Temporal weights in the perception of sound intensity: Effects of sound duration and number of temporal segments

Daniel Oberfeld,[1,a)] Jan Hots,[2] and Jesko L. Verhey[2]
[1]*Institute of Psychology, Section Experimental Psychology, Johannes Gutenberg-Universität Mainz, Wallstrasse 3, 55122 Mainz, Germany*
[2]*Department of Experimental Audiology, Otto von Guericke University Magdeburg, Leipziger Street 44, 39120 Magdeburg, Germany*

(Received 27 July 2017; revised 18 January 2018; accepted 24 January 2018; published online 14 February 2018)

Loudness is a fundamental aspect of auditory perception that is closely related to the physical level of the sound. However, it has been demonstrated that, in contrast to a sound level meter, human listeners do not weight all temporal segments of a sound equally. Instead, the beginning of a sound is more important for loudness estimation than later temporal portions. The present study investigates the mechanism underlying this primacy effect by varying the number of equal-duration temporal segments (5 and 20) and the total duration of the sound (1.0 to 10.0 s) in a factorial design. Pronounced primacy effects were observed for all 20-segment sounds. The temporal weights for the five-segment sounds are similar to those for the 20-segment sounds when the weights of the segments covering the same temporal range as a segment of the five-segment sounds are averaged. The primacy effect can be described by an exponential decay function with a time constant of about 200 ms. Thus, the temporal weight assigned to a specific temporal portion of a sound is determined by the time delay between sound onset and segment onset rather than by the number of segments or the total duration of the sound.
https://doi.org/10.1121/1.5023686

[VMR] Pages: 943–953

I. INTRODUCTION

The sensation *loudness* describes the perceived intensity of sound. It is most closely related to the physical intensity of the sound but is also affected by, for example, the sound's spectral or temporal properties (cf. Scharf, 1978). Among others, loudness is important when assessing environmental noise. While the basic research on loudness has been conducted with steady-state (static) sounds, in recent years there has been increasing interest in the loudness of non-steady (time-varying) sounds because most sounds in our environment vary across time. Several studies consistently showed that not all temporal portions of a sound are weighted equally. The beginning and, in some studies, also the end of an auditory stimulus is more important in loudness perception than the middle portion of the sound (e.g., Namba *et al.*, 1976; Ellermeier and Schrödl, 2000; Plank, 2005; Pedersen and Ellermeier, 2008; Dittrich and Oberfeld, 2009; Rennies and Verhey, 2009; Oberfeld and Plank, 2011; Ponsot *et al.*, 2016). These effects are referred to as the primacy and recency effect, respectively. The present study investigates how the two parameters sound duration and number of segments affect the weighting of the different portions of the stimulus for the overall loudness of the stimulus.

Previous studies measuring temporal weights in loudness judgments observed sizable and consistent primacy effects for sound durations between 250 ms and 1100 ms (e.g., Plank, 2005; Oberfeld and Plank, 2011). Some studies measuring temporal weights in loudness perception also reported a recency effect, i.e., that the most recent portion of the sound was more important than middle temporal portions in the estimate of the overall loudness. However, this was found in only a small subset of studies (Ellermeier and Schrödl, 2000; Pedersen and Ellermeier, 2008; Oberfeld and Plank, 2011; Ponsot *et al.*, 2016), and the recency effects were generally weaker than the primacy effect.

In order to understand the underlying mechanisms of the temporal weighting, it is essential to determine how the weights depend on the stimulus parameters. An important basic property of a sound is its duration, which is one of the two stimulus parameters considered in the present study. Previous studies measuring temporal loudness weights used sound durations below 1.2 s, except for two experiments that presented 2-s and 3-s sounds (Ponsot *et al.*, 2013; Ponsot *et al.*, 2016). The first objective of the present study was to answer the question whether there is an upper temporal boundary for the primacy effect, or more generally, to investigate whether the pattern of temporal weights depends on the sound duration. Based on findings for a frequency discrimination task (Turner and Berg, 2007), a recency effect is expected to become more pronounced with increasing sound duration, while the recency effect should be weak or absent for sound durations below 1.0 s (e.g., Pedersen and Ellermeier, 2008; Rennies and Verhey, 2009). In addition, it is expected that the primacy effect is attenuated at long sound durations (Ponsot *et al.*, 2016).

The other parameter considered here is related to the type of experiment that is typically used to measure temporal

[a)]Electronic mail: oberfeld@uni-mainz.de

weights. In these experiments, listeners are asked to judge the overall loudness of a sound that is divided into several temporal segments. On each trial, the intensities of the segments vary in a random fashion, creating a level-fluctuating sound (see Fig. 1). The two panels of Fig. 1 show stimuli with the same overall sound duration but different number of segments. The present study investigates how the perceptual weights assigned to the different temporal segments of the sound in the loudness-judgment task depend, for a given sound duration, on the number of segments contained in the sound. Put differently, the second question addressed in the present experiment is whether the total duration of the stimulus or the number of separable temporal elements is the critical variable determining the pattern of weights.

Primacy and recency effects are ubiquitous in experiments on short-term memory. In these cases, the serial position curve that plots the probability of correct recall or recognition as a function of the position of the item in the learned list depends on both the number of learned items (list length) and the total duration (i.e., the product of list length and presentation time per item) (e.g., Murdock, 1962; Brodie and Murdock, 1977). Dittrich and Oberfeld (2009) proposed that the primacy effect in temporal loudness weights might originate in a memory system where the sequence of segment levels is processed as serially sorted information. In this case, the weight assigned to a given segment should depend primarily on the serial position of the segment in the sequence of segments (e.g., first segment versus middle segment) rather than on the onset time of the segment on an absolute time scale (e.g., at sound onset versus 500 ms after sound onset).

To answer these two questions, in the present study the sound duration and the number of segments were varied in a factorial within-subjects design. The sound duration was 1.0, 2.5, 5.0, or 10.0 s. The stimuli consisted of contiguous broadband noise segments as in most previous experiments (e.g., Pedersen and Ellermeier, 2008; Rennies and Verhey, 2009; Oberfeld and Plank, 2011), and the sounds contained either 5 or 20 temporal segments with independent level variations.

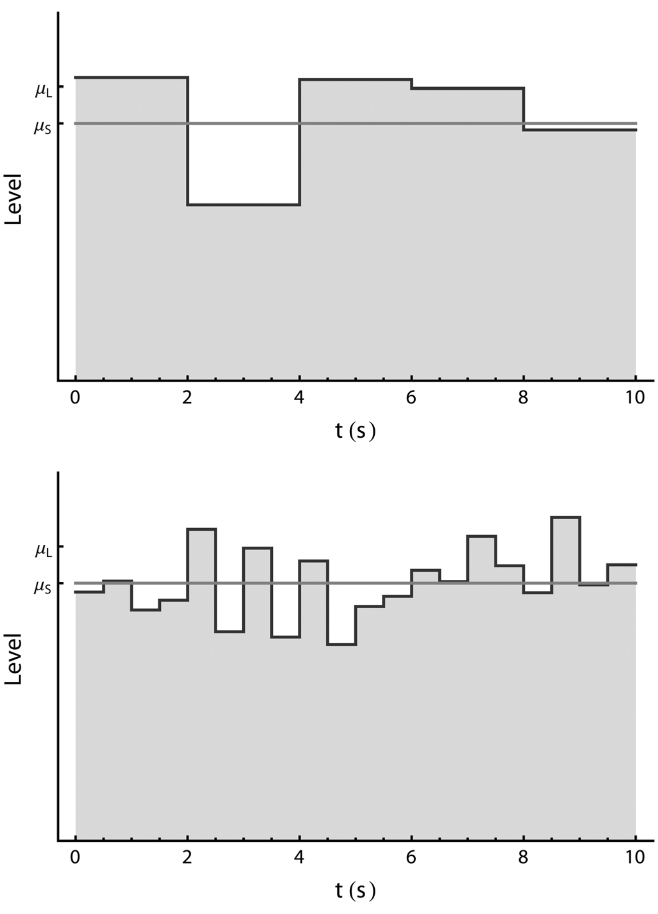

FIG. 1. Schematic plot of the time-varying stimuli. This example shows stimuli with an overall duration of 10 s. Upper panel: randomly varying sound pressure levels of a sound containing five contiguous broadband noise segments, as a function of time. Lower panel: sound with 20 temporal segments. In these examples, all segment levels are independently drawn from a normal distribution with mean μ_S and SD $\sigma = 2.5$ dB. The mean level is represented by the horizontal gray line. In the experiment, all segment levels within a given trial were drawn either from this "soft" distribution with mean μ_S, or from a "loud" distribution with a higher mean (μ_L).

