Atten Percept Psychophys (2011) 73:189–208
DOI 10.3758/s13414-010-0011-8

The temporal weighting of loudness: effects of the level profile

Daniel Oberfeld · Tina Plank

Published online: 20 November 2010

Abstract In four experiments, we studied the influence of the level profile of time-varying sounds on temporal perceptual weights for loudness. The sounds consisted of contiguous wideband noise segments on which independent random-level perturbations were imposed. Experiment 1 showed that in sounds with a flat level profile, the first segment receives the highest weight (primacy effect). If, however, a gradual increase in level (*fade-in*) was imposed on the first few segments, the temporal weights showed a delayed primacy effect: The first unattenuated segment received the highest weight, while the fade-in segments were virtually ignored. This pattern argues against a capture of attention to the onset as the origin of the primacy effect. Experiment 2 demonstrated that listeners adjust their temporal weights to the level profile on a trial-by-trial basis. Experiment 3 ruled out potentially inferior intensity resolution at lower levels as the cause of the delayed primacy effect. Experiment 4 showed that the weighting patterns cannot be explained by perceptual segmentation of the sounds into a variable and a stable part. The results are interpreted in terms of memory and attention processes. We demonstrate that the prediction of loudness can be improved significantly by allowing for nonuniform temporal weights.

D. Oberfeld (✉)
Department of Psychology,
Johannes Gutenberg-Universität Mainz,
55099, Mainz, Germany
e-mail: oberfeld@uni-mainz.de

T. Plank
Department of Psychology, Universität Regensburg,
Regensburg, Germany

Keywords Dynamic loudness · Time-varying sounds · Intensity discrimination · Temporal perceptual weights · Perceptual weight analysis · Molecular psychophysics · Loudness model

Building on the pioneering work of H. Barkhausen, S. S. Stevens, and others (e.g., Barkhausen, 1926; Stevens, 1956), much effort has been devoted to understanding the loudness of simple *laboratory-type* sounds (see Scharf, 1978, for a review of many findings). This effort has led to powerful models for the loudness of stationary sounds (Glasberg & Moore, 2006)—that is, sounds that remain relatively constant in frequency spectrum and waveform amplitude across the duration of presentation (e.g., a sinusoid or a burst of wideband noise). These loudness models encompass a wide variety of psychophysical data, although even fundamental issues such as the form of the loudness function remain subjects of debate (Krueger, 1989). What can be concluded about our understanding of the loudness of time-varying (*dynamic*) sounds changing across time in frequency spectrum, in waveform amplitude, or on both dimensions during presentation, just as many environmental sounds do? For this type of stimuli, a smaller amount of data is available (e.g., Grimm, Hohmann, & Verhey, 2002; Moore, Vickers, Baer, & Launer, 1999; Zhang & Zeng, 1997). Technical measures proposed as estimates of the loudness of fluctuating sounds (e.g., European Parliament and Council of the European Union, 2002; World Health Organization, 1999)—such as, for example, the energy-equivalent level of a steady sound (L_{eq}), or the 95th percentile of the loudness distribution N_5 (see Zwicker & Fastl, 1999)—typically assume that all temporal portions of a sound contribute *equally* to overall loudness (see Ellermeier & Schrödl, 2000). Recent studies using level-fluctuating noise stimuli that remained constant in spectrum but changed in level every 100 ms or so showed, however, that this conjecture is not correct. Listeners' judgments of the global loudness[1] of a level-fluctuating noise with a duration of 1 s are more strongly influenced by the first 100–300 ms of the sound than by its middle portion (Dittrich & Oberfeld, 2009; Ellermeier & Schrödl, 2000; Pedersen & Ellermeier, 2008; Rennies & Verhey, 2009). In other words, the temporal weighting of loudness shows a pattern akin to the primacy effect in short-term memory (e.g., Baddeley, 1966). Higher weights have been observed for the temporal portion at the beginning of the sound, showing that the first part contributes more strongly to the perceived loudness of the sound than does the middle portion of the sound. To a weaker extent, a recency effect has also been observed; that is, higher perceptual weights are placed on the ending portion of the sound than on the middle portion (Dittrich & Oberfeld, 2009; Pedersen & Ellermeier, 2008). This weighting pattern differs from that of an ideal observer, who would apply identical weights to all temporal portions of a sound (Berg, 1989) if each element provided the same amount of information concerning the correct response, as was the case in the experiments above.

Can the overweighting of the beginning of a sound in a loudness judgment task be explained by peripheral mechanisms? For a stimulus of constant sound intensity presented in quiet, the firing rate of auditory nerve neurons is maximum at the stimulus onset and then, within a few milliseconds, decays to a lower, steady-state level (Kiang, Watanabe, Thomas, & Clark, 1965; Nomoto, Katsuki, & Suga, 1964). Thus, if the loudness of the level-fluctuating sound is determined simply by the total firing rate (e.g., Fletcher & Munson, 1933; Howes, 1974; Lachs, Al-Shaikh, Bi, Saia, & Teich, 1984) or by a weighted average of the firing rates of individual neurons (Nizami & Schneider, 1997; Relkin & Doucet, 1997), the level of the first temporal segment should have the greatest impact on perceived intensity. However, because the initial spike in the neural response decays within a few milliseconds, such a mechanism should result in almost exclusive weight being assigned to the first temporal segment. Therefore, the observation that the second and third temporal segments (presented 100 and 200 ms, respectively, after stimulus onset) also receive a higher weight than do the following segments (e.g., Dittrich & Oberfeld, 2009) cannot be accounted for by the neural responses in the auditory periphery. Second, Plank (2005) obtained loudness judgments for a sequence of ten 20-ms noise bursts, separated by pauses of 5, 40, or 100 ms. For all three of these conditions, a primacy effect was observed, with higher weights being assigned to the first three segments of the sequence. This again argues against the initial firing rate in the auditory periphery as an explanation, because, with pauses of 40 or 100 ms, each segment should have elicited a similar neuronal response, due to the fast recovery of the auditory nerve neurons (Harris & Dallos, 1979; Smith, 1977). Furthermore, the primacy effect was most pronounced for the sequences with the longer pauses, of 40 and 100 ms. Thus, the ability of the participants to distinguish segments perceptually from each other and treat them as single events seems to promote the emergence of a primacy effect.

A simple alternative explanation would be that due to the abrupt onset of the noise, attention is captured and directed to the beginning of the stimulus, roughly in the sense of an orienting response (Graham & Hackley, 1991; Pavlov, 1927; Sechenov, 1863/1965). In the visual domain, capture of visual attention by abrupt onsets has frequently been reported (e.g., Desimone & Duncan, 1995; Folk, Remington, & Johnston, 1992; Jonides & Yantis, 1988; Yantis & Jonides, 1990). Whereas, in these studies, attention was captured by one visual object from another object, it seems possible that in the auditory domain, where the temporal evolution of a stimulus is crucial, direction of attention to a specific *temporal portion* of a longer stimulus can occur. For example, if an abrupt onset were to cause an orienting response, attention should be directed to the beginning of the sound.

The aim of Experiment 1 in the present study was to test whether reducing the abruptness of the onset by imposing a gradual increase in level (*fade-in*) on the beginning of a sound would reduce the primacy effect in the pattern of temporal weights. Such a reduction would be evidence that capture of attention to the onset is the cause of the primacy effect. However, the results from Experiment 1 did not demonstrate the expected approximately uniform weighting pattern. Instead, we observed a *delayed primacy effect*. The attenuated temporal segments constituting the fade-in received near-zero weights, while the highest weight was assigned to the first unattenuated segment. Experiments 2–4 were designed to further explore the effects of the level profile of a time-varying sound on temporal weights for loudness and to promote understanding of the origin of these effects.

Experiment 1: effects of a fade-in

In Experiment 1, temporal perceptual weights were estimated in a loudness judgment task. The stimuli were level-

[1] Compatible with previous studies (e.g., Pedersen & Ellermeier, 2008; Susini et al., 2007), we use the term *global loudness* to emphasize that the participants were not required to make separate judgments of the loudness of different temporal portions (e.g., beginning, center, or end of the stimulus). Instead, listeners judged the loudness of a noise in its *entirety* (i.e., over the entire duration). The contribution of each single temporal portion of the sound to this global loudness judgment was then calculated with a specific statistical method (perceptual weight analysis; see Berg, 1989). For details, see the Method section for Experiment 1.

fluctuating noises with three different level profiles. Temporal weights for sounds with a *flat level profile* (i.e., with no changes in mean level across the stimulus duration) were compared with weights for stimuli containing a gradual increase in level (*fade-in*) at the beginning. To estimate the temporal weights, *perceptual weight analysis* was used, just as in other recent studies on the temporal weighting of loudness (e.g., Dittrich & Oberfeld, 2009). The basic principle behind perceptual weight analysis is to present a stimulus consisting of several elements (in the present case, nonoverlapping temporal segments), to introduce trial-by-trial variation in the dimension of interest (here, sound intensity) on each of the elements, and to estimate the impact of the variation of each individual element on a behavioral or neural response. Thus, *molecular*, rather than *molar*, analyses were conducted on the data (Green, 1964) in order to gain insight into the decision process, rather than obtaining molar measures such as loudness or accuracy. Molecular, or perceptual weight, analyses have been used for several decades (Ahumada & Lovell, 1971; Berg, 1989; de Boer & Kuyper, 1968; Gilkey & Robinson, 1986) and have found increasing application in several domains (e.g., Ahumada, 2002; Berg, 2004; Neri, Parker, & Blakemore, 1999; Oberfeld, 2009; Yu & Young, 2000). We applied this technique to a loudness judgment task, in order to estimate the influence of the sound pressure level of different temporal portions of the stimulus on global loudness. Imagine a stimulus consisting of three contiguous wideband noise segments. Each segment has a duration of 100 ms and is presented at a level of 60 dB SPL. Now, if the sound pressure level of one single segment is increased by 2 dB, will the resulting increase in loudness be identical regardless of whether the increment is imposed on the first, second, or third segment? To answer this question, we imposed random and independent level perturbations on the temporal segments. If, now, the participant assigns a high weight to a particular temporal segment of the sound—that is, if attention is directed to this temporal segment (Berg, 1990)—there will be a strong correlation between the random level perturbation imposed on this segment and the response of the participant. If, conversely, the segment is unimportant for the decision, the responses will be statistically independent of the random variation (Oberfeld, 2008a; Richards & Zhu, 1994). In this study, multiple logistic regression was used for estimating the perceptual weights (see below). As compared with older methods for tracking loudness across time, such as continuous ratings of *instantaneous loudness* (Susini, McAdams, & Smith, 2007; Zwicker & Fastl, 1999, p. 322f), perceptual weight analysis has much higher temporal resolution (in the millisecond range; see Plank, 2005) and is less transparent to the participants and, therefore, also less susceptible to biases (see Ellermeier & Schrödl, 2000).

Method participants

Seven volunteers (5 women, 2 men; 20–27 years of age) participated in the experiment for course credit. All the listeners reported normal hearing and had detection thresholds better than 10 dB HL at all octave frequencies between 500 and 8000 Hz, measured in a two-interval forced choice, adaptive procedure with a three-down, one-up rule (Levitt, 1971). The listeners were naïve with respect to the hypotheses under test. Only 2 listeners had experience in comparable tasks.

Apparatus

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). We used no equalization of the headphones' transfer function. The experiment was conducted in a single-walled sound-insulated chamber. Listeners were tested individually.

