Why are damped sounds perceived as shorter than ramped sounds?

Massimo Grassi¹ • Giovanna Mioni¹

Published online: 3 June 2020
© The Psychonomic Society, Inc. 2020

Abstract

Hearing is the most accurate sense for perceiving duration. However, rarely it produces inaccurate estimates of duration, for example when it compares the subjective duration of tones that are increasing in intensity over time (i.e., ramped) with that of tones that are decreasing in intensity over time (i.e., damped). The literature reports that the damped tones are perceived as much being shorter than the ramped tones of the same length. The short subjective duration of damped tones may originate from a decay suppression mechanism that parses the source-informative part of many natural sounds (i.e., the beginning) from the less informative part of them (the decay): listeners may interpret the tail of damped tones like an echo or like the decay portion of an impact sound and exclude it from the account of the duration of the tone. In the natural soundscape, the tail of sounds produced in reverberant environments and the tail of impact sounds have a frequency content that is constant throughout the sound's duration. Here, the carriers used for ramped and damped sounds were a tone constant in frequency and a tone modulated in frequency. The frequency modulation was introduced to prevent the listener from interpreting the tail of these tones as the result of reverberation or the decay portion of an impact sound. Frequency constant damped tones were largely underestimated in duration whereas frequency modulated ones were not (or were only slightly), demonstrating that the decay suppression mechanism is a worthy explanation for the short subjective duration of damped tones.

Keywords Psychoacoustics

Introduction

Time perception is an ubiquitous human experience, and hearing is, among the senses, the most accurate sense for perceiving time and obtaining a subjective estimate of the duration of an event (e.g., Grondin, 2010). However, in some case, hearing too can produce inaccurate estimations of duration. One such condition is when we compare the subjective duration of sounds that are increasing in intensity over time (i.e., ramped) or decreasing in intensity over time (i.e., damped). These types of intensity modulations are very common in our soundscape (e.g., see Gavers, 1993a, 1993b) but they have been rarely investigated in laboratory studies (e.g., Schutz & Vaisberg, 2014). Nonetheless, the research of the last two decades has revealed that the subjective duration of a damped

sound is much shorter than the subjective duration of the same sound when it is played backward in time (DiGiovanni & Schlauch, 2007; Grassi, 2010; Grassi & Darwin, 2006; Grassi & Pavan, 2012; Meunier, Vannier, Chatron, & Susini, 2014; Ries, Schlauch, & DiGiovanni, 2008; Schlauch, Ries, & DiGiovanni, 2001; Vallet, Shore, & Schutz, 2014). This subjective duration asymmetry has been replicated several times, in several conceptual replications, by different laboratories and with different types of sounds and durations. The difference in subjective duration between ramped and damped sounds is modulated by several factors. One such factor is overall duration: the asymmetry is larger for shorter sounds and smaller for longer sounds (e.g., Grassi & Darwin, 2006; Meunier et al., 2014). In addition, the asymmetry is larger when the ramp of the modulator is steeper (i.e., how fast the stimulus changes in decibels per second, Meunier et al., 2014). Finally, the asymmetry is larger for periodic sounds (such as tones) and smaller for aperiodic signals (such as noises, e.g., Grassi & Darwin, 2006; Schlauch et al., 2001). Noticeably, similar results can be observed if ramped and damped sounds are compared in loudness and change in loudness (Canévet, 1986; Canévet &

✉ Massimo Grassi
massimo.grassi@unipd.it

¹ Department of General Psychology, University of Padova, Via Venezia 8, 35131 Padova, Italy

Scharf, 1990; Canévet, Teghtsoonian, & Teghtsoonian, 2003; Neuhoﬀ, 1998, 2001; Olsen & Stevens, 2010; Olsen, Stevens, & Tardieu, 2010; Pastore & Flint, 2011; Ponsot, Susini, & Meunier, 2015; Ponsot, Susini, Saint Pierre, & Meunier, 2013; Schlauch, 1992; Stecker & Hafter, 2000; Susini, McAdams, & Smith, 2002, 2007; Teghtsoonian, Teghtsoonian, & Canévet, 2000, 2005).

Authors have suggested different explanations for this subjective duration asymmetry. The first is an evolutionary one. Ramped sounds are thought to be a salient sound stimulus in comparison to damped sounds. Ramped sounds simulate the proximal auditory pattern of a sound source approaching a listener (i.e., a looming sound), whereas damped sounds simulate the proximal pattern of a sound source moving away from the listener. In practice, there may be some inhibitory effect of damped sounds since, in nature, that might indicate the benign situation of a sound source moving away from the listeners, as opposed to a ramped sound potentially indicating a sound source approaching, thus resulting in a more alerting reaction (see Neuhoﬀ 1998, 2001, for a similar explanation in the context of the perceptual asymmetry in the change in loudness of the two types of sounds). For the same reason, damped sounds could be underestimated in duration in comparison to ramped sounds. However, according to this evolutionary perspective, it is unclear why ramped sounds are underestimated in duration in comparison to a sound that is steady in intensity over time (see Grassi & Darwin, 2006, to observe both results).

