Exploring the role of the amplitude envelope in duration estimation

Guillaume T Vallet[1], David I Shore[2], Michael Schutz[3]§
[1] Multisensory Perception Laboratory, Department of Psychology, Neuroscience and Behaviour, McMaster University, Canada; also Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM), University of Montreal, Canada; [2] Multisensory Perception Laboratory, Department of Psychology, Neuroscience and Behaviour, McMaster University, Canada; [3] Music, Acoustics, Perception and LEarning (MAPLE) Laboratory, School of the Arts, McMaster University, 280 Main Street West, Hamilton, ON L8S 4K1, Canada; e-mail: schutz@mcmaster.ca

Abstract. A sound's duration provides important information about the event producing it. Although many of the sounds we hear every day are 'percussive' in nature (ie resulting from two objects impacting) and therefore exhibit decaying/damped amplitude envelopes, perceptual experiments frequently use tones synthesized with 'flat' or abruptly ending envelopes. Such sounds afford an estimation strategy involving calculating the elapsed time between tone onset and offset—a strategy that would be problematic for ecologically pervasive decaying sounds. Here we compare duration judgments for tones with percussive (ie gradually decaying) and flat (ie abruptly ending) amplitude envelopes, finding evidence for the use of different strategies. This result is discussed in terms of its implications for dominant theories and models of sensory perception that are often assessed using artificial sounds (ie 'flat tones') affording strategies that may not be optimal or even available for everyday listening.

Keywords: amplitude envelope, duration perception, timing, audition, temporal structure

1 Introduction

Imagine being awakened by a suspicious noise in the night: is it an intruder breaking and entering, or a beloved pet jumping from couch to table? Sounds are instructive with respect to understanding our environment, and part of their interpretation rests on duration—a property providing crucial information regarding the materials and actions involved in physical events. As such, our ability to judge relative duration represents an important skill in auditory processing (Grondin, 2010). Here we explore the role of the offset component of a sound's amplitude envelope on duration estimation—specifically, whether the amplitude envelope includes a naturally decaying versus abruptly ending offset. In doing so we complement previous research on differences in duration perception (Grassi & Darwin, 2006; Schlauch, Ries, & DiGiovanni, 2001) by exploring differences in the *underlying strategies* used in duration perception for sounds with natural versus artificial temporal structures. Together, these experiments raise interesting questions about the degree to which research using sounds with artificial temporal structures (offering envelope-specific affordances) generalizes to everyday listening.

The term 'amplitude envelope' describes a sound's temporal structure. Although the acoustic property of *amplitude* (ie energy) governs perceived loudness, the acoustic property of *amplitude envelope* (ie changes in energy over time) governs changes in our perception not only of a sound's loudness over time but also of its timbre. For example, a piano tone played backwards creates a categorically different sound, resembling that of a reed organ rather than a piano (Houtsma, Rossing, & Wagenaars, 1987). This is a dramatic perceptual change, given the preservation of many of the tone's acoustic characteristics (ie peak amplitude,

§ Corresponding author.

frequency, spectral relationships, etc). Amplitude envelope[1] (henceforth, 'envelope') also plays a crucial role in everyday listening, affording knowledge about whether a dropped bottle bounced or broke (Warren & Verbrugge, 1984), indicating the physical properties of a struck object (Lutfi, 2001), and even conveying the hardness of the striking material itself (Freed, 1990; Giordano, Rocchesso, & McAdams, 2010).

In this paper our goal is to explore differences in the underlying strategies used to process sounds with 'flat' sustain periods and abrupt offsets, and the strategies used to process sounds with no sustain and exponentially decaying offsets. Sounds with flat envelopes frequently exhibit a trapezoidal shape with sustain periods of varying duration, ending with abrupt linear offsets (Carlyon, Cusack, Foxton, & Robertson, 2001; Henry & McAuley, 2011; Mondor & Zatorre, 1995; Riecke, van Opstal, & Formisano, 2008; Smits, Sereno, & Jongman, 2006). We recognize that not all sounds with 'flat sustain periods' use this approach—others use cosine-squared (Lutfi, Kistler, Oh, Wightman, & Callahan, 2003), raised-cosine ramps (Roberts, Glasberg, & Moore, 2008), quarter-sine wave (Bregman, Levitan, & Liao, 1990), and/or Gaussian (Cullen & Collins, 1982) ramps. However, in the vast majority of tones surveyed in several hundred experiments from multiple journals (Gillard & Schutz, 2013; Schutz & Vaisberg, 2014), we found that offset ramps were generally abrupt and considerably shorter than their sustain periods. For purposes of our investigation, we believe the underlying strategies useful in processing flat tones with linear offsets would apply equally well to flat tones with other offset shapes. Therefore, for simplicity, throughout this paper we will simply refer to differences in the processing of *flat* and *percussive* envelopes, samples of which appear in figure 1a.

2 Amplitude envelope: an important factor in perception and cognition

The envelope serves as a crucial cue in a variety of auditory tasks. For example, in the *marimba illusion* the perceived duration of percussive sounds is influenced by the gestures used to produce those sounds: this audiovisual integration allows expert musicians to control the perceived (but not acoustic) duration of a note. Specifically, an internationally renowned marimbist was able to 'trick' audiences into hearing notes played with long gestures as 'longer' than notes played with short gestures—even though the acoustic durations of the notes did not differ (Schutz & Lipscomb, 2007). This illusion is robust, replicating with nonmusicians (Schutz & Kubovy, 2009a), simplistic point-light representations of the gestures (Schutz & Kubovy, 2009b), and even a single moving dot (Armontrout, Schutz, & Kubovy, 2009). Intriguingly, this effect runs counter to the widely supported notion that audition dominates estimations of tone duration (Fendrich & Corballis, 2001; Walker & Scott, 1981; Welch & Warren, 1980). The primary reason for this unexpected conflict between previous work suggesting vision has minimal influence on auditory judgments of event duration and repeated findings to the contrary in the marimba illusion paradigm is the amplitude envelope. The envelopes typically employed in duration perception research exhibit flat temporal structures, differing considerably from those produced by striking a marimba. More specifically, this discrepancy can be traced to differences in the processing of tones with abruptly ending versus naturally decaying offsets (Schutz, 2009).

