Surveying the Temporal Structure of Sounds Used in *Music Perception*

Michael Schutz & Jonathan M. Vaisberg
McMaster Institute for Music and the Mind, Hamilton, Ontario, Canada

RECENT WORK FROM OUR LAB ILLUSTRATES AMPLITUDE envelope's crucial role in both perceptual (Schutz, 2009) and cognitive (Schutz & Stefanucci, 2010) processing. Consequently, we surveyed the amplitude envelopes of sounds used in *Music Perception*, categorizing them as either *flat* (i.e., trapezoidal shape), *percussive* (aka "damped" or "decaying"), *other,* or *undefined.* Curiously, the undefined category represented the largest percentage of sounds observed, with 35% lacking definition of this important property (approximately 27% were percussive, 27% flat, and 11% other). This omission of relevant information was not indicative of general inattention to methodological detail. Studies using tones with undefined amplitude envelopes generally defined other properties such as spectral structure (85%), duration (80%), and even model of headphones/speakers (65%) at high rates. Consequently, this targeted omission is intriguing, and suggests amplitude envelope is an area ripe for future research.

THE MOTIVATION FOR THIS SURVEY COMES from our interest in understanding the prevalence of natural vs. artificial sounds in auditory perception research. Although our specific interests are inspired by recent discoveries, general concern regarding the use (and misuse) of artificial sounds has a long history in experimental psychology. For example, Gaver (1993a, 1993b) cautioned that fixating on simplistic synthesized tones carries the risk of overlooking how we process sounds "in the real world" (i.e., outside the laboratory). Part of his concern was that the precise control afforded by contemporary sound synthesis techniques and sophisticated methodological approaches allows researchers to focus surgically on particular aspects of listening (i.e., "pitch" or "timbre") at the expense of other aspects important in natural perceiving – such as recognizing the event producing a sound. Although his argument was focused on what he referred to as "everyday listening" (i.e., listening done in our daily lives), his concern with the impoverished nature of stimuli used to assess the auditory system is in a sense equally relevant to the perception of music.[1] His writings foreshadowed subsequent work in music perception documenting clear differences in the perceptual organization of musical and artificial sounds, differences with significant implications for theories of auditory processing.

Our interest in this issue began with a seemingly unrelated question—whether it was possible to produce long and short notes on the marimba. Marimbists have long debated the role of physical gesture length (i.e., the down-up gesture used to strike a note) in controlling note duration, with strong opinions by well-respected percussionists on both sides of the issue (Schutz & Manning, 2012). To explore this question empirically, a professional marimbist recorded his attempts to create long and short notes by varying striking gesture length. Although these gestures failed to produce notes that differed acoustically, they accidentally created notes differing perceptually through the use of a natural audio-visual illusion. In other words, these notes *sounded* longer when produced by long rather than short striking gestures, even though the sounds of these notes were acoustically indistinguishable (Schutz & Lipscomb, 2007).

Beyond the musical implications of this finding, the illusion is puzzling in that it conflicts with widely held views of audio-visual integration. What is clear from a substantial body of research is that although vision is prioritized over audition in a variety of paradigms (Bertelson, Vroomen, De Gelder, & Driver, 2000; Jackson, 1953; Soto-Faraco, Spence, & Kingstone, 2004) time perception is one area in which the auditory system generally dominates. This can be seen in a variety of temporally sensitive tasks, such as temporal-order-judgments (Morein-Zamir, Soto-Faraco, & Kingstone, 2003). In particular, the auditory system is recognized

---

[1] As his full argument is broad in scope, in the interest of space here we reiterate only those aspects most pertinent to our study. Those interested in his full argument are encouraged to these papers, as well his dissertation (Gaver, 1988).

as dominating judgments of event duration—both in terms of audition's known influence on vision (Walker & Scott, 1981) as well as vision's decided *lack* of influence on audition (Guttman, Gilroy, & Blake, 2005).

