

The subjective duration of ramped and damped sounds

MASSIMO GRASSI

Università di Padova, Padua, Italy

and

CHRISTOPHER J. DARWIN

University of Sussex, Brighton, England

Two experiments demonstrate that the perceived durations of sounds as long as 1 sec are influenced by the sounds' amplitude envelopes, extending Schlauch, Ries, and DiGiovanni's (2001) observations on sounds of 200-msec duration. Sounds with a monotonic decay (i.e., damped sounds) are heard as substantially shorter than both steady sounds and those with a monotonic increase of level (i.e., ramped sounds). Neither a reaction time (Experiments 1 and 2) nor a staircase (Experiment 2) procedure supported a sensory explanation for these different subjective durations. The results are compatible with the suggestion of Stecker and Hafter (2000) that listeners exclude part of the tails of damped sounds in the computation of their subjective durations.

In the present article, we investigate a simple temporal asymmetry in hearing: the difference between sounds that increase in level (i.e., *ramped* sounds) and those that decrease in level (i.e., *damped* sounds). Although pairs of such sounds have identical overall spectral contents, durations, levels, and absolute changes in level, many perceptual attributes of the sounds are strikingly different. Ramped and damped sounds differ in timbre (Akeroyd & Patterson, 1995; Irino & Patterson, 1996; Patterson, 1994a, 1994b), subjective duration (Schlauch, Ries, & DiGiovanni, 2001), overall loudness (Stecker & Hafter, 2000), and possibly changes in loudness (for durations shorter than ~2 sec, see Neuhoff, 1998, 2001; Seifritz et al., 2002; for longer durations, see Canévet, 1986; Canévet & Scharf, 1990; Canévet, Teghtsoonian, & Teghtsoonian, 2003; Schlauch, 1992; Teghtsoonian, Teghtsoonian, & Canévet, 2000, 2005). The explanation for these effects is not yet clear, with differences in sensory coding providing a possible explanation for brief sounds but not for longer ones. With the experiments reported here, we investigated differences in perceived duration for ramped and damped sounds whose durations extend up to 1 sec.

In the first report of asymmetric perception of these temporally reversed sounds, Patterson and his colleagues (Akeroyd & Patterson, 1995; Irino & Patterson, 1996; Patterson, 1994a, 1994b) found that the timbre of a concat-

enated series of brief (≤ 100 -msec) ramped sounds was different from that of the reversed series. The sequence of ramped sounds was perceived as possessing a rhythmic component together with a continuous component, whereas the sequence of damped sounds had only the rhythmic component.

A similar perceptual asymmetry is present for loudness. Stecker and Hafter (2000) found that a 250-msec ramped sound was perceived as louder than the corresponding reversed sound. With ramped and damped sounds of even longer duration (1.8 sec), Neuhoff and his colleagues reported that listeners perceived the change in loudness of ramped sounds to be greater than that for damped sounds (Neuhoff, 1998, 2001; Seifritz et al., 2002).

Schlauch et al. (2001) discovered that a perceptual asymmetry also occurs for subjective duration. Using both a magnitude estimation procedure and a matching procedure, they measured the subjective durations of single ramped and damped sounds 10–200 msec in duration. The two procedures yielded similar results: Ramped sounds were always perceived as much longer than damped sounds of the same physical duration.

In summary, across all three perceptual dimensions (i.e., subjective duration, loudness, and timbre), the ramped sounds are perceived as more salient than the damped ones. In addition, the asymmetry decreases (Akeroyd & Patterson, 1995; Irino & Patterson, 1996; Schlauch et al., 2001) or disappears (Neuhoff, 1998, 2001) when the carrier is a broadband noise rather than a tone, and also decreases when the modulating function is shallow rather than steep (Akeroyd & Patterson, 1995; Irino & Patterson, 1996; Patterson, 1994a, 1994b; Schlauch et al., 2001; Stecker & Hafter, 2000).

There are currently two approaches to explaining such results, one of which is broadly sensory and the other broadly

Experiment 2 was conducted while M.G. was a Marie Curie fellow at the University of Sussex (EC Contract HTMP-CT-2001-00234). The authors thank two anonymous reviewers for helpful comments on an earlier version of the manuscript. Correspondence concerning this article should be addressed to M. Grassi, Dipartimento di Psicologia Generale, Università di Padova, via Venezia 8, 35131 Padova, Italy (e-mail: massimo.grassi@unipd.it).

cognitive. The first approach attributes the asymmetry to peripheral coding. For example, in the auditory image model (AIM; Patterson, Allerhand, & Giguère, 1995; Patterson & Irino, 1998) sounds are first filtered through an auditory filter bank, which converts the sound wave into a simulation of basilar membrane motion. In the second stage, a bank of model hair cells transduces the motion into a simulation of the auditory nerve activity pattern. Both stages of the model suppress the amplitude of the input sound across time, with the suppression progressively weakening as the duration of the sound increases. Furthermore, the abrupt offset of ramped sounds causes a ringing in the filtering process, which extends the duration of their auditory image. In summary, the auditory images of ramped and damped sounds are highly asymmetrical (those of ramped sounds being longer and more powerful than those of damped sounds)¹ because of temporally asymmetric onsets and offsets. Although the model can in principle explain the ongoing timbral differences between repeated ramped and damped sounds and the durational and loudness differences for brief sounds, it is unlikely that the AIM could explain durational or loudness differences for longer sounds.

The second, more cognitive approach attributes the different percepts of ramped and damped sounds to listeners' ignoring the decaying tail of a damped sound. Stecker and Hafter (2000) pointed out that natural sounds (e.g., impact sounds; see Gaver, 1993; Grassi, 2005) are generally characterized by a fast attack and a decay that may be prolonged by environmental reverberation. Listeners may interpret the tail of a damped sound as being due to reverberation rather than intrinsic to the sound source and may therefore exclude it from their computation of the subjective duration or loudness of the sound source. The cognitive mechanism proposed by Stecker and Hafter could enhance asymmetries already present in the sensory representation, allowing perceptual asymmetries at durations beyond the range of purely sensory mechanisms.

The main question addressed by the present experiments is whether or not the longer perceived duration for ramped than for damped sounds holds for sounds that are longer than the 200-msec sounds considered by Schlauch et al. (2001). This question is important because the explanation provided by Stecker & Hafter (2000) can apply to long-duration sounds much more easily than can that provided by the AIM. Although Schlauch et al. used sounds ≤ 200 msec in duration, a simple extrapolation of their results (Figures 2 and 5 of Schlauch et al., 2001) suggests that the subjective shortening of damped sounds that they found should persist for sounds of longer durations.

EXPERIMENT 1

The first experiment had two main aims: (1) to examine, using a matching procedure, whether or not there are differences in perceived duration between long-duration (>200 -msec) ramped and damped sounds; and (2) to estimate the contribution that sensory factors make to these durational asymmetries by measuring listeners' reaction times (RTs) to these sounds' onsets and offsets. Both com-

plex tones and white noise that had been gated by either raised cosine² or, as in Schlauch et al. (2001), exponential functions were used in the experiment. Like Schlauch et al., we also used a steady sound in order to estimate whether differences between ramped and damped sounds are due to overestimation of ramped sounds, underestimation of damped sounds, or both.

Method

Listeners. Ten paid listeners recruited from the population of the University of Sussex participated in the experiment. None of them reported a hearing loss or other difficulties with their hearing. All listeners were naive as to the purposes of the experiment.