A second explanation emerged from the asymmetry in the onset and oﬀset encoding by the auditory system (see Grassi & Darwin, 2006; Grassi & Pavan, 2012). Audition is known to mark more strongly the onset than the oﬀset of auditory events (Philips, Hall, & Boehnke, 2002). In addition, the strength of the mark is a function of the rate of change in intensity of the onset and oﬀset ramp (Phillips et al., 2002; see also Deneux et al., 2016; Lu, Liang, & Wang, 2001; Zhang, et al., 2016). We may observe a subjective counterpart of this difference in loudness perception: the beginning of a sound is the portion of the sound that contributes most to the overall loudness of the sound (Dittrich & Oberfeld, 2009). In ramped sounds, both onset (because of the strong onset response of the auditory system) and oﬀset (because of the high intensity change at oﬀset) are sharply marked. Let's assume that the subjective duration of an auditory event is the difference between a mark signaling the beginning of the event and a mark signaling the end of the event. If we assume this, the subjective duration of a ramped sound should be, in theory, almost veridical and similar to the subjective duration of a sound that is steady in intensity over time. Results support this prediction: If we exclude portions that are too quiet to be perceived, the subjective duration of ramped sounds is veridical (i.e., similar to their physical duration) and very similar to the subjective duration of an intensity-steady sound (e.g., see

the results of Grassi & Darwin, 2006; Grassi & Pavan, 2012). In contrast, the above rationale returns a different outcome when it is applied to the encoding of damped sounds. The onset of damped sounds is sharply marked because of the high intensity-change onset of the sound and the strong onset response of the auditory system. In contrast, the oﬀset is not sharply marked: The weak oﬀset response of the auditory system may occur at any time when the intensity of the sound is low. In other words, there is no clear mark signaling the end of the sound and this mark could occur at any time when the sound is still barely audible, thus, before the physical end of the sound. Once again, if we assume that the subjective duration of an auditory event is the result of the difference between a mark signaling the beginning of the event and a mark signaling the end of the event, the subjective duration of a damped sound is likely to be shorter than its physical duration and quite variable because of the lack of a clear end-point that suggests the auditory system and the subsequent stages of elaboration where the sound ends. Both these results have been observed (e.g., Grassi & Darwin, 2006; Grassi & Pavan, 2012).

There is a third explanation for the asymmetry in subjective duration of ramped and damped sounds. This explanation, hereafter referred to as decay suppression, was originally suggested by Stecker and Hafter (2000) and successively investigated by DiGiovanni and Schlauch (2007). According to Stecker and Hafter (2000), a decay suppression mechanism could be active when listening to sounds such as the damped sounds. This mechanism would parse the sound into the beginning of the sound (informative about the sound source) and the tail of the sound (informative about the reverberation of the environment or the damping of the structure). In other words, listeners may judge the tail of the damped sounds like an echo or like the decay portion of an impact sound or both (Gaver, 1993a, 1993b). For example, the intensity envelope of impact sounds (such as the sound produced by a hammer striking an anvil) is made of two parts: the attack and the decay. The attack is that high-intensity portion of the sound that is generated when one object touches the other and both are set into vibration. The intensity decay occurs when the impact is over and its length depends on the damping properties of the objects involved in the impact (e.g., material and shape) as well as of the characteristics of the environment where the event occurs (that may add echo and further lengthen the duration of the sound). Indeed, listeners need only a few cycles of a sound wave to identify several characteristics of a sound source such as the pitch, the octave, or the identity of the sound source (Robinson & Patterson, 1995a, 1995b). DiGiovanni and Schlauch (2007) investigated this explanation and asked one group of participants to evaluate the duration of the sounds (like in previous experiments) and another group “to include the entire sound in your judgment of its duration.” With the latter instructions, the asymmetry in duration

between ramped and damped sounds was reduced. Indeed, the literature supports the idea that perceivers may spontaneously interpret damped sounds like an impact sound. For example, if damped sounds are presented together with a bistable visual stimulus that can be interpreted as a streaming or a bouncing (i.e., an event compatible with an impact sound), participants preferred the second type of interpretation (Grassi & Casco, 2009), an interpretation that is coherent with the type of sound-envelope they are listening to.

The present study further investigated the asymmetry in perceived duration between ramped and damped sounds by testing the decay-suppression explanation. Here, however, we worked on the stimulus rather than the instructions given to the participant. We do not know, in fact, how participants may interpret instructions that stress taking into account the “entire sound duration.” In the present experiment, we presented a damped sound that, because of its characteristics, could be spontaneously interpreted as the result of reverberation or as an impact sound (or both). Therefore, we used a type of damped sound that, in theory, should prevent the spontaneous decay suppression operated by the participant.

One of the characteristics of impact sounds is that the frequency content of the sound is constant throughout its duration. In other words, each portion of the sound has the same frequency content regardless of whether we are listening to the beginning of the sound (i.e., the attack), the middle of the sound, or the end of the sound (i.e., the decay; see Gaver 1993a, 1993b). By the same token, echo and reverberation just stretch in time the frequency content of the beginning of the sound. For example, in previous experiments, it is possible that listeners judged the tail of damped sounds like an echo because the frequency content of the sound did not change over the sound’s duration (i.e., as in natural impact sounds and/or when reverberation occurs). In the present study ramped and damped sounds were modulated in amplitude (like the ramped and the damped sounds used in previous studies) but also modulated in frequency over time (unlike previous studies, and, above all, unlike the impact sounds we are used to listening to in our soundscape). This manipulation may change the way listeners interpret these sounds, in particular when they are damped in amplitude. In addition, and noticeably, the frequency modulation should have a minimal effect on the forward masking pattern of the ramped sounds as well as on the onset and offset marks that signal the beginning and the end of the ramped and damped sounds. Therefore, if any of these predictions are correct, we should not expect a reduction of the subjective duration asymmetry between ramped and damped sounds. Furthermore, because it is well known that the overall duration of an event affects and interacts with its subjective duration, two target durations were investigated: 300 and 1,200 ms. Although it is yet unclear how humans estimate durations, it is clear that the estimation of relatively long stimuli (i.e., from and over 1.2 s) can

rely on strategies (e.g., counting) that are useless at shorter durations (Grondin, Meilleur-Wells, & Lachance, 1999; Grondin, Ouellet, & Roussel, 2004; Mioni, Stablum, & Grondin, 2014).