II. METHOD

A. Listeners

Eight students (six female; age 20–33 yr) at the Johannes Gutenberg-Universität Mainz participated in the experiment. They either received partial course credit or were paid for their participation. The experiment was conducted according to the principles expressed in the Declaration of Helsinki. All listeners participated voluntarily after providing informed written consent, after the topic of the study and potential risks had been explained to them. They were uninformed about the experimental hypotheses. The Ethics Committee of the Institute of Psychology of the Johannes Gutenberg-Universität Mainz approved the study (reference number 2016-JGU-psychEK-002).

All participants reported normal hearing and no history of hearing disorders. Detection thresholds, measured by Békésy tracking with pulsed 270-ms tones including 10-ms $\cos^2$ on- and off-ramps, were better than 15 dB hearing level (HL) bilaterally between 125 Hz and 8 kHz, except for one participant who showed thresholds between 15 and 22 dB HL in the 3–8 kHz frequency range in the left ear.

B. Stimuli and apparatus

For the estimation of the temporal loudness weights for level-fluctuating stimuli, the established experimental paradigm from previous experiments was adopted (e.g., Pedersen and Ellermeier, 2008; Oberfeld and Plank, 2011). This paradigm uses methods of *behavioral reverse correlation*, also termed perceptual weight analysis (Ahumada and Lovell, 1971; Berg, 1989). On each trial, a level-fluctuating sound was presented and the task was to decide whether it was soft or loud compared to previous trials in a given block. The stimuli were Gaussian broadband noises (20–20 000 Hz) consisting of either 5 or 20 contiguous temporal segments with identical duration (0.1 ms on-and off ramps). The overall duration of the stimuli was 1.0, 2.5, 5.0, or 10.0 s. Figure 1 shows a schematic plot of a 10-s sound with either 5 or 20 segments. Random level perturbations were imposed on the temporal segments, resulting in a level-fluctuating noise that changed in intensity after each temporal segment. On each trial, the sound pressure levels of the 5 or 20 temporal segments were drawn independently from a normal distribution with a mean of either $\mu_S = 55.25$ dB sound pressure level (SPL) ("soft distribution") or $\mu_L = 56.75$ dB SPL ("loud distribution"). Thus, the difference in mean level between the "loud" and "soft" distribution was 1.5 dB. For both distributions, the standard deviation (SD) was $\sigma = 2.5$ dB. To avoid very loud or soft segments, the range of segment levels was restricted to $\mu \pm 3$ SD. Pilot data indicated that this combination of difference in mean level and SD corresponded to sensitivity in the order of $d' = 0.8$. However, for some listeners, the d' measured in the first few sessions was lower than 0.5 in some conditions with the 1.5 dB level difference. For these listeners, the difference in mean level was increased to 2.0 dB ($\mu_S = 55.0$ dB SPL, $\mu_L = 57.0$ dB SPL) in the remaining sessions. Note that the difference in mean level does not affect the estimation of perceptual weights using logistic regression (see Sec. II D), unless the difference is so large that it results in so-called complete separation (Heinze, 2006).

The "louder" and "softer" types of stimuli were presented with equal *a priori* probability. The listeners' task was to decide whether the level-fluctuating noise was soft or loud compared to previous trials in a given block. Thus, a one-interval, two-alternative forced-choice (1I, 2AFC) *absolute identification task* (Braida and Durlach, 1972) with a virtual standard (e.g., Nachmias, 2006) was used. One could also describe it as a *sample discrimination task* (Berg and Robinson, 1987; Sorkin *et al*., 1987; Lutfi, 1989) where the listeners decided whether the segment levels had been drawn from the "loud distribution" or from the "soft distribution".

After each block (containing between 60 and 200 trials; see below) the number of correct responses and the number of incorrect responses were displayed on the screen. Following the usual procedure for sample discrimination tasks, a response was scored as correct if, for example, the segment levels were drawn from the "loud distribution" (with mean μ_L) and the listener responded that the sound was "loud." Note that this can, in rare cases, result in counterintuitive feedback, 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. The listeners were informed about this issue, and trial-by-trial feedback was provided only in the practice blocks. In addition, to reduce biases towards one response category (e.g., a tendency to respond that the sound was "loud"), the number of loud and soft trials and the number of "loud" and "soft" responses were displayed at the end of each block. Thus, the listeners had the opportunity to realize that they selected one of the response options too frequently. In some practice trials, trial-by-trial feedback concerning the correctness of the response was provided to facilitate the understanding of the task. The next trial never started before the response to the preceding trial had been given. The minimum inter-trial interval was 1000 ms.

The stimuli were generated digitally, played back via two channels of an RME ADI/S digital-to-analog converter (sampling frequency 44.1 kHz, 24-bit resolution), attenuated by a TDT PA5 programmable attenuator, buffered by a TDT HB7 headphone buffer, and presented diotically via Sennheiser HDA 200 circumaural headphones calibrated according to IEC 318 (1970). Instructions were presented on a computer monitor. The experiment was conducted in a double-walled sound-insulated chamber.

C. Procedure

In a within-subjects design, each listener received all of the eight factorial combinations of sound duration (1.0, 2.5, 5.0, and 10.0 s) and number of temporal segments (5 or 20). According to our previous experience with this experimental paradigm (e.g., Oberfeld *et al*., 2012; Oberfeld, 2015; Ponsot *et al*., 2016), about 700 trials are necessary per listener and condition to obtain reliable weight estimates for ten temporal segments. For this reason, 420 trials were collected per listener in each condition with five temporal segments, and 1400 trials were collected in each condition with 20 temporal segments. Due to a technical problem, for two of the eight listeners, only 360 trials were collected in the conditions with five segments. These data were included in the data analysis.

Each listener participated in a total of eight experimental sessions. In the first session, audiometric thresholds were measured and practice blocks for all experimental conditions were presented. In the remaining sessions (2–8), one block was presented for each of the eight (sound duration × number of segments) experimental conditions. Only one experimental condition was presented per block. Each block contained 60 and 200 trials for the 5-segment and 20-segment sounds, respectively. The order of conditions was randomized. Each block started with five trials on which trial-by-trial feedback was provided in order to facilitate the adoption of a decision criterion for the new experimental condition. These trials were excluded from the data analysis. The duration of each of the sessions 2–8 was approximately 130 min, including a mandatory pause of at least 15 min.

D. Data analysis

The perceptual weights representing the importance of the 5 or 20 temporal segments for the decision in the sample discrimination task were estimated from the trial-by-trial data via multiple logistic regression (e.g., Gilkey and Robinson, 1986; Alexander and Lutfi, 2004; Oberfeld, 2008; Pedersen and Ellermeier, 2008). The decision model assumed that the listener compares a weighted sum of the segment levels to a fixed decision criterion, and responds that the sound was of the "loud" type if the weighted sum exceeds the criterion (a detailed description of the decision model is provided by Oberfeld and Plank, 2011). If the weighted sum is smaller than the criterion, then it is assumed that the listener classifies the sound as "soft." In the data analysis, the binary responses ("loud" or "soft") served as the dependent variable. The predictors (i.e., the 5 or 20 segment levels) were entered simultaneously. The regression coefficients were taken as the decision weight estimates. For a given level of a segment, a regression coefficient equal to zero means that the segment 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 sound was of the "loud" type increased with the sound pressure level of the segment.