Stimuli. Two timbres served as carriers: a complex tone with the first ten 500-Hz harmonics all at the same amplitude, and a white noise. From these two carriers, ramped, damped, and steady sounds 250, 500, 750, and 1,000 msec in duration were produced. The steady sounds were gated on and off with two 7.5-msec raised cosine ramps but were otherwise constant in level. The ramped and damped sounds were obtained by modulating a carrier with either a raised cosine or an exponential function to yield overall durations of 250, 500, 750, and 1,000 msec. The offset of ramped and the onset of damped sounds were further gated with a 15-msec raised cosine ramp. We will refer to the nonsteady sounds by the acronyms CR, CD, ER, and ED, where C denotes cosine; E, exponential; R, ramped; and D, damped.

The exponential function used in the ERs had the form (adapted from Schlauch et al., 2001)

$$a(t) = \exp\left[5\left(\frac{t}{T} - 1\right)\right],$$

where $a(t)$ is the amplitude of the modulator at time t and T is the overall duration of the sound (250, 500, 750, or 1,000 msec). This function, together with the raised cosine function for the CR sounds, is illustrated in Figure 1. The final 15 msec were given an additional raised cosine ramp to bring the overall amplitude smoothly to zero. ED targets were obtained by reversing the ER targets.

Apparatus. Sounds were synthesized in real time at 44.1 kHz and 16-bit resolution using custom software running on a Macin-

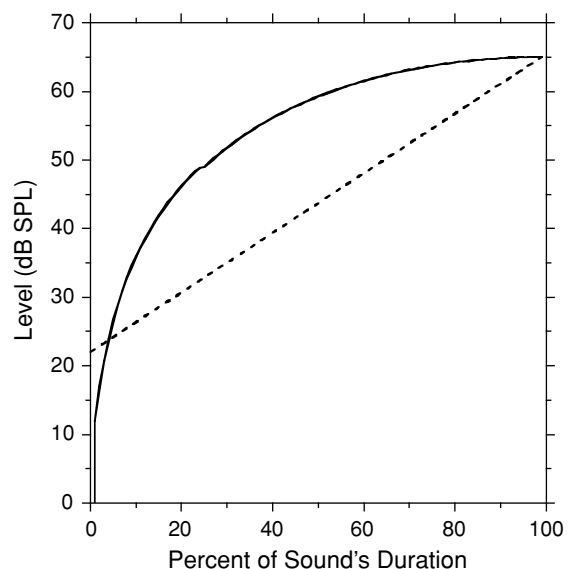


Figure 1. Sound level as a function of percent duration of the two modulators used in Experiment 1. The solid and dashed lines illustrate raised cosine and exponential modulators, respectively.

tosh Power PC that also controlled the experiment. The output of the sound card was passed through Tucker-Davis attenuators and presented diotically through Sennheiser HD414 headphones in a single-skin sound-attenuated IAC booth (Industrial Acoustics Company). Peak level for all sounds used in the experiment was 65 dB SPL.

Procedure. Each listener performed both the matching task and the RT task within a maximum period of 1 week. The order of the tasks was counterbalanced across listeners.

Matching task. The listeners were asked to match the duration of a steady adjustable sound to that of a fixed duration ER, CR, ED, CD, or steady target. The experiment was divided into eight experimental blocks. During each block, a trial consisted of a target sound followed after 500 msec by the adjustable sound, of identical timbre, whose duration the listeners had to match to that of the target. The adjustable sound was similar to the steady target in that it was gated on and off with two raised cosine ramps of 7.5 msec. Each experimental block contained ER, CR, ED, CD, and steady targets sharing the same timbre and duration. Within each experimental block, the listeners matched 50 targets presented in random order: 10 each of ER, CR, ED, CD, and steady sounds. The experimental blocks were presented to the listeners in a nested, randomized order. Experimental blocks with targets of the same overall duration were presented consecutively to obtain more consistent adjustments (Wright, Buonomano, Mahncke, & Merzenich, 1997), and the order in which the different durations were presented was randomized across listeners. Within these four series of two blocks sharing the same target duration, the timbre order was counterbalanced across listeners. Each block lasted about 40 min.

During a trial, the listener adjusted the duration of the target sound by moving a roller ball and could hear the result by pressing the roller ball's button. At the beginning of each trial, the duration of the adjustable sound was chosen randomly from a range of within $\pm 80\%$ of the duration of the target sound. The upper and lower limits of this range also served as the maximum permitted adjustments during each experimental trial.

RT task. In two of the four experimental blocks, the listeners reacted to onsets and in the remaining two blocks to offsets; block order was randomized.

On each trial of an onset block, the listeners heard an alert sound (30-msec, 2-kHz sine wave) followed after a variable silent interval (1,000, 1,250, 1,500, 1,750, or 2,000 msec) by a ramped, a damped,

or a steady sound. The listeners' task was to press a keyboard key as soon as they heard the target sound. The response terminated the sound. Every combination of silent interval and target was randomly presented five times during each block, for a total of 500 trials (5 silent intervals \times 5 target envelopes \times 4 target durations \times 5 presentations), so that each onset block lasted approximately 30 min. The listeners could choose to repeat a trial if they thought they had made an anticipation or had unduly delayed their response. Onset RTs were measured as the temporal difference between the response and the actual onset of the target sound.

The procedure for the offset RT was similar except that the listeners reacted to the sound's offset instead of to its onset. In order to avoid possible anticipations in the reaction to damped sounds, extra (fake-damped) target sounds were added to the target set. Fake-CD and fake-ED targets were obtained by adding a low-level steady tone (or noise) to either ED or CD targets. These manipulations produced a sound with an initial decrement in level followed by a distinctly audible end. The peak level of the fake-damped sound was identical to that of the other targets. In each block, the listeners reacted to a total of 700 sounds (5 silent intervals \times 7 target envelopes \times 4 target durations \times 5 repetitions) presented in random order. Each offset block lasted approximately 40 min. Offset RTs were measured as the temporal difference between the response and the physical offset of the target sound. Data from the fake-damped targets were not analyzed.

Results

Matching task. The average durations of the adjusted sound for each target type are shown in Figure 2. For steady tones, the listeners estimated duration accurately. A simple linear regression between the average matched duration of the steady tone and noise targets and the actual duration of the target had a slope very close to one (0.979, $r^2 = 1$) and an intercept very close to zero (4.78 msec). In contrast, the listeners slightly underestimated the duration of ramped targets and substantially underestimated the duration of damped targets relative to the steady targets [for ER, $F(1,9) = 13.76$, $p = .005$; for CR, $F(1,9) = 23.72$, $p = .001$; for ED, $F(1,9) = 148.68$, $p < .0001$; and for CD, $F(1,9) = 113.69$, $p < .0001$].

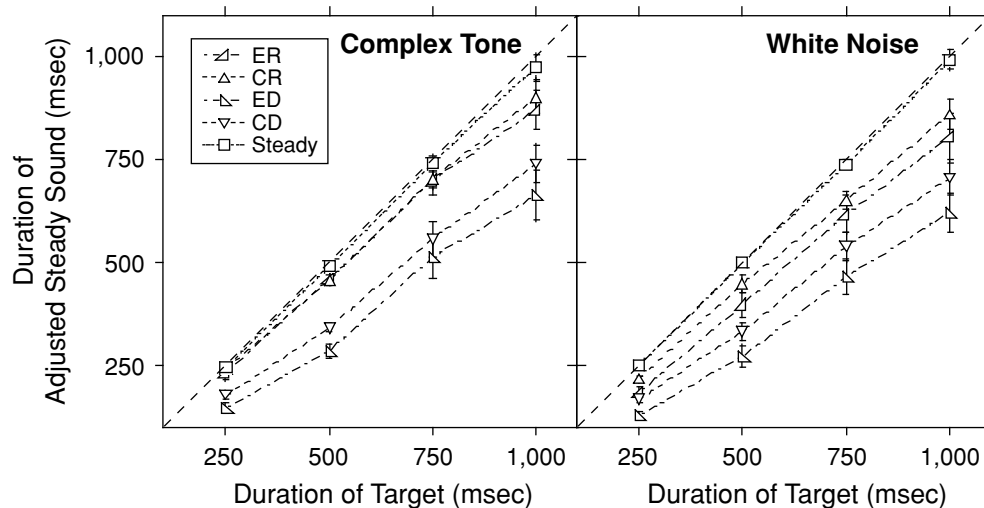


Figure 2. Average matched duration as a function of the duration of the target sound in Experiment 1. In both panels, the diagonal dashed line corresponds to a perfect match. Vertical bars represent ± 1 standard error of the mean.