The puzzling nature of this break with previous findings cannot be explained by the specific visual stimuli used, as the illusion replicates with both point light displays (Schutz & Kubovy, 2009a) and single dots using simplified motion paths (Armontrout, Schutz, & Kubovy, 2009). Furthermore, although the 'optimal integration' hypothesis (the leading theory of multimodal integration) can account for reversals of traditional patterns when the typically dominant mode is ambiguous (Alais & Burr, 2004; Ernst & Banks, 2002), subsequent research testing the ambiguity of percussive tone duration estimates suggests it does not explain this phenomenon. Instead, this discrepancy is largely the result of the perceptual differences triggered by naturally decaying vs. abruptly ending amplitude envelopes (Schutz, 2009).

AMPLITUDE ENVELOPE EXPLAINS A PERCEPTUAL PARADOX

The term "amplitude envelope" (which we will occasionally abbreviate as "envelope" for space considerations) refers to a sound's temporal structure—essentially its shape over time. Marimba notes exhibit a percussive structure, consisting of a natural decay beginning immediately after tone onset. Although pervasive in everyday listening (Gaver, 1993a), such envelopes pose challenges to the kinds of rigorous quantification desirable in research as their gradual decay makes identifying their precise moment of offset difficult. Consequently, auditory perception research frequently employs tones with flat amplitude envelopes with more clearly discerned moments of offset (see Figure 1a, for depictions of these percussive and flat envelopes).

A direct test of the role of envelope in sensory integration using pure tones (i.e., sine waves) synthesized with flat or percussive shapes illustrates the danger of assuming that flat tones serve as accurate assessors of the perceptual system. Consistent with the original illusion (Schutz & Lipscomb, 2007), visual information once again integrated with percussive (i.e., gradually decaying) tones—sounds in the same acoustical family as the marimba. However, the same visual information *failed to integrate* with spectrally matched pure tones synthesized with flat envelopes—the sounds used in previous research on audio-visual integration. Consequently, the "conflict" between the marimba illusion and previous audio-visual integration stems from a previous focus on tones synthesized with flat envelopes (Schutz, 2009).
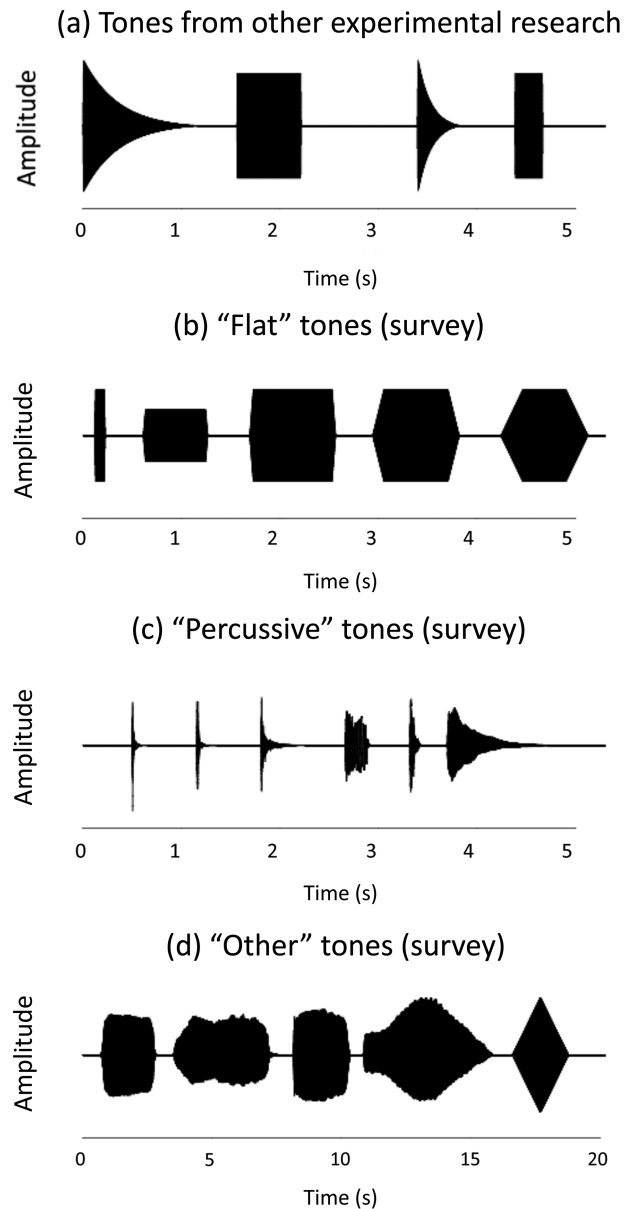