Stimuli and procedure

For the flat level profile, the stimuli were Gaussian wideband noises (20–20000 Hz) consisting of ten contiguous temporal segments. The duration of each segment was 100 ms. Figure 1 shows a schematic depiction of the stimuli. On each trial, the sound pressure levels of the ten temporal segments were drawn independently from a normal distribution, resulting in a level-fluctuating noise. The mean of the distribution was μ = 60.0 dB SPL; the standard deviation was SD = 2 dB. Each of the so-constructed stimuli was then randomly chosen to be a *soft* or a *loud* trial. A fixed level increment of $\Delta L/2$ = 0.5 dB was added to each segment on a *loud* trial, resulting in a mean level of μ_L = 60.5 dB for all segments. The same value of $\Delta L/2$ = 0.5 dB was subtracted from each segment on a *soft* trial, so that the mean level for these trials was μ_S = 59.5 dB. Although the estimation of perceptual weights would be possible without a difference in mean level, we introduced this difference in level mainly to make the task easier for the participants and also to be compatible with previous experiments (e.g., Berg, 1989; Pedersen & Ellermeier, 2008). To avoid overly loud or soft sounds, the range of levels was restricted to $\mu \pm 2.5\ SD$. Therefore, the maximal level difference between the most intense and the least intense segments within a noise was 10 dB.

For the three-step fade-in, the stimuli were first constructed in exactly the same way as for the flat level profile. To produce the fade-in, the levels of the first three segments were subsequently attenuated by subtracting 15, 10, and 5 dB, respectively (see Fig. 1).

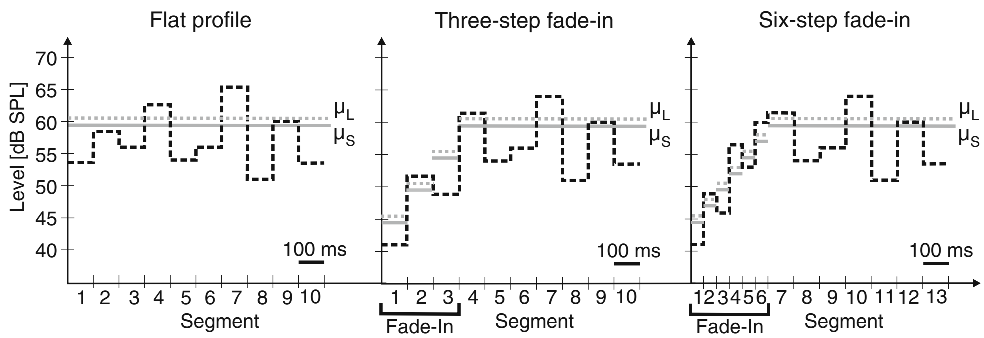

Fig. 1 Schematic depiction of the stimulus configurations used in Experiment 1. Level-fluctuating sounds consisting of 10–13 contiguous wideband noise segments were presented. On each trial, the level of each segment was drawn independently from one of two normal distributions differing in their means (dashed gray line, *loud* distribution; solid gray line, *soft* distribution). The black dashed lines show example segment levels. Participants decided whether the sound had been soft or loud (one-interval absolute identification task). The sounds were presented with a flat level profile (left panel), with a three-step fade-in (middle panel), or with a six-step fade-in (right panel)

For the six-step fade-in, the noise consisted of six 50-ms segments followed by seven 100-ms segments (see Fig. 1). The sound pressure levels of the now 13 temporal segments were again drawn independently from a normal distribution, in the same way as for the flat level profile. An attenuation of 15.0, 12.5, 10.0, 7.5, 5.0, and 2.5 dB was imposed on the first through sixth segments, respectively. Apart from this, the same procedure as that for the flat level profile was used.

The stimuli were presented in a one-interval, two-alternative forced-choice intensity discrimination task (i.e., an absolute identification task; Braida & Durlach, 1972). Each trial was randomly chosen to be a soft or a loud trial with equal probability. The participants decided whether they had been presented a soft or a loud noise. Responses were collected via two buttons on a numeric keypad. As was outlined above, for the listeners, the task was simply to judge each sound as being either soft or loud or, put differently, to evaluate the global loudness of each sound with respect to loudness of the previous sounds presented in a given block. As Pedersen and Ellermeier (2008) pointed out, the task can alternatively be described as intensity discrimination, as we did above. The two alternative descriptions can easily be reconciled by assuming that the subjective quality or sensory continuum (see Durlach & Braida, 1969; Green & Swets, 1966) that listeners base their decisions on is *loudness*.

The next trial followed the response after an intertrial interval of 2 s. No trial-by-trial feedback was provided. Pedersen and Ellermeier (2008) reported that trial-by-trial feedback can alter the decision strategies of listeners, in a comparable loudness judgment task. Since we were interested in "natural" or "spontaneous" judgments of global loudness, we opted against trial-by-trial feedback, which listeners might have used to adjust their "natural" decision weights toward optimal weights.

Design

A repeated measures design was used. Each participant received all of the three level profiles, in separate experimental blocks. After 1 hr of practice, the listeners participated in six experimental sessions, conducted on separate days, with a duration of approximately 60 min each. The experimental sessions were organized as follows. After two practice blocks, three to four 50-trial blocks of one of the three level profiles were presented. Then 3 to 4 blocks of a different level profile followed and, finally, 3 to 4 blocks of the remaining level profile. The reason for presenting 3 or 4 consecutive blocks of the same level profile was to facilitate the adoption of an optimal response strategy for a given level profile. The order of level profiles was varied between sessions. For each level profile, 20 blocks (corresponding to a total of 1,000 trials) were presented.

Data analysis

The trial-by-trial data obtained for each participant in each of the three level profiles were analyzed separately to estimate the relative perceptual weight with which each of the temporal segments had contributed to the decision. Multiple logistic regression (PROC LOGISTIC, SAS 9.2) was used to estimate the influence of the level of each individual temporal segment on the response of the listener (see Agresti, 2002; Alexander & Lutfi, 2004; Oberfeld, 2008a; Pedersen & Ellermeier, 2008).[2] The binary responses served as the dependent variable, and the 10 or

[2] Besides multiple binary logistic regression, there exist other techniques for weight estimation (Ahumada & Lovell, 1971; Berg, 1989; Richards & Zhu, 1994), all of which are based on a similar decision model and produce similar estimates (Plank, 2005; Tang, Richards, & Shih, 2005).

13 segment levels served as predictors, which were entered simultaneously. Listeners' *soft* responses were coded as 0, and *loud* responses as 1. The regression coefficients were taken as the weight estimates. For a given segment, a regression coefficient equal to zero means that the level of the segment had no influence at all on the decision to judge the noise as being either soft or loud. A regression coefficient greater than zero means that the probability to respond that the loud noise had been presented increased with the level of the given segment. A regression coefficient smaller than zero indicates the opposite relation between the level of the segment and the probability to respond that the loud noise had been presented.

This analysis is based on a decision model that assumes that listeners use a decision variable:

$$D(L) = \left(\sum_{i=1}^{k} w_i L_i \right) - c, \tag{1}$$

where L_i is the sound pressure level of segment i, k is number of segments, $\boldsymbol{L}$ is the vector of segment levels, w_i is the perceptual weight assigned to segment i, and c is a constant representing the decision criterion (see Berg, 1989; Pedersen & Ellermeier, 2008). The model assumes that a listener responds that the noise presented on a given trial was loud rather than soft if $D(\boldsymbol{L}) > 0$ and that

$$P(\text{"loud"}) = \frac{e^{D(L)}}{1 + e^{D(L)}}. \tag{2}$$

Due to the difference in mean level between loud and soft trials, the segment levels were correlated. Therefore, to avoid problems with multicollinearity, separate logistic regression analyses were conducted for the trials containing the noise with the higher mean level μ_{L} and for the trials containing the noise with mean level μ_{S} (see Berg, 1989). Thus, a separate logistic regression model was fitted for each combination of participant, level profile, and mean level (μ_{L} or μ_{S}). For each model, the weights w_i were normalized such that the sum of their absolute values was unity (see Kortekaas, Buus, & Florentine, 2003), resulting in a set of relative temporal weights for each listener, level profile, and mean level.

Results

Goodness of fit

Global goodness of fit of the regression models (see Dittrich & Oberfeld, 2009) was assessed with the unweighted residual sum-of-squares test (RSS test; Copas, 1989). This test performs favorably, as compared with some alternative tests (Hosmer, Hosmer, leCessie, & Lemeshow, 1997; Kuss, 2002). An SAS macro (GOFLOGIT; Kuss, 2001) was used to compute the test statistics. The test produced p values smaller than .2, indicating a lack of fit, for only 6 of the 42 (listener × level profile × mean level) fitted multiple logistic regression models.[3]

A summary measure of the predictive power of a logistic regression model is the area under the receiver operating characteristic (ROC) curve (AUC; Agresti, 2002; Swets, 1986b). This measure provides information about the degree to which the predicted probabilities are concordant with the observed outcome (see Dittrich & Oberfeld, 2009, for details). Areas of .5 and 1.0 correspond to chance performance and perfect performance of the model, respectively. Across the 42 fitted logistic regression models, AUC ranged from .67 to .92 ($M = .79$, $SD = .055$), indicating reasonably good predictive power (Hosmer & Lemeshow, 2000).

Perceptual weights

Figure 2 shows the mean relative temporal weights for the three level profiles. For the flat level profile (Fig. 2, left panel), the expected primacy effect was evident in the mean weights. The relative perceptual weights in this condition were analyzed via a repeated measures analysis of variance (ANOVA) using a univariate approach (see Keselman, Algina, & Kowalchuk, 2001). The Huynh–Feldt correction for the degrees of freedom was used (Huynh & Feldt, 1976), and the value of the *df* correction factor $\tilde{\varepsilon}$ is reported. Note that this particular variant of a repeated measures ANOVA performs comparably well even for small samples and nonnormally distributed data (Keselman, Kowalchuk, & Boik, 2000). Partial η^2 is reported as a measure of association strength. An α level of .05 was used for all the tests. The within-subjects factors were segment

[3] Although the proportion of models showing lack of fit that we observed is certainly not ideal, we wish to emphasize that by reporting the goodness-of-fit tests, we make this issue explicit, unlike previous studies using perceptual weight analyses (with the exception of Dittrich & Oberfeld, 2009). We also applied a rather strict criterion ($p \geq .2$) for accepting a model. If a global test of goodness of fit indicates lack of fit, this can have several reasons (e.g., missing covariate, wrong functional form of the covariate, overdispersion, or misspecified link function; see Kuss, 2002). In fact, an inspection of the data did not reveal a simple explanation for the lack of fit of the respective regression models. Recall that for each listener and level profile, we fitted two separate regression models (one for the *loud* trials and one for the *soft* trials). With only one exception, the test indicated a lack of fit ($p < .2$) only for *one* of these two models. Thus, it does not seem that the goodness-of-fit tests indicate a general problem with the decision model for certain level profiles and/or listeners. In an additional analysis, we excluded all models showing a lack of fit and compared the resulting mean weights with the mean weights calculated on the basis of all models. For all experiments and level profiles, the patterns of weights were virtually identical in both cases, indicating that the models exhibiting a lack of fit did not result in a systematic distortion of the estimated weights.