As in the marimba illusion, multisensory integration occurs with pure tones shaped with percussive envelopes, but fails with pure tones shaped with flat envelopes. Curiously, subsequent analysis indicates that the durations of gradually decaying percussive tones are in fact no more ambiguous than the durations of abruptly ending flat tones (Schutz, 2009). This finding replicates and extends similar results involving natural percussive and sustained sounds created by musical instruments (Schutz & Kubovy, 2009a). Additional evidence of the influence of the envelope on audiovisual integration comes from the 'bounce illusion',

[1] As this investigation is focused entirely on amplitude envelope (rather than spectral envelope), for the sake of compactness our subsequent unqualified references to 'envelope' refer to 'amplitude envelope'.

in which the trajectory of two visual disks intersect and are perceived as passing through each other unless accompanied by an auditory signal, which produces a higher proportion of bounce percepts (Sekuler, Sekuler, & Lau, 1997). The envelope of the sound influences the degree to which the disks are seen to 'bounce off of' rather than 'stream through' one another (Grassi & Casco, 2009).

This evidence of vision affecting auditory duration judgments in the context of the marimba illusion is puzzling. It challenges the framework of optimal integration, widely recognized as a coherent explanation of a variety of crossmodal integration phenomena (Ernst & Banks, 2002). Additionally, these findings highlight differences in the perception of sounds with decaying envelopes (common in the natural world—Gaver, 1993a, 1993b), and sounds with flat envelopes. As flat envelope sounds are frequently used in auditory research (Schutz & Vaisberg, 2014), experiments focused heavily on flat stimuli with abrupt offsets may not yield conclusions generalizing to sounds with natural offsets (which are common outside the laboratory).

For instance, the well-established pacemaker–accumulator model (Gibbon, 1977; Wearden, 2003) posits that duration estimations involve monitoring the elapsed time between tone onset and offset. Therefore, the fact that duration estimations of percussive tones (with seemingly ambiguous moments of offset completion) are no worse than estimations of flat tones (Schutz, 2009) or even naturally sustained tones such as those produced by the clarinet (Schutz & Kubovy, 2009a) raises the question of whether these estimations are performed using different underlying strategies. Strategies useful for assessing the durations of flat tones with clearly demarcated offsets may be problematic for percussive tones, given their characteristic decay. Instead, we suspect percussive sounds are processed with an alternative strategy monitoring the slope of the decaying sound to anticipate the moment of acoustic completion, thereby side-stepping the difficulties inherent in detecting the offset explicitly. It is important to note that this strategy would not be viable for flat tones, as their sustain period offers no predictive information about their eventual offset.

3 Exploring the effect of amplitude envelope on duration judgment strategy

Here, our goal is to explore whether abruptly ending flat tones afford a listening strategy that is unavailable with sounds exhibiting more ecologically common decaying (percussive) tones. Previous research has shown that tones with monotonic decay are systematically judged as shorter than tones with flat envelopes, or those with monotonic increases (Grassi & Darwin, 2006; Schlauch et al., 2001). Similarly, sounds with 'rising profiles' are judged longer than reversed versions of the same sounds with 'decaying profiles' (DiGiovanni & Schlauch, 2007). Both of these findings are consistent with more recent research on duration judgments of percussive and reverse-percussive tones in an audiovisual integration paradigm (Schutz, 2009). Here we complement those collective findings by exploring a different question— whether flat and percussive (alternatively, 'steady' and 'decaying') temporal structures afford *fundamentally different listening strategies*.

We suspect the durations of flat tones are estimated using a 'marker strategy' involving calculating the elapsed time between the tone's onset and offset, a notion consistent with the pacemaker–accumulator model (Gibbon, 1977; Wearden, 2003). Although appropriate for flat tones, it would be less suited for (and should lead to worse performance with) percussive tones—whose moment of acoustic completion is more ambiguous (and indeed may in fact be partially ignored—Stecker & Hafter, 2000). Nevertheless, an account of duration perception based on the marker strategy alone is incompatible with previous research on duration estimation using percussive tones which found estimations of their duration were no more variable than flat/sustained tones when using either synthesized (Schutz, 2009) or natural (Schutz & Kubovy, 2009a) sounds. We thus suspect that estimations of percussive

tone duration may be performed using offset slope to 'predict' their acoustic completion. This strategy avoids needing to explicitly identify a tone's moment of completion, which is often difficult to discern. This 'prediction strategy' is consistent with evidence that a decaying slope can be useful in judging other aspects of sounds (Alais & Carlile, 2005). However, it is crucial to note that it would fail when listening to flat tones, as their flat sustain period followed by abrupt offsets would preclude such approaches.

Here, our exploration uses a binary-choice psychophysical procedure, asking observers to judge which of two tones is longer. Each trial employs one 'standard' and one of seven 'comparator' tones (three longer, three shorter, and one equal to the standard). The crucial manipulation concerns the matching of the two tones: on *matched trials* both tones are of the same envelope type (ie consisting of percussive–percussive and flat–flat pairs). These different types of trials were presented either *blocked* (ie observers knew with 100% certainty the type of tone to be presented on any given trial—experiment 1) or *mixed* (ie on any given trial observers could not predict the type of envelope to be presented—experiment 2). When blocked, the observer could select the most envelope-appropriate strategy, whereas, when mixed, the trial-to-trial uncertainty precluded switching, encouraging the use of a single strategy. Because the prediction strategy is unviable for flat tones (their sustain gives no indication of their overall duration), the mixed condition encourages use of a marker strategy for all trials—a strategy ill-suited for percussive tones. By comparing performance for matched trials when blocked separately (experiment 1) versus mixed (experiment 2), we are able to assess whether percussive tone performance decreases when presented in a manner precluding switching to an envelope-specific strategy.

4 Experiment 1

4.1 *Experiment 1a*

4.1.1 *Method*

4.1.1.1 *Participants.* We recruited nine participants (five females and four males) from the McMaster University undergraduate psychology volunteer pool (age $M = 19$ years; SD = 0.83 years). Participants provided written informed consent prior to participation and received one extra course credit in Level 1 Psychology. They all reported normal or corrected-to-normal vision and audition. All procedures were approved by the McMaster Research Ethics Board and complied with the tricouncil (Canada) policy on human participants.