(a) Tones from other experimental research

(b) "Flat" tones (survey)

(c) "Percussive" tones (survey)

(d) "Other" tones (survey)

FIGURE 1. The amplitude envelopes of sounds used in auditory research. The top row (a) depicts two pairs of flat and percussive tones used by Schutz (2009) as well as Schutz and Stefanucci (2010). The remaining three rows depict the types of sounds encountered within our survey: (b) several flat tones (note the final three have different onset/offset ramps but the same duration and peak amplitude); (c) percussive sounds including a hand clap, wood block, rimshot, kick drum, cowbell, piano; and (d) other tones including those produced by a clarinet, violin, bassoon, and oboe, as well as a synthesized tone with a linear onset and offset (which we did not consider flat as it lacked a sustain period).

Although our interest in envelope grew from our investigation of one aspect of audio-visual music perception, general questions surrounding the multimodal

nature of music are widely relevant (i.e., for a review focused on practical applications for musicians, see Schutz, 2008). For example, *Music Perception* has published studies documenting body movement's role in both communicating a performer's emotional intentions (Dahl & Friberg, 2007) and shaping an audience's perception (Nusseck & Wanderley, 2009), as well as a meta-analysis of vision's influence on music listening (Platz & Kopiez, 2012). This illustrates that the topic of audio-visual integration is of broad relevance to the music perception community.

Intriguingly, documentation of categorical differences in the perception of tones differing only in their amplitude envelopes is not limited to this paradigm alone, with envelope known to affect audio-visual integration in the "bounce-or-pass-through" illusion (Grassi & Casco, 2009). Furthermore, subsequent research illustrates significant differences in associative memory tasks involving four note sequences of percussive and flat tones (Schutz & Stefanucci, 2010), as well as differences in the duration judgment strategies employed when estimating their length (Vallet, Shore, & Schutz, 2014). Although this work was not published in *Music Perception*, this journal has published other work on associative memory (Boltz, 2001) as well as duration estimation (Nakajima, Hoopen, & Van Der Wilk, 1991). Consequently, we believe envelope's crucial role in a variety of tasks focused on ostensibly unrelated issues is broadly relevant for those interested in the perception of music (as well as sound in general). Note that in each of the three examples of our lab's findings (sensory integration, associative memory, duration judgment strategies) the manipulations were not between 'natural' and 'artificial' tones, but rather tones synthesized with either 'naturally decaying' or 'artificially abrupt' envelope offsets. As both of these sounds are based on pure tones, neither is strictly natural. For as Gibson (1966) astutely observed, "until the advent of tuning forks and oscillators it is very doubtful that an ear had ever been stimulated by a single frequency and that an organism had ever heard a pure tone"(p. 89). Consequently, these differences come not from comparing the perception of 'natural' and 'artificial' sounds, but rather from using sounds synthesized with either 'natural' or 'artificial' envelopes.

In light of these repeated findings of differences between percussive and flat tones, we began wondering about the general prevalence of flat tones in auditory research. We suspected they were common due to their methodological precision as well as a legacy of frequent use in the early days of sound synthesis. If so, this would raise significant questions about the degree to which experiments performed with flat tones generalize to other contexts.

WHAT TYPES OF SOUNDS ARE USED IN AUDITORY PERCEPTION RESEARCH?

Our suspicions regarding the prevalence of flat tones seem well supported from informal listening at conferences and are generally met with casual agreement from other researchers. However, to the best of our knowledge this suspicion represents an as-of-yet untested hypothesis, albeit one with significant implications for applying "controlled" experimental research to the types of listening performed outside the lab—whether to environmental sounds or music.

Therefore, we set out to survey the sounds used in auditory research using a larger survey of all empirical articles published in *Music Perception* (Tirovolas & Levitin, 2011), which conveniently identified articles using single tones or isolated series of tones. Their classification provided a corpus selected by a third party without consideration of our motivations or assumptions. Although in principle our survey could be based on any prominent journal, the selection of *Music Perception* suggests the articles surveyed share a certain connection to "real world" listening—the perception of music. Consequently, we believe it unlikely that journals less grounded in real-world listening would employ any fewer artificial sounds. Furthermore, as the flagship journal in the field, the selection of *Music Perception* helped ensure our survey was based on high quality, prominent research.