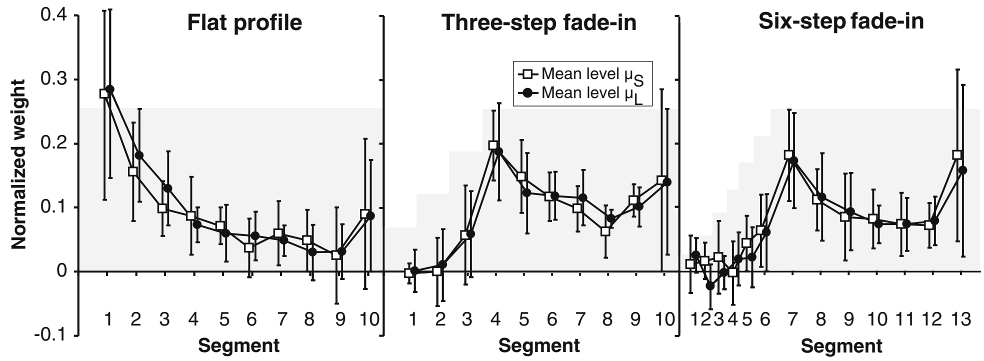

Fig. 2 Experiment 1: Mean relative perceptual weights computed across 7 listeners, as a function of segment number and mean level. The gray areas depict the level profiles. Left panel: Flat level profile. Middle panel: Three-step fade-in. Right panel: Six-step fade-in. Open boxes, mean level μ_S; filled circles, mean level μ_L. Error bars represent 95% confidence intervals

(1–10) and mean level (μ_L or μ_S). The effect of segment was significant, $F(9, 54) = 10.24$, $p < .001$, $\widetilde{\varepsilon} = .38$, partial $\eta^2 = .63$. Thus, as in previous studies (Dittrich & Oberfeld, 2009; Ellermeier & Schrödl, 2000; Pedersen & Ellermeier, 2008; Plank, 2005; Rennies & Verhey, 2009), the temporal weights deviated significantly from the uniform weighting pattern corresponding to the performance of an ideal observer (Berg, 1989). The effect of mean level and the Segment × Mean Level interaction were not significant (all p values >.1). To test for a primacy and recency effect, the weights on the first (1) and on the last (10) segments, respectively, were compared with the average weight on the four middle segments (4–7). Paired-samples t-tests indicated a significant primacy effect, $t(6) = 3.75$, $p = .010$ (all p values for t-tests reported in this article are two-tailed), but no significant recency effect, $t(6) = 0.9$. As the confidence intervals in Fig. 2 show, the weight on the last segment varied considerably across participants. Three listeners showed a clear recency effect; the others did not. In contrast, all but 1 participant showed a primacy effect. This finding is consistent with previous experiments using a similar task, all of which have reported a significant primacy effect but frequently have shown only a weaker or nonsignificant recency effect (Dittrich & Oberfeld, 2009; Pedersen & Ellermeier, 2008; Rennies & Verhey, 2009).

Did the fade-in imposed on the first three or six segments result in the expected approximately uniform temporal weights? As can be seen from the mean weights for the three-step fade-in displayed in Fig. 2, center panel, this was clearly not the case. Instead, the weights assigned to the attenuated fade-in segments (1–3) were not significantly different from zero, as indicated by the error bars showing 95% confidence intervals. In contrast, the weights assigned to the unattenuated segments (4–10) exhibited a *delayed primacy effect,* with the highest weight being assigned to segment 4. The same type of ANOVA as that for the flat level profile showed a significant effect of segment, $F(9, 54) = 8.69$, $p < .001$, $\widetilde{\varepsilon} = .39$, $\eta^2 = .59$, confirming the nonuniform weighting. The effect of mean level and the Segment × Mean Level interaction were not significant (all p values >.4). The average weight assigned to the fade-in segments was significantly smaller than the average weight assigned to the unattenuated segments, $t(6) = 6.38$, $p = .001$. The same type of ANOVA as above, conducted only on the weights on the unattenuated segments (4–10), showed a marginally significant effect of segment, $F(6, 36) = 2.96$, $p = .092$, $\widetilde{\varepsilon} = .32$, $\eta^2 = .33$. The weight on the first unattenuated segment (4) was significantly higher than the average weight on the middle three unattenuated segments (6–8), $t(6) = 4.58$, $p = .004$. The latter two analyses confirm the observed delayed primacy effect. The weight on the last segment was not significantly higher than the average weight on the middle three unattenuated segments, $t(6) = 0.92$. Thus, there was no significant recency effect. The individual data again indicated interindividual variability with respect to the recency effect. Four listeners showed a recency effect; the other 3 listeners did not. In comparison, all the participants showed a delayed primacy effect.

For the six-step fade-in (Fig. 2, right panel), the weights followed approximately the same pattern as that for the three-step fade-in. All the listeners assigned very small weights to the attenuated segments. As can be seen by the confidence intervals in Fig. 2, right panel, only the weight on segment 6, for which the attenuation was only 2.5 dB, was significantly greater than zero. The weights on the unattenuated segments again exhibited a delayed primacy effect, for 6 of the 7 listeners. A repeated measures ANOVA showed a significant effect of segment, $F(9, 54) = 9.60$, $p < .001$, $\widetilde{\varepsilon} = .27$, $\eta^2 = .62$. There was also a significant effect of mean level, $F(1, 6) = 23.07$, $p = .003$, $\widetilde{\varepsilon} = .27$. This effect was not expected. The Segment × Mean Level interaction was not significant, however (all p values >.6). An ANOVA conducted for the weights assigned to the seven unattenuated segments showed a marginally significant effect of segment, $F(6, 36) = 3.87$, $p = .050$, $\widetilde{\varepsilon} = .34$, $\eta^2 = .39$. The weight on the first unattenuated segment (7) was significantly higher than the average weight on the middle three unattenuated segments (9–11), $t(6) = 5.36$, $p = .002$. The weight on the last segment was not significantly higher than the average weight on segments 9–11, $t(6) = 1.93$. Thus, with the six-step fade-in, we found evidence for a delayed primacy effect, but not for a recency effect. The individual weighting curves exhibited patterns similar to those in the three-step fade-in condition. The same 4 listeners showed a clear recency effect, while the other 3 listeners did not.

To compare the pattern of temporal weights between the three different level profiles, it was first necessary to have an equal number of temporal weights for each level profile, in order to be able to conduct the ANOVA. Therefore, in the six-step fade-in condition, in which there were 13, rather than 10, segments, the weights assigned to segments 1 and 2, 3 and 4, and 5 and 6, respectively, were averaged. As a result, there were now exactly ten temporal weights for each level profile. A repeated measures ANOVA with the within-subjects factors of level profile, segment, and mean level showed a significant Level Profile × Segment interaction, $F(18, 108) = 22.33$, $p < .001$, $\widetilde{\varepsilon} = .34$, $\eta^2 = .79$, confirming the differences in temporal weights between the level profiles. To gain a more detailed insight into these differences, three separate ANOVAs with the same factors as above were conducted, for the three pairs of level profiles. The weights for the three-step fade-in and for the six-step fade-in level profiles differed significantly from the weights for the flat level profile [Level Profile × Segment interaction $F(9, 54) = 25.47$, $p < .001$, $\widetilde{\varepsilon} = .56$, and $F(9, 54) = 23.74$, $p < .001$, $\widetilde{\varepsilon} = .45$, respectively]. The two level profiles containing a fade-in differed only insofar as the six-step fade-in contained more level steps and, thus, was "smoother." Did this difference result in different temporal weights? Yes, there was a significant Level Profile × Segment interaction, $F(9, 54) = 2.59$, $p = .015$, $\widetilde{\varepsilon} = 1.0$. One obvious difference between the two weighting patterns is that the final fade-in segment received a weight significantly higher than zero only in the six-step fade-in condition.

As was stated in the introduction, our expectation concerning the effects of a fade-in on the temporal weights had been that an approximately uniform pattern of weights would be adopted, due to the less abrupt onset of the signal. We observed a delayed primacy effect instead. However, it could still be the case that the weights on the unattenuated segments in the fade-in conditions were less variable than those for the flat level profile. To answer this question, the coefficient of variation ($CV = SD/M$) was computed as a measure of the variability of the weights assigned to the ten (flat level profile) or seven (fade-in conditions) unattenuated segments, for each combination of participant, level profile, and mean level. As is shown in Fig. 3, the mean CV for the flat level profile was higher than in the fade-in conditions, indicating that a fade-in results in a less variable pattern of weights on the unattenuated segments. A repeated measures ANOVA, with the within-subjects factors being level profile and mean level, showed a significant effect of level profile, $F(2, 12) = 7.46$, $p = .024$, $\widetilde{\varepsilon} = .62$, $\eta^2 = .55$. As a post hoc test, three separate ANOVAs were conducted, comparing the

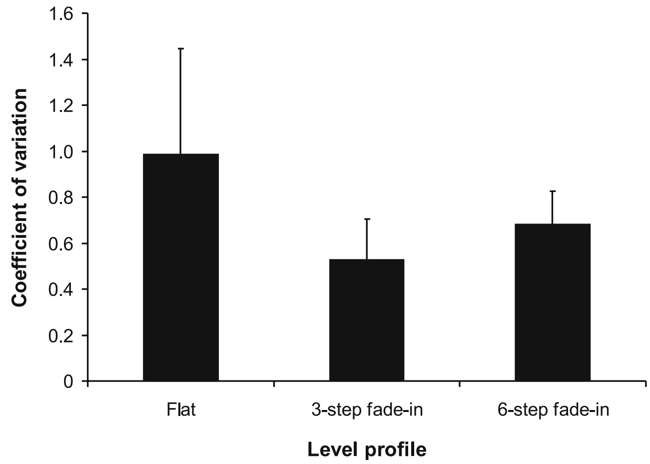

Fig. 3 Experiment 1: The coefficient of variation ($CV = SD/M$) was calculated as a measure of the variability of the weights assigned to the unattenuated segments. The graph shows the mean coefficient of variation as a function of level profile. Error bars represent 95% confidence intervals

CVs for each pair of level profiles. The CV of the unattenuated segments for the three-step fade-in condition was significantly different from the CV for the flat level profile and the six-step fade-in condition, $F(1, 6) = 11.19$, $p = .016$, and $F(1, 6) = 9.80$, $p = .020$, respectively. The difference in CV between the flat level profile and the six-step fade-in condition was only marginally significant, $F(1, 6) = 4.16$, $p = .087$.

Sensitivity

Apart from the effects on the temporal weights, did the level profile have an effect on accuracy in the intensity discrimination task? For example, the near-zero weights assigned to the fade-in segments mean that listeners did not make optimal use of the information available for classifying a given sound as either soft or loud (see Berg, 1990). As a consequence, was the accuracy in the absolute identification task inferior in the fade-in conditions? To answer this question, sensitivity, as indexed by the AUC (see Swets, 1986b), was computed for each participant and each level profile.[4] If, on a given trial, the arithmetic mean of the segment sound pressure levels (SPLs) was greater than 60.0 dB SPL (the midpoint between the *loud* and the

[4] Note that because the data are from a one-interval task using a binary response, d' would be a valid measure of sensitivity only under the assumption of equal-variance Gaussian distributions for *signal* and *noise* on the internal continuum (see Green & Swets, 1966; Swets, 1986a). In contrast, the AUC is a valid measure for sensitivity that does not require strong assumptions about the internal distributions (e.g., Macmillan, Rotello, & Miller, 2004; Swets, 1986b). AUC corresponds to the proportion of correct responses obtained with the same stimuli in a forced choice task (Green & Moses, 1966; Green & Swets, 1966, p. 49; Hanley & McNeil, 1982; Iverson & Bamber, 1997).

soft distribution), a *loud* response was considered as a hit. If the mean SPL was lower than or equal to 60 dB SPL, a *loud* response was considered as a false alarm (Oberfeld & Plank, 2005; Pedersen & Ellermeier, 2008). Note that this classification scheme corresponds directly to Eq. (1). For a given listener and level profile, the hits and false alarms from each session constituted one point on the ROC curve. A maximum-likelihood procedure (Dorfman & Alf, 1968) was used for fitting a binormal model (Hanley, 1988). AUC and its variance were computed from the best-fitting estimates of slope and intercept (Metz, Herman, & Shen, 1998; Swets, 1979).

A repeated measures ANOVA showed that the level profile had no significant effect on AUC, $F(2, 12) = 2.50$ (flat, $M = .70$, $SD = .076$; three-step fade-in, $M = .75$, $SD = .058$; six-step fade-in, $M = .74$, $SD = .051$). Thus, despite the strongly differing patterns of temporal weights, the sensitivity in the absolute identification task did not differ between the three level profiles.