The matched durations collected with amplitude-modulated targets were converted into a derived score, $D\%$, which expresses the under- or overestimation as a percentage of the duration of the ramped or damped target, respectively, relative to that of the matched steady target:

$$D\% = 100 \left(\frac{d_s - d_m}{d} \right),$$

where d_m is the matched duration to the modulated (ramped or damped) target, d_s is the matched duration to the steady target, and d is the actual duration of the target sound (see Figure 3).

For tonal stimuli, the ramped sounds show a constant underestimation of about 5%–10%, which does not change significantly with duration ($F < 1$). The percentage underestimation for damped sounds is much larger and, at least for the exponential envelope, shows a significant reduction with increasing duration [$F(3,27) = 3.91, p = .01$]. Consequently, the difference between the ramped and the damped sounds decreases with increasing duration [$F(1,9) = 6.18, p = .03$].

The pattern of results for the noise stimuli differs from that of the tonal stimuli in the following ways. First, the underestimation of the ramped sounds' duration is much greater than it was for the tonal sounds, especially for the exponential ramp (although, again, overall there is no significant variation with duration). Second, the difference between the ramped and damped sounds is overall less than that for the tonal stimuli.

The small underestimation of the duration of the ramped sounds that is constant with duration could plausibly be due to a simple sensory threshold acting at a fixed level of the ascending ramp. The larger, duration-dependent underestimation of the damped sounds' duration is not so easily explained in this way.

RT task. Onset and offset RTs were collected and averaged for each target separately for each listener. Outliers, defined as individual RTs more than ± 2 SDs from the mean of the listener, were excluded from these averages both here and in Experiment 2; in each experiment, they accounted for less than 5% of the total. The resulting means and standard errors are shown in Figure 4.

Average onset RTs to noises are ~ 7 msec shorter than those to tones [$F(1,9) = 6.03, p = .03$],³ but the pattern of results was otherwise similar for the two timbres. The abrupt onset of the steady and damped sounds gives a constant RT of about 230 msec regardless of the duration of the sound [$F < 1, F < 1$, and $F(1,9) = 1.30, p > .05$ for steady, ED, and CD sounds, respectively]. The ramped sounds result in longer RTs [$F(4,36) = 193.32, p < .0001$], which increase as the duration of the sound increases [$F(12,108) = 42.46, p < .0001$]. RTs to the cosine ramp increase more with duration ($B = 0.08, r^2 = .99$) than do those to the exponential ramp [$B = 0.04, r^2 = .97; F(3,27) = 19.48, p < .0001$].

If the listeners were reacting to an absolute threshold sound level, their RTs would reflect the fact that this level is reached at a constant proportion of the duration of the sound. For example, in Figure 1, 50 dB is reached for the cosine ramp about 30% through the sound. Allowing for a presumed constant additional delay, RTs should increase with duration by that proportion of the sound. This model thus predicts a linear relationship between RT and duration, with the slope giving the percentage of the sound's duration at which the threshold is triggered. For the cosine function, the slope is 0.08 ($r^2 = .99$), giving a threshold (see Figure 1) of about 30 dB. For the exponential ramp, the slope of the RT function is 0.04 ($r^2 = .97$), which brings the threshold very close to the beginning of the sound (first 15 msec), where an additional fixed-duration raised cosine ramp has been applied. The effect of this

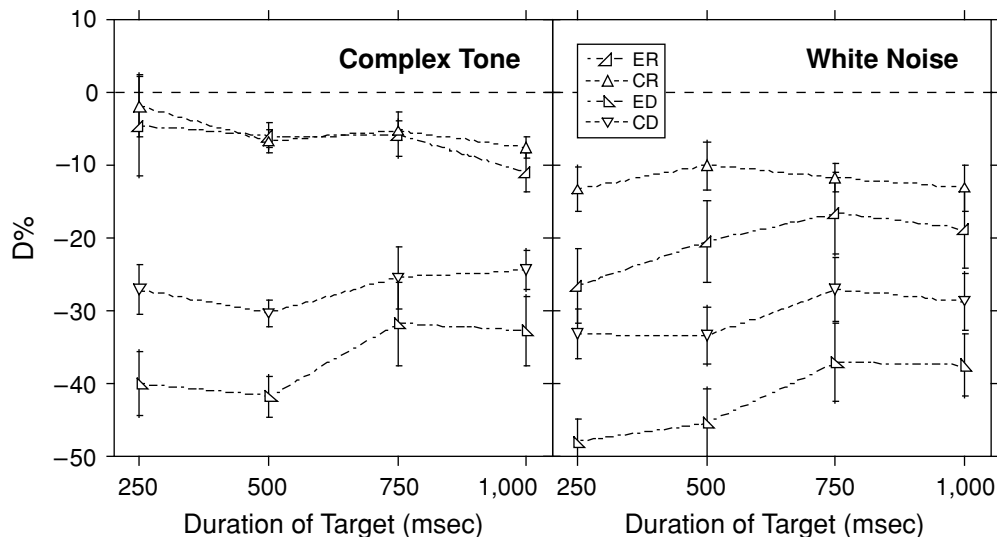


Figure 3. Matching task of Experiment 1: $D\%$ as a function of the duration of the sound. The horizontal dashed line shows a $D\%$ of zero. Vertical bars represent ± 1 standard error of the mean.