Although the present experiment was directly linked to issues related to auditory time perception, its outcome may shed some light on the way we analyze the auditory scene (Bregman, 1990). In particular, it may reveal whether there is a spontaneous “decay suppression” mechanism that rapidly interprets the amplitude envelope and the spectral content of a sound and actively ignores the tail of sounds damped in amplitude when the frequency content of the sound does not change over time. Here, as well as in previous experiments, sounds were presented via headphones and the sounds delivered include no reverberation. If the short subjective duration of damped sounds was due to a decay suppression, results may reveal a quick mechanism that acts like a figure-ground segregation by the auditory system, a figure-ground segregation that occurs immediately and spontaneously, such as, for example, when we observe classic bistable figures like the Rubin vase. In fact, one of the striking facts about the asymmetry in subjective duration between ramped and damped sounds is that the asymmetry is perceptually immediate and clear.

Method

Participants

Participants were recruited among the students of the University of Padova. Although the literature occasionally reports sex differences in the perception of ramped and damped sounds (e.g., Grassi, 2010; Neuhoff, Planisek, & Seifritz, 2009), this factor was not taken into account for recruiting the participants. As far as the asymmetry in duration is concerned, it seems that the sex difference, if existent, is small to negligible and/or may be partly due to the specific method used to evaluate the ramped and the damped sounds. With the adjustment method (see below) we never observed a sex difference (e.g., Grassi & Darwin, 2006; Grassi & Pavan, 2012).

The number of the participants was calculated starting from the data of Experiment 1 of Grassi and Pavan (2012).¹ We estimated the number of participants with a conservative approach. We calculated the effect size starting from the difference in subjective duration observed in Grassi and Pavan (2012) in the auditory condition between the 1,000-ms steady

¹ Noticeably, in that study, in Experiment 2, for target sounds of 1 s of duration, authors did not observe a difference in subjective duration between damped and steady targets when the dynamic range of the target was 20 dB. For targets of 1 s, the difference was observed only when the dynamic range was 40 dB.

sound and the ramped sounds (40 dB dynamic range condition). This difference was the smallest difference in subjective duration between any two target sounds observed in those data. For targets of 1 s duration, the subjective duration of the steady target was 1,026 ms ($SD = 58$ ms) whereas the subjective duration of the damped target was 870 ms ($SD = 100$ ms). The effect size of this difference resulted in $d=1.68$. The lower limit of the 95% confidence intervals of this difference is equal to $d=1.00$. If we take this lower limit as representative of a possible difference, a power of 90%, and an alpha level of 0.05/45 (i.e., assuming comparison with only paired-sample t-tests of all the subjective durations of all the target sounds that will be presented in the current experiment) the resulting $N = 25.69$, which we rounded up to 30.

The code and the data we used to calculate power are available from: https://osf.io/bg49u/?view_only=e1aea4fcc5eb4784b9dd98426918f547

Thus, 30 students (16 males) participated in the experiment (age range, 18–28 years, $M=22$ years). All participants passed an audiometric screening test assessing the ability to detect tones of 30 dB HL for frequencies 250, 500, 1,000, 1,500, and 2,000 Hz. In addition, none reported being aware of hearing problems.

Apparatus

The experiment was written in Octave using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) and the PSYCHOACOUSTICS extensions (Soranzo & Grassi, 2014). The software was implemented with an ASUS computer (Cpu Intel i5 650 3.20 GHz, Motherboard Asus P7H55-V RAM 4 GB, Graphic Card AMD Radeon HD 5700 Series, OS Windows 7 Professional 64 bit). The computer was connected to a monitor (NEC MultiSync FE950p) and M-AUDIO FastTrack Pro sound card. The output of the sound card was delivered to a pair of Sennheiser HDA 300 headphones. The audiometer for the audiometric screening of the participants was an Interacoustics AD229b. Audiometric screening and experiment were run inside a single walled IAC sound proof booth. The code for running the experiment is available from:

https://osf.io/bg49u/?view_only=e1aea4fcc5eb4784b9dd98426918f547

Stimuli and procedure

Each participant took two blocks. In one block s/he was asked to match the duration of the 300-ms target, in the other s/he was asked to match the duration of the 1,200-ms target. The blocks' order was counterbalanced across the participants. In each block, the participant was asked to match the duration of a duration-adjustable amplitude-steady noise to that of a fixed-duration amplitude-ramped, amplitude-damped or amplitude-