Method

In choosing a corpus for this survey we aimed to select literature representative of research on auditory perception to explore the prevalence of sounds synthesized with naturally decaying versus abruptly ending amplitude envelopes.

IDENTIFYING THE CORPUS OF SOUNDS TO SURVEY

Tirovolas and Levitin (2011) classified articles ($n = 384$) published between 1983 (inaugural year) and 2010, organizing them according to criteria such as stimulus, department, topic, subject, materials, and outcome measure. We focused on their category of "stimulus," identifying studies using isolated sounds and/or sequential sounds ($n = 120$). We omitted one article involving sound production rather than perception (Schneider, Münte, Rodriguez-Fornells, Sailer, & Altenmüller, 2010) and another primarily reviewing previously published data (D'Amato, 1988) A third article included

separate visual and auditory experiments (Breckler, Allen, & Konecni, 1985); and so we included the auditory but omitted the visual.

Therefore, our final corpus consisted of 118 empirical papers identified by Tirovolas & Levitin, (2011) as those using single tones or a series of tones. To accurately reflect all sounds employed, we separately analyzed individual experiments ($n = 222$), distributing one "point" in proportion to each sound's functional usage. For example, we assigned a half point to each of the two sounds used by Mikumo (1992), and one seventh of a point to each of the seven sounds used by Nielzén & Olsson (1993). The sum of these fractional points totaled 222—the number of experiments contained in the 118 articles.

CLASSIFYING TEMPORAL STRUCTURE

We initially planned to classify the sounds' temporal structures using three categories: (1) *flat*, (2) *percussive*, and (3) *other*. However, after beginning the project we recognized the need for an *undefined* category to account for sounds lacking sufficient envelope information. We considered sounds to be flat when they were described as having a period of sustain between (typically linear) onsets and offsets. For example, the temporal structure employed by Demany and Semal (1993) had "a total duration of 200 msec and was gated on and off with 10-msec linear amplitude ramps" (implying a sustain period of 180 ms). Other examples of flat descriptions include "All tones were 500 msec in length with smooth, rapid onset and offset" (Benguerel & Westdal, 1991) as well as those described as having a duration that "was always 100 msec including 10 msec rise-fall times" (Nakajima, Tsumura, Matsuura, & Minami, 1988). Depictions of sounds classified as flat appear in Figure 1b.

We classified sounds with a rapid onset followed by an immediate decay as percussive. Although rarely stated as such, it was often implied through description of the instruments and/or materials involved. For example, Marvin and Brinkman (2000) mentioned using piano notes without supplying explicit information about the notes' temporal structures. Because notes on the piano are produced by percussive (i.e., impact) events, we classified these sounds as percussive. Similarly, we placed sounds described as having a "cowbell-character" (Madison & Merker, 2005) or sounding "like a rim shot on small drum" (Nielzén & Olsson, 1993) in this group. This category also included impact sounds that were not produced by traditional percussion instruments, such as hand claps (Levitin & Bellugi, 1998). Depictions of sounds categorized as percussive appear in Figure 1c.

We classified sounds with defined temporal structures that were neither percussive nor flat in a third category—other. This included sounds produced by string instruments (Marvin & Brinkman, 2000), wind instruments (Kendall, Carterette, & Hajda, 1999), and the human voice (Patel, Wong, Foxton, Lochy, & Peretz, 2008). As with the piano tones, any instrumental description sufficed for inclusion even without explicit articulation of the tones' amplitude envelopes. Consequently, in addition to instrumental sounds this category also included several synthesized sounds, such as those with a "50 ms rise followed by linear decrease to zero" (Krumhansl & Schmuckler, 1986) as well as sounds described by diagrams, such the one provided by Kim and Iwamiya (2008) illustrating a 1000 ms linear onset followed by a 1000 ms linear offset. Figure 1d depicts several sounds classified in this category.