A separate logistic regression model was fitted for each combination of listener, sound duration, and number of segments. Since the *relative* contributions of the different segments to the decision were of interest rather than the absolute magnitude of the regression coefficients, the 5 or 20 regression coefficients were normalized for each fitted model such that the mean of their absolute values was 1.0. This resulted in a set of relative perceptual weights for each combination of listener, sound duration, and number of segments. Note that we normalized to a mean of 1.0, rather than to a sum of 1.0 as some previous studies, in order to facilitate the comparison of the pattern of weights across the two numbers of segments.

A summary measure of the predictive power of a logistic regression model is the area under the Receiver Operating Characteristic (ROC) curve (Swets, 1986; Agresti, 2002). This measure provides information about the degree to which the predicted probabilities are concordant with the observed outcome (for details see Dittrich and Oberfeld, 2009). Areas of 0.5 and 1.0 correspond to chance performance and perfect performance of the model, respectively. Across the 64 fitted logistic regression models, the area under the ROC curve ranged between 0.63 and 0.86 ($M = 0.77$, $\mathrm{SD} = 0.061$), indicating on average reasonably good predictive power (Hosmer and Lemeshow, 2000).

The individual normalized temporal weights estimated by the multiple logistic regressions were analyzed with repeated-measures analyses of variance (rmANOVAs) using a univariate approach with Huynh-Feldt correction for the degrees of freedom (Huynh and Feldt, 1976). The correction factor $\tilde{\varepsilon}$ is reported, and partial η^2 is reported as measure of association strength. An α-level of 0.05 was used for all analyses.

III. RESULTS

Figure 2 shows the mean normalized temporal weights for sound durations of 1.0 s (upper left panel), 2.5 s (upper right panel), 5 s (lower left panel), and 10 s (lower right panel), and for stimuli divided into five temporal segments (red squares) and 20 segments (blue circles). To facilitate the comparison between the different sound durations, time units were normalized in this plot, i.e., the segment onset is expressed relative to the sound duration (i.e., segment onset divided by sound duration). Thus, 0.0 represents the sound onset, and 1.0 represents the sound offset.

For the five-segment sounds (red squares in Fig. 2), the weights showed a (relatively weak) primacy effect (higher weight on first segment than on the middle segments), as it was expected. At the longer sound durations, the weight on the final segment was also higher than for the middle segments (recency effect).

An rmANOVA with the within-subjects factors sound duration (1.0, 2.5, 5.0, and 10 s) and segment number (1–5)

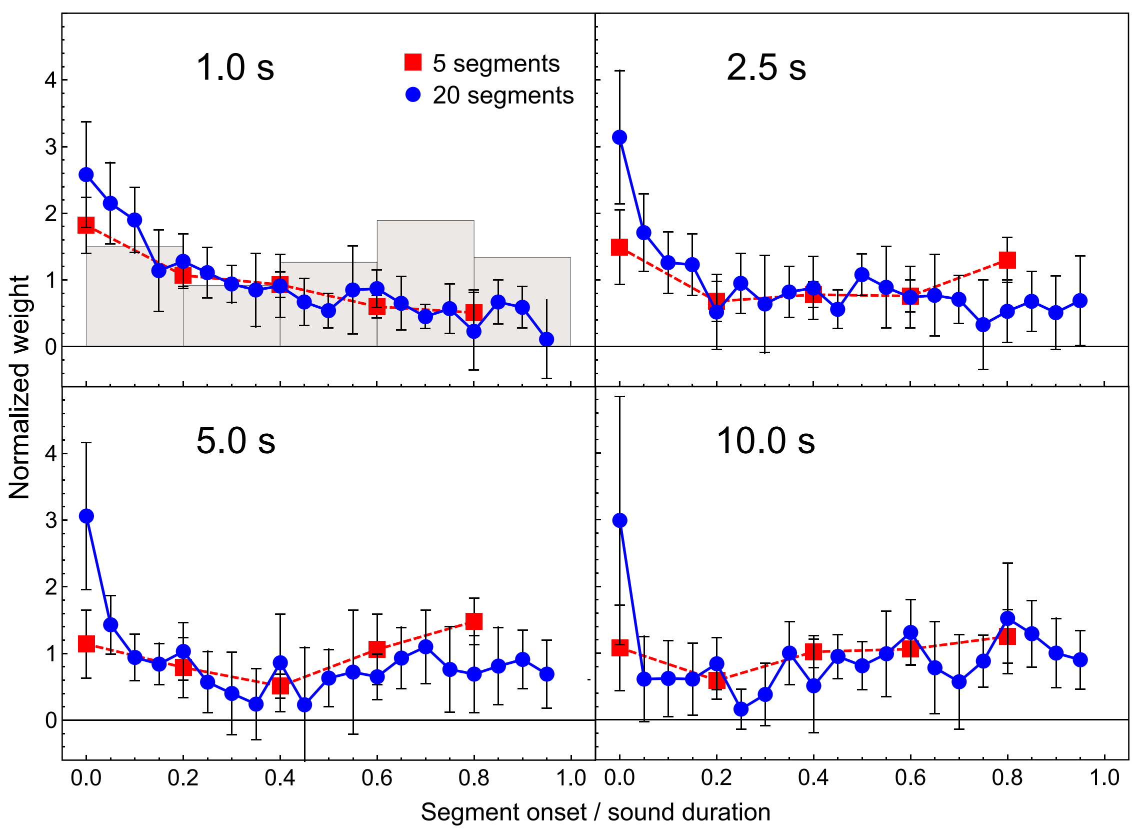

FIG. 2. (Color online) Mean normalized weights as a function of segment onset relative to the sound duration (0: sound onset, 1: sound offset), for different sound durations (panels) and number of segments (5 segments: red squares; 20 segments: blue circles). The gray bars in the upper left panel represent example temporal segments of a five-segment sound. Error bars show 95% confidence intervals (CIs).

showed a significant effect of segment number, $F(4, 28) = 5.45$, $\tilde{\varepsilon} = 0.451$, $p = 0.022$, $\eta_p^2 = 0.44$. Separate *post hoc* rmANOVAs showed a significant effect of segment number at all sound durations except for 10 s. The segment number × sound duration interaction was significant, $F(12, 84) = 4.78$,. $\tilde{\varepsilon} = 0.552$, $p = 0.001$, $\eta_p^2 = 0.41$. While at a sound duration of 1 s the weights decreased monotonically with the segment number (showing only a primacy effect), a u-shaped pattern was observed at the longer durations (showing both a primacy and a recency effect).

For the 20-segment sounds (blue circles in Fig. 2), the weights showed a strong primacy effect at each sound duration. No clear recency effects were observed. An rmANOVA with the within-subjects factors sound duration and segment number (1–20) showed a significant effect of segment number, $F(19, 133) = 10.66$, $\tilde{\varepsilon} = 0.183$, $p < 0.001$, $\eta_p^2 = 0.60$. Separate *post hoc* rmANOVAs indicated a significant effect of segment number for each of the sound durations (all p-values < 0.006). Inspection of the individual weighting patterns revealed that all listeners showed a primacy effect at sound durations of 1.0, 2.5, and 5.0 s. At the 10 s duration, three of the eight listeners showed no clear primacy effect. Only one listener showed an additional recency effect at the 1.0-s duration, another listener showed a recency effect at the 2.5-s duration, and yet another listener showed a recency effect at the 5.0-s duration. At the 10-s duration, two of the three listeners who did not show a primacy effect produced a recency effect in the sense of an increase in the weights from the beginning to the end of the sound. However, the variability of the 20 weights was relatively high for these three listeners at this sound duration. The segment number × sound duration interaction was significant, $F(57, 399) = 2.52$, $\tilde{\varepsilon} = 0.446$, $p < 0.001$, $\eta_p^2 = 0.27$. Whereas at a sound duration of 1.0 s the weights decreased gradually across the sound duration, at the longer durations only the weights assigned to the first few segments were higher than the weights on the following segments.