steady target. A trial consisted of the fixed-duration target stimulus followed after 500 ms by the duration-adjustable stationary noise, the duration of which was to be matched by listeners to that of the target. The duration of the amplitude-steady noise could be manipulated with the mouse. At the beginning of each trial, the duration of the adjustable amplitude steady noise was set randomly within a range of $\pm 80\%$ of the duration of the target stimulus (this range was also the maximum permitted adjustment during each experimental trial). The target-adjustable pair was then presented to the participant. At this point, the participant could manipulate the duration of the adjustable stimulus with the mouse. A mouse-click in the upper half of the screen incremented the duration of the adjustable stimulus whereas a click in the lower half of the screen decremented the duration of the adjustable stimulus. After the click, the fixed-duration target was presented to the participant followed by the duration-adjusted stimulus, so that the participant could assess whether target and adjustable stimulus were similar in duration. If the participant was satisfied, s/he could save the trial and move to the next trial. If not, this process was repeated until the participant was satisfied. Within each experimental block participants matched the duration of 50 targets (i.e., five target types each presented ten times during the experiment; see below for the target-type description) presented in random order. Participants took the experimental blocks in a counterbalanced order. Each block lasted about 50 min, totaling almost 2 h of experiment for each participant. The first experimental block taken by the participant was preceded by a short session where the participant could familiarize him/herself with the procedure. Data collected in this session were not saved. The adjustment method used returns a direct estimate of the participant's point of subjective equality (PSE). The reader can find a video example of one trial from:

https://osf.io/bg49u/?view_only=e1aea4fcc5eb4784b9dd98426918f547

Instructions given to the participant

Because instructions seem to be able to modulate the subjective duration of damped sounds, in the present experiment we used an Italian translation and adaptation of the instructions given by DiGiovanni and Schlauch (2007). In particular, the instructions were to “match duration as best as possible” but did not mention “to include the entire sound in your judgment of its duration.” Here we report the text in English: “You will hear two sounds played one after the other through the headphones. The sounds will be separated by brief silent period. Your task is to match the duration of the two sounds as closely as possible. In order to match the duration of the two sounds you will use the computer's mouse that will increase or decrease the duration of the second of the two sounds. During a matching session, one sound will always remain fixed in

duration while the second is adjustable. Click the upper part of the screen to make the adjustable sound longer and the lower part of the screen to make the sound shorter. If you click along the midline of the screen you can listen to sound without making it longer or shorter. Please consider the whole range of adjustments available to you and bracket your adjustments several times around the point of equal duration before making your final decision. For example, click the mouse in the very upper part of the screen or in the very lower part of the screen, and listen to the sounds at each of these points. Then move and click the mouse several times above and below the point of equal duration, listening to the results of each adjustment. Finally, move and click the mouse to the point where the duration of the two sounds is equal. When the mouse-click is at the point of equal duration for both sounds, press the “OK” area on the right to save your result and move to the next trial.” The Italian version of the instructions is available from:

https://osf.io/bg49u/?view_only=e1aea4fcc5eb4784b9dd98426918f547

Target sounds

Target sounds were of five types. In each block there were five target sounds: two ramped, tonal complexes, two damped, tonal complexes, and one amplitude-steady noise. This last target served as a control sound to evaluate the subjective duration of remaining target sounds. Ramped and damped sounds were obtained by modulating in amplitude two types of carriers. The first carrier was a harmonic complex frequency sweep starting from a fundamental frequency of 250 Hz and modulated exponentially in frequency over its duration to arrive at a fundamental frequency of 500 Hz. All remaining harmonics (i.e., 500, 750, 1,000, 1,250 Hz) were added in phase, had the same amplitude of the fundamental frequency and were also modulated in frequency like the fundamental frequency. This carrier is referred to as “sweep” in the rest of the paper. The second type of carrier was a five-harmonics complex tone with a fundamental frequency of 353.5 Hz (i.e., six semitones higher/lower than the starting/ending frequency of fundamental of the frequency sweep carrier) and upper partials at 707.1, 1,060.6, 1,414.21, and 1,767.7 Hz. This carrier is referred to as “fixed” in the rest of the paper. The two carriers were modulated in amplitude with an exponential ramp spanning 40 dB over the duration of the sound. The ramp was “increasing” to obtain the ramped sounds or “decreasing” to obtain the damped sounds. The carrier for the amplitude steady target was white noise generated from a uniform distribution. All targets were onset and offset modulated by two raised cosine ramps of 10 ms to avoid onset and offset clicks. Sounds were synthesized at a sample rate of 44,100 Hz and 16-bit resolutions. Target sounds

were of two durations: 300 ms and 1,200 ms. In the experiment, the peak level of ramped and damped target was 80 dB SPL. The RMS level of the steady noise was 60 dB SPL. One example of each target sound used in the experiment is available from: https://osf.io/bg49u/?view_only=e1aea4fcc5eb4784b9dd98426918f547

Data analysis

Two main analyses were performed, on two slightly different dependent measures. In the first one, predictors were the type of target sound (five levels) and the target duration (two levels) and the analysis was performed on the estimation error (E%) for each target sound. E% is the percent difference between the average participants’ PSE versus the physical duration of the target sound. To calculate the E%, no outliers were removed.

A 2 (durations) $\times$ 5 (target type) within-subjects, repeated-measures ANOVA was calculated on E%. Successively, individual comparisons between each of the target sounds for a given target duration was performed by means of ten paired-samples t-tests (see Tables 1 and 2 respectively for the 300 ms target and the 1200 ms target). The alpha level of these t-tests was Bonferroni corrected.