Discussion

Reducing the abruptness of the signal onset by imposing a gradual increase in level on the beginning of a level-fluctuating noise did not result in the expected approximately uniform temporal weights. Instead, we found evidence for a delayed primacy effect. Thus, the primacy effect observed with a flat level profile cannot be attributed to capture of attention to the onset of the stimulus. The near-zero weights assigned to the fade-in segments are especially surprising, because they indicate that listeners virtually ignored the information provided by this temporal part of the stimulus. Weber's law holds for intensity discrimination of wideband stimuli at levels at least 30 dB above the detection threshold (e.g., Houtsma, Durlach, & Braida, 1980; Miller, 1947), so that intensity resolution is independent of level. Given this fact, and given that the difference in mean level and the standard deviation of the distributions were identical for all the segments, the fade-in segments should have provided the same amount of information as the unattenuated segments. Thus, the observed nonuniform temporal weighting reflects a suboptimal use of the available information about global intensity (Berg, 1989). The potential alternative explanation—that due to the use of a one-interval task, intensity resolution was inferior for the attenuated segments—was addressed in Experiment 3.

Note that we obtained relative judgments of the global loudness of level-fluctuating sounds in order to estimate perceptual weights. We did not obtain absolute loudness judgments in the sense of how loud the different sounds were on a loudness scale. We also did not measure whether, on average, sounds with different level profiles differed in loudness. A discussion of the latter question can be found in several studies comparing the loudness of *ramped* (i.e., containing a fade-in) and *damped* (i.e., decreasing in level across time) sounds (e.g., Neuhoff, 1998; Ries, Schlauch, & DiGiovanni, 2008; Stecker & Hafter, 2000; Susini et al., 2007). Finally, note that the task was not to discriminate between different level profiles, in the sense of deciding (for example) whether the sound had a flat level profile or contained a fade-in. Sounds with one and only one level profile were presented in each experimental block, and the task was to judge them as being either soft or loud in a one-interval task.

Experiment 2: are the temporal weights adjusted on a trial-by-trial basis?

Experiment 1 showed that a gradual increase in level at the beginning of a stimulus results in a delayed primacy effect, with zero weights on the fade-in part and the highest weight assigned to the first unattenuated temporal element. Thus, the temporal weights differed considerably between stimuli with different level profiles. Experiment 2 was designed to answer the question of whether listeners adapt their weights to the level profile on a trial-by-trial basis. In Experiment 1, the level profiles were presented blockwise, so that the listener encountered 150–200 successive trials with the same level profile. Thus, listeners might have consciously adopted a (suboptimal) strategy of ignoring the attenuated segments. In Experiment 2, we introduced trial-by-trial variation in the level profile. Specifically, we presented two types of fade-ins randomly interleaved within each experimental block. The *fast fade-in* was an increase in level across the first three temporal segments, whereas the *slow fade-in* comprised the first six segments (see Fig. 4). If, now, the listeners adopted a *fixed* set of temporal weights effective for the complete experimental block, the perceptual weights would be identical for the two different level profiles. If, however, the listeners adjusted their weights on a trial-by-trial basis, a delayed primacy effect with the highest weight assigned to segment 4 would be observed in the fast fade-in condition. In contrast, in the slow fade-in condition, the highest perceptual weight would be expected on segment 7. Our hypothesis was that the latter pattern of results indicating trial-by-trial adjustment of the temporal weights would be observed in Experiment 2.

Method

Participants

Five volunteers (4 women, 1 man; 21-41 years of age) participated in the experiment for course credit. Two of them had already participated in Experiment 1. All listeners

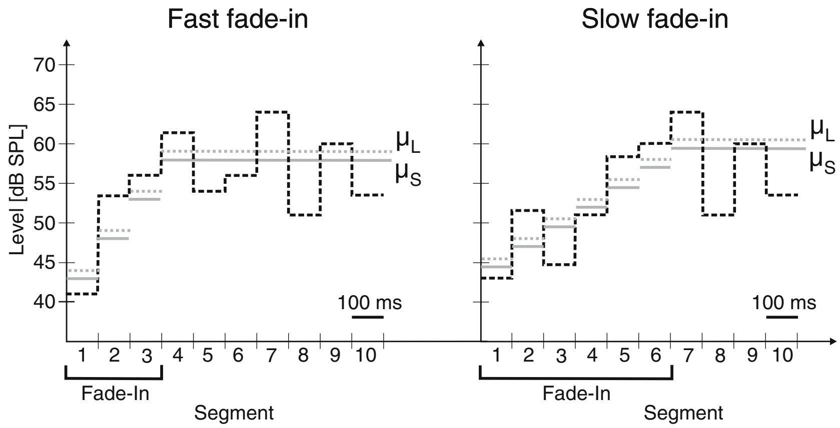

Fig. 4 Schematic depiction of the stimulus configurations in Experiment 2. On each trial, a level-fluctuating noise with a fast fade-in (left panel) or with a slow fade-in (right panel) was presented. The level profile varied on a trial-by-trial basis. The level of each segment was drawn independently from one of two normal distributions differing in their means (represented by solid and dashed gray lines). The black dashed lines show example segment levels. The participants' task was again to decide whether the sound had been soft or loud

reported normal hearing and had detection thresholds better than 10 dBH at all octave frequencies between 500 and 8000 Hz, measured in a two-interval forced choice, adaptive procedure with a three-down, one-up rule. The listeners were naïve with respect to the hypotheses under test.

Apparatus

The apparatus was the same as that in Experiment 1.

Stimuli and procedure

As in Experiment 1, the stimuli were level-fluctuating wideband noises containing ten contiguous temporal segments with a duration of 100 ms each. The construction of the stimuli by drawing from normal distributions was exactly as in Experiment 1. Two level profiles were presented. In the fast fade-in condition, the level of the first through third segments was attenuated by 15, 10, and 5 dB, respectively (see Fig. 4). In the slow fade-in condition, the levels of the first through sixth segments were attenuated by 15.0, 12.5, 10.0, 7.5, 5.0, and 2.5 dB, respectively.

After 1 hr of practice, each listener completed 2,000 trials in 40 blocks, 1,000 trials in the slow fade-in condition and 1,000 trials in the fast fade-in condition. The two level profiles were presented randomly interleaved in each block of 50 trials. To prevent trial-by-trial changes in mean level beyond the difference between loud and soft trials (see the Method section for Experiment 1), 2.25 dB were subtracted from each of the ten segment levels in the fast fade-in condition, in order to equalize the mean level in the fast fade-in and slow fade-in conditions. No trial-by-trial feedback was provided. The experiment comprised five sessions.

Results

Goodness of fit

The relative perceptual weights were estimated from the trial-by-trial data, using the same analyses as those in Experiment 1. The RSS test indicated a lack of fit ($p < .2$) for only 5 of the 20 (Listener × Level Profile × Mean Level) fitted multiple logistic regression models. The AUC ranged from .70 to .83 ($M = .76$, $SD = .039$), indicating reasonably good predictive power.

Perceptual weights

Figure 5 shows the mean relative temporal weights for the two level profiles.

The relative perceptual weights were analyzed via a repeated measures ANOVA. The within-subjects factors were level profile (fast fade-in, slow fade-in), segment (1–10), and mean level (μ_L or μ_S). A delayed primacy effect was evident for both level profiles, indicating that the listeners adjusted their weights on a trial-by-trial basis. The attenuated segments received weights close to zero, except for segment 6 in the slow fade-in condition, for which the attenuation was only 2.5 dB. In both conditions, the highest weight was assigned to the first unattenuated segment. Thus, the position of the highest weight varied as a function of level profile. All the listeners except 1 showed this pattern. There was no evidence for a recency effect. The participant who showed no delayed primacy effect assigned high weights to segment 9 in both conditions. This listener had also participated in Experiment 1 and had shown a recency effect there.

The effect of segment was significant, $F(9, 36) = 11.35$, $p < .001$, $\tilde{\varepsilon} = .61$, $\eta^2 = .74$, confirming the nonuniform

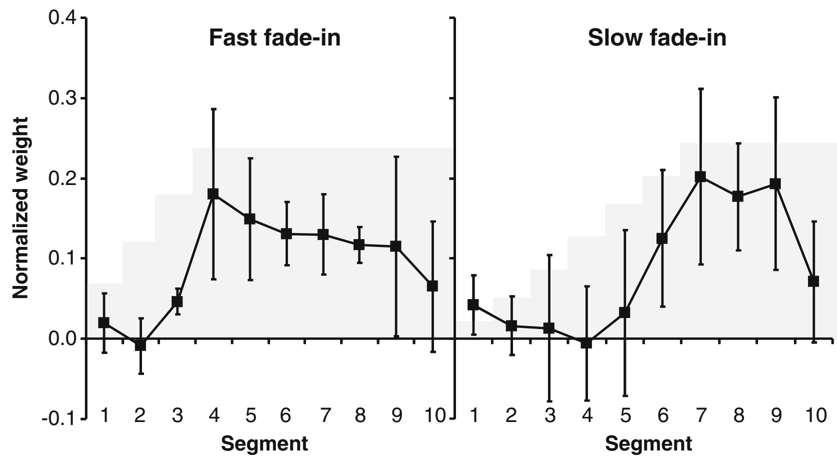

Fig. 5 Experiment 2: Mean relative perceptual weights for 5 participants, averaged across mean level (μ_L and μ_S). The horizontal axis shows the segment number. The gray areas depict the level profiles. Left panel: Fast fade-in. Right panel: Slow fade-in. Error bars represent 95% confidence intervals

temporal weighting. Two post hoc ANOVAs separately analyzing the weights for the two level profiles showed a significant effect of segment in both the fast fade-in and the slow fade-in conditions, $F(9, 36) = 10.76$, $p < .001$, $\tilde{\varepsilon} = .77$, $\eta^2 = .73$, and $F(9, 36) = 9.63$, $p = .001$, $\tilde{\varepsilon} = .37$, $\eta^2 = .71$, respectively. The Segment × Level Profile interaction was also significant, $F(9, 36) = 8.86$, $p = .002$, $\tilde{\varepsilon} = .33$, $\eta^2 = .69$, confirming the observation that the listeners adjusted their temporal weights on a trial-by-trial basis.

Sensitivity

The sensitivity did not differ significantly between the two level profiles (slow fade-in, AUC = .72, *SD* = .028; fast fade-in, AUC = .73, *SD* = .089), $t(4) = 0.34$.

Discussion

Our observation of a delayed primacy effect for both level profiles and the clear differences between the two weighting patterns demonstrate that presenting a given level profile for several consecutive trials is not a necessary condition for listeners to adopt a specific pattern of temporal weights. Instead, the participants adjusted their weights on a trial-by-trial basis, virtually ignoring the attenuated segments and assigning the highest weight to the first unattenuated segment. This shows that the delayed primacy effect does not result from a strategy consciously adopted for the complete block.[5] Rather, the participants appear to have decided in a *stimulus-driven* fashion, basing their judgments automatically on particular perceptual characteristics of the sound. The idea of capture of attention could be taken up again at this point, assuming that the delayed primacy effect is due to the saliency of the segment with the highest level after a varying number of softer segments. This first unattenuated segment could also be interpreted as the peak of a sequence of short sound events rising in level. An alternative explanation for listeners ignoring the attenuated part of the sequence could arise from a reduced intensity resolution for the low-level segments. This question was addressed in Experiment 3.

We did not present a flat level profile in this experiment, and doing so would be an interesting follow-up study. However, the two level profiles we presented allowed us to positively answer the question of whether listeners adjust their weights on a trial-by-trial basis. Interleaving a fade-in profile and a flat profile could even be viewed as a weaker test of this idea, because the latter two profiles differ more strongly than the two different fade-ins we used in the present experiment.

Experiment 3: is the delayed primacy effect confined to a one-interval task?

Experiments 1 and 2 showed that a gradual increase in level at the beginning of a stimulus results in a delayed primacy effect. We argued that the nonuniform temporal weighting reflects a suboptimal use of the available information about overall intensity (Berg, 1989). Especially the near-zero weights to the fade-in segments mean that listeners virtually ignored the information provided by this temporal part of the stimulus. However, this interpretation of the observed weighting patterns rests on the assumption that each temporal segment provides an identical amount of information concerning the overall intensity of the stimulus.