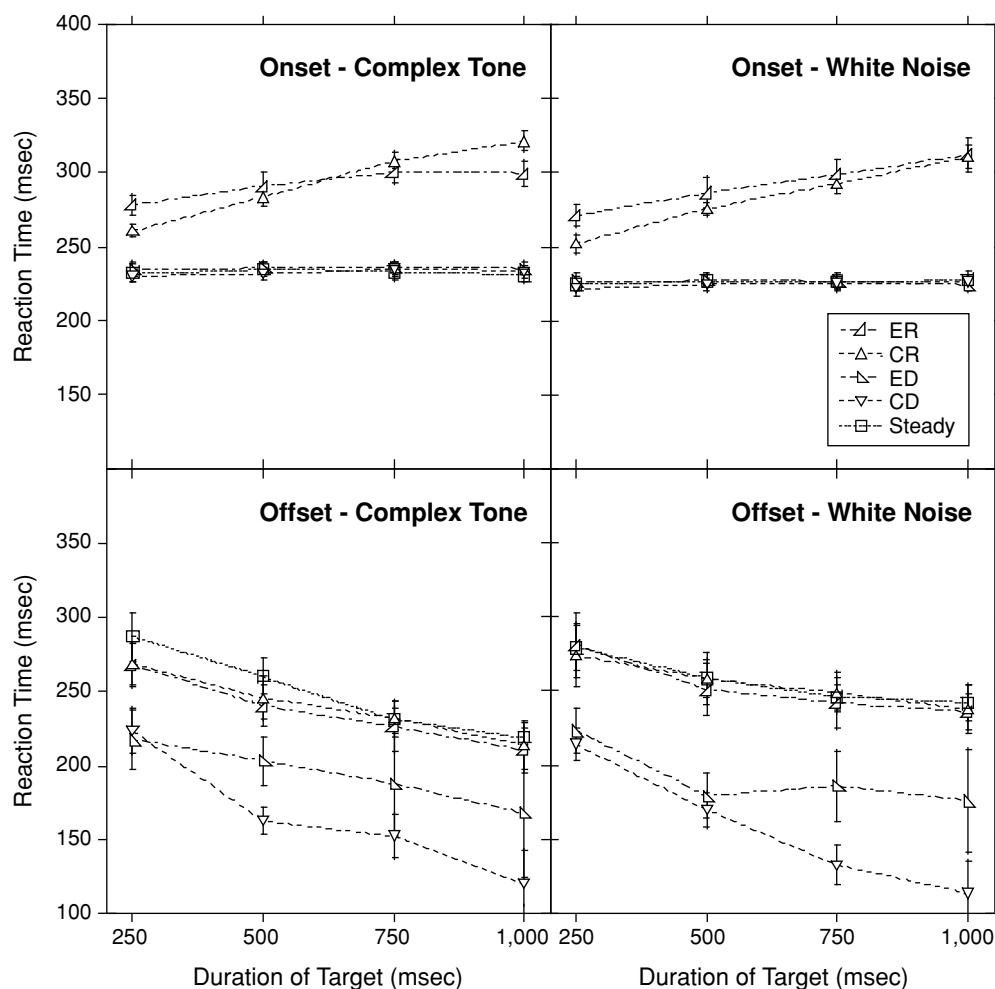


Figure 4. Onset and offset reaction times as a function of the duration of the target sound in Experiment 1. Vertical bars represent ± 1 standard error of the mean.

additional fixed-duration ramp would be to reduce the effect of the sound's duration on the RT, as we have found. Although precise predictions are difficult, the general pattern of the results is broadly compatible with a fixed, low threshold triggering the onset RTs.

The overall pattern of offset RTs is rather different. First, the timbre of the carrier influenced neither the average RT ($F < 1$) nor its variation with duration [$F(3,27) = 1.11, p > .05$] or with envelope [$F(4,36) = 1.99, p > .05$]. A simple threshold model would predict that RTs to the steady and the ramped sounds should be similar to each other and constant across durations, whereas RTs to the damped sounds should be anticipated, with anticipation increasing linearly as the overall duration of the sound increases.

However, although RTs to ramped and steady sounds are similar to each other [$F(1,9) = 1.35, p > .05$ and $F < 1$ for ER and CR targets, respectively], all RTs decreased with increasing duration of the sound [$F(3,27) = 7.46, p = .001$]. The most obvious explanation for this decrease is that listeners are more efficient at anticipating the ends

of longer sounds. Such a decrease could either affect all targets of a given duration in a similar fashion (i.e., all RTs would be translated down by a fixed factor) or affect damped sounds more than ramped and steady sounds. In fact, a further anticipation could have occurred with damped sounds, since the ends of these sounds can be more easily predicted⁴ than those of ramped and steady targets (in that the former sounds invariably dissolve into silence). Nonetheless, the minimal requirement for a simple threshold model should predict, at least, that offset RTs to damped sounds should *decrease* relative to those to steady and ramped sounds with increasing duration. Such a decrease is apparent overall for the cosine damped sounds, but less so for the exponential damped sounds [$F(1,9) = 5.74, p = .04$]. As with the onset RTs, the exponential ramps gave somewhat flatter functions than did the cosine. The linear regression between the durations of ramped and steady targets and those of actual targets shows a negative slope of -0.06 ($r^2 = .97$). If a fixed threshold is operating, slopes calculated for RTs to damped targets should be more negative than those calculated for RTs to

ramped and steady targets. This prediction is confirmed for RTs to CD targets, which decreased in comparison with those to ramped and steady targets with increasing duration ($-0.13, r^2 = .96$), but not for RTs to ED targets, which showed a constant anticipation in comparison with ramped and steady sounds ($-0.06, r^2 = .91$).

In summary, offset RTs (and subsequent manipulations) might be biased by listeners' anticipations, particularly for long-duration targets.

Comparison of the results of the two tasks. We now calculate the perceived duration in the RT task as the difference between offset and onset RTs (with the addition of target duration) normalized (i.e., divided) by the steady sound duration calculated in a similar fashion.⁵ Figure 5 shows the estimated durations of the four different ramped and damped sounds relative to the steady sound, calculated for the RT task. The overall pattern of results is that the RT data provide estimated relative durations that are broadly similar to the matched durations of the ramped sounds [with only CR tones showing a significantly shorter matched duration— $F(1,9) = 6.67, p < .02$] but not to those of the damped sounds, which are much shorter than the durations estimated from the RT data (for all comparisons, $p < .005$).

Discussion

To the extent to which the RT data provide an estimate of the sensory contribution to estimated duration, we can conclude that, whereas sensory factors could well provide an adequate explanation for the small underestimate of the ramped sounds relative to the steady sounds in the matching task, they do not explain the substantial matching underestimate of the damped sounds. Moreover, this substantial underestimate of the duration of damped sounds is found even for sounds as long as 1 sec.

We have discussed the possibility that listeners may anticipate the offset of targets in general and those of

damped sounds in particular. If the anticipation affected all targets of a given duration in an identical manner, the normalization performed on RT scores should be sufficient to limit the inaccuracy of subsequent estimated durations. On the contrary, if the anticipation affected damped sounds in particular, the subsequent estimated durations would only exaggerate the role of an early sensory offset in the durational underestimation of damped sounds. Specifically, the damped sound durations, estimated by means of the RT data, would result in a shorter estimate than would a theoretically bias-free model. Thus, even if a strong anticipation on damped sounds is hypothesized, the results of our experiment show a substantial difference in the damped durations estimated from RT data or collected with the matching paradigm. In conclusion, although durations calculated with RTs have some inaccuracies, they indicate that factors other than purely sensory ones are responsible for the underestimation of damped sounds.

EXPERIMENT 2

The second experiment provides a partial replication of the first and introduces a different paradigm for estimating the sensory contribution to the perceived shorter duration of damped sounds that does not suffer from the RT experiment's anticipation problem.

The new paradigm aims for a bias-free measure of the perceptual position of ramped sensory onsets and damped sensory offsets. In this paradigm, listeners judge which of two similar ramped or damped sounds is shorter. One of the sounds is a normal ramped or damped sound, and the other is similar but has had varying amounts of its onset (for ramped sounds) or offset (for damped sounds) removed. How much has been removed is varied adaptively to provide a threshold. The aim is to find out how much of the initial/final part of the sound does not contribute to perceived duration.