4.1.1.2 *Materials.* We created fourteen different tones (seven each of percussive and flat) in MAX/MSP (Cycling 74), using a patch designed for a previous study (Paul & Schutz, 2010). The flat tones exhibited trapezoidal shapes with constant rise/fall times and variable sustain periods such that longer tones had longer sustain times but identical onset/offset times. The percussive tones exhibited constant linear rises followed immediately by exponential decays of varying length—that is, longer tones had longer decay times but identical onset times. These percussive tones are similar to those created by SuperCollider, and mimic the envelopes of sounds created by impact events. The sounds were based on a 1000 Hz pure tone and sampled at 44.1 kHz. Keeping with previous approaches to roughly equate the perceived loudness of the two tone types (Schutz, 2009), we set the peak amplitude for the flat tones at 80% of the peak amplitude for the percussive (about 1.93 dB difference). In order to match onset slope, percussive tone onset lasted 5 ms, whereas flat tone onset (as well as offset) lasted 4 ms.[2] We subtracted the combined onset/offset time from the sustain (flat) and decay (percussive) periods, thereby matching flat and percussive tones in acoustic duration. Examples of the basic percussive and flat tones used in these experiments can be found in figure 1a.

[2] The peak amplitude of the flat tones was 80% that of the percussive tones. A rise from 0 to 0.8 over 4 ms is identical in slope to a rise of 0 to 1.0 over 5 ms (ie 0.2 unit ms^{-1}).

These duration pairs ranged from 300 ms to 900 ms in 100 ms steps (see figure 1 for canonical flat and percussive tones). Each trial consisted of two tones: standard (always 600 ms) and comparator, with the comparator selected randomly from amongst the entire set. We ran the experiment in MatLab R2007b (MathWorks, Inc.) on an Apple MacPro PowerPC G5, using a 17″ CRT monitor (resolution: 1024 × 768; refresh rate: 75 Hz) and Sennheiser HDA-200 stereo headphones with a rated impedance of 40 Ω and a frequency range of 20–20 000 Hz. Tones were presented at a comfortable level, and that level was identical for all subjects.

4.1.1.3 *Procedure and design.* This experiment assessed performance under 'optimal' conditions with respect to strategy switching, in which a blocked approach afforded selection of the best strategy for each type of envelope. For each trial participants heard both a standard and comparator tone (presented in a random order) and indicated which sounded longer. We presented tone pairs under two conditions within this blocked approach. The condition of primary interest is *blocked (matched)*, in which two tones were always of the same envelope shape (either percussive–percussive or flat–flat). This condition contained two blocks—one with percussive–percussive trials and the other with flat–flat trials (see figure 1b for details). Additionally, this experiment also included a second condition: *blocked (mismatched)*,[3] with tones differing in envelope shape (either flat–percussive or percussive–flat). We randomized the order of these blocks (as well as the order of trials within all blocks) anew for each participant. Therefore, the experiment employed a 2 (standard tone: flat or percussive) × 2 (test tone: flat or percussive) × 7 (duration: 300 to 900 by steps of 100 ms) within-participant design with 20 trials per condition, leading to a total of 560 trials.

Participants sat approximately 60 cm in front of a computer screen in a dimly lit quiet room. They used their right index finger to hold down the '/?' key and their left index finger to hold down the 'z' key on the keyboard. We asked participants to lift their left finger ('z' key)
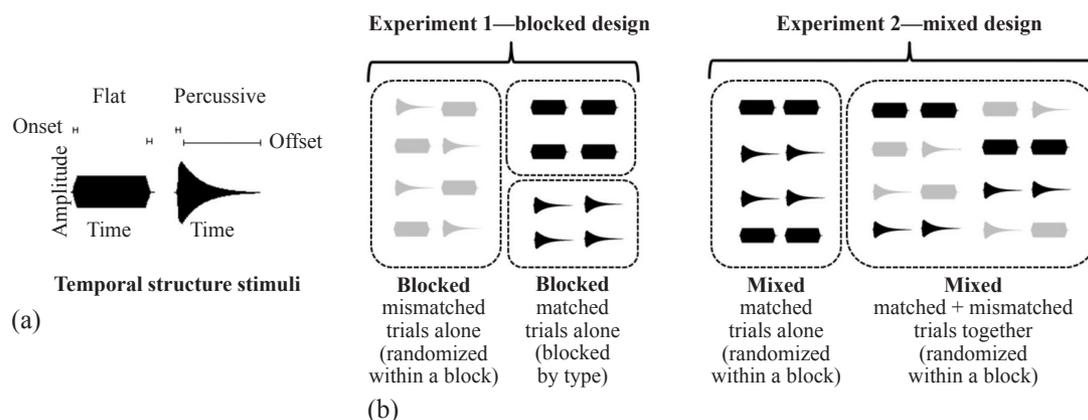
Figure 1. (a) Although both tone types share abrupt onsets, the abrupt offsets of flat tones (left) afford duration evaluation using a marker strategy which would be problematic for the gradual offsets of percussive tones (right). This type of offset could be effectively evaluated using a prediction strategy, which would be problematic for flat tones since their offsets do not begin until near the tone's completion. (b) Experiment 1 presented matched trials (percussive–percussive; flat–flat) blocked separately. Experiment 2 presented matched trials mixed in one of two ways: in either blocks containing only matched trials, or blocks containing both matched and mismatched trials. Mismatched trials are shown in light gray, indicating they were included in the experiment but are not discussed extensively for reasons detailed in the text.

[3] Initially, we were also interested in exploring ratings on mismatched trials mixing percussive and flat tones. However, differences in the perceived durations of percussive and flat tones of equivalent acoustic duration rendered this exploration inconclusive. Therefore, we decided not to include these conditions in the analyses.

when the first tone sounded longer than the second, and requested they keep their eyes open during the whole experiment to follow on-screen instructions. The experimenter then left the room, and participants completed the two blocks of trials. To start a trial, the participant had to press and keep pressed the two response keys.

Trials consisted of a visual fixation point (500 ms) followed by the first tone, a 500 ms gap, the second tone, and then a yellow fixation point prompting for a response. If participants tried to respond before the yellow fixation point, they received a warning message indicating recognition of a premature response. In the same way, if the participants did not respond within 5 s, they received a 'time-out' warning message. The next trial began as soon as the participant pressed both response keys. The participant could take a break at any time by leaving one or both buttons unpressed. During breaks, a message on the screen gave the number of completed and remaining trials in the block.