IV. DISCUSSION

A. Do the weights for five-segment sounds represent average temporal weights?

As shown in Fig. 2, the weighting profiles for the five-segment sounds were flatter than for the 20-segment sounds. A simple explanation for this pattern could be that the weight assigned to, e.g., the first segment of a five-segment sound is equal to the average weight assigned to the first four segments of a 20-segment sound with identical total duration. In other words, the weight assigned to a given segment in a five-segment sound might represent the average weight assigned to the temporal sub-parts of this segment. To test this hypothesis, the average weights for the five groups of four consecutive segments in a 20-segment sound were computed, for each listener and each sound duration. This resulted in one average weight for each of the five time windows corresponding to the five segments in a five-segment sound.

Figure 3 shows the four-segment averaged weights for the 20-segment sounds (gray circles) displayed together with the weights for the five-segment sounds (red squares; replotted from Fig. 2), as a function of relative segment

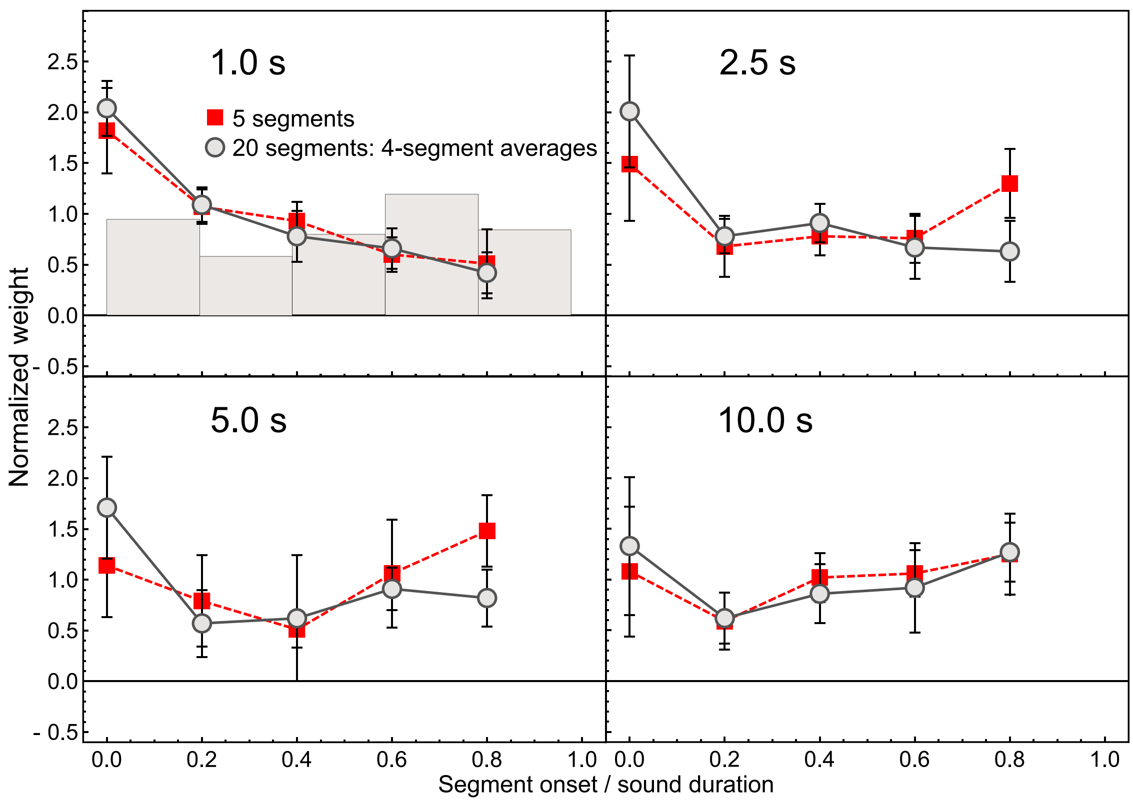

FIG. 3. (Color online) Mean normalized weights as a function of segment onset relative to sound duration (0: sound onset, 1: sound offset), sound duration (panels), and sound type. Red squares: five-segment sounds. Light gray circles: four-segment averages for 20-segment sounds. The gray bars in the upper left panel represent the temporal segments of a five-segment sound. Error bars show 95% CIs.

onset. The four-segment averaged weights were relatively similar to the weights for the five-segments stimuli, especially at sound durations of 1.0 and 10.0 s, supporting our hypothesis on the relation between the weights for five-segment and 20-segment sounds. An rmANOVA with the within subject factors sound type (five segments, four-segment average for 20-segment sounds), sound duration, and segment number (1–5) showed a significant effect of segment number, $F(4,\ 28)=9.47$, $\tilde{\varepsilon}=0.335$, $p=0.009$, $\eta_p^2=0.58$. As shown in Fig. 3, the temporal weights were not uniform but depended on the segment onset. There was a significant interaction of sound duration and segment number, $F(12,\ 84)=9.29$, $\tilde{\varepsilon}=0.417$, $p<0.001$, $\eta_p^2=0.57$. The pattern of temporal weights depended on sound duration. The sound type $\times$ segment number interaction was also significant, $F(4,\ 28)\ =\ 6.85$, $\tilde{\varepsilon}=0.774$, $p=0.002$, $\eta_p^2=0.50$. Thus, across the four sound durations, the pattern of weights for the four-segment averages differed significantly from the weights for the five-segment sounds. Separate *post hoc* rmANOVAs conducted at each sound duration showed that the sound type $\times$ segment number interaction was not significant at sound durations of 1 s and 10 s (p-values >0.448), where Fig. 3 shows that the four-segment averages were close to the weights for the five-segment sounds. At sound durations of 2.5 and 5 s, the four-segment averages showed a stronger primacy and a weaker recency effect than the weights for the five-segment sounds, and the sound type $\times$ segment number interaction was significant, $F(4,\ 28)\ =\ 6.47$, $\tilde{\varepsilon}=0.825$, $p=0.002$, $\eta_p^2=0.48$, and not significant, $F(4,\ 28)\ =\ 2.93$, $\tilde{\varepsilon}=0.812$, $p=0.052$, $\eta_p^2=0.30$, respectively. Taken together, our hypothesis that the temporal weight assigned to each segment in a five-segment sound is identical to the average weight assigned to the temporal sub-parts of this segment received only partial support. It is interesting to note that the four-segment averages and the weights for the five-segment sounds in Fig. 3 look more similar than the weights for five-segment and 20-segment sounds in Fig. 2. One reason for this is that, somewhat arbitrarily, we plotted the weights for the five-segment sounds as a function of segment onset in Fig. 2. This choice ensured that all stimuli had the same onset time (0 ms). An alternative way (where the onset time depends on the segment duration) is to plot the weights as a function of the segment midpoint, which is compatible with the idea of weight averaging across the segment duration.[1] This would shift the red symbols in Fig. 2 to the right, resulting in better agreement with the 20-segment weights.