Finally, although not anticipated at the outset we found it necessary to create an "undefined" category for use when a sound's amplitude envelope was not described in sufficient detail to allow for classification either as flat, percussive, or other. Although we suspect many of these sounds were in fact flat, it would be inappropriate to make this assumption in the absence of sufficient detail (i.e., descriptions of tones falling in this category do not constrain the range of possible envelopes used).

ADDITIONAL STIMULUS INFORMATION

Surprised by the number of sounds lacking definitions of their envelopes, we began wondering whether this oversight indicated a broader lack of methodological detail. To explore this question we revisited each article using sounds with undefined envelopes in order to code the presence or absence of other information. This included stimulus characteristics such as spectral structure, duration, and intensity, as well as other seemingly less-relevant properties such as the exact model of tone generator/computer equipment used to synthesize the sounds and the exact model of headphones/speakers used to deliver them. To do so, we used the following criteria.

1. *Spectral information*: A short description of the sound's spectral information (i.e., pure tone/sine tone, white noise, square wave) followed by a range of frequencies or musical notes used for tones or harmonics, if applicable. If an instrument was used, then the instrument name was used as the descriptor of the spectral content. The range of pitches or frequencies played was included as well. Common sounds with a similar spectral envelope across
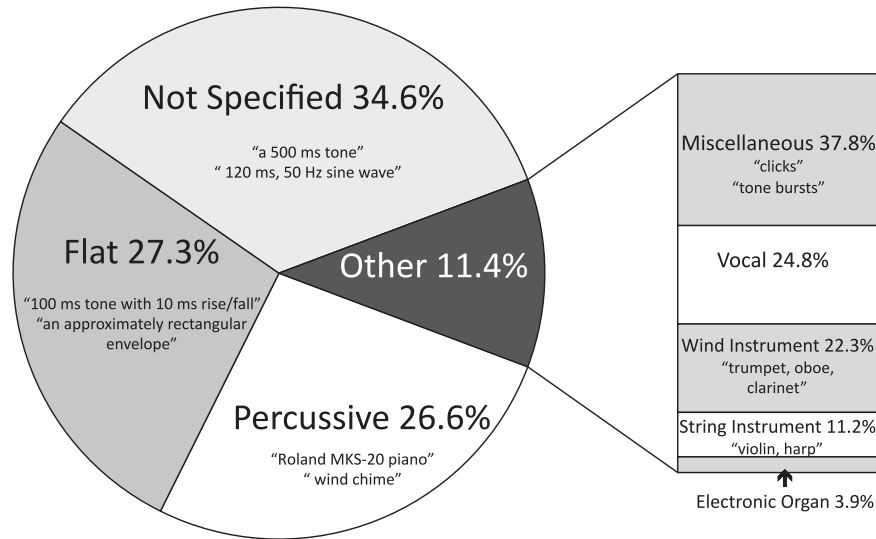
FIGURE 2. Pie chart of percentages of distribution of different auditory stimuli classifications. The other category is further subdivided in the rectangle to the right.

multiple studies were labeled with the name of that sound. For example, we used this approach for "Shepard tones," with bell shaped spectral envelopes. We also created a new category of *undefined (spectral information)* for situations in which this information was not given.

2. *Duration*: The temporal length of the sound. For experiments in which multiple sounds of the same category were used, a range of durations was included. If insufficient information was provided, then this category was labeled as *undefined (duration)*.

3. *Intensity*: We identified the sound pressure level as measured in decibels when specified, noting that some studies mentioned sounds being presented at a "comfortable loudness level." If insufficient information was provided, then this category was labeled as *undefined (intensity)*.

Results

Although the most interesting outcome of this survey is the high proportion of *undefined* sounds (see Figure 2 for a complete breakdown), we start this section first with a discussion of the defined sounds initially motivating this work.

SUMMARY OF THE "DEFINED" SOUNDS

We found "flat" sounds used in 27.3% of our 222 experiments (we assigned fractional weights for experiments involving multiple sounds). Their durations ranged from 7 ms to 2000 ms, and their fundamental frequencies from

65 Hz to 2960 Hz. We note that the majority of these sounds were pure tones (i.e., sine waves); others were complex tones with various combinations of harmonics.