4.1.2 *Results and discussion.* We used MatLab for initial data treatment and R version 2.13.1 (R Foundation for Statistical Computing) for full analysis. We excluded premature and time-out responses from the analyses ($<2\%$ of the data), converting the raw responses into proportions of 'comparator longer than standard' responses (see figure 2). To assess relative differences in performance, we then computed the slopes for each trial type by converting each proportion of 'comparator longer than standard' responses to its associated z-score from a normal distribution (figure 3) and then using a least-squares fit to a linear function (ie a probit analysis—Finney, 1971).

Within the blocked (matched) condition containing percussive–percussive and flat–flat trials separately blocked, slopes did not differ significantly ($t_8 = -1.03$, $p = 0.33$). This null finding (ie meaning the slope of the psychometric curves for flat–flat and percussive–percussive conditions did not differ significantly) indicates that participants performed the task equally well with matched trials of percussive and flat tones.

Assessment of slope was more problematic within the blocked (mismatched) condition (flat–percussive and percussive–flat), as participants systematically judged percussive tones to sound shorter than their flat counterparts. Indeed, the majority of participants judged all of the percussive tones as sounding shorter than all of the flat tones (consistent with Grassi & Darwin, 2006; Schlauch et al., 2001). Consequently, we could not compute a point of subjective equality (PSE) value, since for some participants the slope was near zero and so the PSE would be an unrealistic extrapolation. Therefore, we decided to run a similar experiment (1b) using a wider range of durations (ie a bigger step size between duration values). We hoped this might serve two goals: (a) assessing the robustness of our findings from experiment 1a by attempting a replication using a wider range of duration values, and (b) allowing us to obtain a PSE for the blocked (mismatched) condition as the greater step size might render some percussive tones equivalent in duration to the shortest flat tone.
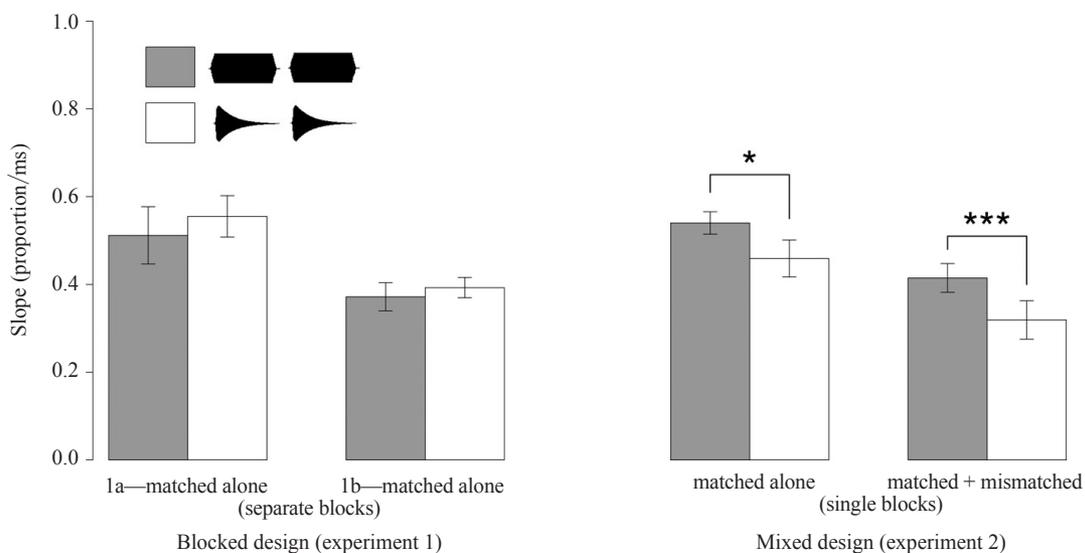
4.2 *Experiment 1b*
4.2.1 *Method*
4.2.1.1 *Participants.* We recruited twelve participants from the McMaster University undergraduate psychology subject pool. This included nine females and three males with an average age of 19.78 years (SD = 3.98 years). None of them participated in experiment 1a.
4.2.1.2 *Materials and procedure.* We used the same design, materials, and procedures as in experiment 1a, changing only the range of tone durations. Here we used a 750 ms standard, and comparator tones ranging from 300 to 1200 ms by steps of 150 ms. We increased the step size in order to include a wider range of durations with more overlap in equivalent perceived duration between the two categories of tones.
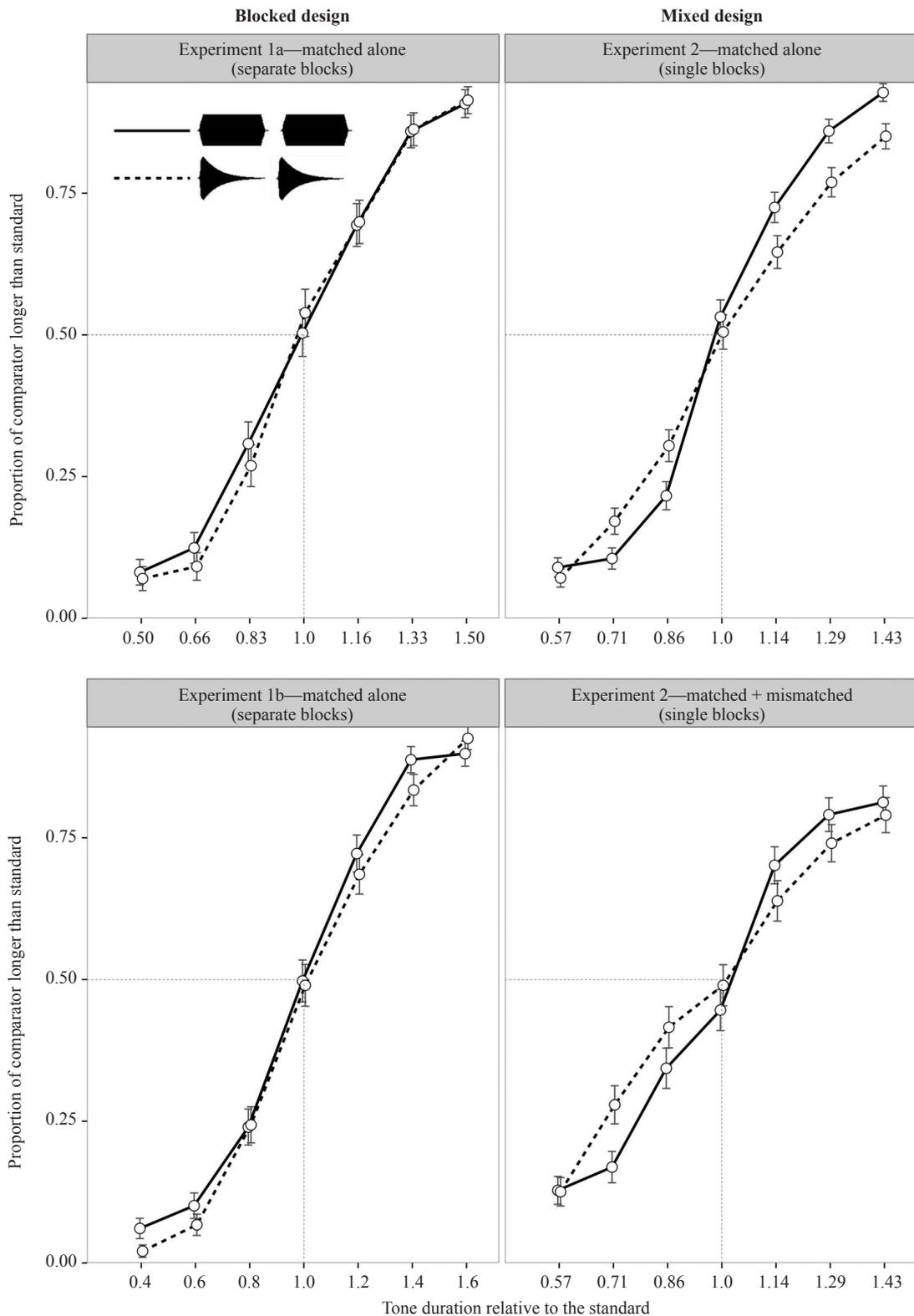
Figure 2. Psychometric curves for indicating mean proportion of 'comparator sounds longer than standard' responses as a function of comparator duration across all experiments for matched trials (percussive–percussive and flat–flat tone combination). The left panels show results for the first experiment's blocked design with the duration ratios used in experiment 1a (top) and experiment 1b (bottom). The right panels show performance when matched trials were mixed—either only with one another (top) or together with mismatched trials (bottom). Error bars represent standard errors, corrected for within-subject design.

