**ACTA ACUSTICA UNITED WITH ACUSTICA**
**Vol. 101 (2015) 1105 – 1115**

DOI 10.3813/AAA.918904

# Are Temporal Loudness Weights Under Top-Down Control? Effects of Trial-By-Trial Feedback

Daniel Oberfeld
Department of Psychology, Section Experimental Psychology, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany. oberfeld@uni-mainz.de

**Summary**
Previous studies showed that listeners apply non-uniform temporal weights when judging the overall loudness of a time-varying sound. The loudness judgments are influenced more strongly by level changes on the initial temporal portion of the sound than on later temporal portions (primacy effect). In additional, higher weights are assigned to stimulus components with a higher average loudness (loudness dominance). The present study investigated whether the temporal weights in a loudness-judgment task are under top-down control. Specifically, the experiment studied whether trial-by-trial feedback helps listeners in adjusting their temporal weights to approximate the uniform weighting pattern that would maximize the accuracy in the task. The stimuli were time-varying sounds with a duration of 1 s, with either a flat level profile or with a gradual increase or decrease in level imposed on the first 300 ms. A clear primacy effect was observed for sounds with a flat level profile, with the first temporal segment receiving the highest weight. For the two other level profiles, the weight showed strong level dominance effects. The sensitivity was higher when trial-by-trial feedback rather than only block feedback was provided. The patterns of temporal weights did not differ significantly between the two types of feedback. However, trial-by-trial feedback resulted in the adoption of more efficient weighting strategies. Thus, the characteristic nonuniform patterns of temporal perceptual weights observed for loudness judgments of dynamic sounds are not removed by trial-by-trial feedback, but are under limited top-down control.

**PACS no. 43.64.Bt, 43.66.Ba, 43.66.Cb, 43.66.Fe, 43.66.Mk, 43.66.Lj**

## 1. Introduction

Sounds in our environment typically change across time in acoustic intensity, as for example the sound of a car passing by or the sound of a power drill. Often, the intensity even evolves differently in different frequency regions, leading to dynamic changes in spectral configuration. Thus, these “dynamic” or time-varying sounds are rather different from the static sounds (e.g., pure tones, broadband noise) used in many experiments on loudness, which in turn resulted in powerful models for the loudness of static sounds (e.g. [1, 2, 3, 4]). There are attempts to formulate models for the loudness of time-varying sounds, most important are the time-varying loudness model (TVL) by Glasberg and Moore [5] and the dynamic loudness model (DLM) by Chalupper and Fastl [6], the latter being the basis of a recent German standard [7]. However, further research seems to be required to gain a better understanding of the mechanisms underlying the loudness of time-varying sounds.

Received 13 March 2015,
accepted 10 August 2015.

One important finding in this context is that listeners apply strongly non-uniform temporal weights when judging the overall loudness (*global loudness*) of a longer, time varying sound (e.g. [8, 9, 10, 11]). These studies used *perceptual weight analysis* or *behavioral reverse correlation* [12, 13], which are methods of *molecular psychophysics* [14], to measure temporal weights quantifying the influence of the sound pressure level individual temporal portions of a sound on the loudness of the sound as a whole (*global loudness*). The weights show how strongly the global loudness changes when the sound pressure level of a temporal portion of the sound is changed by a certain amount. Studies on the temporal weighting of loudness consistently showed that the first 100–300 ms receive a higher weight than later portions of the stimulus [9, 10, 11, 15, 16, 17, 18, 19, 20]. This means that, for example, a 5 dB increase in the level of the *first* 100 ms of the sound causes a stronger increase in global loudness than a 5 dB increase in the level at the temporal center of the sound. This *primacy effect* (highest weight assigned to the beginning of a sound) is not predicted by common technical measures used for noise assessment, as for example the A-weighted energy-equivalent continuous sound pressure level $L_{\mathrm{Aeq}}$. Two temporal portions with identical fre-

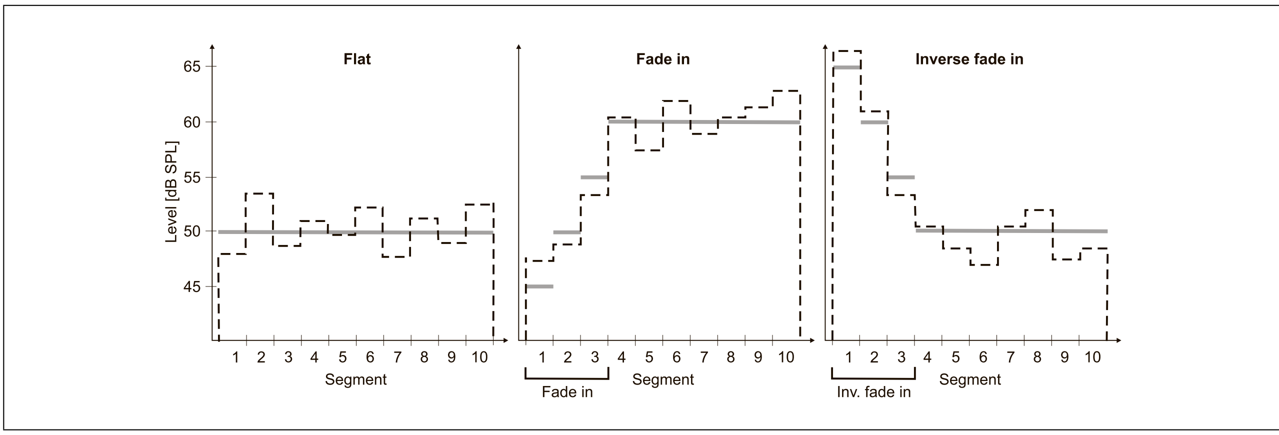


Figure 1. Level profiles. Left: Flat. Center: fade in. Right: inverse fade in. The thick gray lines show the mean segment levels for the "soft" level distribution (with mean $\mu_S$). All level profiles were presented in a 2I task (see Figure 2), second interval not shown here.

quency spectrum and level have the same impact on $L_{\mathrm{Aeq}}$ and similar measures, regardless of their temporal position within the sound (e.g., beginning versus middle versus end). Preliminary analyses also indicate that the two dynamic loudness models mentioned above (TVL and DLM) do not predict a primacy effect [17, 21]. Some experiments also observed a recency effect, that is, higher weights assigned to the final temporal portion of a sound than to the middle portion. The recency effect in the temporal weights for loudness was consistently found to be smaller than the primacy effect [8, 11, 20].

A second very consistent finding is that the *relative level* (or relative loudness) of the stimulus components strongly affects the perceptual weights, with higher weights being assigned to the most intense elements [15, 16, 20, 22, 23, 24, 25, 26, 27]. This phenomenon has been termed *loudness dominance* or *level dominance*. For example, if a gradual increase in level is imposed on the first 300 ms of a 1000 ms sound (see center panel of Figure 1), then only very small weights are assigned to the attenuated portion of the signal that constitutes the fade in [15].