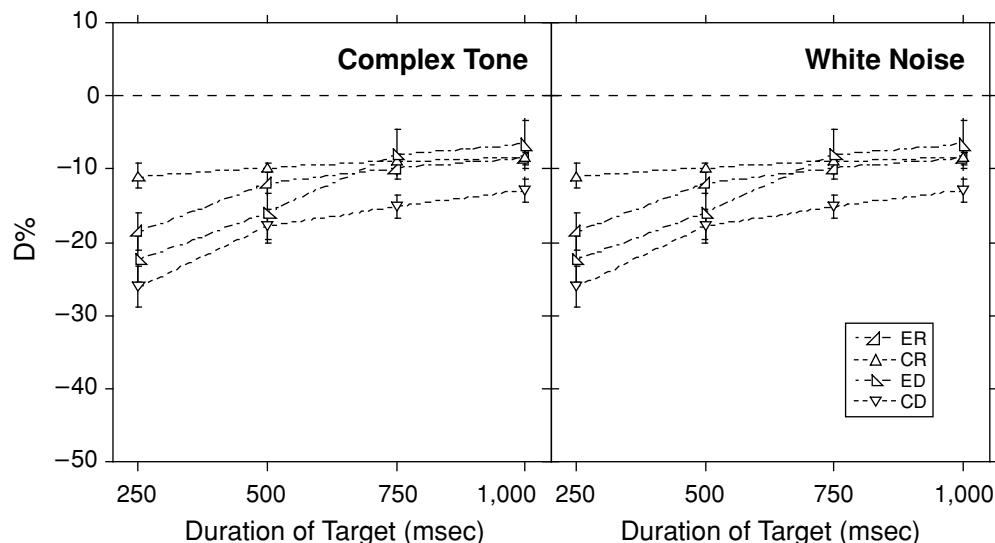


Figure 5. Reaction time task of Experiment 1: D% as a function of the duration of the sound. The horizontal dashed line shows a D% of zero. Vertical bars represent ± 1 standard error of the mean.

The experiment includes the matching and RT tasks from Experiment 1, together with the new adaptive task, in which the truncated ramped/damped comparison sound is used. Only the exponential modulator was used, since this envelope had given the largest difference between the behavioral measures collected with the matching and RT tasks. To reduce the duration of the testing, only tones were used as carriers, since they had previously given the largest difference between ramped and damped sounds.

Method

Listeners. Twelve paid listeners were recruited from the University of Sussex student population and tested individually. They had normal audiometric thresholds, and none of them had participated in the previous experiment.

Stimuli. The stimuli were ramped, damped, and steady sounds of four durations: 250, 500, 750, and 1,000 msec. The carrier was a complex tone consisting of the first ten 500-Hz harmonics at equal amplitude. The modulating function of ramped sounds was

$$a(t) = \exp\left[k\left(\frac{t}{T} - 1\right)\right],$$

where $a(t)$ is the amplitude of the waveform as a function of time t , T is the overall duration of the sound (250, 500, 750, or 1,000 msec), and k determines the slope of the exponential (here, $k = 5.18$). When $t = 0$, the level of the sound is 45 dB below peak level regardless of T . In the adaptive task, varying amounts of the initial part of the rise were set to zero. In addition, a 10-msec raised cosine ramp smoothed the onsets and offsets of all sounds. Damped targets were obtained by reversing the ramped targets.

Apparatus and Procedure. Sounds were generated as in Experiment 1. The experiment consisted of three tasks: a matching task, an RT task, and an adaptive task. The order in which the tasks were performed was counterbalanced across listeners. The experiment was performed in six 1-h sessions, and all the listeners completed the experiment within 10 days.

Matching task. This task was divided into four blocks. Each block consisted of matches performed with a single target duration. The listeners performed the four blocks in a random order. Within each

block, the listeners performed the same task as in Experiment 1, with a total of 30 matches for each block: 10 with ramped sounds, 10 with damped sounds, and 10 with steady sounds.

RT task. In one block, the listeners reacted to the onset of the target sounds, and in the other they reacted to its offset, in a counterbalanced order across listeners. Within each block, the procedure was identical to that of the exponential modulation conditions of Experiment 1, including the fake-damped sounds (offset block only).

Adaptive task. The adaptive task was divided into four blocks presented in a random order. Within each block, the listeners performed the adaptive procedure on a single target duration with ramped, damped, and steady sounds counterbalanced across listeners. Thresholds for sensitivity in difference of duration between target sound and variable sound were measured 10 times for each target. A two-interval, forced-choice, 2-down 1-up adaptive procedure was used to estimate 70.7% of the psychometric function (Levitt, 1971). The procedure involved three step sizes: 1/10, 1/20, and 1/50 of the duration of the target sound. The numbers of reversals needed to change step size were 2, 2, and 8, respectively. The listeners received feedback after responding to each pair. At the beginning of each trial, the duration of the variable sound was half that of the target. The threshold was calculated as the arithmetic mean of the duration of the variable sound over the final eight reversals.

Results

Matching task. The matching data shown in the left panel of Figure 6 largely replicate those of Experiment 1. The durations of the steady tones are matched veridically ($B = 0.95$, $r^2 = .99$, intercept slightly above zero: 44.8 msec). The ramped tones are slightly underestimated [$F(1,11) = 10.80$, $p = .007$] whereas the damped tones are substantially underestimated [$F(1,11) = 37.67$, $p < .0001$].

The D% scores (left panel of Figure 7) again show a pattern similar to that of Experiment 1. The small underestimate for ramped sounds decreases proportionately with increasing duration [$F(3,33) = 2.86$, $p = .051$], whereas the large underestimate for damped tones is roughly constant at 30%–40% ($F < 1$).

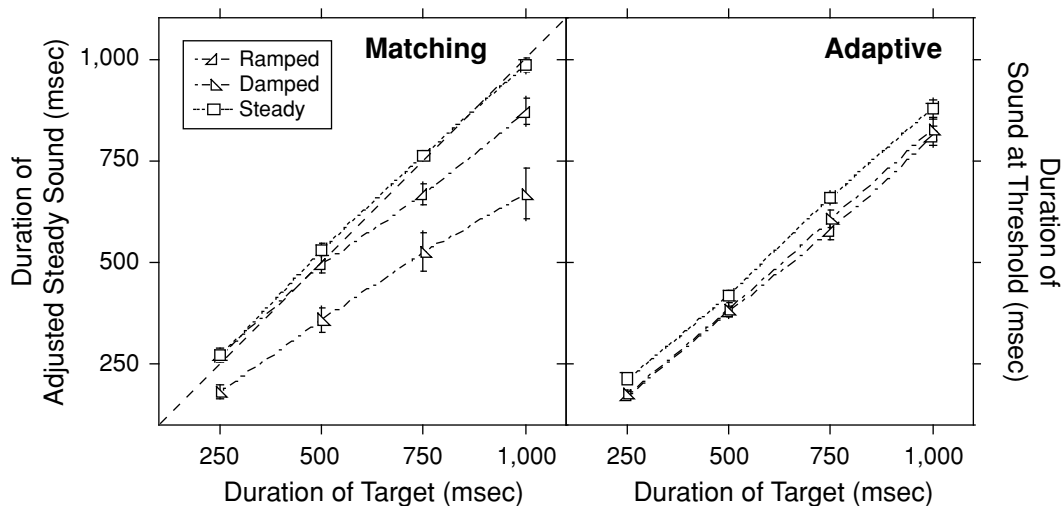


Figure 6. Results of Experiment 2. The left panel shows average matched duration as a function of the duration of the target sound (the diagonal dashed line shows a perfect match). The right panel shows threshold for duration discrimination as a function of the duration of the target sound. Vertical bars represent ± 1 standard error of the mean.

RT task. The listeners reacted more quickly to the onset of steady and damped tones than to the onset of ramped sounds [$F(2,22) = 111.58, p < .0001$], and this difference increased with increasing overall duration of the sound [$F(6,66) = 19.59, p < .0001$]. The offset RTs were shorter for damped sounds, whereas the RTs to the offset of steady and ramped sounds were longer [$F(2,22) = 10.08, p = .008$]. As in Experiment 1, with all sounds RT decreased as the overall duration of the sound increased [$F(3,33) = 26.08, p < .0001$]. However, this effect was more evident for damped than for steady and ramped tones [$F(6,66) = 4.48, p = .001$; see Figure 8].