Figure 3. Mean slopes of matched trials (percussive–percussive, flat–flat) when presented within a blocked design (experiments 1a and 1b) or mixed design (experiment 2) mixing either matched trials alone or matched trials and mismatched trials together. Error bars represent standard errors corrected for within-subject design. $*p = 0.05$; $***p < 0.05$.

4.2.2 *Results*. The blocked (mismatched) condition also showed patterns of results similar to experiment 1a, with participants again judging all of the percussive tones as shorter than nearly all of the flat—despite the larger range of durations (ie they perceived the 1200 ms percussive tone as shorter than the 650 ms flat standard 55% of the time). This again precluded meaningful statistical testing of the differences. However, we did replicate the key finding of the first experiment; the lack of difference in performance between the flat–flat and percussive–percussive blocks ($t_{10} = -0.99$, $p = 0.35$) in the blocked (matched) condition (see figures 2 and 3).

4.2.3 *Discussion*. Despite the larger step sizes, participants still rated percussive tones as sounding shorter than their flat counterparts to a degree precluding meaningful analysis of the blocked (mismatched) condition. In all likelihood, the gradual decay of a percussive envelope renders the last portion of the tone so soft as to be imperceptible. This is consistent with previous findings regarding differences in the perceived durations of sounds with temporal structures similar to our flat and percussive tones (Schlauch et al., 2001). Because this does not interfere with analysis of performance in the block (matched) condition, we will focus in the remainder of this manuscript on the primary issue of interest—performance on matched trials.

The most important outcome emerging from both versions of experiment 1 is the equivalent performance for the matched (ie flat–flat and percussive–percussive) pairs in the blocked (matched) condition. There are two ways to interpret this result. One is that precision in judging duration is the same for percussive and flat tones, consistent with previous findings (Schutz, 2009; Schutz & Kubovy, 2009a). However, the use of a blocked design affords a strategy optimization that might explain this outcome. Because we blocked the percussive–percussive and flat–flat trials separately within the block (matched) condition, participants were able to anticipate the envelope of each tone prior to its onset. This compelled investigation of whether participants might have used different strategies for judging the durations of percussive and flat tones, with each strategy yielding similar precision. In order to explore this possibility, we conducted a second experiment using a mixed design constraining participants' ability to switch strategies.

5 Experiment 2

The primary goal of the second experiment was to compare the precision of participants' duration estimations when unable to reliably switch strategies. To fully explore the consequences of this manipulation, here we used two types of mixed approaches. The durations of flat tones simply cannot be assessed using the prediction strategy, as their moment of acoustic termination cannot be discerned until just before their completion. Therefore, we suspect participants would be forced to consistently apply the marker strategy even when listening to percussive tones, a type of sound for which it is ill-suited.

5.1 *Method*

Here we used two types of mixed conditions. The first randomized matched trials (flat–flat and percussive–percussive) within a single block. The second went further, randomizing both matched (flat–flat, percussive–percussive) and mismatched (flat–percussive, percussive–flat) trials within a single block. Crucially, both of these conditions differed from those in the first experiment in that they employed mixed designs precluding the strategy switching possible in experiment 1.

We implemented two versions of this experiment in order to control the number of trials per block and the number of trials per condition. The number of trials per block was controlled in experiment 2a, whereas the number of trials per condition was controlled in experiment 2b. Because there were no effects and no interactions of this design feature, we collapsed the data into one experiment to more clearly illustrate its key outcomes.

5.1.1 *Participants.* We recruited nineteen participants from the McMaster University undergraduate psychology subject pool: eleven (six females) in experiment 2a and eight (four females) in experiment 2b. The average age was 18.82 years (SD = 0.75 years) in experiment 2a and 19.25 years (SD = 1.58 years) in experiment 2b. Three participants (experiment 2a) and four participants (experiment 2b) reported a previous experience with music such as playing piano or guitar, and none had participated in the previous experiment.