"Percussive" sounds with naturally decaying envelopes were nearly as common; used in 26.6% of these experiments. The vast majority (80%) of these percussive sounds were piano tones, either produced by an explicitly specified synthesizer (Roland RD-250, Rhodes 760) or by an acoustic piano. This category also contained numerous percussive sounds such as a bass drum, conga, cowbell, gangsa, gender wayang, hand clap, hi-hat, Japanese style drum, snare drum, tom-tom, wind chime, and woodblock. While these instruments are acoustically distinct, all involve impacts and consequently exhibit decaying envelopes beginning immediately after onset completion.

Finally, despite its broad nature only 11.4% of the sounds in our survey fell into the "other" category. Vocal sounds were the most common (24.8% of sounds in the "other" category), followed by sounds produced by wind (22.3%), string (11.2%), and organ (3.9%) instruments. The remaining 37.8% of the "other" tones consisted of synthesized sounds generated with specified profiles that were neither flat nor percussive (i.e., clicks, "tone bursts" with non-flat envelopes, click trains, etc.).

SUMMARY OF THE "UNDEFINED" SOUNDS

The most interesting (and surprising) outcome of this survey is that 34.6% of the sounds encountered used envelopes that were undefined. That is not to say that

these experiments lacked methodological detail—many contained other descriptive details such as "pure tone" or "three-tone harmonic complex," offering useful stimulus information. In fact our additional classifications indicate that authors defined the spectral structure of approximately 85% of tones for which the temporal structure was undefined. Similarly, authors defined the intensity of 70%[2] and the durations of nearly 80% of tones with undefined envelopes. Although these studies contained varying levels of detail regarding intensity, duration, and spectral information, what they share in common is that they lacked any definition of their temporal structures.

It is important to note that this high level of detail for attributes other than amplitude envelope was not confined merely to stimulus characteristics such as spectral structure or duration, characteristics widely recognized as holding the possibility of meaningfully altering an experiment's outcome. Intriguingly, about 65% of tones with undefined envelopes were played over precisely specified headphones and/or speakers (Grason-Stadler TDH-49, Sennheiser 424, or Avant 2AX, etc.). Moreover, nearly 70% were generated with precisely specified equipment such as a Commodore 64, Roland JV-30, NeXT Computer, or VAX 11/780, etc. (curiously, many of these sounds consisted of pure tones or summations of pure tones). This impressive level of specificity regarding stimulus characteristics and equipment indicates that the lack of amplitude envelope information was not indicative of general inattentiveness to detail.

Discussion

Our lab's previous research illustrates that a sound's amplitude envelope (in particular its amount of decay) can qualitatively change outcomes of perceptual tasks ostensibly exploring unrelated issues (e.g., sensory integration, associative memory). Although many sounds heard in everyday listening are produced by impact events (Gaver, 1993a) and consequently have naturally decaying offsets, our intuition was that research on auditory perception frequently employs sounds with abruptly ending offsets—i.e., flat tones. This survey was motivated by a desire to formally test this intuition.

Although we found that percussive sounds were only slightly less numerous than flat (26.6% vs. 27.3%) in our survey, we were surprised to discover that 34.6% of the 222 experiments surveyed employed undefined temporal structures—in fact the most prominent of our four

categories. This finding raises important questions regarding interpretations of these experiments in light of work showing qualitatively different patterns of processing for pure tones synthesized with either flat or percussive envelopes, described below.

AMPLITUDE ENVELOPE AFFECTS EXPERIMENTAL OUTCOMES

Recent research using percussive and flat tones illustrates strikingly different outcomes on a variety of perceptual tasks in which envelope was not previously thought to play an important role. For example, they lead to qualitatively different patterns of audio-visual integration, an observation that challenges long-held assumptions about the proper theoretical framework for understanding multimodal interactions. For example, participants instructed to judge the durations of percussive and flat tones paired with point light representations of a marimba player were able to ignore the visual influence when listening to flat tones (Schutz, 2009). This finding is consistent with previous research demonstrating that visual information does not affect auditory judgments of event duration (Walker & Scott, 1981). However, these same participants were unable to ignore visual information when judging the duration of percussive tones, a finding consistent with earlier sensory integration experiments involving percussive sounds (Schutz & Kubovy, 2009b; Schutz & Lipscomb, 2007). This finding suggests that patterns of integration widely agreed upon (previously assessed primarily with flat tones) might not apply to real world sounds with natural decays.