B. Time course of the primacy effect

Figure 4 shows the average temporal weights for the 20-segment sounds plotted on an absolute time scale (milliseconds), rather than on a relative time scale as in Fig. 2. This demonstrates that the primacy effects were very similar across the four sound durations. To quantify the magnitude and time course of the primacy effect, exponential decay functions were fitted to the mean weights at each of the four sound durations. The weight assigned at the time t was assumed to be

$$w(t) = D \cdot e^{-t/\tau} + c = c(D/c \cdot e^{-t/\tau} + 1)$$
$$= c(D_r \cdot e^{-t/\tau} + 1), \tag{1}$$

where $t=0$ corresponds to the sound onset, c is the asymptotic weight at $t \to \infty$, D_r is the weight at sound onset ($t=0$) relative to the asymptotic weight $w(\infty) = c$ (i.e., D_r is the "dynamic range" of the weights), and the time constant τ quantifies the time needed for the weight to decay to a value of $1/e$ of the weight range between $w(0)$ and the asymptotic weight c. Thus, the function $w(t)$ can be used to quantify the primacy effect in terms of (a) the relative change in the weight between the onset of the signal and the "steady state," asymptotic value of the weights, $(w(0) - w(\infty))/w(\infty) = D_r$ (magnitude/"dynamic range"), and (b) in terms of the rate of decay, τ (time constant). The weight assigned to a temporal segment with onset at t_{on} and duration d was assumed to be the integral of $w(t)$ across the segment duration,

$$\bar{w}(t_{on}, d) = \int_{t=t_{on}}^{t_{on}+d} w(t)dt. \tag{2}$$

The function $\bar{w}(t_{on}, d)$ was fitted to the mean weights at each sound duration, using the Mathematica function *NonlinearModelFit*, with the weight for a given data point w_{i} proportional to $1/\text{SD}^2_{w_i}$, where $\text{SD}^2_{w_i}$ is the variance of the eight individual estimated weights for segment i. Since the time course of the primacy effect was of interest, the function was fitted only to the first ten weights at the 10-s sound duration. Here, the remaining mean weights showed a weak recency effect.

Figure 5 shows the fits of the exponential decay function $\bar{w}(t_{on}, d)$ (red lines) and 95% confidence bands (red shaded area) together with the mean normalized weights of the

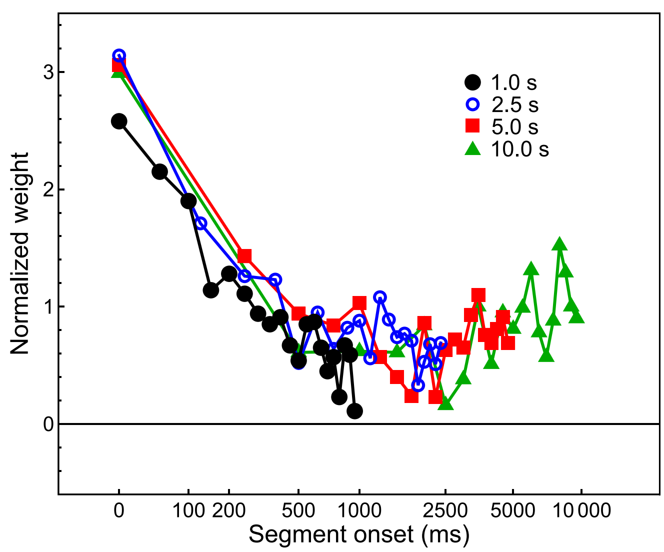

FIG. 4. (Color online) Mean normalized weights for the 20-segment sounds as a function of segment onset (absolute time scale in milliseconds, plotted on a log axis) and sound duration. Filled circles, open circles, squares, and triangles represent sound durations of 1.0, 2.5, 5.0, and 10.0 s, respectively.

20-segment sounds (black squares, replotted from Fig. 2), as a function of segment onset for the different sound durations. The estimated parameters of the decay function are shown in Table I. The decay function provided an excellent fit to the mean weights ($R^2 \geq 0.96$) except at the 10-s duration, where the variance accounted for by the decay function was lower than desirable. In the latter condition, a very short time constant was estimated. This can be attributed to a problem with too broadly spaced data points. In a 20-segment sound with 10 s duration, the duration of the first segment is 500 ms (see Fig. 1). The mean weights at the shorter durations show that the weights have already dropped to their asymptotic value after 500 ms. Thus, when fitting the function at the 10-s duration, it is not possible to obtain an exact estimate of the time constant, because there are no data points between 0 and 500 ms. At the three shorter sound durations, the estimated time constants were very similar, with an average estimated value of $\tau = 198.0$ ms (SD = 20.8 ms). The estimated values of D_r were also relatively similar across the three shorter sound durations ($M = 4.83$, SD = 0.50). Thus, the "dynamic range" of the primacy effect does not show a strong dependence on sound duration, and on average the onset of the sound receives a 4.83 times higher weight than temporal portions located 500 ms or more after sound onset.

The dashed blue lines in Fig. 5 show the average decay function with $\tau = 198.0$ ms and $D_r = 4.83$. This average function was fitted to the mean weights for each of the sound durations using only the multiplicative constant c as a free parameter. As seen in Table I, the goodness of fit (R^2_{avg}) of this mean decay function was only minimally lower than for the specific decay function fitted at a given sound duration. Thus, the average decay function provides a reasonably good description of the primacy effect at sound durations between 1.0 and 10 s. The implication is that the temporal position of the segment, measured on an absolute time scale, rather than the segment onset relative to the sound duration (put differently, the segment number) determines the weight assigned to a given segment.

This stands in contrast to serial position curves in memory experiments, where the primacy effects encompass approximately the first 3–5 items in the list, are relatively independent of the presentation rate, and are also observed when each item is presented for several seconds. For instance, Anderson and Burns (1973) presented auditory lists of 12 random digits with a presentation rate of one item per second, three items per second, or four items per second (i.e., item durations of 1.0, 0.33, or 0.25 s). After each list, the experimenter visually presented the list with one item