5.1.2 *Materials.* Our procedures and method were similar to experiment 1, however, here we used tone durations ranging from 400 ms to 1000 ms by steps of 100 ms, with a 700 ms standard tone.

5.1.3 *Procedure and design.* We used a procedure and design similar to experiment 1, with the exception that here we mixed the trials in one of two ways—either randomizing matched trials alone (first condition) or randomizing all four trial types together (second condition)—see figure 1b for details. In experiment 2a we roughly matched the number of trials per condition: the first block contained 16 repetitions of each of the 14 (matched) pairs for a total of 224 trials. The second condition contained 10 repetitions of the 28 (matched + mismatched) pairs for a total of 280 trials (504 total experimental trials). In experiment 2b we matched the number of trials per stimulus (in this case 14) across the two conditions: the first (matched trials only) contained 196 trials, and the second (all four trial types) contained 392 trials—for a total of 588 experimental trials.

5.2 *Results*

We excluded one participant in experiment 2b due to poor performance making it impossible to fit a psychometric curve. We focused our slope analysis on the matched trials only, as the difference in perceived duration for percussive and flat tones of the same acoustic duration precluded a meaningful assessment of the mismatched trials. However, their inclusion in the experiment was essential in order to manipulate possible duration judgment strategies.

Experiments 2a and 2b did not differ with respect to slope results ($F_{1,17} = 0.50$, $p = 0.49$) for the interaction of experiment × condition × block, and so the data were collapsed for

subsequent analysis (see figure 2 for the psychometric curves of the collapsed data). The slopes of the matched trials were steeper (0.50) when randomized alone (first condition) rather than when randomized along with the mismatched trials (0.37) in the second condition ($F_{1,17} = 32.17$, $p < 0.001$, $\eta_g^2 = 0.16$). This reflects decreased performance when removing certainty about the second tone within each trial. However, the most important outcome of experiment 2 is that, in contrast to experiment 1 (where flat–flat and percussive–percussive were blocked separately), here performance on the percussive–percussive trials was less precise than performance on the flat–flat trials ($F_{1,17} = 7.31$, $p = 0.02$, $\eta_g^2 = 0.05$). There was no interaction of condition by trial type ($F_{1,17} = 0.32$, $p = 0.58$). See figure 3 for a side-by-side comparison of performance across different conditions and experiments.

5.3 *Discussion*

The primary result of interest is the inferior performance of the percussive–percussive (relative to flat–flat) trials when presented in a mixed fashion in the second experiment. This inferior performance when unable to anticipate envelope shape contrasts to equal performance in the first experiment where participants were able to optimize their estimation strategies. It appears that forcing participants to adopt the marker strategy for all trials by mixing the two types of tones led to worse performance for the types of tones ill-suited for the marker strategy—percussive tones. It is important to note that this difference in performance does not merely reflect differences in the difficulty of judging mixed (ie percussive–flat, flat–percussive) versus matched (ie flat–flat, percussive–percussive) trials. Here we are interested in the performance on matched trials alone (ie the darker trials depicted in figure 1b)—exploring how the quality of their evaluation differs depending upon the context in which they were blocked (and therefore the listener's ability to switch between different strategies).

The inferior performance on matched percussive trials presented within a block structure precluding strategy switching (flat matched trials show no such difference) is consistent with the idea that they are processed with a strategy that is not available for flat tones. Consequently, we believe these data represent evidence for the use of different strategies when judging the durations of percussive and flat tones. This builds upon parallel research on qualitative differences in the perception of percussive versus flat tones in multimodal tasks, such as audiovisual integration (Schutz, 2009) or sound–object memory associations (Schutz & Stefanucci, 2010). Additionally, it extends previous work documenting the role of envelope in judgments of duration magnitude (Grassi & Pavan, 2012; Schlauch et al., 2001) by exploring differences in the underlying strategies used to make these judgments.

These results also connect with another theme in research on duration perception—speculation that ignoring the tails of decaying sounds leads to an underestimation of their duration (Stecker & Hafter, 2000). To explore that hypothesis, participants judging tone duration were asked to either (a) "include the entire sound in your judgment" or "ignore the echo in your judgment". Curiously, these instructions affected falling-intensity sounds (similar to our percussive tones) more so than rising-intensity sounds (DiGiovanni & Schlauch, 2007). The rising-intensity sounds used in that study are just as ill-suited to a prediction strategy for duration estimation as our flat tones. Their continual rise before an abrupt onset means that their moment of acoustic completion could not be predicted in advance. Consequently, findings that explicit instructions regarding attention to a tone's characteristics affect decaying sounds (which could be processed using the prediction strategy) more so than sounds presumably processed using a marker strategy are consistent with our ideas with respect to the use of envelope-specific strategies.

6 Conclusions

Here, our goal was to explore an envelope's shape on duration estimation *strategy*, extending previous research exploring its effect on duration estimation magnitude. We pursued this in two experiments, broken into subexperiments to address potential methodological confounds. The most interesting outcome is the difference in performance on the matched (percussive–percussive; flat–flat) trials as a function of blocking structure. The first experiment blocked flat–flat and percussive–percussive trials separately, permitting participants to use the 'marker strategy' for the flat–flat trials and the 'predictor strategy' for percussive–percussive trials. We observed no difference in duration judgment precision in this situation. However, the uncertainty with respect to a priori knowledge of a tone's envelope inherent in the mixed approach of experiment 2 precluded participants' ability to optimally select an envelope-specific strategy. As the prediction strategy is not viable for flat tones, we suspect participants were forced to employ the marker strategy for all tones—a strategy ill-suited for percussive tones which decay gradually and lack a clear moment of acoustic completion.

The structure of the second experiment made it impossible to choose an optimal strategy prior to the beginning of a trial. In principle it might be possible for participants to dynamically switch strategies midtrial (from marker to predictor) after recognizing a tone as percussive. However, we do not yet know whether such a dynamic shifting of strategy after tone onset is possible, how long it would take to switch strategies, or the perceptual cost of this approach (which might alternatively explain the decreased performance in experiment 2). Although a great deal of future research would be required to explore that possibility, finding evidence of a 'dynamic shift' would serve to only strengthen our assertion that a tone's temporal structure affects the underlying processes used to estimate its duration.