In the second analysis, the PSE of the participant for ramped and damped targets was normalized by the PSE of the control, stationary noise. This index (En%) was used to take into account possible individual differences among the skills of listeners to judge durations that may be a source of unpredictable individual variability. En% is the percent of under- or overestimation of the ramped (or damped) target in comparison to the PSE of the stationary target. Positive values of En% indicate that the stimulus was overestimated in comparison to the subjective duration of the stationary-control target, whereas negative values indicate that the stimulus was underestimated in duration in comparison to the subjective duration of the stationary-control target. In this second analysis, En% was used to calculate a 2 (durations) $\times$ 2 (target types) $\times$ 2 (carriers) within-subjects, repeated-measures ANOVA. Successively, individual comparisons between each of the amplitude-modulated targets for a given target duration were performed by means of six paired-samples t-tests. The alpha level of these t-tests was Bonferroni corrected (see Tables 3 and 4 for respectively for the 300 ms target and the 1200 ms target).

Results

The 2 (durations) $\times$ 5 (target type) within-subjects, repeated-measures ANOVA was calculated on E%. The En% gathered for each target sound is represented in Fig. 1. The main effect of duration [$F(1, 116) = 4.46, p = .043, \eta^2 = .13$] and target type [$F(4, 116) = 41.11, p < .001, \eta^2 = .59$] were significant. The interaction was also significant: $F(4, 116) = 4.57, p = .002$,

Table 1 *t*-tests comparing the E% gathered for the 300-ms duration target sounds used in the experiment. The *p*-values are Bonferroni corrected for the number of tests (i.e., ten)

300 ms Df=29	Damped fixed	Damped sweep	Ramped fixed	Ramped sweep
Control	$t=11.18, p<.001, d=2.47$	$t=3.78, p=.007, d=0.86$	$t=1.89, p=.682, d=0.45$	$t=1.60, p=.99, d=0.37$
Damped fixed		$t=6.33, p<.001, d=0.81$	$t=8.39, p<.001, d=1.28$	$t=7.95, p<.001, d=1.19$
Damped sweep			$t=2.72, p=.1, d=0.35$	$t=3.33, p=.023, d=0.34$
Ramped fixed				$t=0.16, p=.99, d=0.16$

$\eta^2 = .14$. We also compared the E% of each target, separately for each target duration, with ten Bonferroni-corrected paired-sample *t*-tests. The results of these tests are reported in Tables 1 and 2 and the E% gathered for each target sound are represented in Fig. 1.

Successively, we calculated the 2 (durations) $\times$ 2 (envelope type) $\times$ 2 (carrier type) within-subjects, repeated-measures ANOVA on En%. The ANOVA returned the following results. The main effect of duration was significant, $F(1, 29) = 4.58, p < .041, \eta^2 = .14$. The main effect of the envelope type was also significant: $F(1, 29) = 68.03, p < .001, \eta^2 = .70$ and the main effect of carrier type was also significant: $F(1, 29) = 35.61, p < .001, \eta^2 = .55$. As far as interactions are concerned, the interaction between target duration and envelope type was significant [$F(1, 29) = 6.55, p < .016, \eta^2 = .18$] and also the interaction between envelope type and carrier type was significant, $F(1, 29) = 58.52, p < .001, \eta^2 = .67$, but not the interaction between carrier type and target duration: $F(1, 29) = .059, p = .810, \eta^2 = .01$. The three-way interaction was not significant $F(1, 29) = .057, p = .813, \eta^2 = .01$. Here too we compared the En% of each target, separately for each target duration, with six Bonferroni-corrected paired-sample *t*-tests (see Table 3 and 4). Data, analysis scripts, and scripts to generate the figures are available from:

https://osf.io/bg49u/?view_only=e1aea4fcc5eb4784b9dd98426918f547

Discussion

In the current study, we investigated why damped sounds are perceived as shorter than ramped sounds. Although this

asymmetry has been investigated extensively over the past two decades, its origin remains unclear. Here, we investigated the explanation originally suggested by Stecker and Hafter (2000) and successively expounded on by Schlauch and colleagues (DiGiovanni & Schlauch, 2007; Ries et al., 2008; Schlauch et al., 2001). According to Stecker and Hafter (2000), when we listen to a sound that is constant in the spectral content but that is damped in amplitude over time, we interpret the tail of the sound like an echo or a reverb and we ignore this portion of the sound. Because listeners ignore the tail of the sound (Stecker & Hafter, 2000, mention a possible decay suppression mechanism) damped sounds are perceived as shorter than ramped ones. For the same reason, damped sounds are also perceived as softer (Stecker & Hafter, 2000). We may discuss whether this explanation extends to the perception of change in loudness (ramped sounds that are perceived as extending over a larger dynamic range than damped sounds, e.g., Neuhoof, 1998), but the amplitude modulation used in those studies is substantially smaller than those used here and in studies comparing duration of loudness.

In the current study, listeners overestimated slightly (10%) the duration of the control sound at 300 ms and they estimated its duration veridically at 1,200 ms. As far as the other stimuli are concerned, the subjective duration of the ramped sound is only slightly shorter (i.e., ~10%) than that of the control sound, regardless of the carrier timbre and the carrier duration. In contrast, the subjective duration of the damped fixed tone was much shorter than its physical duration and much shorter than the subjective duration of all remaining sounds. This underestimation was slightly smaller for the 1,200-ms carrier.