[5] On a brief questionnaire administered at the end of each experiment, we asked the participants about the "strategy" they had used when deciding whether a sound was soft or loud. They generally reported that they found it difficult to describe how they had actually made their decisions. Several listeners reported, in fact, that they had concentrated on the beginning of the stimulus, a strategy that would be compatible with the observed primacy effect. One participant who was one of the few showing a strong recency effect reported having paid attention to the end of the sound. However, across all participants, the statements seemed rather unsystematic, probably owing to the fact that we asked only a single open question.

While, as was discussed above, Weber's law holds for wideband noise (Houtsma et al., 1980) and thus, in general, intensity resolution should be identical for the unattenuated and the attenuated parts of the stimulus, the peculiarities of the one-interval, absolute identification task used in Experiments 1 and 2 might have resulted in a difference in intensity resolution between the two parts of the noise. According to the model of intensity discrimination by Braida, Durlach, and colleagues (e.g., Braida & Durlach, 1972; Durlach & Braida, 1969), listeners use *context coding* in a one-interval task (see Oberfeld, 2008b, for an in-depth discussion of context coding). In this mode, the representation of intensity is based on a comparison with internal or external references (Braida et al., 1984), such as, for example, the edges of the intensity range. The model predicts intensity resolution to be superior near sharp edges in the distribution of intensities presented during the course of the experiment (Braida et al., 1984). Now, if one considers the distribution of segment levels presented, for example, in the three-step fade-in condition in Experiment 1, Fig. 6 shows that there was a sharp peak at higher intensities around 60 dB SPL, corresponding to the mean level of the seven unattenuated segments. In contrast, due to the different amounts of attenuation imposed on the three fade-in segments, the intensity distribution exhibited a much flatter shape at lower intensities. Thus, according to the Braida and Durlach model, intensity resolution in the context-coding mode might have been inferior for the fade-in part of the stimulus. If this were the case, placing smaller weights on the fade-in segments would be compatible with optimal information integration, according to which the

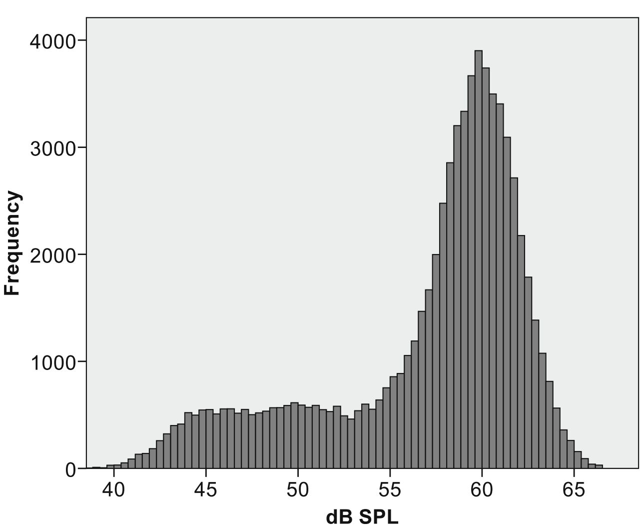

Fig. 6 Histogram of the distribution of segment levels in the three-step fade-in condition of Experiment 1. A sharp peak can be seen around 60 dB SPL, corresponding to the seven unattenuated segments, whereas the distribution is flatter at lower intensities, due to the different amounts of attenuation for the three fade-in segments

weight should be proportional to intensity resolution (integration model; Green, 1958). Experiment 3 was designed to decide whether context coding and the resulting difference in intensity resolution between unattenuated and attenuated segments caused the underweighting of the fade-in segments. To this end, a two-interval task was used, instead of a one-interval task. According to the Braida and Durlach model, listeners use the *trace mode,* rather than the context-coding mode, when comparing two sounds presented in rapid succession (Berliner & Durlach, 1973). In this mode of intensity processing, the position of the to-be-judged intensity within the intensity distribution plays no role. Thus, if the delayed primacy effect were to be found again using the two-interval task, this would be evidence *against* a role of context coding.

Method

Participants

Seven volunteers (6 women, 1 man; 21–29 years of age) participated in the experiment for payment or course credit. One listener had already participated in Experiment 1. All the listeners reported normal hearing and had detection thresholds better than 20 dB HL at all octave frequencies between 500 and 4000 Hz, measured in a two-interval forced choice, adaptive procedure with a three-down, one-up rule. The listeners were naïve with respect to the hypotheses under test.

Apparatus

The apparatus was the same as that in Experiment 1.

Stimuli and procedure

The flat level profile and the three-step fade-in level profile from Experiment 1 were presented. As in Experiment 1, the stimuli were level-fluctuating wideband noises containing ten contiguous temporal segments with a duration of 100 ms each. Two noises were presented on each trial, as depicted in Fig. 7. On each trial and for each of the two intervals, the sound pressure levels of the ten temporal segments were drawn independently from a normal distribution with a mean of μ_S = 60 dB SPL and a standard deviation of σ = 2.0 dB for the interval containing the softer noise. For the interval containing the louder noise, the mean was μ_L = 61 dB SPL, also with a standard deviation of σ = 2.0 dB. The louder noise was presented in interval 1 or interval 2 with identical probability. The two noises were presented with a silent interstimulus interval of 500 ms. Listeners selected the interval containing the louder noise.

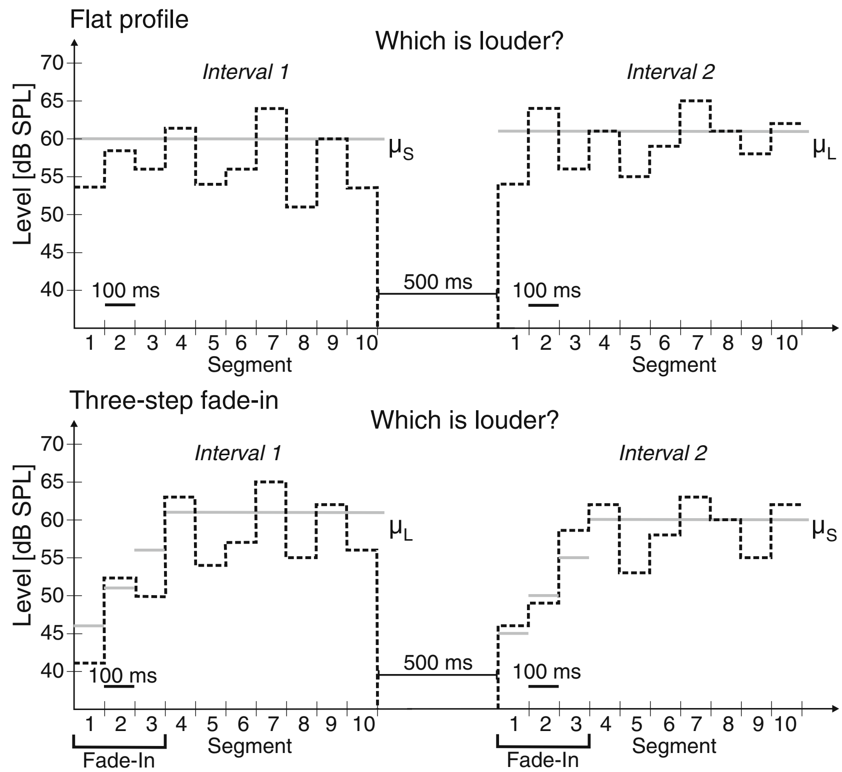

Fig. 7 Schematic depiction of the stimulus configurations in Experiment 3. On each trial, two level profiles were presented in a two-interval forced choice task. Listeners decided whether the first or the second interval had contained the louder noise. On each trial, the level of each segment was drawn independently from one of two normal distributions differing in their means (see gray lines). The dashed lines show example segment levels. The noise with levels drawn from the *loud* distribution was presented in interval 1 or 2 with identical probability. Two types of level profiles were presented: a flat level profile (upper panel) and a level profile with a three-step fade-in (lower panel)

No trial-by-trial feedback was provided. In the fade-in condition, the levels of the first through third segments were attenuated by 15.0, 10.0, and 5.0, respectively. The next trial followed the response after an intertrial interval of 2 s.

After 1 hr of practice, each listener completed 2,000 trials in 40 blocks, 1,000 trials in the fade-in condition and 1,000 trials for the flat level profile. The two level profiles were presented in a blockwise manner.

Data analysis

Multiple logistic regression was used to estimate the relative perceptual weights from the trial-by-trial data. The binary responses served as the dependent variable. Listeners' *interval 2 louder* responses were coded as 2 and *interval 1 louder* responses as 1. Due to the two-interval task, the ten within-trial *segment level differences* served as predictors (see Dittrich & Oberfeld, 2009). For each trial and each segment ($i = 1 \ldots 10$), the difference ΔL_i between the level of segment i in interval 2 and the level of segment i in interval 1 was computed. This analysis is based on a decision model similar to Eq. 1, but using segment level differences (ΔL_i) rather than segment levels (L_i):

$$D(\Delta L) = \left(\sum_{i=1}^{10} w_i \Delta L_i \right) - c. \tag{3}$$

The model assumes that a listener responds that the louder noise was presented in the second interval if $D(\Delta \boldsymbol{L}) > 0$, and that

$$P(\text{"Interval 2 louder"}) = \frac{e^{D(\Delta L)}}{1 + e^{D(\Delta L)}}. \tag{4}$$

Due to the difference in mean level between the two intervals, the within-trial segment level differences were correlated. Therefore, separate logistic regression analyses were conducted for the trials on which the noise with the higher mean level (μ_L) occurred in interval 1 and for the trials on which the position of the noise with mean level μ_L was interval 2.

Results

Goodness of fit

The RSS test indicated a lack of fit ($p < .2$) for only 7 of the 28 (listener × level profile × position μ_L) fitted multiple logistic regression models. The AUC ranged from .69 to .92 ($M = .82$, $SD = .058$).

Perceptual weights

Figure 8 shows the mean relative temporal weights for the two level profiles. The patterns of weights were similar to

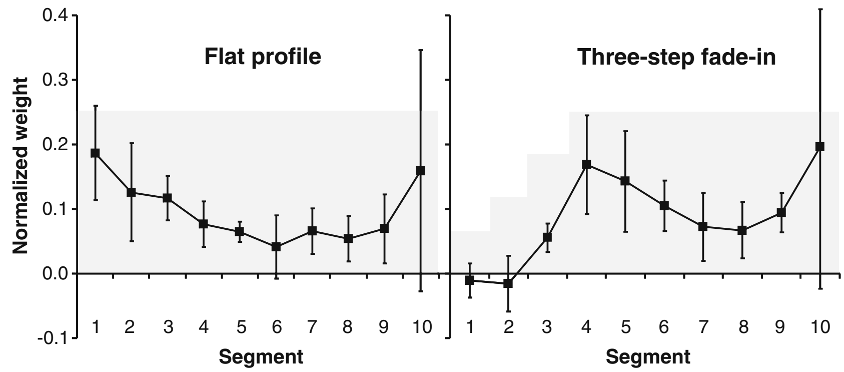

Fig. 8 Experiment 3: Mean relative perceptual weights across 7 listeners as a function of segment number. Left panel: Flat level profile. Right panel: Fade-in. Same format as Fig. 5

the weights obtained in Experiment 1 with a one-interval procedure (see Fig. 2). For the flat level profile, the weights followed a primacy/recency pattern, while a delayed primacy effect was evident for the fade-in condition. Thus, the small weights on the fade-in segments cannot be attributed to the use of a one-interval task.