RTs were transformed into expected durations, as in Experiment 1 (see middle panel of Figure 7). A 2 (envelopes) \times 4 (durations) ANOVA was performed on these data. The magnitude of delays and anticipations measured with ramped and damped sounds, respectively, were identical ($F < 1$). Two new ANOVAs revealed that the delay in the sensory onset of ramped sounds decreased as the overall duration of the sound increased [$F(3,33) = 38.13, p < .0001$]. On the contrary, the second analysis revealed that the anticipation of the sensory offset of damped sounds was constant across durations ($\sim 15\%$, $F < 1$).

Adaptive procedure. The right panel of Figure 6 shows the durations that are just noticeably shorter than their originals. The ramped and damped tones need more shortening than the steady tones do [$F(2,22) = 20.52, p < .0001$] and do not differ from each other ($F < 1$). The percentage shortening duration of these threshold sounds relative to that of the steady sound decreased as the duration of the sound increased [$F(3,33) = 7.76, p = .0005$; see Figure 7, right panel].

Comparison of the results of the three procedures. The D% values measured with the three procedures were compared by means of a 3 (tasks) \times 2 (envelopes) \times 4 (durations) ANOVA (see Figure 7). Behavioral data were dependent on the experimental task [$F(2,22) = 8.82, p = .001$]. Furthermore, whereas estimates collected with

ramped and damped sounds were similar for the RT and adaptive tasks, they were strikingly different for the duration matching task [$F(2,22) = 18.98, p < .0001$].

The underestimation measured for ramped sounds in the matching task was comparable to the delay in their sensory onsets measured with the adaptive procedure ($F < 1$), whereas the delay measured for the RT task was larger [$F(1,11) = 10.71, p = .007$]. By contrast, the anticipation measured for damped sounds with the matching task was greater than either that in the adaptive task [$F(1,11) = 26.92, p < .0001$] or that in the RT task [$F(1,11) = 12.38, p = .005$].

Discussion

This experiment replicated the main features of the first experiment for the exponential modulating function applied to the complex tone: The substantially larger underestimation of damped than of ramped tones in the matching task cannot be explained by the difference between the appropriate onset and offset RTs. Experiment 2 presents additional evidence against a sensory basis of the underestimation of damped tones in that the adaptive procedure applied to curtailed tones gives thresholds for both the ramped and the damped sounds that are comparable to those from the RT data and from the ramped matching data, but substantially lower than those of the damped matching data. Furthermore, as was hypothesized in the discussion of the previous experiment, the bias-free task (i.e., the staircase procedure) shows a smaller influence of sensory factors in underestimation of damped (but also of ramped) sounds than does the RT task.

GENERAL DISCUSSION

In the present study, we attempted to find a sensory explanation for the perceptual asymmetry between ramped and damped sounds with durations exceeding 200 msec. Specifically, we investigated whether the difference in

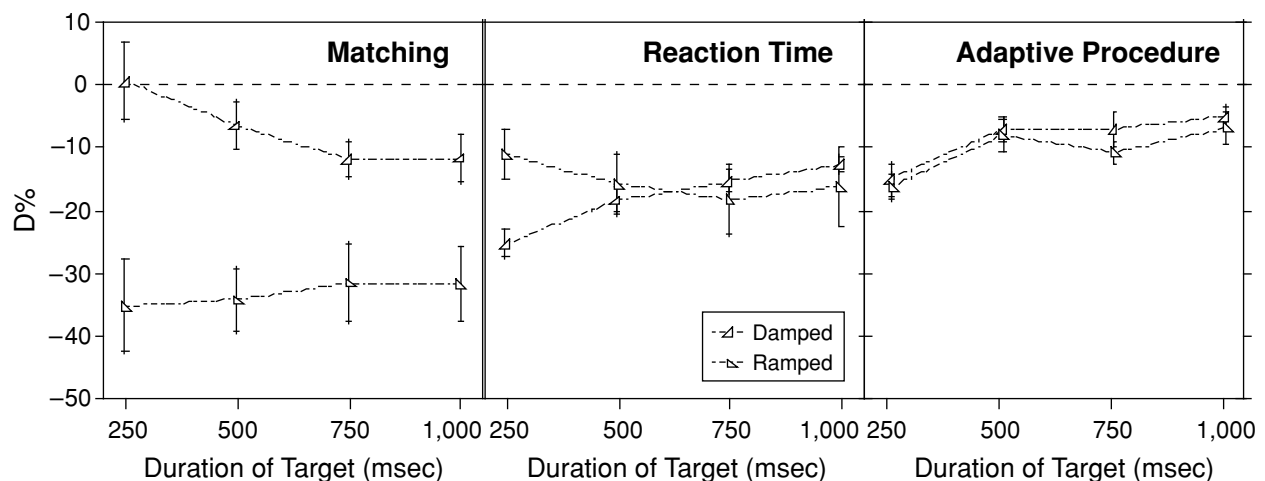


Figure 7. D% measured in the three tasks as a function of the duration of the target sound in Experiment 2. In all panels, the horizontal dashed line corresponds to a D% of zero. Vertical bars represent ± 1 standard error of the mean.

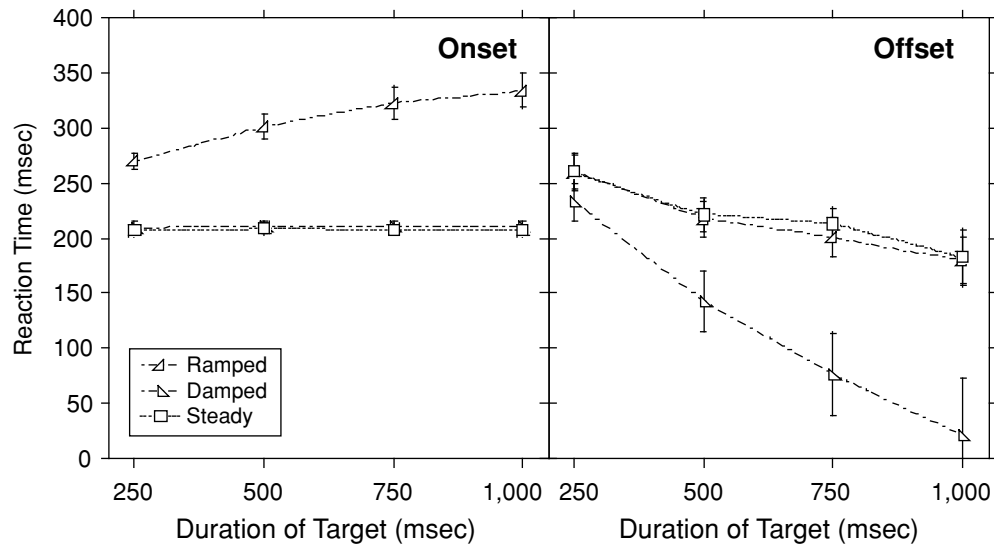


Figure 8. Onset and offset reaction times as a function of the duration of the sound in Experiment 2. In both panels, vertical bars represent ± 1 standard error of the mean.

subjective duration between these sounds might arise from delays and anticipations in their sensory onsets and offsets, as suggested by the AIM for shorter durations. Overall, the results of the RT experiments suggest only a small role for delays and anticipation on subjective duration. Moreover, at least for the two longest stimuli of the experiments, the data suggest that listeners should perceive ramped and damped sounds as having similar durations. For the two shortest durations, the RT data of Experiment 1 are consistent with the findings of Patterson and colleagues (Akeroyd & Patterson, 1995; Irino & Patterson, 1996; Patterson, 1994a, 1994b), who found that the longest half-life that yielded a timbre difference between ramped and damped sounds was ~ 60 msec.⁶

Nevertheless, the results of the duration matching experiments do show that damped sounds are perceived as much shorter than ramped sounds across all the durations investigated. Such a result does not arise from inaccuracy in duration matching, since the results of the steady target control condition show that the listeners' estimations were in fact highly accurate. In summary, the results of our experiments are more suitable to the cognitive approach of Stecker and Hafter (2000), henceforth referred to as the *echo hypothesis*, rather than to the sensory approach of the AIM (Patterson et al., 1995; Patterson & Irino, 1998): Although listeners can potentially perceive the two types of sounds as having similar durations, they exclude the tails of damped sounds from the computation of subjective duration. In consistency with the more cognitive approach of the echo hypothesis, Wang, Lu, and Liang (2003) found that the neural response to ramped and damped sounds becomes more and more asymmetrical at more central sites of the auditory system.