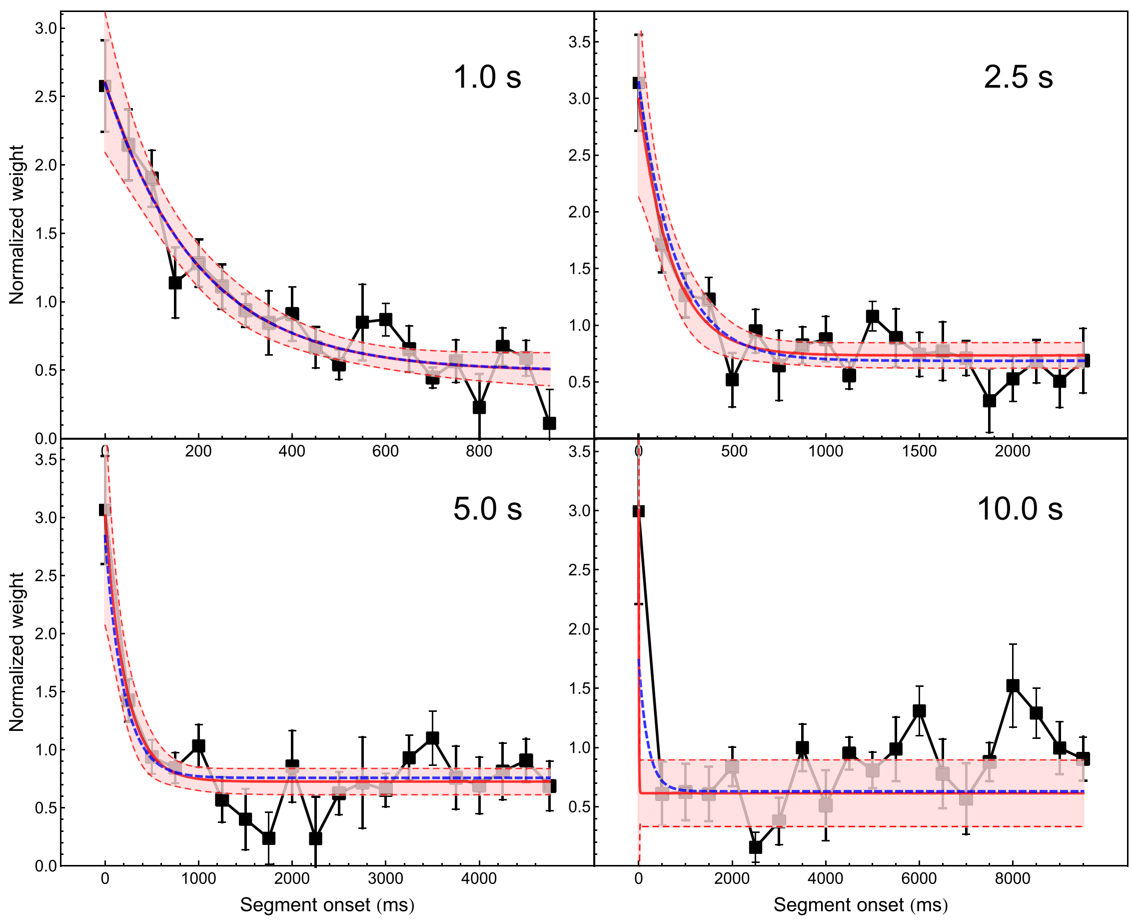

FIG. 5. (Color online) Mean normalized weights (black squares) for the 20-segment sounds, as a function of segment onset (absolute time scale in milliseconds) and sound duration (panels). Error bars show ±1 standard error of the mean (SEM). The solid red line in each panel indicates the exponential decay function $\bar{w}(t_{on}, d)$ that was fitted to the data for this sound duration. Note that for the 10-s duration, only the first ten segments (onsets between 0 and 4500 ms) were used for fitting the decay function. The red shaded areas represent 95% confidence bands for the mean. The dashed blue line shows the mean decay function [Eq. (2)], using the average values of $\tau = 198.0$ ms and $D_r = 4.83$.

TABLE I. Fits of the exponential decay function $\bar{w}(t_{on}, d)$ [Eq. (2)] to the observed primacy effects (20-segment sounds). For each of the sound durations, the estimated parameters τ, D_r, and c are displayed, together with the standard error of the estimate, the p-value for a test of the estimated parameter against 0, the 95% CI, R^2 for the fitted function, and R^2_{avg} for the decay function with average parameters (see text).

Sound duration	Parameter	Estimate	SE	p	95% CI lower	95% CI upper	R^2	R^2_{avg}
1.0 s	τ	198.31	36.56	0.000	121.17	275.45	0.97	0.97
	D_r	4.89	0.79	0.000	3.21	6.56		
	c	0.49	0.07	0.000	0.34	0.63		
2.5 s	τ	177.02	42.30	0.001	87.78	266.27	0.96	0.95
	D_r	4.30	1.02	0.001	2.16	6.45		
	c	0.73	0.05	0.000	0.62	0.85		
5.0 s	τ	218.72	49.64	0.000	113.99	323.45	0.96	0.95
	D_r	5.29	1.42	0.002	2.31	8.28		
	c	0.73	0.05	0.000	0.61	0.84		
10.0 s	τ	6.35	4.30	0.184	−3.83	16.52	0.81	0.79
	D_r	307.07	0.09	0.000	306.86	307.28		
	c	0.61	0.12	0.001	0.33	0.90		

missing, and the task was to write down the missing item. The serial position curves showed a (relatively weak) primacy effect and were virtually identical across the three presentation rates when proportion correct was plotted as a function of missing-item number (serial position). In contrast, the weights for the 20-segment sounds of the present study, when plotted as a function of segment number (i.e., relative segment onset; Fig. 2), differ considerably between presentation rates of 20/s, 8/s, 4/s, and 2/s. Murdock (1962) plotted serial position curves for auditory lists of 20 items as a function of item number and even found a slower decay of the primacy effect when the duration of each item was 2 s rather than 1 s. The data of the present study show the opposite pattern: when plotted as a function of segment number, the weight for the 20-segment sounds decayed faster when the segment duration was long (see Fig. 2). Thus, the primacy effect in term short-term memory depends mainly on the serial position (i.e., item number) within the list, while the primacy effect in loudness depends on time (i.e., the delay between the sound onset and the segment onset). In addition, the primacy effect in memory does not decay within less than 1 s as it does for the temporal loudness weights. Apart from these differences, serial position curves in memory often additionally show a clear recency effect, or even only a recency but no primacy effect (e.g., Murdock, 1962; Waugh and Norman, 1965; Norman, 1966; Wickelgren, 1970; Anderson and Burns, 1973; McElree and Dosher, 1989). In conclusion, the primacy effect in loudness appears to be very different from serial position effects in memory.

The solid black line in Fig. 6 shows the weight function $w(t)$ from Eq. (1) representing the primacy effect, with the average estimated parameters $\tau = 198.0$ ms and $D_r = 4.83$, and c set to 1.0. According to this function, the beginning of the stimulus receives the highest weight, and the primacy effect decays within about 500 ms. According to our model function [Eq. (2)], the weight assigned to a given temporal segment is the integral of $w(t)$ from segment onset to segment offset. Note that a resulting prediction is that for a given sound duration, the segment weights will show a weaker primacy effect for longer segment durations (e.g., when the sound is divided into five rather than 20 equal-duration segments), compatible with Fig. 2 and roughly compatible with the analysis visualized in Fig. 3. The normalized segment weights $\bar{w}(t_{on}, d)/d$ for segment durations of 20, 200, and 2000 ms are shown by the long-dashed, short-dashed, and dashed-dotted line in Fig. 6, respectively. For each of these functions, the value on the y-axis is the predicted weight for a segment with onset at time t_{on} and segment duration d. Note that as d approaches 0, the predicted normalized segment weight approaches $w(t)$.

C. Time course of primacy effects in previous studies

Eight previous studies reported temporal weights for sounds with a duration of approximately 1 s (Ellermeier and Schrödl, 2000; Plank, 2005; Pedersen and Ellermeier, 2008; Dittrich and Oberfeld, 2009; Rennies and Verhey, 2009; Oberfeld and Plank, 2011; Oberfeld *et al.*, 2012; Oberfeld, 2015). Two of these studies reported data from more than

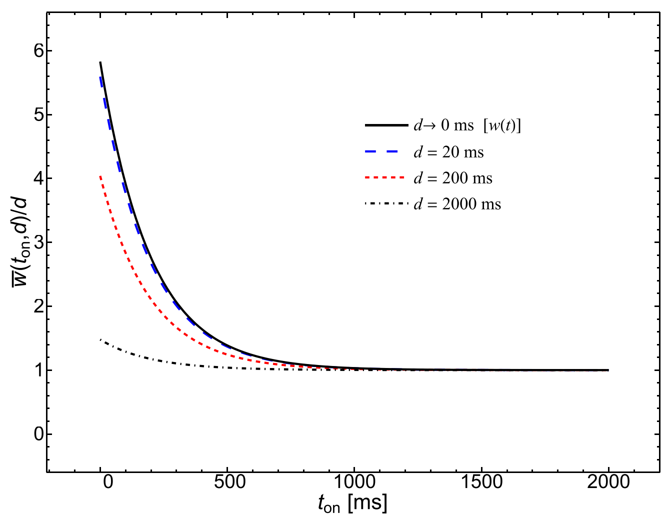

FIG. 6. (Color online) Primacy effect. Long-dashed blue line: predicted normalized segment weights $\bar{w}(t_{on}, d)/d$ for a segment duration of $d = 20$ ms, using the average estimated decay-function parameters $\tau = 198.0$ ms and $D_r = 4.83$, and c set to 1.0. Short-dashed red line: $d = 200$ ms. Dot-dashed black line: $d = 2000$ ms. Solid black line: decay function $w(t)$ [see Eq. (1)] with the average estimated parameters.