The relative perceptual weights were analyzed via a repeated measures ANOVA with the within-subjects factors of level profile (flat, fade-in), segment (1–10), and position of the noise with mean level μ_L. The effect of segment was not significant, $F(9, 54) = 1.93$. However, the segment × level profile interaction was significant, $F(9, 54) = 28.87$, $p < .001$, $\widetilde{\varepsilon} = .47$, $\eta^2 = .80$. Thus, the level profile had a significant effect on the pattern of weights. Two post hoc ANOVAs separately analyzing the weights for the two level profiles showed a significant effect of segment in the fade-in condition, $F(9, 54) = 5.56$, $p = .040$, $\widetilde{\varepsilon} = .146$, $\eta^2 = .48$, but only a marginally significant effect for the flat level profile, $F(9, 54) = 3.41$, $p = .084$, $\widetilde{\varepsilon} = .18$, $\eta^2 = .36$.

To test for a primacy and recency effect for the flat level profile, the weight on the first (1) and last (10) segments, respectively, was compared with the average weight on the four middle segments (4–7). These tests indicated a significant primacy effect, $t(6) = 6.10$, $p = .001$, but no significant recency effect, $t(6) = 1.43$. For the fade-in condition, the weight on the first unattenuated segment (4) was significantly higher than the average weight on the middle three unattenuated segments (6–8), $t(6) = 4.82$, $p = .003$. This result confirms the observed delayed primacy effect. The weight on the last segment was not significantly higher than the average weight on the middle three unattenuated segments, $t(6) = 1.33$.

Large confidence intervals on the last segment stand out in both conditions, again pointing to considerable interindividual differences in the weighting patterns, as reported in a previous study (Pedersen & Ellermeier, 2008). For both level profiles, 2 listeners showed a strong recency effect, 3 listeners showed a moderate recency effect, and the 2 remaining listeners showed no recency effect. One listener with a strong recency effect had also participated in Experiment 1 and had shown a clear recency effect there, too. All the listeners except the participant with the strongest recency effect showed a primacy effect for the flat level profile, and a delayed primacy effect in the fade-in condition.

Sensitivity

In a two-interval task, the ROC curve can be expected to be symmetric about the negative diagonal (Green & Swets, 1966; Markowitz & Swets, 1967; Norman, 1964), and thus d', computed from a single point on the ROC curve, is a valid measure of sensitivity. The sensitivity was significantly higher for the flat level profile ($d' = 0.99$, $SD = 0.26$) than for the fade-in condition ($d' = 0.85$, $SD = 0.30$), $t(6) = 2.69$, $p = .036$.

Discussion

Using a two-interval task, we observed approximately the same temporal weights as in the corresponding conditions in Experiment 1. This clearly indicates that the delayed primacy effect is not due to the potentially inferior intensity resolution at lower sound pressure levels in a one-interval task. Thus, the weighting patterns observed in the one-interval tasks cannot be explained by a context-coding effect (Braida et al., 1984). An alternative explanation could be related to the temporal structure of the sounds containing a fade-in. Experiment 4 addressed this question.

Experiment 4: is the delayed primacy effect caused by perceptual segmentation?

Experiment 3 demonstrated that the delayed primacy effect—specifically, the small weights on the attenuated segments—cannot be attributed to inferior intensity resolution for the attenuated segments due to the use of context coding (Braida et al., 1984). An alternative explanation would be that the listeners perceived the stimuli with fade-in as consisting of a *variable* part and a *stable* part and ignored the variable part.[6] In the fade-in stimuli presented in Experiment 1–3, the level changed systematically within a range of 15 dB during the fade-in, while the level remained relatively constant for the unattenuated segments. Therefore, it is possible that the listeners found the increase in level at the beginning of the sound distracting, or less reliable, and therefore selectively attended to the stable part. Experiment 4 was designed to test whether such a perceptual segmentation can explain the small weights on the variable part of the stimulus. To this end, the temporal weights for a flat level profile were compared with the weights for a stimulus starting with an *inverse fade-in*—that is, a gradual *decrease* in level during the first three segments (see Fig. 9). If perceptual segmentation into a variable and a stable part was the cause of the ignorance of the segments constituting the gradual change in level, the same pattern of weights (delayed primacy effect) as that for the fade-in condition should be observed.

Method

Participants

Eight volunteers (7 women, 1 man; 19–29 years of age) participated in the experiment for payment or course credit. One listener had already participated in Experiment 3, and 1 listener had participated in Experiment 1 and 3. All the listeners reported normal hearing and had detection thresholds better than 10 dB HL at all octave frequencies between 250 and 8000 Hz, measured in a two-interval forced choice, adaptive procedure with a three-down, one-up rule. The listeners were naïve with respect to the hypotheses under test.

Apparatus

The apparatus was the same as that in Experiment 1, but the stimuli were presented to the right ear only.

Stimuli and procedure

Two level profiles were presented in a one-interval absolute identification task, just as in Experiment 1 and 2. The stimuli for the flat level profile were constructed exactly like the corresponding stimuli in Experiment 1, except that the grand mean level was 50 dB SPL rather than 60 dB SPL in order to avoid overly loud sounds in the inverse fade-in condition. To create the stimuli for the latter condition, the levels of the first through third segments were amplified by 15, 10, and 5 dB, respectively (see Fig. 9).

[6] We thank Stefan Berti for suggesting this possibility.

Results

Goodness of fit

The same type of data analysis as that in Experiment 1 was used. The RSS test indicated a lack of fit ($p < .2$) for only 8 of the 32 (listener × level profile × mean level) fitted multiple logistic regression models. The AUC ranged from .63 to .85 ($M = .73$, $SD = .058$).

Perceptual weights

Figure 10 shows the mean relative temporal weights for the two level profiles. For the flat level profile, a primacy effect was again evident. In the inverse fade-in condition, an extraordinarily high weight was observed on the first segment—that is, the segment with the highest mean level. As the confidence intervals in Fig. 10, right panel, show, only the weights to the first two segments were significantly higher than 0. Thus, the data provide clear evidence against the segmentation hypothesis.

The relative perceptual weights were analyzed via a repeated measures ANOVA with the within-subjects factors of level profile (flat, inverse fade-in), segment (1–10), and mean level. The effect of segment was significant, $F(9, 63) = 43.67$, $p < .001$, $\tilde{\varepsilon} = .77$, $\eta^2 = .86$. The Segment × Level Profile interaction was also significant, $F(9, 63) = 22.48$, $p < .001$, $\tilde{\varepsilon} = .52$, $\eta^2 = .76$.

For the flat level profile, comparisons of the weights on the first and last segments, respectively, with the average weight on the middle segments, as in Experiment 1, indicated a significant primacy effect, $t(7) = 4.38$, $p = .003$, but only a marginally significant recency effect, $t(7) = 1.92$, $p = .095$.There were no large differences between the individual weighting patterns in this experiment. For the flat level profile, only 1 participant showed no primacy effect, and only the same participant showed a recency effect. This participant had exhibited clear recency effects in Experiment 1 and 3, pointing to a high temporal stability of the individual weighting patterns. In the inverse fade-in condition, all the listeners showed a pronounced primacy effect.

Sensitivity

Sensitivity did not differ significantly between the two level profiles (flat, AUC = .67, $SD = .089$; inverse fade-in, AUC = .61, $SD = .052$), $t(7) = 1.71$.

Discussion

The rather extreme pattern of weights for the inverse fade-in condition, with almost exclusive weight assigned to the

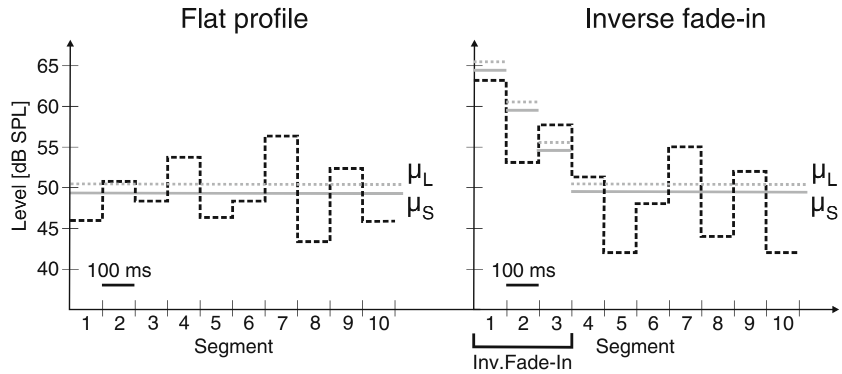

Fig. 9 Schematic depiction of the stimulus configurations in Experiment 4 (one-interval absolute identification task). A noise with flat level profile (left panel) and a noise with a three-step inverse fade-in (right panel) were presented. The stimuli were constructed as in Experiment 1 (see Fig. 1). Gray lines represent mean levels; black dashed lines show example segment levels

first segment, shows that the effects of the level profile on the temporal weights cannot be explained by perceptual segmentation of the level-fluctuating noise into a variable and a stable part. If the variability of the first part of the sound caused the participants to ignore those segments, this should have resulted in a delayed primacy effect once again. Instead, the segment with the highest mean level received the highest weight. This indicates that the mean level of a segment is a strong predictor of the importance of this segment for a participant's loudness judgment.

General discussion

In four experiments, we presented time-varying wideband noise stimuli fluctuating in level and estimated temporal weights in loudness judgment tasks by means of perceptual weight analysis. We found a consistent and robust effect of the level profile on the pattern of temporal weights.

For a flat level profile where the mean level remained constant, we observed a primacy effect; that is, the highest weight was assigned to the first temporal segment. This observation is consistent with the results of previous studies (e.g., Pedersen & Ellermeier, 2008), and a primacy effect was recently also observed for judgments of annoyance (Dittrich & Oberfeld, 2009).

The new experimental manipulation in our experiments was to alter the level profile by imposing a gradual increase in level (fade-in) on the first 300 ms of the noise and, thus, making the stimulus onset less abrupt. Contrary to expectation, the fade-in did not result in a uniform pattern of weights but, instead, gave rise to a delayed primacy effect. The attenuated temporal segments constituting the fade-in received near-zero weights, while the highest weight was assigned to the first unattenuated segment. Thus, the results from Experiment 1 and 2 are evidence against a capture of attention to the onset of the stimulus as the origin of the primacy effect. The effect of the level profile was observed with a blockwise presentation of the different level profiles in Experiment 1, as well as in Experiment 2, where the level profile changed from trial to trial, indicating that the participants adjusted their temporal weights on a trial-by-trial basis. This result also speaks against the possibility that the observed weighting patterns were due to a decision strategy that the listeners adopted at the beginning of a block of trials with the same level profile.

In Experiment 3, we tested the hypothesis that the near-zero weights on the fade-in segments might be due to inferior intensity resolution for the attenuated segments, owing to the use of context coding in the one-interval task applied in Experiment 1 and 2 (Braida & Durlach, 1972).

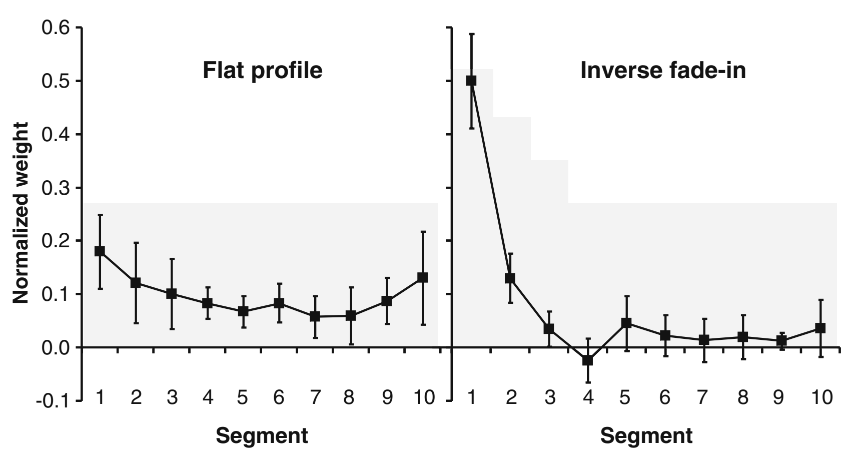

Fig. 10 Experiment 4: Mean relative perceptual weights across 8 participants as a function of segment number. Left panel: Flat level profile. Right panel: Inverse fade-in. Same format as Fig. 5

However, in the two-interval task used in Experiment 3, where context coding should play no role and, thus, the intensity resolution should be identical for unattenuated and attenuated segments, we found the same patterns of temporal weights as in the corresponding conditions in Experiment 1. Thus, the ignoring of the attenuated segments cannot be explained by reduced intensity resolution at lower sound pressure levels. Were the listeners able to use the level information from the fade-in segments in the two-interval task but, nevertheless, disregarded it? A potential cause for such a behavior would be that they did not judge the fade-in part to be reliable enough, due to its variability and, thus, complexity. This question was addressed in Experiment 4.