Table 2 *t*-tests comparing the E% gathered for the 1,200-ms duration target sounds used in the experiment. The *p*-values are Bonferroni corrected for the number of tests (i.e., ten)

1,200 ms Df=29	Damped fixed	Damped sweep	Ramped fixed	Ramped sweep
Control	$t=6.59, p<0.001, d=1.52$	$t=2.67, p=.121, d=0.63$	$t=1.52, p=.99, d=0.35$	$t=2.49, p=.186, d=0.53$
Damped fixed		$t=-5.63, p<.001, d=0.98$	$t=6.76, p<.001, d=0.99$	$t=5.80, p<.001, d=1.03$
Damped sweep			$t=0.76, p=.99, d=0.12$	$t=0.40, p=.99, d=0.06$
Ramped fixed				$t=0.49, p=.99, d=0.06$

Table 3 t-tests comparing the En% gathered for the 300-ms target sounds used in the experiment. The p-values are Bonferroni corrected for the number of tests (i.e., six)

300 ms Df=29	Damped sweep	Ramped fixed	Ramped sweep
Damped fixed	$t=6.46, p<.001, d=0.89$	$t=7.98, p<.001, d=1.35$	$t=8.16, p<.001, d=1.30$
Damped sweep		$t=2.83, p=.049, d=0.35$	$t=3.24, p=.049, d=0.38$
Ramped fixed			$t=0.06, p=.99, d=0.01$

Finally, the subjective duration of the damped sweep was just slightly underestimated with the shortest carrier and not underestimated with the longest carrier. Overall, a higher similarity between subjective duration and objective duration of sounds was observed for the longer carrier than for the shorter carrier. Along the same line, smaller differences in subjective durations among the stimuli were observed at the longer duration, a result that may be explained by the persistence of excitation. The persistence of excitation may have contributed to the smaller difference for the shorter duration because any persistence that exists represents a smaller percentage of a longer stimulus (the contribution of persistence has been explored in prior studies using forward masking; Ries et al., 2008). In brief, the present results largely support the decay suppression explanation (Stecker & Hafter, 2000). Let's revisit the question posed by Stecker and Hafter 20 years ago. What is a decay suppression mechanism good for? Let's begin with a simple observation: damped fixed tones do sound short in duration and damped sweep tones do sound longer. Noticeably (and in addition), if we listen to the damped sweep tone we hardly perceive any modulation in amplitude, despite a 40 dB reduction in level over the duration of the sound. All these impressions are strong and immediate. The reason why damped fixed tones are perceived as being short (and damped sweeps tones are not) lays, we believe, is in the way humans perceive auditory objects. In everyday listening, auditory objects coincide with the sound source: we listen to the sound of cars, footsteps, the human voice and its components (words, syllables, phonemes) and so on (Gaver, 1993a). In our soundscape, there is a large class of sounds that is similar, acoustically speaking, to a damped fixed tone. In the *Introduction* we mentioned impact sounds but also any impulsive-like sound that is produced into an acoustically reverberant environment will look like a damped sound

(e.g., a gunshot). Altogether, these sounds have short-time spectra that largely coincide with the long-term spectrum of the sound. In other words, the frequency content of these sounds does not change over time, only amplitude does, and it does so in a highly predictable way: the sound dissolves into silence. In order to recognize the auditory object behind the sound, humans extract properties (often invariant) that enable us to attribute the sound to a specific sound source: the frequency content of the sound or its timbre but also some acoustical characteristics of the attack of the sound (Giordano, Rocchesso, & McAdams, 2010). The extraction of these properties is rather fast. For example, we need only a few cycles of a sound wave in order to recognize the pitch of a tone and its timbre (Robinson & Patterson, 1995a, 1995b). The rest of the acoustic information carried by the sound is redundant and can be ignored in a sort of auditory amodal competition: the object does not change even if a substantial portion of it, although available, can be ignored. Of course, we need to speculate that this decay suppression mechanism is automatic. In other words, it is active when we listen to our everyday natural soundscape as well as when our participants listen to the duration of the synthetic tones used in our study. We also need to assume that our perception of auditory objects is driven by fast-mapping and predictive mechanisms, mechanisms that rapidly extract the acoustical properties of the incoming wave and that rapidly make predictions on whether the processing of the sound can be interrupted (and part of the acoustic information ignored) because the invariant characteristics of it are already encoded. In contrast, when we listen to a damped sweep tone, the decay suppression mechanisms do not act because the sound is not constant in frequency (it cannot be classified as an impact sound or an impulsive sound produced into a reverberant environment) and we listen to the tone over its entire duration. Note that the amplitude

Table 4 t-tests comparing the En% gathered for the 1,200-ms target sounds used in the experiment. The p-values are Bonferroni corrected for the number of tests (i.e., six)

1,200 ms Df=29	Damped sweep	Ramped fixed	Ramped sweep
Damped fixed	$t=5.68, p<.001, d=1.03$	$t=6.75, p<.001, d=1.05$	$t=5.73, p<.001, d=1.10$
Damped sweep		$t=0.72, p=.99, d=0.11$	$t=0.33, p=.99, d=0.05$
Ramped fixed			$t=0.52, p=.99, d=0.07$

Fig. 1 Adjustment percent error (E%, top graphs) and adjustment percent error normalized to the control sound adjustment (En%, bottom graphs) for the target sounds used in the experiment. In the graph, “D” represents damped targets (either fixed or sweep), “R” represents ramped target (either fixed or sweep), and “Control” is the noise steady in intensity over time. In each box, the midline is the median. The edges of the box are the 25th and 75th percentiles. The whiskers are the interquartile range (i.e., Q3–Q1) augmented by 50% and compressed to the first real data

point close to the 50% augmented value. Plus symbols are outliers. On each box we also show the corresponding mean (i.e., the cross symbol) and the corresponding ± 1 standard error of the mean. On the top graph, the horizontal dashed line corresponds to no error (i.e., veridical estimation) in the subjective duration of the target. In the bottom graph, the line represents no difference between the subjective duration of the target sound and the control target sound

modulation of sounds is not, per se, a characteristic that can be ignored. In several contexts, the amplitude modulation enables us to recognize and distinguish auditory objects, such as when we distinguish the sound of a bottle bouncing on the floor or breaking on the floor (Warren & Verbrugge, 1984).