one group of subjects (Pedersen and Ellermeier, 2008; Oberfeld and Plank, 2011). The average temporal weights for each of the 11 experiments (i.e., independent groups of subjects) are shown as a function of segment onset in Fig. 7, representing a total of 88 subjects. We compared the primacy effect found in the present experiment to these previous data by fitting the exponential decay function [Eq. (2)]. To account for the different numbers of subjects per experiment and the different within-experiment variability of the temporal weights, each data point was weighted by $1/SE^2$, where SE is the standard error of the weight estimate. As for the present data, a few temporal weights at the end of the stimulus were excluded from the analysis for three experiments that showed a weak recency effect (Pedersen and Ellermeier, 2008, condition without feedback; Oberfeld and Plank, 2011, experiments 3 and 4). The fitted exponential decay function is shown by the gray line in Fig. 7. The gray shaded area shows the 95% confidence band. The estimated parameters were $\tau = 272$ ms and $D_r = 4.2$, $R^2 = 0.90$. Thus, in the previous data for sound durations of about 1 s, the primacy effect was slightly weaker and the rate of decay slightly slower than in the data from the present experiment. Still, the average decay function from the present study (blue line in Fig. 7) is similar to the fitted decay function for the 11 previous experiments.

Only one previous study (Ponsot *et al.*, 2016) measured temporal loudness weights for sound durations considerably longer than 1 s. This study found a relatively weak primacy effect when the global loudness was judged in a magnitude estimation task, and an even smaller primacy effect when the global loudness was judged in a one-interval absolute identification task similar to the task used in the present experiment. We currently have no explanation for these diverging results. It would be desirable to collect more data for sound durations longer than 1 s to gain further insights regarding this effect.

It should be noted that the results of the present study are also compatible with the previous data in showing no recency effect in the average temporal weights in most conditions. The only exceptions are a relatively weak recency effect for 20-segment sounds at a sound duration of 10 s, and weak recency effects for some of the five-segment sounds (see Fig. 2). Also, as in a previous study (Oberfeld and Plank, 2011), inspection of the individual data suggested stronger individual differences for the recency than for the primacy effect. While all listeners showed a clear primacy effect for the 20-segment sounds at sound durations of 1.0, 2.5, and 5.0 s, only three listeners showed a recency effect at these sound durations. Even at the 10-s duration, only three of the eight listeners showed a recency effect. In conclusion, it remains unclear under which conditions temporal loudness weights show a recency effect.

D. Primacy effect in the light of current loudness models

Several models have been proposed to predict loudness perception (e.g., Zwicker, 1977; Chalupper and Fastl, 2002; Glasberg and Moore, 2002; Glasberg and Moore, 2006). Some of them are included in current standards on loudness. The models can be divided into stationary models that analyze the long-term spectrum of a sound and dynamic models that are sensitive to the temporal properties of the sound. Since stationary models use the long-term spectrum as a basis for the loudness estimate, these models consider each portion of the sound in the same way, i.e., they assign uniform temporal weights. In contrast, dynamic models could in theory include a different weighting of the different temporal portions of a sound since they process the time signal of the sound (Zwicker, 1977; Chalupper and Fastl, 2002; Glasberg and Moore, 2002). The main aims of the dynamic models were to simulate (a) temporal integration of loudness, i.e., that a short sound is softer than an equal-intensity long sound with the same spectral characteristics (cf. Zwislocki, 1969; Hots *et al.*, 2014), and (b) the effect of amplitude modulation on the loudness, i.e., that a modulated sound is usually louder than an equal-intensity unmodulated sound, at least if the modulation frequency is only a few Hz (e.g., Glasberg and Moore, 2002; Grimm *et al.*, 2002). Both aspects are accounted for by temporal integration stages such as a low-pass filter, combined with a decision stage. This latter stage usually computes the maximum, a percentile, or sometimes also the average of the (smoothed) loudness-time function as an estimate of the overall loudness of the sound. Due to this

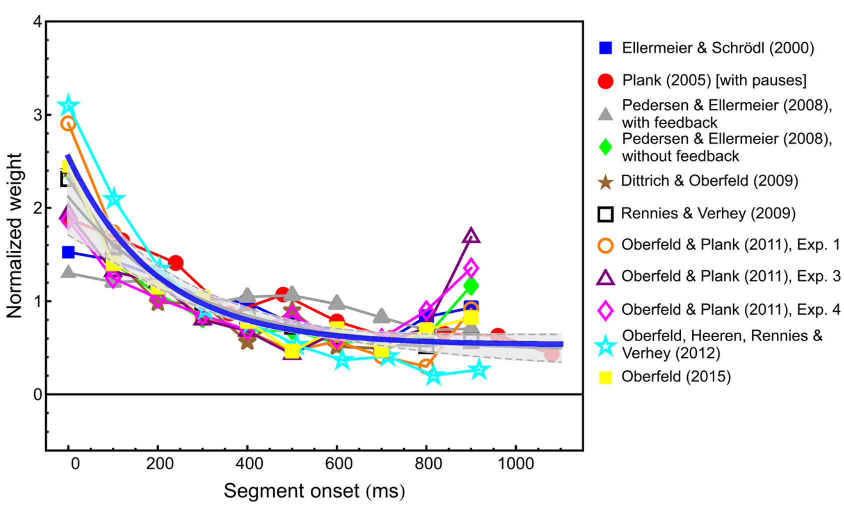

FIG. 7. (Color online) Average normalized weights from previous experiments as a function of segment onset for sound durations between 900 and 1100 ms. Each line represents one experiment (i.e., an independent group of subjects). In total, the data represent temporal weights for 88 subjects. The gray line shows the best-fitting exponential decay function $\bar{w}(t_{on}, d)$ [Eq. (2); $\tau = 272$ ms, $D_r = 4.2$, $R^2 = 0.90$]. The shaded gray area is the 95% confidence band. The blue line shows the average decay function from the present experiment ($\tau = 198.0$ ms, $D_r = 4.83$).

model structure, the temporal position of loud portions of the stimulus should hardly affect the predicted loudness. However, the temporal integration stages in Glasberg and Moore (2002) include a faster attack time constant than release time constant. This could yield to a slight emphasis of the first segment of a signal. Chalupper and Fastl (2002) used the same release and attack time constant in the integration stage. However, this model also has a slight asymmetry due to a simulated forward masking effect included in the model (Heeren *et al.*, 2011). Simulations by Pedersen (2006) and our own unpublished simulations indicate that these dynamic models hardly predict the primacy effect observed in the studies mentioned above.

Rennies and Verhey (2009) proposed a model that assumes a higher spectral loudness summation at the beginning of the sound. This was motivated by experiments that showed a larger spectral loudness summation for short noise bursts than for long signals. This model predicts a higher loudness at the beginning of the sound than at later portions in time. However, the time constants for this process are too short to account for the primacy effects observed here. The model can also not account for primacy effects observed with pure-tone stimuli (Oberfeld *et al.*, 2013; Ponsot *et al.*, 2013), where spectral summation plays no role.

Thus, current models seem to be unable to predict the effect, suggesting that an additional processing stage is required to account for the primacy effect.