In Experiment 4, the fade-in was replaced by an inverse fade-in (i.e., a gradual decrease in level across the first 300 ms), in order to test the hypothesis that the effect of the fade-in could be attributed to a perceptual segmentation of the noise into a variable part (fade-in) and a stable part (unattenuated segments). The temporal weights provided clear evidence against this hypothesis, however, with almost exclusive weight being assigned to the first segment in the inverse fade-in condition. Thus, listeners do not, on principle, ignore the variable or complex part of a time-varying auditory stimulus.

What could be the cause of the observed effects of the level profile on the temporal weights? We propose that two independent processes are in effect. As Dittrich and Oberfeld (2009) suggested, the primacy effect observed for the flat level profile could be explained by assuming that the segment levels are processed as serially sorted information, analogous to item lists in working memory (e.g., Postman & Phillips, 1965), where the characteristic serial position curve is observed. It has been suggested that the primacy and the recency effects observed in such experiments can be explained by temporal distinctiveness (e.g., Mondor & Morin, 2004; Murdock, 1960; Neath, 1993; Surprenant, 2001). According to this concept, beginning and end items of the stimuli have only one neighboring item and, therefore, are more distinct, resulting in better discriminability. In fact, Neath, Brown, McCormack, Chater, and Freeman (2006) reported that distinctiveness predicted accuracy in an absolute identification task—that is, the psychophysical procedure used in Experiments 1, 2, and 4 of the present study. One of the alternative accounts for serial position effects would be an attentional primacy gradient (Farrell & Lewandowsky, 2002; Lewandowsky & Murdock, 1989; Page & Norris, 1998). This explanation assumes that during encoding, the amount of attention devoted to each additional list item decreases (for a discussion, see Oberauer, 2003). Thus, attentional mechanisms might underlie the primacy effect observed in the present study, even though our data are evidence against capture of attention toward the first element. Models like the attentional primacy gradient also include additional mechanisms accounting for a recency effect (see Oberauer, 2003), although it is unclear how the strong individual differences concerning the presence or absence of a recency effect observed, for example, in our Experiment 3 could be explained.

The second process explains the effects of attenuating or amplifying certain segments, thus introducing differences in mean level between segments. We propose that the higher weight on segments with a higher mean level could be due to attention to the most physically intense elements. In fact, it has been reported that more intense elements receive higher weight even if they provide less information about the correct response than do less intense elements (Berg, 1990; Lutfi & Jesteadt, 2006; Turner & Berg, 2007) or, alternatively, and in many cases equivalently, that *louder* elements receive higher weight (e.g., Rennies & Verhey, 2009). A potential explanation might be that more intense elements have a higher saliency than do less intense elements. Concerning a potential role of saliency, Pedersen and Ellermeier (2008) showed that a spectral change in a level-fluctuating sequence from a wide- to a narrow-band sound leads to a second "primacy" effect in the middle of the temporal sequence. Here, the segment immediately after the spectral change received a high weight comparable to that for the first segment in the sequence. This suggests that perceptual weighting is guided by saliency. In the case of a fade-in, the first unattenuated segment not only is the first segment with a regularly high level, but also could be perceived as the top of a sequence of rising intensity. Noting that saliency and distinctiveness are virtually exchangeable concepts, it may be possible to integrate these ideas into the distinctiveness framework discussed above. Following this rationale, the first unattenuated segment in the fade-in level profile possesses high distinctiveness that improves the sensory memory or short-term memory representation of the loudness of this segment, ultimately determining how much weight it receives in the judgment of loudness. However, a critical test of these hypotheses would at least require a valid and independent measure of saliency.

An alternative description of the ideas discussed in the preceding paragraph would be that the fade-in might have acted as an attention cue, directing attention to the first unattenuated segment. What effects could, for example, a visual attention signal presented before the onset of the noise be expected to have?[7] Would this attention signal have removed the delayed primacy effect? Such a finding seems unlikely, for two reasons. First, the sounds with a flat level profile were also presented without a warning signal,

[7] We are grateful to Lance Nizami for raising this question.

and nevertheless, the high weight assigned to the first segment is compatible with the idea that attention was directed to this segment. Second, Oberfeld (2008a) presented a sequence of auditory signals before a noise containing a fade-in. This rhythmic context should have been even more effective than a visual cue for directing attention to the onset of the noise. Nevertheless, a delayed primacy effect was evident in the temporal weights.

It remains for future experiments to identify the minimum deviation in mean segment level from a flat profile for which the weights start to deviate from the pattern observed with a flat level profile, to explore the dependence of the temporal weighting patterns on overall sound duration and to answer the question as to whether the effect of the level profile on the temporal weights can be modified by trial-by-trial feedback (Pedersen & Ellermeier, 2008), which would suggest that top-down control is possible. It also remains to be shown to what extent models for the loudness of time-varying sounds (e.g., Chalupper & Fastl, 2002; Glasberg & Moore, 2002) can account for some effects of the level profile on the temporal weights. Previous studies have concluded, however, that existing models cannot explain the primacy effect observed with a flat level profile (Pedersen, 2006; Rennies, Verhey, & Fastl, 2010). Note that a recent model of spectral loudness summation predicts an onset accentuation (Rennies, Verhey, Chalupper, & Fastl, 2009), but due to its short effective time constants, the model probably cannot explain the primacy effects observed with sound durations of 1 s, as in the present study.

As was discussed in the introduction, technical measures such as L_{eq} or one of its frequency-weighted variants are frequently used as estimates of the loudness of a time-varying sound, even though it is rather well accepted that these technical measures may deviate substantially from perceived loudness in certain conditions. To test whether a model allowing for a nonuniform temporal weighting of the segment levels (e.g., the model specified in Eq. 1) represents a gain in information, two different multiple logistic regression models were fitted to the data from the four experiments, separately for each participant, level profile, and mean level. The restricted model contained as a predictor only the L_{eq} and, thus, corresponded to the traditional approach for estimating the loudness of time-varying sounds, while the full model contained the L_{eq} plus the individual segment levels (i.e., Eq. 1 with L_{eq} added to the weighted sum of segment levels). A likelihood-ratio test was used for model comparison. Because the full model containing both the segment levels and the L_{eq} has additional free parameters equal to the number of segments, the test statistic is distributed as χ^2_k, where k is number of segments. Table 1 (Column 4) shows that the fit of the full model was significantly better than the fit of the simpler model ($p < .05$) in a large proportion of cases, compatible with the hypothesis that the prediction of loudness is improved by allowing for a nonuniform weighting of individual temporal portions of the signal. These results are compatible with the findings by Dittrich and Oberfeld (2009). The only exception was Experiment 4. For the inverse fade-in, the fact that the restricted model performed nearly as well as the full model could be explained by assuming that for this level profile, the judgments were dominated by the first segment (see Fig. 10, right panel)—that is, by the segment with the highest mean level. Energy-based measures such as L_{eq} are also dominated by components with the highest sound pressure level, because

Table 1 Comparison of Goodness-of-Fit and Predictive Power of the Full Model (Containing As Predictors the L_{eq} and the Segment Levels) and the Restricted Model (Containing as Predictor only the L_{eq}). The Fourth Column Shows the Number of Cases Where Goodness-of-Fit of the Full Model Was Significantly Higher Than That For the Restricted Model (Likelihood-Ratio Test; See Text). The Two Rightmost Columns Show the Area Under the Receiver Operating Characteristic Curve (AUC; With *SD*s In Parentheses) of the Restricted and of the Full Models

				AUC	
Experiment	Level Profile	No. of Fitted Models	No. of Better-Fitting Full Models ($p < .05$)	Restricted Model	Full Model
1	Flat	14	14	.76 (0.06)	.80** (0.06)
	3-Step Fade-in	14	12	.73 (0.05)	.79** (0.05)
	6-Step Fade-in	14	12	.73 (0.05)	.82** (0.06)
2	Slow Fade-in	10	8	.74 (0.06)	.78** (0.06)
	Fast Fade-in	10	6	.75 (0.05)	.78** (0.04)
3	Flat	14	14	.76 (0.06)	.84** (0.04)
	Fade-in	14	11	.77 (0.08)	.83* (0.06)
4	Flat	16	7	.66 (0.04)	.71** (0.06)
	Inverse Fade-in	16	2	.74 (0.04)	.76** (0.04)

Note. Asterisks indicate full models with a significantly higher AUC than the restricted model (see text). *$p < .05$. **$p < .01$.

the energy is a quadratic function of the sound pressure. It remains unclear, however, why, for the flat level profile in Experiment 4, the full model did not show the clear advantage in performance as in the other experiments. The superior performance of the full model indicated by the goodness-of-fit tests was corroborated by an analysis of the predictive power of the two alternative models in terms of AUC. For each experiment and each level profile, separate repeated measures ANOVAs with the within-subjects factors of model (restricted vs. full) and mean level were conducted. As the two rightmost columns in Table 1 show, the AUC for the full model was, in all cases, significantly higher than the AUC for the restricted model.

To conclude, the level profile of time-varying sounds has a systematic effect on temporal perceptual weights for loudness. Additionally, we confirmed the nonuniform temporal-weighting patterns for sounds with a flat level profile. Taken together, our results suggest that models and technical measures of loudness not taking into account temporal weights are oversimplifications. The prediction of loudness can be improved significantly by considering the dependence of the perceptual weights on the temporal position within a stimulus.

Author Note We are grateful to Kerstin Dittrich, Bernadette Hachgenei, and Maike Rathgeber for their assistance in data collection. Felicitas Klöckner helped with figure preparation.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Ahumada, A. J. (2002). Classification image weights and internal noise level estimation. *Journal of Vision, 2*(1), 121–131.

Ahumada, A. J., & Lovell, J. (1971). Stimulus features in signal detection. *The Journal of the Acoustical Society of America, 49*, 1751–1756.

Alexander, J. M., & Lutfi, R. A. (2004). Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights. *The Journal of the Acoustical Society of America, 116*, 2234–2247.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic semantic and formal similarity. *The Quarterly Journal of Experimental Psychology, 18*, 362–365.

Barkhausen, H. (1926). Ein neuer Schallmesser für die Praxis. *Zeitschrift fur Technische Physik, 7*, 599–601.

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

Berg, B. G. (1990). Observer efficiency and weights in a multiple observation task. *The Journal of the Acoustical Society of America, 88*, 149–158.

Berg, B. G. (2004). A molecular description of profile analysis: Decision weights and internal noise. *The Journal of the Acoustical Society of America, 115*, 822–829.

Braida, L. D., & Durlach, N. I. (1972). Intensity perception: II. Resolution in one-interval paradigms. *The Journal of the Acoustical Society of America, 51*, 483–502.

Berliner, J., & Durlach, N. I. (1973). Intensity perception: IV. Resolution in roving-level discrimination. *The Journal of the Acoustical Society of America, 53*, 1270–1287.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., & Purks, S. R. (1984). Intensity perception. XIII: Perceptual anchor model of context-coding. *The Journal of the Acoustical Society of America, 76*, 722–731.

Chalupper, J., & Fastl, H. (2002). Dynamic loudness model (DLM) for normal and hearing-impaired listeners. *Acta Acustica United Acustica, 88*, 378–386.