The decay suppression mechanism may explain the underestimation of the damped fixed tones (and the no underestimation of damped sweep) at the longest duration. In contrast, at the shortest duration, the underestimation of the damped tone is overall larger than at the longer duration and, in particular, the damped sweep is slightly underestimated in duration in comparison to the ramped fixed tone and the ramped sweep. Therefore, at the shortest duration, the decay suppression mechanism does not explain the entire subjective

duration of the damped tones. We cannot exclude the possibility that, at short durations, the periphery of the auditory system further shortens the duration of damped sounds and, as a consequence, that of the damped sweep. Models of the periphery of the auditory system (Auditory Image Model, AIM model; Patterson, Allerhand, & Giguère, 1995) do predict a small durational difference between ramped and damped sounds when the carrier is short (i.e., much shorter than 300 ms). Perhaps, the small underestimation of damped sweeps at the shortest duration has its origin in the periphery of the auditory system. Further investigations are needed to explore the possible double origin of the short subjective duration of damped sounds when the overall duration of the carrier is short.

In sum, the present study expands on the origin of the difference in subjective duration between ramped and damped sounds. The origin can be found in a decay suppression mechanism that shortens the perception of damped tones that have constant frequency in comparison to the same sound reversed in time.

Acknowledgements The information in this article and the article itself has not been published either electronically or in print previously. There are no financial or other relationships that could be interpreted as a conflict of interest with regard to this article. This research received no specific grant from any funding agency, commercial or not-for-profit sectors. This work was carried out within the scope of the project “use-inspired basic research,” for which the Department of General Psychology of the University of Padova has been recognized as “Dipartimento di Eccellenza” by the Ministry of University and Research. We thank the reviewers for their time, effort, and helpful comments.

Data Availability Statement Digital and supporting materials are available from: https://osf.io/bg49u/?view_only=e1aea4fcc5eb4784b9dd98426918f547

The present study works within an “open everything” framework: data, statistical analysis scripts, and experiment scripts are shared and available to the scientific community.