V. CONCLUSIONS

The present study measured temporal weights in a task where listeners judged the global loudness of a level-fluctuating broadband noise. The sound duration was varied across a much larger range than in previous studies (1.0–10.0 s), and the number of temporal segments with independent level variations (5 or 20 segments) was varied independently of the sound duration. The data showed pronounced primacy effects at all sound durations, compatible with previous studies that predominantly presented sound durations of about 1 s. The time course of the primacy effect was well described by an exponential decay function with a time constant of about 200 ms. The temporal weighting curve was flatter when the sound consisted of only five rather than 20 temporal segments, compatible with the idea that the weight assigned to a temporal segment is the integral of this exponential decay function from segment onset to segment offset. Plotted on an absolute time scale, the primacy effects at the four sound durations were very similar and were well accounted for by the same exponential decay function with an averaged parameter set. Thus, the temporal weight assigned to a specific temporal portion of a sound is determined by the time delay between sound onset and segment onset (e.g., 500 ms), rather than by the relative temporal position within the sound (e.g., the middle segment).

ACKNOWLEDGMENTS

This study was supported by a grant from Deutsche Forschungsgemeinschaft (www.dfg.de) to D.O. (OB346/6-1) and J.L.V. (VE373/2-1). We are grateful to Annika Grotjohann and Marius Frenken for their assistance in data collection, figure preparation, and contributions to the analysis of the results from previous studies. We thank Walt Jesteadt and an anonymous reviewer for thoughtful comments on an earlier version of this article.

[1]We are grateful to Walt Jesteadt for this suggestion.

Agresti, A. (**2002**). *Categorical Data Analysis* (Wiley, New York).

Ahumada, A. J., and Lovell, J. (**1971**). "Stimulus features in signal detection," J. Acoust. Soc. Am. **49**, 1751–1756.

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

Anderson, N. S., and Burns, V. D. (**1973**). "Comparison of presentation rates using a missing item probe test of immediate memory," Bull. Psychonom. Soc. **2**, 200–202.

Berg, B. G. (**1989**). "Analysis of weights in multiple observation tasks," J. Acoust. Soc. Am. **86**, 1743–1746.

Berg, B. G., and Robinson, D. E. (**1987**). "Multiple observations and internal noise," J. Acoust. Soc. Am. **81**, S33.

Braida, L. D., and Durlach, N. I. (**1972**). "Intensity perception: II. Resolution in one-interval paradigms," J. Acoust. Soc. Am. **51**, 483–502.

Brodie, D. A., and Murdock, B. B. (**1977**). "Effect of presentation time on nominal and functional serial-position curves of free recall," J. Verbal Learn. Verbal Behav. **16**, 185–200.

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

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

Ellermeier, W., and Schrödl, S. (**2000**). "Temporal weights in loudness summation," in *Fechner Day 2000. Proceedings of the 16th Annual Meeting of the International Society for Psychophysics*, edited by C. Bonnet (Université Louis Pasteur, Strasbourg), pp. 169–173.

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

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

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

Grimm, G., Hohmann, V., and Verhey, J. L. (**2002**). "Loudness of fluctuating sounds," Acta Acust. united Acust. **88**, 359–368.

Heeren, W., Rennies, J., and Verhey, J. L. (**2011**). "Spectral loudness summation of nonsimultaneous tone pulses," J. Acoust. Soc. Am. **130**, 3905–3915.

Heinze, G. (**2006**). "A comparative investigation of methods for logistic regression with separated or nearly separated data," Stat. Med. **25**, 4216–4226.

Hosmer, D. W., and Lemeshow, S. (**2000**). *Applied Logistic Regression* (Wiley, New York).

Hots, J., Rennies, J., and Verhey, J. L. (**2014**). "Modeling temporal integration of loudness," Acta Acust. united Acust. **100**, 184–187.

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

IEC 318 (**1970**). *An IEC Artificial Ear, of the Wide Band Type, for the Calibration of Earphones Used in Audiometry* (International Electrotechnical Commission, Geneva, Switzerland).

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

McElree, B., and Dosher, B. A. (**1989**). "Serial position and set size in short-term-memory: The time course of recognition," J. Exp. Psychol. Gen. **118**, 346–373.

Murdock, B. B. (**1962**). "The serial position effect of free recall," J. Exp. Psychol. **64**, 482–488.

Nachmias, J. (**2006**). "The role of virtual standards in visual discrimination," Vision Res. **46**, 2456–2464.

Namba, S., Kuwano, S., and Kato, T. (**1976**). "Loudness of sound with intensity increment," Jpn. Psychol. Res. **18**, 63–72.

Norman, D. A. (**1966**). "Acquisition and retention in short term-memory," J. Exp. Psychol. **72**, 369–381.
Oberfeld, D. (**2008**). "Does a rhythmic context have an effect on perceptual weights in auditory intensity processing?," Can. J. Exp. Psychol.-Rev. Can. Psychol. Exp. **62**, 24–32.
Oberfeld, D. (**2015**). "Are temporal loudness weights under top-down control? Effects of trial-by-trial feedback," Acta Acust. united Acust. **101**, 1105–1115.
Oberfeld, D., Heeren, W., Rennies, J., and Verhey, J. (**2012**). "Spectro-temporal weighting of loudness," PLoS One **7**, e50184.
Oberfeld, D., Kuta, M., and Jesteadt, W. (**2013**). "Factors limiting performance in a multitone intensity-discrimination task: Disentangling non-optimal decision weights and increased internal noise," PLoS One **8**, e79830.
Oberfeld, D., and Plank, T. (**2011**). "The temporal weighting of loudness: Effects of the level profile," Atten. Percept. Psychophys. **73**, 189–208.
Pedersen, B. (**2006**). "Auditory temporal resolution and integration. Stages of analyzing time-varying sounds," in *Department of Acoustics, Sound Quality Research Unit* (Aalborg University, Aalborg, DK).
Pedersen, B., and Ellermeier, W. (**2008**). "Temporal weights in the level discrimination of time-varying sounds," J. Acoust. Soc. Am. **123**, 963–972.
Plank, T. (**2005**). "Auditive Unterscheidung von zeitlichen Lautheitsprofilen" ("Auditory discrimination of temporal loudness profiles"), in *Philosophische Fakultät II* (Universität Regensburg, Regensburg).
Ponsot, E., Susini, P., and Oberfeld, D. (**2016**). "Temporal weighting of loudness: Comparison between two different psychophysical tasks," J. Acoust. Soc. Am. **139**, 406–417.
Ponsot, E., Susini, P., Saint Pierre, G., and Meunier, S. (**2013**). "Temporal loudness weights for sounds with increasing and decreasing intensity profiles," J. Acoust. Soc. Am. **134**, EL321–EL326.
Rennies, J., and Verhey, J. L. (**2009**). "Temporal weighting in loudness of broadband and narrowband signals," J. Acoust. Soc. Am. **126**, 951–954.
Scharf, B. (**1978**). "Loudness," in *Handbook of Perception, Volume IV. Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic Press, New York), pp. 187–242.
Sorkin, R. D., Robinson, D. E., and Berg, B. G. (**1987**). "A detection theory method for the analysis of auditory and visual displays," in *Proceedings of the 31st Annual Meeting of the Human Factors Society*, October 19–22, New York, pp. 1184–1188.
Swets, J. A. (**1986**). "Indices of discrimination or diagnostic accuracy: Their ROCs and implied models," Psychol. Bull. **99**, 100–117.
Turner, M. D., and Berg, B. G. (**2007**). "Temporal limits of level dominance in a sample discrimination task," J. Acoust. Soc. Am. **121**, 1848–1851.
Waugh, N. C., and Norman, D. A. (**1965**). "Primary memory," Psychol. Rev. **72**, 89–104.
Wickelgren, W. A. (**1970**). "Time, interference, and rate of presentation in short-term recognition memory for items," J. Math. Psychol. **7**, 219–235.
Zwicker, E. (**1977**). "Procedure for calculating loudness of temporally variable sounds," J. Acoust. Soc. Am. **62**, 675–682.
Zwislocki, J. J. (**1969**). "Temporal summation of loudness: An analysis," J. Acoust. Soc. Am. **46**, 431–441.