The present study investigates whether the temporal weights in a loudness-related task are under top-down control. Specifically, the experiment studied whether trial-by-trial feedback helps listeners in adjusting their temporal weights to approximate the uniform weighting pattern that, according to our knowledge of auditory intensity processing, would maximize the accuracy in the task (for a detailed explanation see [15]). In virtually all previous studies on temporal weights in loudness no trial-by-trial feedback was provided, because the motivation of these studies was to measure "natural" or "spontaneous" judgments of loudness (e.g. [15]). Only one study compared temporal weights between a condition with trial-by-trial feedback and a condition with block feedback, where only the proportion of correct responses was indicated to the listener after completion of an experimental block containing about 100 trials [11]. As noted by Pedersen and Ellermeier [11], trial-by-trial feedback should be more efficient than block feedback in helping the listeners to adopt a set of decision weights that maximizes the percentage of correct responses. In fact, in a between-subjects design, on average the five listeners receiving trial-by-trial feedback showed a more uniform pattern of temporal weights than the group of five listeners that received only block feedback [11]. However, there was a considerable variation of the weighting patterns between subjects, within each of the two groups, compatible with the relatively strong individual differences typically found in perceptual weights [25, 28]. Given the rather small sample size, it is therefore somewhat difficult to decide on the basis of the study by Pedersen and Ellermeier [11] how strong the effect of trial-by-trial feedback on the temporal weights is in general.

To address this question, a within-subjects design was used in the present study. Each subject was tested on the same loudness-judgment task in two different feedback conditions, one with trial-by-trial feedback and the other with block feedback. This allowed for a direct assessment of the effect of trial-by-trial feedback, uncontaminated by potential individual differences in the temporal weighting patterns.

A second important question is whether the loudness dominance effect is also reduced by trial-by-trial feedback. In the stimuli presented by Pedersen and Ellermeier [11], all temporal portions of the sounds had the same mean level (flat level profile), so that only the primacy effect, but not loudness dominance, played a role. As will be explained below, it is likely that the primacy effect and the loudness dominance effect can be attributed to different mechanisms. For this reason, trial-by-trial feedback might differentially affect the primacy effect and loudness dominance. To answer this question, in the present experiment the effect of trial-by-trial feedback was investigated for three different level profiles. For the flat level profile (left panel in Figure 1), a reduced primacy effect, or more generally more uniform temporal weights, were expected with trial-by-trial feedback compared to block feedback [11]. The second type of level profile was a *fade-in condition*, where the level increased gradually during the first 300 ms of the 1000-ms sound (center panel in Figure 1). For this level profile, a "delayed primacy effect" was observed in two previous studies [15, 16]. The attenuated temporal segments constituting the fade in received near-zero weights, the highest weight was assigned to the first unattenuated segment, and then the weights leveled off across the remaining segments. The third level profile was the *inverse fade-in condition*, where a gradual decrease in level was imposed on the first 300 ms of the sound (right panel in Figure 1). Previous data showed a near-exclusive weight on the first (loudest) temporal portion in this condition [15]. Relatively strong level dominance effects were observed in experiments providing trial-by-trial feedback [22, 25], although the size of the effects was never compared between conditions with and without trial-by-trial feedback. In view of these results, a smaller effect of trial-by-trial feedback on the temporal weights was expected for the fade in and inverse fade in condition than for the flat level profile.

Before turning to the description of the experiment, potential explanations for the primacy effect and for the loudness dominance effect are briefly discussed. It will be argued that the existing data suggest that both the primacy effect and the loudness dominance effect can be attributed to higher-level effects like memory and attention, rather than to mechanisms in the auditory periphery.

First, can the primacy effect can be attributed to a peripheral mechanism such as the initial peak in the firing rate of auditory nerve neurons at sound onset (cf. [29])? This is unlikely because the primacy effect is also observed for a sequence of noise bursts or tones separated by pauses of 100 ms [18, 25]. With this silent interval between the sounds, *each* noise burst would have elicited a similar neuronal response of the auditory nerve, due to the fast recovery of the majority of auditory nerve fibers [30, 31]. Data of Oberfeld and Plank [15] also argue against a capture of attention due to the abrupt onset of a sound [32, 33, 34] as an explanation for the primacy effect. Oberfeld and Plank attenuated the abruptness of the onset by imposing a gradual increase in level ("fade in") across the first 300 to 700 ms of a sound with 1 s duration. This did not result in uniform temporal weights, however, but in a *delayed primacy effect*, with very small weights assigned to the attenuated segments constituting the fade in, and the highest weight assigned to the first unattenuated segment. We have proposed that the primacy effect is caused by a memory process [8], assuming that the levels of the different temporal portions of a sound are processed as serially sorted information, thus linking the results to experiments on working memory (e.g. [35]) and auditory sensory memory [36], where the characteristic serial position curve also showing a primacy effect is observed [15]. The assumption that the primacy effect is caused by a higher-level process rather than by a loudness-specific sensory mechanism seems reasonable because non-uniform temporal weights are found not only for loudness judgments, but also for frequency discrimination [13, 24, 37], and for localization/lateralization [38, 39]. One model for the primacy effect that explains a sizeable portion of the data [40] is an attentional primacy gradient [41]. According to this model, the first item of a list (i.e., the sequence of temporal segments in a time-varying sound) receives the highest attention, and less and less attention is devoted to each additional item.

The loudness dominance effect has also been attributed to a higher mechanism, attention to the loudest elements [24]. This attention-based explanation is compatible with data by Lutfi and Jesteadt [22] who found that listeners virtually ignored the softer tones in a multitone sequence where loud and soft tones alternated, even if the level information from the soft tones was rendered more reliable by presenting a larger level increment on the soft than on the loud tones. However, when the loud elements were wideband noise bursts rather than pure tones, listeners placed the higher weights on the more reliable soft tones. A plausible explanation of this result is that the spectral difference between soft and loud sounds facilitated the direction of attention to the soft elements.

## 2. Method

### 2.1. Subjects

Eight students at Johannes Gutenberg-Universität Mainz participated in the experiment voluntarily (6 women, 2 men, age between 20 and 30 years). They received partial course credit. All listeners reported normal hearing. For the right ear (the ear tested in the experiment), detection thresholds for 720-ms pure tones (including 10-ms $\cos^2$ on- and off-ramps) measured with a two-interval task and a two-down, one-up adaptive procedure [42] were better than 12 dB HL at octave frequencies between 250 Hz and 4 kHz. Once the topic of the study and potential risks had been explained to them, all participants gave written informed consent according to the Declaration of Helsinki. They were uninformed about the experimental hypotheses.

### 2.2. Stimuli and apparatus

Temporal loudness weights for level-fluctuating broadband noises of 1 s duration were measured in a two-interval sample discrimination task [43, 44, 45]. The stimuli were similar to those presented in Oberfeld and Plank [15].

The stimuli were Gaussian wide-band noises (range 20–20,000 Hz) consisting of ten contiguous temporal segments. The duration of each segment was 100 ms. Random level perturbations were imposed on the ten temporal segments, resulting in a level fluctuating noise that changed in intensity every 100 ms. Two such noises were presented on each trial (see Figure 2), with a silent inter-stimulus interval of 700 ms. The stimuli were generated digitally, played back via an RME ADI/S D/A converter ($f_s$ = 44.1 kHz, 24-bit resolution), attenuated by a TDT PA5 programmable attenuator, buffered by a TDT HB7 headphone buffer, and presented to the right ear via Sennheiser HDA 200 circumaural headphones calibrated according to IEC 318 [46]. The experiment was conducted in a double-walled sound-insulated chamber.