These differences are not confined to sensory integration tasks, but are also documented in higher-level cognitive tasks involving associative memory. Participants listening to a series a 4-note melodies paired with various household objects (keys, alarm clocks, watches, etc.) remembered 60% more of the sequence-object associations when listening to sequences of percussive rather than flat tones (Schutz & Stefanucci, 2010). We suspect this difference in memory may stem from difficulties in the perceptual processing of tones with abruptly ending offsets that are fundamentally different than the kinds of sounds with natural decays heard frequently in our everyday listening. This idea has been tested formally in a third paradigm involving judgments of relative duration, suggesting different strategies may be employed for flat and percussive tones (Vallet et al., 2014). These findings are consistent with subsequent work suggesting differences in the neural processing of pure tones shaped with percussive and flat envelopes (Sandhu, Schutz, & Dyson, 2013). Because the full explanation for these findings remains undetermined, other

---

[2] About 50% were specified explicitly, with an additional 20% described as being at a "comfortable level."

members of our research team are currently exploring this issue to further understand the underlying differences in the perceptual processing of tones differing only in their amplitude envelopes. This work will complement and extend previous studies demonstrating the effect of amplitude envelope in a variety of auditory perception tasks (Grassi & Casco, 2009; Neuhoff, 2001; Schlauch, Ries, & DiGiovanni, 2001).

BROADER IMPLICATIONS

We surveyed articles published in *Music Perception* in part because of its focus on listening to natural (i.e., "acoustic") sounds rather than tone beeps or white noise as commonly employed in psychophysical studies. Consequently, we suspect research surveyed in this journal is, if anything, *less* inclined to feature sounds with artificially abrupt envelopes than articles in other journals publishing research on auditory perception (i.e., *Attention, Perception and Psychophysics*, *Hearing Research*, etc.). Yet flat tones appeared here slightly more often than percussive, and we suspect that the balance of these two envelope types in other journals would lean even more heavily towards flat.[3] We are unsure whether other journals will have a higher proportion of sounds with undefined amplitude envelopes; however, the high level of attention to methodology and the rigorous review process inherent in a journal of *Music Perception*'s caliber makes it unlikely other journals would exhibit significantly fewer such classifications.

Due to the ease of precise quantification/control available with flat tones, we suspect the majority of sounds with undefined envelopes in fact used flat temporal structures. However, it would be irresponsible to make this assumption based on the information provided, as researchers attempting to replicate the reported results would not be able to recreate the original stimuli from these descriptions. This implies that amplitude envelope may have been regarded as less important than the exact model of headphones/speakers used to deliver these sounds, and/or the precise model of computer used to synthesize sine waves and complex tones. This previous lack of interest[4] in a simple-to-manipulate property now known to affect experimental outcomes strongly suggests that amplitude envelope is a parameter ripe for

future exploration. Such work would inform our understanding of the processing of natural sounds heard outside the laboratory. Consequently, we strongly concur with earlier observations pointing to the value of employing paradigms and sounds "that are closer to real-world tasks faced by the auditory system" (Joris, Schreiner, & Rees, 2004). To aid such investigations, we are now sharing our software for generating percussive and flat tones (the same tool used to generate sounds in our lab's experiments – Figure 1, top panel) online at www.maplelab.net/software.

In conclusion, we hope this survey will help inspire greater interest in amplitude envelope, research with broad implications for listening in a wide variety of domains outside the laboratory. As sounds with different amplitude envelopes can lead to qualitatively different outcomes on a variety experiments (Grassi & Casco, 2009; Schutz & Kubovy, 2009b; Schutz & Lipscomb, 2007; Schutz & Stefanucci, 2010; Vallet et al., 2014), it is worthwhile to ponder the degree to which features of the sounds used in research should parallel features of the sounds heard in natural situations. Amplitude envelope is a parameter no less important than other stimulus characteristics more commonly defined within our survey (i.e., spectral information, duration and/or sound intensity), and likely more important than others (i.e., the specific model of tone generators and headphones/speakers used to deliver a pure tone). Therefore, we hope this survey illustrates the potential for fruitful future research exploring this important topic.