References

- Brainard, D. H. (1997). *The Psychophysics Toolbox*. Spatial Vision, 10, 433–436.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- 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.
- Deneux, T., Kempf, A., Daret, A., Ponsot, E., & Bathellier, B. (2016). Temporal asymmetries in auditory coding and perception reflect multi-layered nonlinearities. *Nature communications*, 7, 12682.
- DiGiovanni, J. J., & Schlauch, R. S. (2007). Mechanisms responsible for differences in perceived duration for rising-intensity and falling-intensity sounds. *Ecological Psychology*, 19, 239–264. doi:<https://doi.org/10.1080/10407410701432329>
- Dittrich, K., & Oberfeld, D. (2009). A comparison of the temporal weighting of annoyance and loudness. *The Journal of the Acoustical Society of America*, 126(6), 3168–3178. doi:<https://doi.org/10.1121/1.3238233>
- Gaver, W. W. (1993a). How do we hear the world? Explanations in ecological acoustics. *Ecological Psychology*, 5, 285–313. doi:https://doi.org/10.1207/s15326969eco0504_2
- Gaver, W. W. (1993b). What in the word do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5(1), 1–29. doi:https://doi.org/10.1207/s15326969eco0501_1
- Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of Experimental Psychology: Human Perception & Performance*, 36(2), 462–476.
- Grassi, M. (2010). Sex difference in subjective duration of looming and receding sounds. *Perception*, 39, 1424–1426. doi:<https://doi.org/10.1068/p6810>
- Grassi, M., & Casco, C. (2009). Audiovisual bounce-inducing effect: Attention alone does not explain why the discs are bouncing. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 235–243. doi:<https://doi.org/10.1037/a0013031>
- Grassi, M., & Darwin, C. J. (2006). The subjective duration of ramped and damped sounds. *Perception & Psychophysics*, 68, 1382–1392. doi:<https://doi.org/10.3758/BF03193737>
- Grassi, M., & Pavan, A. (2012). The subjective duration of audiovisual looming and receding stimuli. *Attention, Perception, & Psychophysics*, 74(6), 1321–1333. doi:<https://doi.org/10.3758/s13414-012-0324-x>
- Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception & Psychophysics*, 72, 561–582. doi:<https://doi.org/10.3758/APP.72.3.561>
- Grondin, S., Meilleur-Wells, G., & Lachance, R. (1999). When to start explicit counting in a time-intervals discrimination task: A critical point in the timing process of humans. *Journal of Experimental Psychology: Human Perception & Performance*, 25, 993–1004. doi:<https://doi.org/10.1037/0096-1523.25.4.993>
- Grondin, S., Ouellet, B., & Roussel, M.-E. (2004). Benefits and limits of explicit counting for discriminating temporal intervals. *Canadian Journal of Experimental Psychology*, 58, 1–12. doi:<https://doi.org/10.1037/h0087436>
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, 36, 1–16. doi:<https://doi.org/10.1068/v070821>
- Lu, T., Liang, L., & Wang, X. (2001). Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. *Journal of Neurophysiology*, 85(6), 2364–2380. doi:<https://doi.org/10.1152/jn.2001.85.6.2364>
- Meunier, S., Vannier, M., Chatron, J., & Susini, P. (2014). Asymmetry in perceived duration between up-ramp and down-ramp sounds as a function of duration. *The Journal of the Acoustical Society of America*, 136(2), EL166–EL172. doi:<https://doi.org/10.1121/1.4890175>
- Mioni, G., Stablum, F., & Grondin, S. (2014). Interval discrimination across different duration ranges with a look at spatial compatibility and context effects. *Frontiers in Psychology*, 5, 717. doi:<https://doi.org/10.3389/fpsyg.2014.00717>
- 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. doi:https://doi.org/10.1207/S15326969ECO1302_2
- Neuhoff, J. G., Planisek, R., & Seifritz, E. (2009). Adaptive sex differences in auditory motion perception: Looming sounds are special. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 225–234. doi:<https://doi.org/10.1037/a0013159>
- Olsen, K. N., & Stevens, C. J. (2010). Perceptual overestimation of rising intensity: is stimulus continuity necessary? *Perception*, 39, 695–704. doi:<https://doi.org/10.1068/p6592>
- Olsen, K. N., Stevens, C. J., & Tardieu, J. (2010). Loudness change in Response to dynamic acoustic intensity. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1631–1644. doi:<https://doi.org/10.1037/a0018389>
- Pastore, R. E., & Flint, J. (2011). Magnitude judgments of loudness change for discrete, dynamic, and hybrid stimuli. *Attention, Perception & Psychophysics*, 73, 886–907. doi:<https://doi.org/10.3758/s13414-010-0056-8>
- 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.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. doi:<https://doi.org/10.1163/156856897X00366>
- 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. doi:[https://doi.org/10.1016/S0378-5955\(02\)00393-3](https://doi.org/10.1016/S0378-5955(02)00393-3)
- Ponsot, E., Susini, P., Saint Pierre, G., & Meunier, S. (2013). Temporal loudness weights for sounds with increasing and decreasing intensity profiles. *The Journal of the Acoustical Society of America*, 134(4), EL321–EL326. doi:<https://doi.org/10.1121/1.4819184>
- Ponsot, E., Susini, P., & Meunier, S. (2015). A robust asymmetry in loudness between rising-and falling-intensity tones. *Attention, Perception, & Psychophysics*, 77(3), 907–920. doi:<https://doi.org/10.3758/s13414-014-0824-y>
- Ries, D. T., Schlauch, R. S., & DiGiovanni, J. J. (2008). The role of temporal-masking patterns in the determination of subjective duration and loudness for ramped and damped sounds. *The Journal of the Acoustical Society of America*, 124(6), 3772–3783. doi:<https://doi.org/10.1121/1.2999342>
- Robinson, K., & Patterson, R. D. (1995a). The duration required to identify the instrument, the octave, or the pitch chroma of a musical note. *Music Perception: An Interdisciplinary Journal*, 13(1), 1–15. doi:<https://doi.org/10.2307/40285682>
- Robinson, K., & Patterson, R. D. (1995b). The stimulus duration required to identify vowels, their octave, and their pitch chroma. *The Journal of the Acoustical Society of America*, 98(4), 1858–1865. doi:<https://doi.org/10.1121/1.414405>
- 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. doi:<https://doi.org/10.1121/1.403999>
- 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. doi:<https://doi.org/10.1121/1.1372913>
- Schutz, M., & Vaisberg, J. M. (2014). Surveying the temporal structure of sounds used in Music Perception. *Music Perception: An Interdisciplinary Journal*, 31(3), 288–296. doi:<https://doi.org/10.1525/mp.2014.31.3.288>
- Soranzo, A., & Grassi, M. (2014). PSYCHOACOUSTICS: a comprehensive MATLAB toolbox for auditory testing. *Frontiers in Psychology*, 5, 712. doi:<https://doi.org/10.3389/fpsyg.2014.00712>
- Stecker, G. C., & Hafer, E. R. (2000). An effect of temporal asymmetry in loudness. *Journal of the Acoustical Society of America*, 107, 3358–3368. doi:<https://doi.org/10.1121/1.429407>
- Susini, P., McAdams, S., & Smith, B. K. (2002). Global and continuous loudness estimation of time-varying levels. *Acta Acustica United with Acustica*, 88, 536–548.
- Susini, P., McAdams, S., & Smith, B. K. (2007). Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings. *Acta Acustica United with Acustica*, 93, 623–631.
- Teghtsoonian, R., Teghtsoonian, M., & Canevet, 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.
- Vallet, G. T., Shore, D. I., & Schutz, M. (2014). Exploring the role of the amplitude envelope in duration estimation. *Perception*, 43(7), 616–630. doi:<https://doi.org/10.1068/p7656>
- Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception & Performance*, 10(5), 704–712.
- Zhang, Y., Cheng, B., Koerner, T. K., Schlauch, R. S., Tanaka, K., Kawakatsu, M., ... Imada, T. (2016). Perceptual temporal asymmetry associated with distinct on and off responses to time-varying sounds with rising versus falling intensity: A magnetoencephalography study. *Brain Sciences*, 6(3), 27. doi:<https://doi.org/10.3390/brainsci6030027>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.