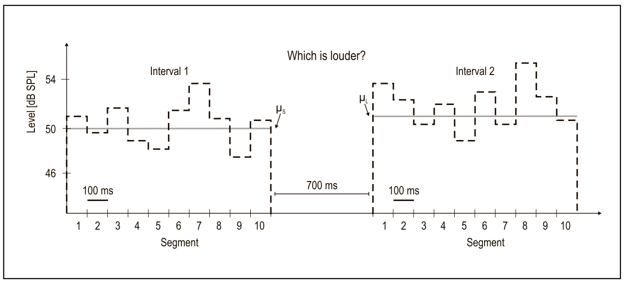


Figure 2. Example trial from the two-interval sample-discrimination task, presenting the flat level profile. Each interval contained a broadband noise that consisted of 10 contiguous 100-ms temporal segments. The sound pressure levels of the segments were randomly sampled from normal distributions. The thick gray lines show the mean segment levels. The dashed lines represent the segment levels presented on this example trial. The noise sampled from the "louder" distribution (with mean $\mu_L$) was presented in interval 1 or interval 2 with identical probability. The task was to identify the interval that contained the louder sound.

### 2.3. Procedure and experimental conditions

Three different level profiles were presented in the experiment, and two different types of feedback were provided. Both experimental factors were varied within subjects.

For the *flat level profile*, shown in Figure 2, on each trial, in one of the intervals the sound pressure levels of the ten temporal segments were drawn independently from a normal distribution with mean $\mu_S = 50$ dB SPL and standard deviation $\sigma = 2.0$ dB ("softer noise"). In the other interval, the mean of the level distribution was 1 dB higher, $\mu_L = 51$ dB SPL, also with standard deviation $\sigma = 2.0$ dB ("louder noise"). Put differently, a level increment of $\Delta\mu_L = \mu_L - \mu_S = 1$ dB was placed on each segment in one of the two observation intervals. To avoid overly loud or soft sounds, the range of levels was restricted to $\mu \pm 3\sigma$ for each interval. The "louder" noise with segment levels drawn from the distribution with mean $\mu_L$ was presented in interval 1 or interval 2 with identical probability. The two noises were presented with a silent inter-stimulus interval of 700 ms. The task was to select the interval containing the louder noise.

In the *inverse fade-in condition*, displayed in right panel of Figure 1, each sound contained a gradual decrease in level during the first three segments. The levels of the first through third segment were amplified by 15.0, 10.0, and 5.0 dB, respectively, after the 10 segment levels had been drawn from the same distributions as for the flat level profile.

In the *fade-in condition* (center panel of Figure 1), the level of the first through third segment was attenuated by 15.0, 10.0, and 5.0 dB, respectively. Thus, each sound started with a gradual increase in level. In this condition, the means of the level distributions were increased ($\mu_S = 60$ dB SPL, $\mu_L = 61$ dB SPL) so that the mean level of the first, softest segment (45 dB SPL) was at least 30 dB above the detection threshold. The detection threshold for 100-ms broadband noise segments was measured with a 3-down, 1-up adaptive procedure in a two-interval task. Across the eight listeners, the mean threshold was 14.1 dB SPL (SD = 2.75 dB).

In blocks presenting *trial-by-trial feedback*, the listener received visual feedback concerning the correctness of the response immediately after pressing a response button. In the remaining blocks, the number of correct responses and the number of incorrect responses was displayed on the screen after each block of 105 trials (block feedback), but no trial-by-trial feedback was provided.

In a sample discrimination task, a response is typically scored as correct if for example the segment levels in interval 2 were drawn from the distribution with the higher mean ($\mu_L$) and the listener responds that the louder sound was presented in interval 2. This can result in counterintuitive feedback on some trials, because with a small probability all segment levels drawn from the "loud" distribution can be lower than the segment levels drawn from the "soft" distribution. For this reason, the feedback was based on the trial-by-trial mean sound pressure level of the ten sound segments in interval 1 compared to the mean sound pressure level of the segments in interval 2 [11]. If the listener responded that the loud sound had been presented in the interval that contained the higher mean segment level, the response was scored as correct. As noted by Pedersen and Ellermeier [11], this scoring variant favors uniform temporal weights, and therefore maximized the probability that the trial-by-trial feedback would cause the listeners to adopt more uniform temporal weights.

### 2.4. Sessions

Each listener participated in a total of ten experimental sessions, each with a duration of approximately 55 minutes. In sessions 1 and 2, audiometric detection thresholds were measured and practice blocks were run for all conditions.

In sessions 3 to 10, the sample discrimination task was presented. Sessions with and without trial-by-trial feedback alternated. Four listeners received trial-by-trial feedback in sessions with odd numbers, and four listeners in sessions with even numbers. In each session, two blocks containing 105 trials each were presented for each of the three level profiles (flat, fade in, inverse fade in). The order of conditions within a session was randomized, with the restriction that a given condition was not presented in two consecutive blocks. Across the eight sessions presenting the sample discrimination task, 840 trials were collected per listener for each combination of level profile (flat, fade in, inverse fade in) and feedback type (trial-by-trial, block).

### 2.5. Statistical data analysis

The data were analyzed with repeated-measures analyses of variance (rmANOVAs) using a univariate approach with Huynh-Feldt correction for the degrees of freedom [47]. The correction factor $\tilde{\varepsilon}$ is reported, and partial $\eta^2$ is reported as measure of association strength. An $\alpha$-level of .05 was used for all analyses.

# 3. Results

## 3.1. Sensitivity

The sensitivity in the sample-discrimination task was analyzed using the signal-detection theory index $d'$. Just as for the feedback (see above), the analysis was based on the mean sound pressure level of the 10 temporal segments presented in interval 1 and in interval 2. Trials on which the mean of the 10 segment levels was higher in the second than in the 230 first interval were taken as "signal", and the remaining trials were taken as "noise". Figure 3 shows mean $d'$ as a function of level profile and feedback. An rmANOVA with the within-subjects factors level profile (flat, fade in, inverse fade in) and feedback type (block, trial-by-trial) showed a significant effect of level profile, $F(2, 14) = 11.9$, $p = .004$, $\tilde{\varepsilon} = .72$, $\eta_p^2 = .63$. The sensitivity ($d'$) was similar for the flat level profile and the fade in condition, but lower with the inverse fade in. The effect of feedback was significant, $F(1, 7) = 9.37$, $p = .018$. The mean sensitivity was higher with trial-by-trial feedback, Cohen's [48] $d_z = 1.08$. According to Cohen's [48] classification of effect sizes, this is a large effect. For comparison, in a between-subjects design, Pedersen and Ellermeier [11] found a small, non-significant increase in $d'$ with trial-by-trial feedback, for a flat level profile. In the present data, the level profile × feedback interaction was not significant, $F(2, 14) = 0.25$, indicating that the effect of feedback on sensitivity did not differ strongly between the three level profiles.

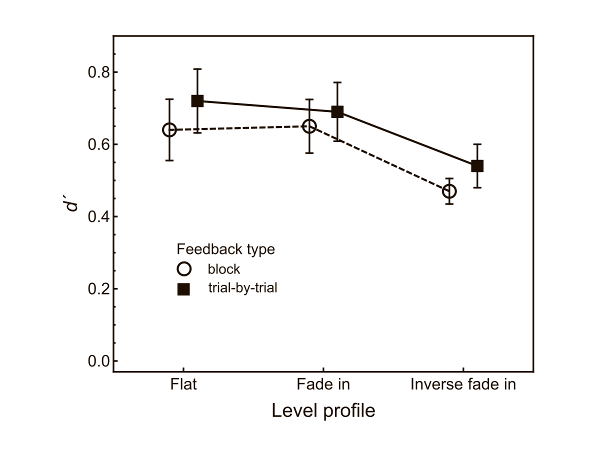


Figure 3. Mean $d'$ as a function of level profile and feedback type. Circles: block feedback. Squares: trial-by-trial feedback. Error bars show ±1 standard error of the mean.