Therefore, we conclude that participants perform best when given an opportunity to choose the most envelope-appropriate strategy, consistent with the idea that a tone's envelope affords certain duration estimation approaches. These findings complement recent studies showing qualitatively different patterns of processing when using pure tones synthesized with either percussive or flat envelopes (Sandhu, Schutz, & Dyson, 2013; Schutz, 2009; Schutz & Stefanucci, 2010). Additionally, we note that all of the percussive tones were generally judged shorter than all of the flat tones, despite a large difference in absolute acoustic duration, a finding consistent with previous studies reporting flat tones are consistently shown to sound longer than 'damped'[4] tones (Grassi & Darwin, 2006; Schlauch et al., 2001).

7 Implications for research and assessment of auditory perception

We find the possibility that duration estimation may be performed using different strategies intriguing, given that findings from studies using flat tones are typically assumed to generalize to real-word listening. For example, flat tones are common in a wide range of tasks such as evaluating frequency sensitivity (Gockel, Moore, Carlyon, & Plack, 2007), determining difference levels (Oxenham & Buus, 2000), and exploring auditory streaming (Bendixen, Denham, Gyimesi, & Winkler, 2010; Bregman, 1990). Flat tones are also used frequently when assessing the performance of adaptive devices such as exploring stream segregation for cochlear implant users (Cooper & Roberts, 2007). Indeed, the standards in assessment devices laid forth by the American National Standards on Acoustics in "Specification for audiometers" specify precise tolerances for devices synthesizing sounds—tones with flat envelopes (ANSI, 2010). If the strategies used for processing naturally decaying sounds differs from those used to process abruptly ending tones, then further explorations using

---

[4] Their damped tones were similar to our percussive tones, which decay exponentially. However, we have intentionally avoided using the term 'damped' since it is used in the literature to refer to either tones with either exponentially or linear decay shapes.

sounds with different envelopes could be of significant benefit in assessing hearing devices designed for use in everyday listening situations.

These results also have significant implications for models of optimal integration—the dominant framework for explaining multimodal cue integration (Ernst & Banks, 2002). This powerful framework gracefully accounts for both traditional cross-modal illusions such as the 'ventriloquist effect' in which vision influences auditory localization (Bertelson, Vroomen, de Gelder, & Driver, 2000; Jackson, 1953), as well as 'reversals' of traditional patterns such as when audition influences the localization of low-quality visual information (Alais & Burr, 2004). Optimal integration posits that conflicts are resolved by weighting sensory information in proportion to its quality, which is operationally defined by the variability/uncertainty of the perception of a given stimulus (Ernst & Banks, 2002). If uncertainty is related in part to strategy, then a core aspect of this framework is indirectly influenced by blocking structure (insofar as it cues different processing strategies). Therefore the implications of using different strategies for judging duration extend beyond those directly related to temporal processing and are relevant to a range of tasks.

Consequently, these results hold implications for auditory research ranging from auditory scene analysis (Bregman, 1990) to statistical learning (Saffran, Johnson, Aslin, & Newport, 1999) to multimodal integration (Walker & Scott, 1981)—much of which relies heavily on flat tones. Indeed, a survey of auditory stimuli based on a review of research published in *Music Perception* (Tirovolas & Levitin, 2011) illustrates that flat tones are commonly employed—even in a journal explicitly devoted to listening outside the laboratory (ie the perception of music; Schutz & Vaisberg, 2014). A similar survey of the journal *Attention, Perception, & Psychophysics* also found widespread use of flat tones (Gillard & Schutz, 2013). Therefore, these results are of broad relevance, given that everyday listening often involves percussive sounds (Gaver, 1993a, 1993b). It is possible that the amplitude envelope plays a role in some of these tasks but not others, and consequently some outcomes and models generalize better than others to real-world listening.

8 Final remarks

It is important to note that the widespread use of flat tones in auditory research is not necessarily problematic. These tones offer certain benefits for perceptual research due to their 'controlled' nature. The durations of tones with abrupt offsets can be more precisely quantified than durations of tones with gradually decaying offsets, and in many cases this precise control is necessary in order to draw strong conclusions about certain types of processing. However, we undertook this investigation out of interest in exploring whether it is possible that these abrupt offsets afford not only precise quantification of their durations but also certain perceptual processing strategies not available when listening to the sounds with natural decay profiles we so often hear outside the laboratory. Our results are consistent with this idea, suggesting that amplitude envelope is an important property of auditory perception holding great potential for fruitful future research. Indeed, these data represent the latest step in a series of experiments exploring differences in the perception and cognition of sounds with these two kinds of envelopes—building upon previous work showing differences in sensory integration (Grassi & Casco, 2009; Schutz, 2009) as well as judgments of duration (Grassi & Pavan, 2012; Sandhu et al., 2013; Schlauch et al., 2001), loudness (Ries, Schlauch, & DiGiovanni, 2008), and time-to-contact (Neuhoff, 1998, 2001).

Consequently, we suspect that the role of envelope in perceptual organization transcends particular domains/paradigms and holds relevance for a broad range of auditory research. Researchers interested in exploring these issues can take advantage of the software used to create these sounds (which can easily create tones of varying durations, rise times, exponential decays, etc at any frequency), available freely online at http://www.maplelab.net/software.

We hope this work will help inspire other explorations of envelope in a variety of tasks to clarify the degree to which theories, models, and clinical tests frequently performed using flat tones apply to our everyday listening—where sounds with naturally decaying envelopes figure prominently.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, **14**, 257–262.

Alais, D., & Carlile, S. (2005). Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound. *Proceedings of the National Academy of Sciences of the USA*, **102**, 2244–2247.

ANSI (2010). *American National Standard: Specification for Audiometers* (ANSI/ASA S3.6-2010).

Armontrout, J., Schutz, M., & Kubovy, M. (2009). Visual determinants of a cross-modal illusion. *Attention, Perception, & Psychophysics*, **71**, 1618–1627.

Bendixen, A., Denham, S. L., Gyimesi, K., & Winkler, I. (2010). Regular patterns stabilize auditory streams. *The Journal of the Acoustical Society of America*, **128**, 3658–3666.

