar pattern of weights observed for both loudness and annoyance would be that primacy and recency effects are no special psychoacoustical phenomena but can also be found in the domain of cognitive psychology (cf. Anderson, 2001). Possibly, the segments of the stimuli used here can be viewed as serially sorted information in which the first elements could be remembered better than the middle of the sequence.

According to the results of the present study conventional measures of annoyance should consider temporal components. Particularly the beginning and ending of a noise should be taken into account more strongly. A limitation of the present study is that the stimuli used here were very short compared to aircraft noise for example. Additional experiments are necessary to decide whether these results could be replicated for longer stimuli.

References

Anderson, J.R. (2001). *Kognitive Psychologie.* Heidelberg: Spektrum Akad. Verl., 3. Aufl. 2001.

Berg, B.G. (1989). Analysis of weights in multiple observation tasks. *Journal of the Acoustical Society of America*, **86**, 1743-1746.

Dittrich, K., Hachgenei, B. & Oberfeld (2006). Führt die graduelle Einblendung eines Geräusches zu einer gleichmäßigeren zeitlichen Lautheitsgewichtung? In Hecht, H., Berti, S., Meinhardt, G. & Gamer, M. (Eds.), *Beiträge zur 48. Tagung experimentell arbeitender Psychologen*. Lengerich: Pabst Science Publishers, 2006.

Ellermeier, W. & Schrödl, S. (2000). Temporal weights in loudness summation. In Bonnet, C. (Ed.), *Fechner Day 2000*. Université Louis Pasteur, Strasbourg, 169-173.

Kalivoda, M.T. & Steiner, J.W. (1998). *Taschenbuch der angewandten Psychoakustik.* Wien: Springer.

Kortekaas, R., Buus, S. & Florentine, M. (2003). Perceptual weights in auditory level discrimination. *Journal of the Acoustical Society of America*, **113**, 3306-3322.

Levitt, H. (1971). Transformed up-down procedures in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.

Lutfi, R.A. (1995). Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks. *Journal of the Acoustical Society of America*, **97**,1333-1334.

Oberfeld, D. & Plank, T. (2005). Temporal weighting of loudness: Effects of a Fade In. Deutsche Gesellschaft für Akustik (Ed.), *Fortschritte der Akustik - DAGA 2005*. Berlin: DEGA, 227-228.

Richards, V. & Zhu, S. (1994). Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients. *Journal of the Acoustical Society of America*, **95**, 423-434.

Widmann, U. (1994). Zur Lästigkeit von amplitudenmodulierten Breitbandrauschen. In Daga 94 (Ed.), *Fortschritte der Akustik*. DPG-GmbH.

Zwicker, E. & Fastl, H. (1999). *Psychoacoustics - Facts and models (2nd edition).* Heidelberg: Springer-Verlag.