## 3.2. Temporal weights

The perceptual weights representing the importance of the 20 temporal stimulus components (segments) presented on each trial for the decision in the sample discrimination task were estimated from the trial-by-trial data via multiple logistic regression [11, 16, 49, 50]. The decision model assumed that the listener compares a weighted sum of the 10 segment levels presented in interval 2 to a weighted sum of the segment levels presented in interval 1, and responds that the louder level-fluctuating noise was presented in interval 2 rather than in interval 1 if the difference between these weighted sums exceeds a certain decision criterion. In more formal terms, the decision variable underlying the analysis is given by

$$D(\boldsymbol{L}) = \left(\sum_{i=1}^{10} w_{2,i} L_{2,i}\right) - \left(\sum_{i=1}^{10} w_{1,i} L_{1,i}\right) + c, \tag{1}$$

where $\boldsymbol{L}$ is the vector of 20 component levels, $L_{1,i}$ denotes the level of $i$th segment ($i = 1, \ldots, 10$) in interval 1, $w_{1,i}$ is the decision weight assigned to the level of this component, $L_{2,i}$ and $w_{2,i}$ denote the segment levels and decision weights, respectively, in interval 2, and $c$ is a constant representing the decision criterion [11, 13]. In other words, $D(\boldsymbol{L})$ is a weighted sum of the 20 (interval × segment) independent component levels. The decision model assumes that the listener responds that the louder level-fluctuating noise was presented in interval 2 if $D(\boldsymbol{L}) > 0$. If $D(\boldsymbol{L}) \leq 0$, the listener is assumed to respond that the louder sound was presented in interval 1. According to the logistic model,

$$P(\text{“Louder sound in interval 2”}) = \frac{e^{D(\boldsymbol{L})}}{1 + e^{D(\boldsymbol{L})}}. \tag{2}$$

In the data analysis, the binary responses ("Louder noise in interval 1" or "Louder noise in interval 2") served as the dependent variable. The predictors (i.e., 20 component levels) were entered simultaneously. The regression coefficients were taken as the decision weight estimates. For a given component (e.g., the level of the first segment in interval 2), a regression coefficient equal to zero means that the component had no influence at all on the decision. For the same segment, a regression coefficient greater than zero means that the probability of responding that the louder sound was presented in interval 2 increased with the sound pressure level of the first segment in interval 2. A regression coefficient smaller than zero indicates the opposite relation.

A separate logistic regression model was fitted for each combination of listener, level profile, and feedback type. As we were interested in the *relative* contributions of the different components to the decision rather than in the absolute magnitude of the regression coefficients, the 20 decision weights $w_{1,i}$ and $w_{2,i}$ were normalized for each fitted model such that the sum of their absolute values was 1.0 [15], resulting in a set of relative perceptual weights for each listener, level profile, and feedback type.

A summary measure of the predictive power of a logistic regression model is AUC, the area under the receiver operating characteristic (ROC) curve [51, 52]. This measure provides information about the degree to which the predicted probabilities are concordant with the observed outcome (for details see [8]). Areas of 0.5 and 1.0 correspond to chance performance and perfect performance of the model, respectively. Across the 48 fitted logistic regression models, AUC ranged between 0.66 and 0.93 ($M$

Table I. Results of the repeated-measures ANOVA on the normalized weights (displayed in Figure 4). Note: Bold font indicates significant effects ($p < .05$). Num. $df$: numerator degrees of freedom. Den. $df$: denominator degrees of freedom. $\tilde{\varepsilon}$: Huynh-Feldt correction factor for the degrees of freedom. $\eta_p^2$: partial $\eta$-squared.

| Factor | num. $df$ | den. $df$ | $F$ | $p$ | $\tilde{\varepsilon}$ | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| Segment | **9** | **63** | **41.32** | **<.001** | **.36** | **.86** |
| Level profile | **2** | **14** | **7.80** | **.005** | **1.0** | **.53** |
| Feedback | 1 | 7 | 2.70 | .145 | | .28 |
| Interval | **1** | **7** | **15.21** | **.006** | | **.69** |
| Segment × Level profile | **18** | **128** | **66.04** | **<.001** | **.53** | **.90** |
| Segment × Feedback | 9 | 63 | 1.56 | .147 | .43 | .18 |
| Segment × Interval | **9** | **63** | **6.94** | **<.001** | **1.0** | **.50** |
| Level profile × Feedback | 2 | 14 | 1.60 | .237 | 1.0 | .19 |
| Level profile × Interval | 2 | 14 | 3.12 | .076 | .87 | .31 |
| Feedback × Interval | 1 | 7 | 4.98 | .061 | | .42 |
| Level profile × Segment × Feedback | 18 | 126 | 1.16 | .302 | 1.0 | .14 |
| Level profile × Segment × Interval | **18** | **126** | **8.76** | **<.001** | **.57** | **.56** |
| Level profile × Feedback × Interval | 2 | 14 | .23 | .801 | 1.0 | .03 |
| Segment × Feedback × Interval | 9 | 63 | .79 | .627 | .97 | .10 |
| Segment × Level profile × Feedback × Interval | 18 | 126 | .82 | .676 | 1.0 | .11 |

= 0.80, SD = 0.074), indicating on average reasonably good predictive power [53].

Figure 4 shows the mean normalized perceptual weights. The normalized weights were analyzed with an rmANOVA with the within-subjects factors segment number (1...10), interval (1 or 2), level profile (flat, fade in, inverse fade in), and feedback type (block, trial-by-trial). The ANOVA results are displayed in Table I. The effect of segment was significant. For the flat level profile, the temporal weights showed the expected primacy effect. In the fade in condition, the attenuated temporal segments constituting the fade-in received near-zero weights, while the highest weight was assigned to the first unattenuated segment (delayed primacy effect; [16]). In the inverse fade in condition, only the first two segments (with the highest mean levels) received a significant weight, as can be seen from the confidence intervals in Figure 4. A significant segment × level profile interaction confirmed that the temporal weights depend on the level profile. As a post-hoc test, separate ANOVAs were computed per level profile. For all level profiles, there was a significant effect of segment number. Taken together, these results are very similar to findings from previous studies [8, 10, 11, 15, 16], showing a primacy effect (flat level profile), a strong level dominance effect (fade in and inverse fade in conditions), and a delayed primacy effect (fade in condition).

On average, the segments in the second interval received significantly higher weights than the segments in the first interval, $d_z = 1.38$. A higher reliance on the second observation interval in 2I discrimination tasks was reported by several previous studies [17, 54, 55, 56]. The segment × interval and the segment × level profile × interval interactions were also significant. However, Figure 4 suggests that the patterns of weights in interval 1 followed the same shape as the weights in interval 2, but attenuated by a constant multiplicative factor representing the higher average weight assigned to interval 2. To gain further insight into this phenomenon, an additional ANOVA was conducted with the weights normalized (sum of the absolute values = 1.0) *per interval* rather than across the two intervals. The average weights obtained with this normalization are shown in Figure 5. It is evident that the weighting patterns are very similar for the two intervals, and in fact with the per-interval normalization the segment × interval interaction was no longer significant ($p = .16$).

Returning to the analysis with normalization across intervals, despite the normalization the effect of level profile and the level profile × interval interaction were significant. This can be attributed to the frequent near-zero weights in the inverse fade in and the fade in condition, which resulted in several slightly negative weights. The latter contribute to the sum of the absolute values of the 20 weights used for normalizat ion, explaining the difference in mean weight between level profiles. In fact, when all weights that were not significantly different from 0 (Wald $p$-value $> .2$) were set to 0, the effect of level profile and the level profile × interval interaction were no longer significant.