In summary, two separate explanations for the perceptual asymmetry of time-reversed stimuli are necessary.

For short sounds and high rates of change in level (see, e.g., Akeroyd & Patterson, 1995; Irino & Patterson, 1996; Patterson, 1994a, 1994b; Schlauch et al., 2001), the asymmetry has a sensory basis, whereas for longer sounds and lower rates of change of level (see, e.g., Neuhoff, 1998, 2001; Stecker & Hafter, 2000) more cognitive factors must also be invoked. For sounds longer than 200 msec, the echo hypothesis is supported by additional empirical data: Schlauch and colleagues (personal communication, May 2005) asked listeners to perform a duration matching experiment. Listeners were divided into three groups, the first of which received no specific instruction, the second of which was told that the tail of the damped sounds was an echo, and the third of which was instructed to evaluate the damped sounds in their entirety (i.e., they were informed that the tails of the damped sounds were not artifacts). Results showed similar underestimations for damped sounds in the first and second groups, whereas the underestimation of the third group was smaller.

Nonetheless, Experiment 2 failed to corroborate one of the predictions suggested by Stecker and Hafter (2000). According to these authors, the asymmetry between ramped and damped sounds may originate as listeners switch from "literal mode" perception to "constancy mode" perception. Initially, listeners act in the literal mode and thus perceive the two sounds as having similar levels of loudness (or, as in our case, similar durations). Subsequently, they switch from literal to constancy mode, interpreting the damped sound as being composed of two parts (attack plus reverberation), and so they neglect the tail of the damped sounds. Such a switch in the perceptual behavior of the listener would require some time to be actuated (~ 200 msec; Stecker & Hafter, 2000).

In the present study, two different perceptual mechanisms could explain the difference between damped sound

scores in the RT task and those in the matching task. In the RT task, listeners perceive in literal mode because of time pressure, whereas in the matching task listeners perceive in constancy mode because there is no time pressure. Since there is no time pressure with the staircase procedure, listeners should again perceive in constancy mode, and the subsequent ramped and damped estimates should be substantially different from each other. However, ramped and damped estimates gathered with the staircase procedure were in fact very similar. These observations also do not support Stecker & Hafter's (2000) two modes.

Overall, however, the echo hypothesis can explain well the perceptual asymmetry observed for subjective duration and loudness. Moreover, it can be easily extended to explain the perception of a change in loudness of ramped and damped sounds. If we assume that listeners exclude the tails of damped sounds from the computation of subjective estimates, the dynamic range of damped sounds (in comparison with that of ramped sounds) will be reduced.

The echo hypothesis contrasts with Neuhoff's (1998, 2001) claim, based on auditory looming judgments, that the perceptual system exaggerates looming sounds. Whether or not this discrepancy reflects different perceptual mechanisms for auditory movement than for loudness and duration judgments will have to be clarified by subsequent work.

We believe that, although different perceptual continua have been investigated in studies on time-reversed stimuli, the overall consistency of the results shows that what was actually being studied were different aspects of a common perceptual phenomenon. In this sense, there is a clear convergence of the results gathered for the different perceptual continua. The perceptual asymmetry (whether what is estimated in the experiment is timbre, loudness, change in loudness, or subjective duration) always has the same properties: Ramped sounds are more salient than damped ones (or, more correctly, damped sounds are less salient than ramped ones); there is greater perceptual asymmetry for steep than for shallow amplitude modulators; and there is a partial or total reduction of the asymmetry passing from tone to noise carriers. The only real inconsistency is the lack of reduction of the asymmetry when noise carriers are used, as Stecker and Hafter (2000) observed. However, their experiment is the only one in which the comparison between tones and noises was made between rather than within subjects. Furthermore, in that experiment the estimates collected for ramped and damped noises showed a large variability (see Figure 4 of Stecker & Hafter, 2000).

Two other properties seem to modify the magnitude of the perceptual asymmetry. The first was observed by Neuhoff (1998): The difference in change in loudness increases with the overall level of the sound. According to Teghtsoonian et al. (2005), however, this effect might be an artifact: Such an asymmetry could originate as Neuhoff's (1998) listeners evaluated the high-level end of ramped sounds *per se* rather than the change in level of the sound. Nonetheless, such a criticism does not explain why Neuhoff (1998, 2001)

observed the perceptual asymmetry with tones but not with noises. The second property is that the magnitude of the asymmetry depends on the temporal order of events within a trial. Stecker and Hafter (2000) reported that the difference in loudness between ramped and damped sounds was greater when the damped sound preceded the ramped sound (i.e., when the damped sound was presented as the first auditory event within a sequence) than vice versa. A similar effect has been observed twice for subjective duration. Grassi (2003) asked listeners to compare the duration of a ramped sound concatenated to a damped sound (or vice versa) and asked listeners to judge which portion of the two-stage sound was longer in duration. A reanalysis of Grassi's data shows that, in keeping with the findings of Stecker and Hafter, the subjective duration of the two sounds was more different when the damped sound preceded the ramped one. Furthermore, the outcome of a duration matching experiment (whose results are not reported here) showed that the asymmetry in subjective duration was greater when the ramped (or damped) sound preceded the adjustable steady sound than vice versa. Taken together, these results suggest that memory might play an important role in the subjective estimates of damped sounds: The longer the time elapsing before a judgment is made on a damped sound, the greater is the echo suppression effect.

There are currently two problems that require attention and remain unanswered. The first is the issue of why there is a reduction of asymmetry with noise carriers. Unfortunately, the present study does not suggest a solution. Nonetheless, Experiment 1 shows the origin of such a reduction: a larger underestimation of ramped noises than of ramped tones. The second problem was originally addressed by Stecker and Hafter (2000), who asked, What is an echo suppression mechanism good for? One hypothesis arises from the difference in predictability of the two stimuli. The time course of ramped sounds is unpredictable, whereas that of damped sounds is usually predictable because the sound will dissolve, at some point, into silence. The perceptual system may neglect events that are highly predictable to ensure efficient coding and to enable a prompter response to new events. In addition, the lack of salience of the end of an auditory stimulus is a well-known phenomenon in hearing perception. Physiological findings highlight a greater onset encoding relative to the offset encoding: The onset response is clearly visible, whereas the offset response is scarcely observed and often reported anecdotally (Phillips, Hall, & Boehnke, 2002). Furthermore, listeners are more accurate in making discriminations on onsets than on offsets (Zera & Green, 1993). The onset of an auditory event seems somehow more important than its offset (e.g., it influences the grouping of the frequency components of different sound sources—Darwin, 1984; Darwin & Sutherland, 1984) for auditory scene analysis. In summary, the experiments reported here provide some support for the involuntary suppression (or neglect) of the ends of sounds, especially when the stimulus (e.g., a damped sound) is highly predictable.