Copas, J. B. (1989). Unweighted sum of squares test for proportions. *Applied Statistics-Journal of the Royal Statistical Society Series C, 38*, 71–80.

de Boer, E., & Kuyper, P. (1968). Triggered Correlation. *IEEE Transactions on Biomedical Engineering, BM15*, 169–179.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual-attention. *Annual Review of Neuroscience, 18*, 193–222.

Dittrich, K., & Oberfeld, D. (2009). A comparison of the temporal weighting of annoyance and loudness. *The Journal of the Acoustical Society of America, 126*, 3168–3178.

Dorfman, D. D., & Alf, E. (1968). Maximum likelihood estimation of parameters of signal detection theory–A direct solution. *Psychometrika, 33*, 117–124.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *The Journal of the Acoustical Society of America, 46*, 372–383.

Ellermeier, W., & Schrödl, S. (2000). Temporal weights in loudness summation. In C. Bonnet (Ed.), *Fechner day 2000. Proceedings of the 16th Annual Meeting of the International Society for Psychophysics* (pp. 169–173). Strasbourg: Université Louis Pasteur.

European Parliament and Council of the European Union (2002). Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*, 59–79.

Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *The Journal of the Acoustical Society of America, 5*, 82–108.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 1030–1044.

Gilkey, R. H., & Robinson, D. E. (1986). Models of auditory masking: A molecular psychophysical approach. *The Journal of the Acoustical Society of America, 79*, 1499–1510.

Glasberg, B. R., & Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society, 50*, 331–342.

Glasberg, B. R., & Moore, B. C. J. (2006). Prediction of absolute thresholds and equal-loudness contours using a modified loudness model. *The Journal of the Acoustical Society of America, 120*, 585–588.

Graham, F. K., & Hackley, S. A. (1991). Passive and active attention to input. In J. R. Jennings & M. G. H. Coles (Eds.), *Handbook of cognitive psychophysiology: Central and autonomic nervous system approaches* (pp. 251–356). Chichester, U.K.: Wiley.

Green, D. M. (1958). Detection of multiple component signals in noise. *The Journal of the Acoustical Society of America, 30*, 904–911.

Green, D. M. (1964). Consistency of auditory detection judgments. *Psychological Review, 71*, 392–407.

Green, D. M., & Moses, F. L. (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin, 66*, 228–234.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Grimm, G., Hohmann, V., & Verhey, J. L. (2002). Loudness of fluctuating sounds. *Acta Acustica united with Acustica, 88*, 359–368.

Hanley, J. A. (1988). The robustness of the binormal assumptions used in fitting ROC curves. *Medical Decision Making, 8*, 197–203.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29–36.

Harris, D. M., & Dallos, P. (1979). Forward masking of auditory-nerve fiber responses. *Journal of Neurophysiology, 42*, 1083–1107.

Hosmer, D. W., Hosmer, T., leCessie, S., & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine, 16*, 965–980.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.

Houtsma, A. J. M., Durlach, N. I., & Braida, L. D. (1980). Intensity perception: XI. Experimental results on the relation of intensity resolution to loudness matching. *The Journal of the Acoustical Society of America, 68*, 807–813.

Howes, W. L. (1974). Loudness function derived from data on electrical discharge rates in auditory-nerve fibers. *Acustica, 30*, 247–259.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics, 1*, 69–82.

IEC 318 (1970). *An IEC artificial ear, of the wide band type, for the calibration of earphones used in audiometry*. Geneva: International Electrotechnical Commission

Iverson, G., & Bamber, D. (1997). The generalized area theorem in signal detection theory. In R. D. Luce & A. A. J. Marley (Eds.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (p. xv). Mahwah, N.J: L. Erlbaum. 472 p.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics, 43*, 346–354.

Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: A review. *The British Journal of Mathematical and Statistical Psychology, 54*, 1–20.

Keselman, H. J., Kowalchuk, R. K., & Boik, R. J. (2000). An examination of the robustness of the empirical Bayes and other approaches for testing main and interaction effects in repeated measures designs. *The British Journal of Mathematical and Statistical Psychology, 53*, 51–67.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., & Clark, L. F. (1965). *Discharge patterns of single fibers in the cat's auditory nerve*. Cambridge, MA: MIT Press.

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

Krueger, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *The Behavioral and Brain Sciences, 12*, 251–267.

Kuss, O. (2001). *A SAS/IML macro for goodness-of-fit testing in logistic regression models with sparse data.* Paper presented at the 26th Annual SAS Users Group International Conference.

Kuss, O. (2002). Global goodness-of-fit tests in logistic regression with sparse data. *Statistics in Medicine, 21*, 3789–3801.

Lachs, G., Al-Shaikh, R., Bi, Q., Saia, R. A., & Teich, M. C. (1984). A neural-counting model based on physiological-characteristics of the peripheral auditory-system: V. Application to loudness estimation and intensity discrimination. *IEEE Transactions on Systems, Man, and Cybernetics, 14*, 819–836.

Levitt, H. (1971). Transformed up–down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*(Suppl. 2), 467–477.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review, 96*, 25–57.

Lutfi, R. A., & Jesteadt, W. (2006). Molecular analysis of the effect of relative tone level on multitone pattern discrimination. *The Journal of the Acoustical Society of America, 120*, 3853–3860.

Macmillan, N. A., Rotello, C. M., & Miller, J. O. (2004). The sampling distributions of Gaussian ROC statistics. *Perception & Psychophysics, 66*, 406–421.

Markowitz, J., & Swets, J. A. (1967). Factors affecting the slope of empirical ROC curves: Comparison of binary and rating responses. *Perception & Psychophysics, 2*, 91–100.

Metz, C. E., Herman, B. A., & Shen, J. H. (1998). Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine, 17*, 1033–1053.

Miller, G. A. (1947). Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *The Journal of the Acoustical Society of America, 19*, 609–619.

Mondor, T. A., & Morin, S. R. (2004). Primacy, recency, and suffix effects in auditory short-term memory for pure tones: Evidence from a probe recognition paradigm. *Canadian Journal of Experimental Psychology, 58*, 206–219.

Moore, B. C. J., Vickers, D. A., Baer, T., & Launer, S. (1999). Factors affecting the loudness of modulated sounds. *The Journal of the Acoustical Society of America, 105*, 2757–2772.

Murdock, B. B. (1960). The distinctiveness of stimuli. *Psychological Review, 67*, 16–31.

Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory & Cognition, 21*, 689–698.

Neath, I., Brown, G. D. A., McCormack, T., Chater, N., & Freeman, R. (2006). Distinctiveness models of memory and absolute identification: Evidence for local, not global, effects. *The Quarterly Journal of Experimental Psychology, 59*, 121–135.

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature, 401*, 695–698.

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature, 395* (6698), 123–124.

Nizami, L., & Schneider, B. (1997). Auditory dynamic range derived from the mean rate-intensity function in the cat. *Mathematical Biosciences, 141*, 1–28.

Nomoto, M., Katsuki, Y., & Suga, N. (1964). Discharge pattern and inhibition of primary auditory nerve fibers in the monkey. *Journal of Neurophysiology, 27*, 768–787.

Norman, D. A. (1964). Sensory thresholds, response biases, and the neural quantum-theory. *Journal of Mathematical Psychology, 1*, 88–120.

Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language, 49*, 469–483.

Oberfeld, D. (2008a). Does a rhythmic context have an effect on perceptual weights in auditory intensity processing? *Canadian Journal of Experimental Psychology, 62*, 24–32.

Oberfeld, D. (2008b). The mid-difference hump in forward-masked intensity discrimination. *The Journal of the Acoustical Society of America, 123*, 1571–1581.

Oberfeld, D. (2009). The decision process in forward-masked intensity discrimination: Evidence from molecular analyses. *The Journal of the Acoustical Society of America, 125*, 294–303.

Oberfeld, D., & Plank, T. (2005). Temporal weighting of loudness: Effects of a fade in. In H. Fastl & M. Frühmann (Eds.), *Fortschritte der Akustik: Plenarvorträge und Fachbeiträge der 31. Deutschen Jahrestagung für Akustik DAGA 2005, München* (pp. 227–228). Berlin: Deutsche Gesellschaft für Akustik.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761–781.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. (G. V. Anrep, Trans.)*. London: Oxford University Press: Humphrey Milford.

Pedersen, B. (2006). *Auditory temporal resolution and integration: Stages of analyzing time-varying sounds.* Unpublished doctoral thesis, Aalborg University, Aalborg, DK

Pedersen, B., & Ellermeier, W. (2008). Temporal weights in the level discrimination of time-varying sounds. *The Journal of the Acoustical Society of America, 123*, 963–972.

Plank, T. (2005). *Auditive Unterscheidung von zeitlichen Lautheitsprofilen.* Unpublished doctoral thesis, Universität Regensburg, Regensburg. Retrieved from www.opus-bayern.de/uni-regensburg/volltexte/2006/620

Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free-recall. *The Quarterly Journal of Experimental Psychology, 17*, 132–138.

Relkin, E. M., & Doucet, J. R. (1997). Is loudness simply proportional to the auditory nerve spike count? *The Journal of the Acoustical Society of America, 101*, 2735–2740.

Rennies, J., & Verhey, J. L. (2009). Temporal weighting in loudness of broadband and narrowband signals. *The Journal of the Acoustical Society of America, 126*, 951–954.

Rennies, J., Verhey, J. L., Chalupper, J., & Fastl, H. (2009). Modeling temporal effects of spectral loudness summation. *Acta Acustica United Acustica, 95*, 1112–1122.

Rennies, J., Verhey, J. L., & Fastl, H. (2010). Comparison of loudness models for time-varying sounds. *Acta Acustica United Acustica, 96*, 383–396.

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

Ries, D. T., Schlauch, R. S., & DiGiovanni, J. J. (2008). The role of temporal-masking patterns in the determination of subjective duration and loudness for ramped and damped sounds. *The Journal of the Acoustical Society of America, 124*, 3772–3783.

Scharf, B. (1978). Loudness. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (*Hearing* (pp, Vol. IV, pp. 187–242). New York: Academic Press.

Sechenov, I. M. (1965). *Reflexes of the brain*. Cambridge, MA: MIT Press. (Original work published 1863).

Smith, R. L. (1977). Short-term adaptation in single auditory nerve fibers: Some poststimulatory effects. *Journal of Neurophysiology, 40*, 1098–1112.

Stecker, G. C., & Hafter, E. R. (2000). An effect of temporal asymmetry on loudness. *The Journal of the Acoustical Society of America, 107*, 3358–3368.

Stevens, S. S. (1956). The direct estimation of sensory magnitudes: Loudness. *The American Journal of Psychology, 69*, 1–25.

Surprenant, A. M. (2001). Distinctiveness and serial position effects in tonal sequences. *Perception & Psychophysics, 63*, 737–745.

Susini, P., McAdams, S., & Smith, B. K. (2007). Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings. *Acta Acustica United Acustica, 93*, 623–631.

Swets, J. A. (1979). ROC Analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology, 14*, 109–121.

Swets, J. A. (1986a). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99*, 181–198.

Swets, J. A. (1986b). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99*, 100–117.

Tang, Z., Richards, V. M., & Shih, A. (2005). Comparing linear regression models applied to psychophysical data [Abstract]. *The Journal of the Acoustical Society of America, 117*, 2597.

Turner, M. D., & Berg, B. G. (2007). Temporal limits of level dominance in a sample discrimination task. *The Journal of the Acoustical Society of America, 121*, 1848–1851.

World Health Organization (1999). *Guidelines for community noise.* Retrieved from www.who.int/docstore/peh/noise/guidelines2.html

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 121–134.

Yu, J. J., & Young, E. D. (2000). Linear and nonlinear pathways of spectral information transmission in the cochlear nucleus. *Proceedings of the National Academy of Science, 97*, 11780–11786.

Zhang, C. Y., & Zeng, F. G. (1997). Loudness of dynamic stimuli in acoustic and electric hearing. *The Journal of the Acoustical Society of America, 102*, 2925–2934.

Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: Facts and models* (2nd ed.). Berlin: Springer.