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, **62**, 321–332.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Bregman, A. S., Levitan, R., & Liao, C. (1990). Fusion of auditory components: Effects of the frequency of amplitude modulation. *Perception & Psychophysics*, **47**, 68–73.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, **27**, 115–127. doi:10.1037/0096-1523.27.1.115

Cooper, H. R., & Roberts, B. (2007). Auditory stream segregation of tone sequences in cochlear implant listeners. *Hearing Research*, **225**, 11–24.

Cullen, J. K., & Collins, M. J. (1982). Audibility of short-duration tone-glides as a function of rate of frequency change. *Hearing Research*, **7**, 115–125.

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:10.1080/10407410701432329

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, **415**(6870), 429–433.

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, **63**, 719–725.

Finney, D. J. (1971). *Probit analysis* (3rd ed.). London: Cambridge University Press.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *The Journal of the Acoustical Society of America*, **87**, 311–322.

Gaver, W. W. (1993a). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, **5**, 1–29.

Gaver, W. W. (1993b). How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology*, **5**, 285–313.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, **84**, 279–325.

Gillard, J., & Schutz, M. (2013). The importance of amplitude envelope: Surveying the temporal structure of sounds in perceptual research. In *Proceedings of the Sound and Music Computing Conference* (pp. 62–68). Berlin: Logos Verlag. http://smcnetwork.org/node/1850

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 and Performance*, **36**, 462– 476.

Gockel, H. E., Moore, B. C. J., Carlyon, R. P., & Plack, C. J. (2007). Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency. *The Journal of the Acoustical Society of America*, **121**, 373–382.

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**, 235–243. doi:10.1037/a0013031

Grassi, M., & Darwin, C. J. (2006). The subjective duration of ramped and damped sounds. *Perception & Psychophysics*, **68**, 1382–1392.

Grassi, M., & Pavan, A. (2012). The subjective duration of audiovisual looming and receding stimuli. *Attention, Perception, & Psychophysics*, **74**, 1321–1333.

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings. *Attention, Perception, & Psychophysics*, **72**, 561–582.

Henry, M. J., & McAuley, J. D. (2011). Velocity perception for sounds moving in frequency space. *Attention, Perception, & Psychophysics*, **73**, 172–188.

Houtsma, A. J. M., Rossing, T. D. &Wagenaars, W. M. (1987). Auditory demonstrations (Philips Compact Disc). Melville, NY: Acoustical Society of America.

Jackson, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, **5**, 52–65.

Lutfi, R. A. (2001). Auditory detection of hollowness. *The Journal of the Acoustical Society of America*, **110**, 1010–1019.

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., & Callahan, M. R. (2003). One factor underlies individual differences in auditory informational masking within and across age groups. *Perception & Psychophysics*, **65**, 396–406.

Mondor, T. A., & Zatorre, R. J. (1995). Shifting and focusing auditory spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 387–409.

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

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, **13**, 87–110. doi:10.1207/S15326969ECO1302_2

Oxenham, A. J., & Buus, S. (2000). Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions. *Journal of the Acoustical Society of America*, **107**, 1605–1614.

Paul, G., & Schutz, M. (2010). Using percussive sounds to improve the efficacy of auditory alarms in medical devices. In *Poster session presented at the Society of Music Perception and Cognition.* (p. 3). Rochester, NY.

Riecke, L., van Opstal, A. J., & Formisano, E. (2008). The auditory continuity illusion: A parametric investigation and filter model. *Perception & Psychophysics*, **70**, 1–12.

Ries, D. T., Schlauch, R. S., & DiGiovanni, J. J. (2008). The role of temporal-masking patterns in the determination of subjective duration and loudness for ramped and damped sounds. *The Journal of the Acoustical Society of America*, **124**, 3772–3783. doi:10.1121/1.2999342

Roberts, B., Glasberg, B. R., & Moore, B. C. J. (2008). Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, **34**, 992–1006. doi:10.1037/0096-1523.34.4.992

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, **70**, 27–52.

Sandhu, R., Schutz, M., & Dyson, B. J. (2013). Duration estimations of tones with naturally decaying vs abruptly ending amplitude envelopes: An ERP investigation. In *Poster presentation at the 20th Annual Meeting of the Cognitive Neuroscience Society, 13–16 April.* San Francisco, CA.

Schlauch, R. S., Ries, D. T., & DiGiovanni, J. J. (2001). Duration discrimination and subjective duration for ramped and damped sounds. *The Journal of the Acoustical Society of America*, **109**, 2880–2887.

Schutz, M. (2009). *Crossmodal integration: The search for unity* (doctoral thesis). University of Virginia, Charlottesville, VA.

Schutz, M., & Kubovy, M. (2009a). Causality and cross-modal integration. *Journal of Experimental Psychology: Human Perception and Performance*, **35**, 1791–1810.

Schutz, M., & Kubovy, M. (2009b). Deconstructing a musical illusion: Point-light representations capture salient properties of impact motions. *Canadian Acoustics*, **37**(1), 23–28.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, **36**, 888–897.

Schutz, M., & Stefanucci, J. (2010). Amplitude envelope and auditory alarms. In *Proceedings of the International Conference on Music Perception and Cognition*. Seattle, WA.

Schutz, M., & Vaisberg, J. M. (2014). Surveying the temporal structure of sounds used in *Music Perception*. *Music Perception*, **31**, 288–296.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, **385**, 308.

Smits, R., Sereno, J., & Jongman, A. (2006). Categorization of sounds. *Journal of Experimental Psychology: Human Perception and Performance*, **32**, 733–754. doi:10.1037/0096-1523.32.3.733

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

Tirovolas, A., & Levitin, D. (2011). Music perception and cognition research from 1983 to 2010: A categorical and bibliometric analysis of empirical articles in *Music Perception*. *Music Perception*, **29**, 23–36.

Walker, J., & Scott, K. (1981). Auditory–visual conflicts in the perceived duration of lights, tones and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 1327–1339.

Warren, W., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 704–712.

Wearden, J. (2003). Chapter 2: Applying the scalar timing model to human time psychology: Progress and challenges. In H. Helfrich (Ed.), *Time and mind II: Information processing perspectives* (pp. 21–39). East York, ON: Hogrefe & Huber.

Welch, R., & Warren, D. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, **88**, 638–667.