The central research question of this study was whether trial-by-trial feedback would result in a change in the perceptual weights. As suggested by Figure 4, this effect was absent or very weak. Most important, in the ANOVA the segment × feedback interaction was not significant. Thus, the pattern of temporal weights did not change when listeners received trial-by-trial feedback rather than only block feedback. All other interactions involving feedback were also not significant. The only exception was a trend ($p = .061$) towards a feedback × interval interaction. Somewhat surprising, the sound presented in the first interval received even slightly lower weights when trial-by trial feedback was provided.

### 3.3. Efficiency analysis

In a multiple observation task, like the intensity discrimination task for stimuli consisting of several temporal com-

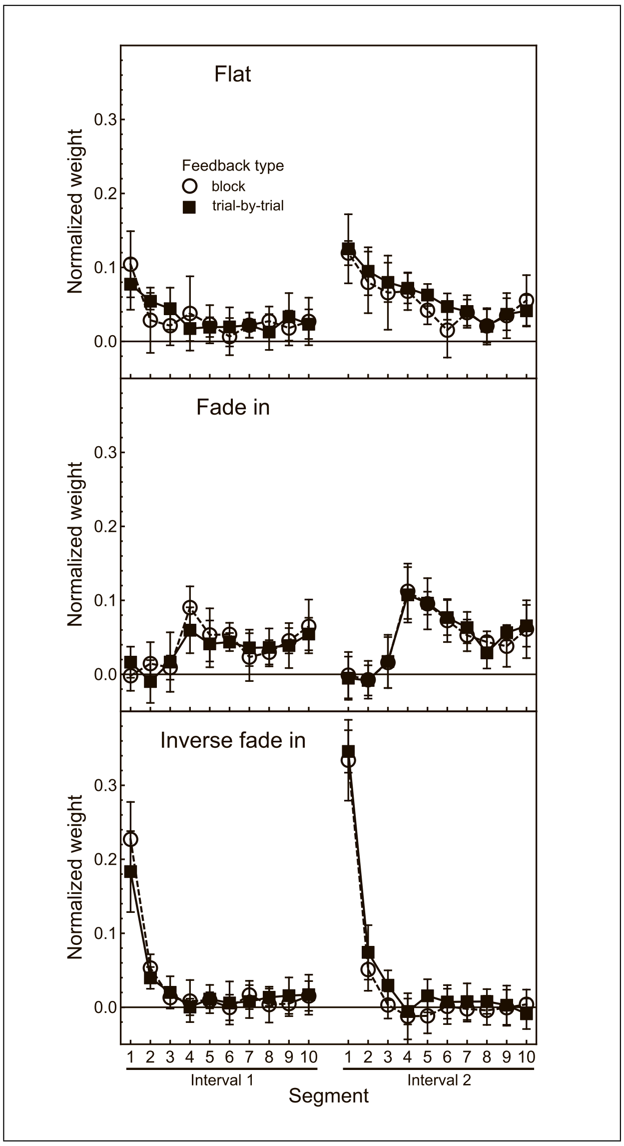


Figure 4. Mean normalized perceptual weights for the 20 temporal segments (10 segments in interval 1, 10 segments in interval 2), as a function of level profile and feedback type. Top: Flat level profile. Center: fade in. Bottom: inverse fade in. Circles: block feedback. Squares: trial-by-trial feedback. Weight normalization: sum of the absolute values of the 20 weights = 1.0. Error bars show 95% confidence intervals.

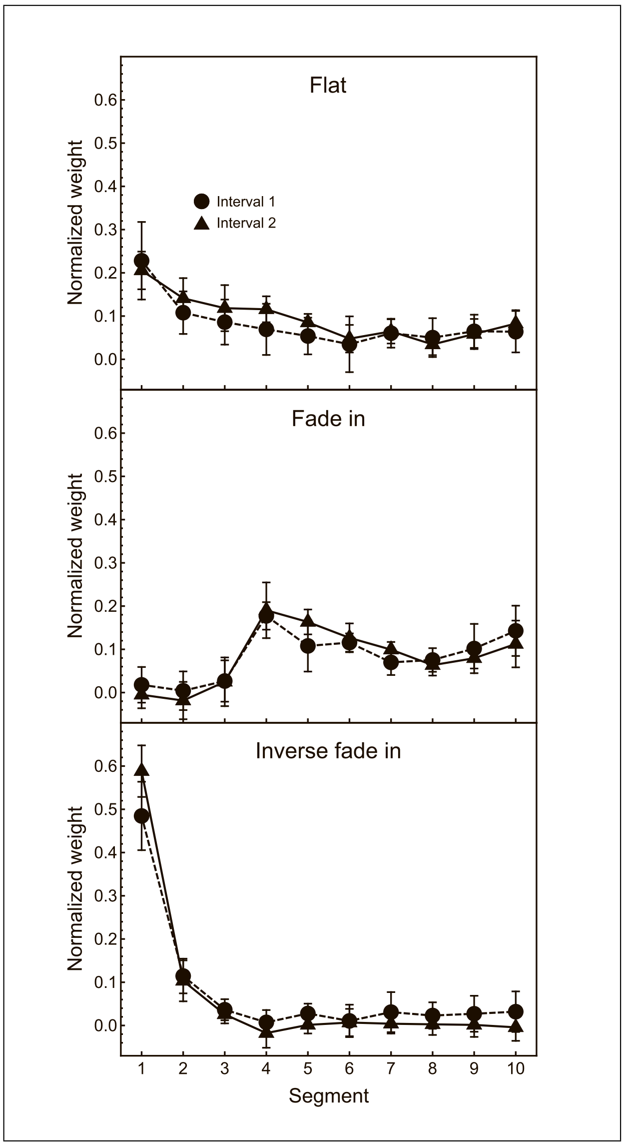


Figure 5. Mean normalized perceptual weights for the temporal segments, as a function of level profile and interval. Weight normalization: sum of the absolute values of the 10 weights in each interval = 1.0. Top: 691 Flat level profile. Center: fade in. Bottom: inverse fade in. Circles: interval 1. Triangles: interval 2. Error bars show 95% CIs.

ponents as in our experiment, at least two different factors limit the performance [24, 57]. First, the information provided by the different stimulus components might not be combined in an optimal fashion. For the present data, this is indicated by the strongly non-uniform and therefore suboptimal temporal weights. Second, "internal noise" in the sense of inaccuracies and inherent fluctuations in the sensory systems or at higher processing stages is another factor that limits the sensitivity. The present data showed no significant effects of trial-by-trial feedback on the temporal weights, but significantly higher sensitivity if trial-by-trial feedback was provided. This suggests that the loss in efficiency resulting from the use of nonoptimal weights might not be reduced by trial-by-trial feedback, but that trial-by-trial feedback may have reduced the internal noise [58]. However, both effects might play a role. To quantify the effects of trial-by-trial feedback on both factors, an analysis of observer efficieny [59] can be used [24, 25, 60]. As explained by Berg [24], the first step of this analysis is to compute the sensitivity of an ideal observer who applies optimal decision weights and whose performance is not limited by internal noise. Because all temporal segments provided the same amount of information concerning the correct response in the present discrimination task (see [15]), the proportion of correct responses should be maximal if the listener assigns equal weight to each of the 10 segments and in the two intervals [13, 15]. The ideal sensitivity is thus

$$d'_{\text{ideal}} = \left( \sum_{k=1}^{2} \sum_{i=1}^{10} \tilde{w}_{k,i} \Delta L \right) \Big/ \left( \sum_{k=1}^{2} \sum_{i=1}^{10} \tilde{w}_{k,i}^2 \sigma_{\text{ext}}^2 \right) \quad (3)$$
$$= 2.24,$$

where $\Delta L = 1$ dB is the level increment added to each segment in one of the observation intervals, $\sigma_{\text{ext}} = 2$ dB is the SD of the random level perturbations imposed on all segments, and $\tilde{w}_{k,i}$ is the optimal weight for segment $i$ in interval $k$ (all $\tilde{w}_{k,i}$ are identical).