REFERENCES

- AKERROYD, M. A., & PATTERSON, R. D. (1995). Discrimination of wideband noises modulated by a temporally asymmetric function. *Journal of the Acoustical Society of America*, **98**, 2466-2474.
- BRESSAN, P. (2006). The place of white in a world of grays: A double-anchoring theory of lightness perception. *Psychological Review*, **113**, 526-553.
- CANÉVET, G. (1986). Estimation de sonie pour de sons purs à variation monotone de niveau. *Acustica*, **62**, 256-264.
- CANÉVET, G., & SCHARF, B. (1990). The loudness of sounds that increase and decrease continuously in level. *Journal of the Acoustical Society of America*, **88**, 2136-2142.
- CANÉVET, G., TEGHTSOONIAN, R., & TEGHTSOONIAN, M. (2003). A comparison of loudness change in signals that continuously rise and fall in amplitude. *Acta Acustica United With Acustica*, **89**, 339-345.
- DARWIN, C. J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Society of America*, **76**, 1636-1647.
- DARWIN, C. J., & SUTHERLAND, N. S. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic? *Quarterly Journal of Experimental Psychology*, **36A**, 193-208.
- DI LOLLO, V., ENNS, J. T., YANTIS, S., & DECHIEF, L. G. (2000). Response latencies to the onset and offset of visual stimuli. *Perception & Psychophysics*, **62**, 218-225.
- GARNER, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology*, **46**, 373-380.
- GAVER, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, **5**, 1-29.
- GILCHRIST, A., KOSSYFIDIS, C., BONATO, F., AGOSTINI, T., CATALIOTTI, J., LI, X., ET AL. (1999). An anchoring theory of lightness perception. *Psychological Review*, **106**, 795-834.
- GRASSI, M. (2003). Differenza nella durata percettiva di suoni crescenti o calanti in intensità: Permanenza o decurtamento? *Giornale Italiano di Psicologia*, **30**, 659-663.
- GRASSI, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & Psychophysics*, **67**, 274-284.
- IRINO, T., & PATTERSON, R. D. (1996). Temporal asymmetry in the auditory system. *Journal of the Acoustical Society of America*, **99**, 2316-2331.
- LEVITT, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.
- NEUHOFF, J. G. (1998). Perceptual bias for rising tones. *Nature*, **395**, 123-124.
- NEUHOFF, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, **13**, 87-110.
- PATTERSON, R. D. (1994a). The sound of a sinusoid: Spectral models. *Journal of the Acoustical Society of America*, **96**, 1409-1418.
- PATTERSON, R. D. (1994b). The sound of a sinusoid: Time-interval models. *Journal of the Acoustical Society of America*, **96**, 1419-1428.
- PATTERSON, R. D., ALLERHAND, M. H., & GIGUÈRE, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America*, **98**, 1890-1894.
- PATTERSON, R. D., & IRINO, T. (1998). Modeling temporal asymmetry in the auditory system. *Journal of the Acoustical Society of America*, **104**, 2967-2979.
- PHILLIPS, D. P., HALL, S. E., & BOEHNKE, S. E. (2002). Central auditory onset responses, and temporal asymmetries in auditory perception. *Hearing Research*, **167**, 192-205.
- SCHLAUCH, R. S. (1992). A cognitive influence on the loudness of tones that change continuously in level. *Journal of the Acoustical Society of America*, **92**, 758-765.
- SCHLAUCH, R. S., RIES, D. T., & DIGIOVANNI, J. J. (2001). Duration discrimination and subjective duration for ramped and damped sounds. *Journal of the Acoustical Society of America*, **109**, 2880-2887.
- SEIFRITZ, E., NEUHOFF, J. G., BILECEN, D., SCHEFFLER, K., MUSTOVIC, H., SCHÄCHINGER, H., ET AL. (2002). Neural processing of auditory looming in the human brain. *Current Biology*, **12**, 2147-2151.
- STECKER, G. C., & HAFTER, E. R. (2000). An effect of temporal asymmetry on loudness. *Journal of the Acoustical Society of America*, **107**, 3358-3368.
- TEGHTSOONIAN, R., TEGHTSOONIAN, M., & CANÉVET, G. (2000). The perception of waning signals: Decruitment in loudness and perceived size. *Perception & Psychophysics*, **62**, 637-646.
- TEGHTSOONIAN, R., TEGHTSOONIAN, M., & CANÉVET, G. (2005). Sweep-induced acceleration in loudness change and the "bias for rising intensities." *Perception & Psychophysics*, **67**, 699-712.
- WANG, X., LU, T., & LIANG, L. (2003). Cortical processing of temporal modulations. *Speech Communication*, **41**, 107-121.
- WRIGHT, B. A., BUONOMANO, D. V., MAHNCKE, H. W., & MERZENICH, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *Journal of Neuroscience*, **17**, 3956-3963.
- ZERA, J., & GREEN, D. M. (1993). Detecting temporal onset and offset asynchrony in multicomponent complexes. *Journal of the Acoustical Society of America*, **93**, 1038-1052.

NOTES

1. Neuhoff (1998, 2001) proposed that such a sensory bias toward sounds that increase in intensity has an evolutionary advantage, since it emphasizes the increase in intensity of steady sound sources that are approaching the listener. The fact that the perception of loudness increase is more marked for tones than for noise indicates to Neuhoff (2001) the system's preference for motion of biological sources. According to Neuhoff (2001), tones are normally generated by biological sources, whereas noises, although they can also be produced by biological sources, are usually the result of multiple sources sounding together. However, Neuhoff's (2001) basic finding may not be robust, since Canévet et al. (2003) played ramped and damped sounds identical to those used by Neuhoff (1998) and asked listeners to evaluate the initial and final loudness of the sounds. When the change in loudness is measured as the difference between initial and final loudness, results show no perceptual asymmetry between ramped and damped sounds.
2. The time course of the raised cosine modulator is similar to that of the custom modulator used by Stecker and Hafter (2000): It increases rapidly in level during the first half of the sound's duration and then increases more gradually.
3. Such a difference might be purely incidental. In fact, in an experiment (the results of which are not reported here) involving matching and RT tasks of similar duration, in which only raised cosine modulators were used, onset reactions to tonal and noise stimuli were identical.
4. Theoretically, in the present experiment the ends of ramped sounds are also predictable because these sounds invariably reach a fixed level of 65 dB SPL. In order to make this prediction, however, listeners would require the long-term storage of the final level of the sound. It is well-known that human memory for magnitude of stimulations within an intensive continuum (i.e., those continua in which the signal involves a change in stimulus energy—e.g., loudness, lightness) is rather poor (see Garner, 1953, for an example involving lightness). On the contrary, maximal or minimal (e.g., silence) stimulations can validly serve as anchors for estimations (see, e.g., Bressan, 2006, and Gilchrist et al., 1999).
5. The normalization for the steady sound duration prevents the calculation of excessively large measures because of the so-called onset advantage. In fact, offset RTs are usually longer than onset RTs (see, e.g., Di Lollo, Enns, Yantis, & Dechief, 2000). Moreover, this normalization reduces the influence of the anticipation hypothesized for offset RTs.
6. In the present study, the exponential modulator used in Experiment 1 produced envelopes with half-lives of 34, 69, 103, and 138 msec for the four respective durations.

(Manuscript received March 24, 2004;
revision accepted for publication February 16, 2006.)