Next, the sensitivity $d'_{\text{wgt}}$ of an observer who is not affected by internal noise but applies non-optimal decision weights is computed. For each listener and each experimental condition, the corresponding sensitivity is given by

$$d'_{\text{wgt}} = \left( \sum_{k=1}^{2} \sum_{i=1}^{10} \hat{w}_{k,i} \Delta L \right) \Big/ \left( \sum_{k=1}^{2} \sum_{i=1}^{10} \hat{w}_{k,i}^2 \sigma_{\text{ext}}^2 \right) \quad (4)$$
$$= 2.24,$$

where the $w$-terms are the observed rather than the ideal weights. The efficiency measure $\eta_{\text{wgt}} = (d'_{\text{wgt}})^2/(d'_{\text{ideal}})^2$ quantifies the loss in efficiency due to the assignment of nonoptimal weights, with a value of 1.0 representing no loss in sensitivity and a value of 0 corresponding to a complete loss in sensitivity. As can be seen in Figure 6, on average the loss in sensitivity due to nonoptimal weights was reduced by trial-by-trial feedback, except for the fade-in level profile. An rmANOVA with the within-subjects factors level profile (flat, fade in, inverse fade in) and feedback type (block, trial-by-trial) showed a significant effect of feedback type, $F(1, 7) = 15.76$, $p = .005$, $d_z = 1.40$. This means that despite that non-significant effect of trial-by-trial feedback on the pattern of temporal weights, it on average helped the listeners to combine the information from the 20 stimulus components in a more efficient fashion. It should be noted, however, that the assumption of absolutely no internal noise in the ideal observer amplifies and probably somewhat exaggerates the effects of non-optimal weighting strategies on the efficiency [25]. In particular, the sometimes slightly negative estimated weights result in a strong reduction in $\eta_{\text{wgt}}$ while in the presence of internal noise they would have only a small negative effect on sensitivity. The effect of level profile was also significant, $F(2, 14) = 51.5$, $p < .001$, $\tilde{\varepsilon} = 1.0$, $\eta_p^2 = .88$, with particularly low values of $\eta_{\text{wgt}}$ in the inverse fade-in condition where listeners almost exclusively used information from only the first two temporal segments of the sounds. The level profile × feedback interaction was not significant ($p = .30$).

In the third step, $d'_{\text{wgt}}$ is compared to the observed sensitivity ($d'_{\text{obs}}$). The efficiency measure

$$\eta_{\text{noise}} = \left(d'_{\text{obs}}\right)^2 \Big/ \left(d'_{\text{wgt}}\right)^2$$

quantifies the additional loss in efficiency due to internal noise. As Figure 7 shows, trial-by-trial feedback had almost no effect on $\eta_{\text{noise}}$. An rmANOVA showed neither a significant effect of feedback nor a significant level profile × feedback interaction (both $p$-values > .8). There was a significant effect of level profile, $F(2, 14) = 24.5$, $p < .001$, $\tilde{\varepsilon} = 1.0$, $\eta_p^2 = .77$. The higher values of $\eta_{\text{noise}}$ can be attributed to the fact that the strongly non-uniform weights in this condition resulted in very small values of $\eta_{\text{wgt}}$, which already accounted for most of the difference between the observed and the ideal sensitivity.

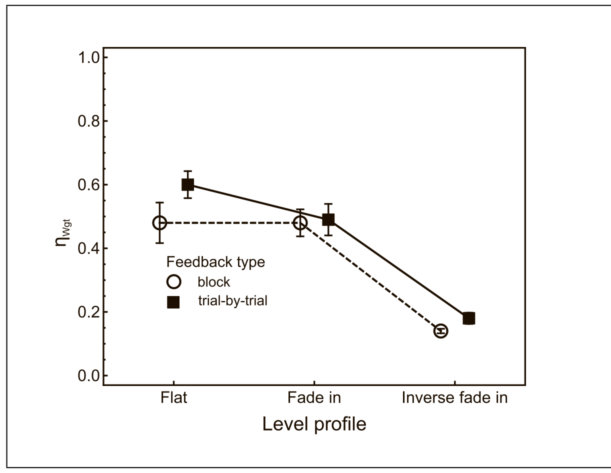


Figure 6. Mean $\eta_{\text{wgt}}$ as a function of level profile and feedback type. Circles: block feedback. Squares: trial-by-trial feedback. Error bars show ±1 standard error of the mean.

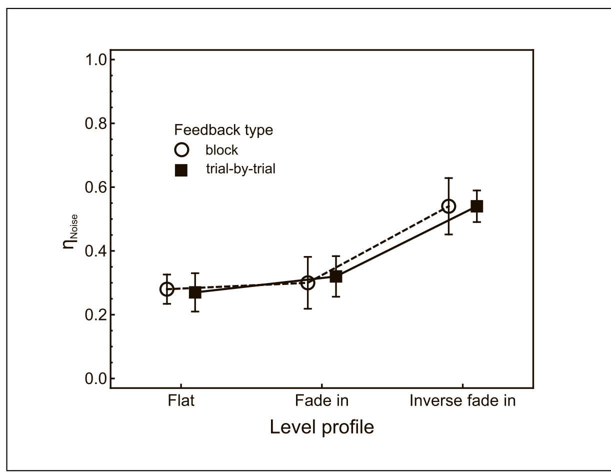


Figure 7. Mean $\eta_{\text{noise}}$ as a function of level profile and feedback type. Same format as Figure 6.

## 4. Discussion

The experiment measured the temporal weights listeners apply in loudness comparisons between two time-varying sounds of 1 s duration. The sounds had either a flat level profile, or contained a gradual increase or decrease in level across the first 300 ms. Compatible with previous studies, strongly non-uniform temporal weights were observed, with a primacy effect for the flat level profile and the inverse fade-in condition, and a delayed primacy effect for the fade-in condition.

Listeners' sensitivity ($d'$) was significantly higher when receiving trial-by-trial feedback rather than only block feedback. However, the patterns of temporal weights assigned to the 10 temporal segments in the first and in the second observation interval did not differ significantly between the two feedback conditions. Even for the flat level profile, there was no clear reduction of the primacy effect , which is at odds with the earlier findings of Pedersen and Ellermeier [11]. Qualitatively, the strongly non-uniform weights remained even if trial-by-trial feedback was provided. Still, efficiency analyses showed that trial-by-trial feedback helped the listeners to combine the information from the 20 stimulus components in a more efficient fashion. Thus, the temporal weights in the conditions with trial-by-trial feedback were on average somewhat closer to the optimal, uniform weights. This suggests that the temporal weights listeners apply when judging the global loudness of a level-fluctuating sound are under limited top-down control, although trial-by-trial feedback did not remove the higher reliance on loudness information from the beginning of a 1-s sound compared to later temporal portions of the sound. This observation is compatible with the fact that in short-term memory clear primacy and recency effects in are often observed even if trial-by-trial feedback is provided [40]. The trial-by-trial feedback also only slightly reduced the difference between the weights assigned to the softer and to the louder segments in the fade in and the inverse fade in conditions. In the same line of reasoning as for the primacy effect, this is compatible with attention being directed automatically to the loudest elements, as proposed by [24].

It would be interesting to examine whether subjects had difficulty in inferring the optimal weights from the feedback. It remains to be shown whether alternative strategies are more efficient in helping subjects to adopt more uniform weights. For example, one could explicitly instruct participants to equally consider all temporal portions of the sounds. Alternatively, a different task could be presented before the actual intensity discrimination task, to first train subjects to pay equal attention to all segments (e.g., by detecting a tone that could appear on any temporal position within the noise).

The observation that trial-by-trial feedback resulted in more efficient decision weights is consistent with studies on perceptual learning that found improvements in the decision strategies due to trial-by-trial feedback (e.g. [61, 62, 63]).

One potential limitation of the present study arises because, following Pedersen and Ellermeier [11], the condition with trial-by-trial feedback was compared to a condit ion with block feedback. Interestingly, some studies in the visual domain found almost identical effects of trial-by-trial feedback and block feedback on the sensitivity in a perceptual learning framework [64, 65]. However, these studies focused on accuracy (percent correct) rather than on the decision strategy. For this reason, it is difficult to predict whether the absence of an effect of feedback type on the perceptual weights in the present experiment can be explained by block feedback having the same effect as trial-by-trial feedback. In addition, the present data show a significant advantage of trial-by-trial feedback in terms of sensitivity and of efficiency of the weighting strategies, which contradicts the cited studies. Still, it would be interesting for future experiments to compare the sensitivity and weights between a condition with trial-by-trial feedback and a condition in which absolutely no feedback is provided.

In the present experiment, each participant received sessions with and without trial-by-trial feedback in alternating order. If one assumes that trial-by-trial feedback caused a change in the temporal weights relative to the "spontaneous" weights, there might have been some transfer of this effect from a session with trial-by-trial to a subsequent session where only blockwise feedback was provided. Averaged across all sessions, such a transfer would have reduced the observed differences in the weights between the two feedback conditions. However, the fact that there was a significant effect of feedback type on $d'$ and $\eta_{\mathrm{wgt}}$ is evidence against complete transfer, i.e., the participants arguably performed the task differently in sessions with and without trial by-trial feedback. Additional analyses of only the data obtained in the first two sessions of the main experiment also showed no significant effects of order of feedback type. Four subjects received blockwise feedback in session 3 and trial-by-trial-feedback in session 4. The remaining four subjects received the reverse order of feedback conditions. In the former group, according to the "transfer" argument, the weights might have been "spontaneous" in session 3 (blockwise feedback) but altered by trial-by-trial feedback in session 4. In the latter group, the weights in session 4 (blockwise feedback) might have been influenced by the trial-by-trial feedback provided in session 3.

However, the estimated temporal weights were very similar for the two orders, and there were no significant interactions of order of feedback type with feedback condition and segment number. Thus, the data do not indicate that the weights in the blockwise feedback condition were altered when the subject received trial-by-trial feedback throughout the previous session.

The limited top-down control indicated by the absence of a significant effect of trial-by-trial feedback on the temporal weights does not imply a peripheral or "early sensory" origin of the primacy effect and of the level dominance effect. As discussed above, both effects are compatible with an attentional mechanism. While in some circumstances the direction of attention to certain stimulus components is under voluntary control, attention shifts are often automatic. For example, in the seminal experiments by Posner [66], subject were unable to ignore misleading spatial cues in a visual detection task.

Taken together, the characteristic patterns of temporal perceptual weights observed for loudness judgments of dynamic sounds, namely the primacy effect (which can be viewed as attention directed to the beginning of the sound) and the loudness dominance effect (attention directed to the loudest elements), are not removed by trial by-trial feedback, but are under limited top-down control.

## Acknowledgement

I am grateful to Wolfgang Ellermeier and an anonymous reviewer for helpful comments on a previous version of this paper.

## References

[1] B. R. Glasberg, B. C. Moore: Prediction of absolute thresholds and equal-loudness contours using a modified loudness model. J. Acoust. Soc. Am. **120** (2006) 585–588.

[2] B. C. J. Moore: Development and current status of the "Cambridge" loudness models. Trends in Hearing **18** (2014) 1–29.

[3] E. Zwicker: Ein Verfahren zur Berechnung der Lautstärke. Acustica **10** (1960) 304–308.

[4] E. Zwicker, B. Scharf: A model of loudness summation. Psychol. Rev. **72** (1965) 3–26.

[5] B. R. Glasberg, B. C. J. Moore: A model of loudness applicable to time-varying sounds. J. Audio Eng. Soc. **50** (2002) 331–342.

[6] J. Chalupper, H. Fastl: Dynamic loudness model (DLM) for normal and hearing-impaired listeners. Acta Acust. united Acust. **88** (2002) 378–386.

[7] DIN 45631/A1: Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum - Verfahren nach E. Zwicker - Änderung 1: Berechnung der Lautheit zeitvarianter Geräusche (Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculation of the loudness of time-varying sounds). Deutsches Institut für Normung, Berlin, Germany, 2010.

[8] K. Dittrich, D. Oberfeld: A comparison of the temporal weighting of annoyance and loudness. J. Acoust. Soc. Am. **126** (2009) 3168–3178.

[9] W. Ellermeier, S. Schrödl: Temporal weights in loudness summation. – In: Fechner Day 2000. Proceedings of the 16th Annual Meeting of the International Society for Psychophysics. C. Bonnet (ed.). Université Louis Pasteur, Strasbourg, 2000, 169–173.

[10] J. Rennies, J. L. Verhey: Temporal weighting in loudness of broadband and narrowband signals. J. Acoust. Soc. Am. **126** (2009) 951–954.

[11] B. Pedersen, W. Ellermeier: Temporal weights in the level discrimination of time-varying sounds. J. Acoust. Soc. Am. **123** (2008) 963–972.

[12] A. J. Ahumada, J. Lovell: Stimulus features in signal detection. J. Acoust. Soc. Am. **49** (1971) 1751–1756.

[13] B. G. Berg: Analysis of weights in multiple observation tasks. J. Acoust. Soc. Am. **86** (1989) 1743–1746.

[14] D. M. Green: Consistency of auditory detection judgments. Psychol. Rev. **71** (1964) 392–407.

[15] D. Oberfeld, T. Plank: The temporal weighting of loudness: Effects of the level profile. Atten. Percept. Psychophys. **73** (2011) 189–208.

[16] D. Oberfeld: Does a rhythmic context have an effect on perceptual weights in auditory intensity processing? Can. J. Exp. Psychol.-Rev. Can. Psychol. Exp. **62** (2008) 24–32.

[17] B. Pedersen: Auditory temporal resolution and integration. Stages of analyzing time-varying sounds. Ph.D., Department of Acoustics, Sound Quality Research Unit, Aalborg University, Aalborg, DK, 2006.

[18] T. Plank: Auditive Unterscheidung von zeitlichen Lautheitsprofilen (Auditory discrimination of temporal loudness profiles). Dr. Phil., Philosophische Fakultät II, Universität Regensburg, Regensburg, 2005.

[19] D. Oberfeld, W. Heeren, J. Rennies, J. Verhey: Spectro-temporal weighting of loudness. PLoS One **7** (2012) e50184.

[20] E. Ponsot, P. Susini, G. Saint Pierre, S. Meunier: Temporal loudness weights for sounds with increasing and decreasing intensity profiles. J. Acoust. Soc. Am. **134** (2013) EL321–EL326.

[21] D. Oberfeld: The temporal weighting of the loudness of time-varying sounds reflects both sensory and cognitive processes. – In: Fechner Day 2009. Proceedings of the 25th Annual Meeting of the International Society for Psychophysics. M. A. Elliott, S. Antonievi, S. M. Berthaud, P. B. Bargary, C. Martyn, H. Schmidt (eds.). International Society for Psychophysics, Galway, Ireland, 2009, 261–266.

[22] R. A. Lutfi, W. Jesteadt: Molecular analysis of the effect of relative tone level on multitone pattern discrimination. J. Acoust. Soc. Am. **120** (2006) 3853–3860.

[23] M. D. Turner, B. G. Berg: Temporal limits of level dominance in a sample discrimination task. J. Acoust. Soc. Am. **121** (2007) 1848–1851.

[24] B. G. Berg: Observer efficiency and weights in a multiple observation task. J. Acoust. Soc. Am. **88** (1990) 149–158.

[25] D. Oberfeld, M. Kuta, W. Jesteadt: Factors limiting performance in a multitone intensity-discrimination task: Disentangling non-optimal decision weights and increased internal noise. PLoS One **8** (2013) e79830.

[26] R. A. Lutfi, C. J. Liu, C. N. J. Stoelinga: Auditory discrimination of force of impact. J. Acoust. Soc. Am. **129** (2011) 2104–2111.

[27] V. M. Richards, Y. Shen, C. Chubb: Level dominance for the detection of changes in level distribution in sound streams. J. Acoust. Soc. Am. **134** (2013) EL237–EL243.

[28] R. A. Lutfi, C. J. Liu, C. N. J. Stoelinga: A new approach to sound source segregation. – In: Basic Aspects of Hearing: Physiology and Perception, Volume 787. B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, H. E. Gockel (eds.). 2013, 203–211.

[29] N. Y. S. Kiang, T. Watanabe, E. C. Thomas, L. F. Clark: Discharge patterns of single fibers in the cat's auditory nerve. M.I.T. Press, Cambridge, Mass., 1965.

[30] R. L. Smith: Short-term adaptation in single auditory nerve fibers: some poststimulatory effects. J. Neurophysiol. **40** (1977) 1098–1112.

[31] D. M. Harris, P. Dallos: Forward masking of auditory-nerve fiber responses. J. Neurophysiol. **42** (1979) 1083–1107.

[32] I. P. Pavlov: Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. translated and edited by G. V. Anrep. Oxford University Press, London, 1927.

[33] J. Jonides, S. Yantis: Uniqueness of abrupt visual onset in capturing attention. Percept. Psychophys. **43** (1988) 346–354.

[34] I. M. Sechenov: Reflexes of the brain (original publication 1863). M.I.T. Press, Cambridge, Mass., 1965.

[35] L. Postman, L. W. Phillips: Short-term temporal changes in free recall. Q. J. Exp. Psychol. **17** (1965) 132–138.

[36] D. J. McFarland, A. T. Cacace: Aspects of short-term acoustic recognition memory: Modality and serial position effects. Audiology **31** (1992) 342583.

[37] R. A. Lutfi: Informational processing of complex sound. III: Interference. J. Acoust. Soc. Am. **91** (1992) 3391–3401.

[38] K. Saberi: Observer weighting of interaural delays in filtered impulses. Percept. Psychophys. **58** (1996) 1037–1046.

[39] G. C. Stecker, E. R. Hafter: Temporal weighting in sound localization. J. Acoust. Soc. Am. **112** (2002) 1046–1057.

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

[41] M. P. A. Page, D. Norris: The primacy model: A new model of immediate serial recall. Psychol. Rev. **105** (1998) 761–781.

[42] H. Levitt: Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. **49** (1971) Suppl 2:467–477.

[43] B. G. Berg, D. E. Robinson: Multiple observations and internal noise. J. Acoust. Soc. Am. **81** (1987) S33–S33.

[44] R. A. Lutfi: Informational processing of complex sound. I: Intensity discrimination. J. Acoust. Soc. Am. **86** (1989) 934–944.

[45] R. D. Sorkin, D. E. Robinson, B. G. Berg: A detection theory method for the analysis of auditory and visual displays. Proceedings of the 31st Annual Meeting of the Human Factors Society, 1987, 1184–1188.

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

[47] H. Huynh, L. S. Feldt: Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. J. Educ. Stat. **1** (1976) 69–82.

[48] J. Cohen: Statistical power analysis for the behavioral sciences, 2nd edition. L. Erlbaum Associates, Hillsdale, N.J., 1988.

[49] J. M. Alexander, R. A. Lutfi: Informational masking in hearing impaired and normal-hearing listeners: Sensation level and decision weights. J. Acoust. Soc. Am. **116** (2004) 2234–2247.

[50] R. H. Gilkey, D. E. Robinson: Models of auditory masking: a molecular psychophysical approach. J. Acoust. Soc. Am. **79** (1986) 1499–1510.

[51] A. Agresti: Categorical data analysis, 2. edition. Wiley, New York, NY, 2002.

[52] J. A. Swets: Indices of discrimination or diagnostic accuracy: their ROCs and implied models. Psychol. Bull. **99** (1986) 100–117.

[53] D. W. Hosmer, S. Lemeshow: Applied logistic regression, 2nd edition. Wiley, New York, 2000.

[54] D. Oberfeld, M. Kuta, P. Stahn: Why do forward maskers affect auditory intensity discrimination? Evidence from "molecular psychophysics". PLoS One **9** (2014).

[55] O. Dyjas, K. M. Bausenhart, R. Ulrich: Trial-by-trial updating of an internal reference in discrimination tasks: Evidence from effects of stimulus order and trial sequence. Attention Perception & Psychophysics **74** (2012) 1819–1841.

[56] Y. Yeshurun, M. Carrasco, L. T. Maloney: Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. Vision Res. **48** (2008) 1837–1851.

[57] J. A. Swets, E. F. Shipley, M. J. McKey, D. M. Green: Multiple observations of signals in noise. J. Acoust. Soc. Am. **31** (1959) 514–521.

[58] B. A. Dosher, Z. L. Lu: Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. Proc. Natl. Acad. Sci. U. S. A. **95** (1998) 13988–13993.

[59] W. P. Tanner, T. G. Birdsall: Definitions of $d'$ and $\eta$ as psychophysical measures. J. Acoust. Soc. Am. **30** (1958) 922–928.

[60] D. Oberfeld: An objective measure of auditory stream segregation based on molecular psychophysics. Atten. Percept. Psychophys. **76** (2014) 829–851.

[61] S. G. Kuai, D. Levi, Z. Kourtzi: Learning optimizes decision templates in the human visual cortex. Curr. Biol. **23** (2013) 1799–1804.

[62] J. Dobres, A. R. Seitz: Perceptual learning of oriented gratings as revealed by classification images. Journal of Vision **10** (2010).

[63] J. Gold, P. J. Bennett, A. B. Sekuler: Signal but not noise changes with perceptual learning. Nature **402** (1999) 176–178.

[64] M. H. Herzog, M. Fahle: The role of feedback in learning a Vernier discrimination task. Vision Res. **37** (1997) 2133–2141.

[65] L. P. Shiu, H. Pashler: Improvement in line orientation discrimination is retinally local but dependent on cognitive set. Percept. Psychophys. **52** (1992) 582–588.

[66] M. I. Posner: Orienting of attention. Q. J. Exp. Psychol. **32** (1980) 3–25.