Author Note

Correspondence concerning this article should be addressed to Michael Schutz, McMaster University MAPLE Lab (School of the Arts), 1280 Main Street West, Hamilton, ON L8S 4M2. E-mail: schutz@mcmaster.ca

---

[3] This represents another testable hypothesis, and our lab is currently working on parallel surveys of other journals in order to explore this issue.

[4] Despite the lack of envelope definition found within this survey, we are not claiming its importance has gone completely without notice. Envelope has long been to recognized as playing an important role in the discrimination of timbre (Gordon, 1987; Strong & Clark, 1967), and more recently in predicting reading deficits (Goswami, 2011). However it is important to note that in these cases interest has traditionally focused on the *onset* rather than *offset* component of envelope—the key aspect on which the percussive and flat tones used in our lab's experiments differ.

References

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

ARMONTROUT, J., SCHUTZ, M., & KUBOVY, M. (2009). Visual determinants of a cross-modal illusion. *Attention, Perception, and Psychophysics, 71*, 1618-1627.

BENGUEREL, A. P., & WESTDAL, C. (1991). Absolute pitch and the perception of sequential musical intervals. *Music Perception, 9*, 105-119.

BERTELSON, P., VROOMEN, J., DE GELDER, B., & DRIVER, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception and Psychophysics, 62*, 321-332.

BOLTZ, M. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Perception, 18*, 427-454.

BRECKLER, S. J., ALLEN, R. B., & KONECNI, V. J. (1985). Mood-optimizing strategies in aesthetic-choice behavior. *Music Perception, 2*, 459-470.

DAHL, S., & FRIBERG, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception, 24*, 433-454.

DEMANY, L., & SEMAL, C. (1993). Pitch versus brightness of timbre: Detecting combined shifts in fundamental and formant frequency. *Music Perception, 11*, 1-14.

D'AMATO, M. (1988). A search for tonal pattern perception in cebus monkeys: Why monkeys can't hum a tune. *Music Perception, 5*, 453-480.

ERNST, M. O., & BANKS, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429-433.

GAVER, W. (1988). *Everyday listening and auditory icons.* (Unpublished doctoral dissertation). University of California, San Diego.

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

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

GIBSON, J. J. (1966). *The senses considered as perceptual systems.* Boston, MA: Houghton Mifflin.

GORDON, J. W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America, 82*, 88-105.

GOSWAMI, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences, 15*, 3-10.

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.

GUTTMAN, S. E., GILROY, L. A., & BLAKE, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological Science, 16*, 228-235.

JACKSON, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology, 5*, 37-41.

JORIS, P. X., SCHREINER, C. E., & REES, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews, 84*, 541-577.

KENDALL, R. A., CARTERETTE, E. C., & HAJDA, J. M. (1999). Perceptual and acoustical features of natural and tones synthetic orchestral instrument. *Music Perception, 16*, 327-363.

KIM, K. H., & IWAMIYA, S. I. (2008). Formal congruency between Telop patterns and sound effects. *Music Perception, 25*, 429-448.

KRUMHANSL, C. L., & SCHMUCKLER, M. A. (1986). The Petroushka chord: A perceptual investigation. *Music Perception, 2*, 153-184.

LEVITIN, D., & BELLUGI, U. (1998). Musical abilities in individuals with Williams Syndrome. *Music Perception, 15*, 357–389.

MADISON, G., & MERKER, B. (2005). Timing of action during and after synchronization with linearly changing intervals. *Music Perception, 22*, 441-459.

MARVIN, E. W., & BRINKMAN, A. R. (2000). The effect of key color and timbre on absolute pitch recognition in musical contexts. *Music Perception, 18*, 111-137.

MIKUMO, M. (1992). Encoding strategies for tonal and atonal melodies. *Music Perception, 10*, 73-81.

MOREIN-ZAMIR, S., SOTO-FARACO, S., & KINGSTONE, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17*, 154-163.

NAKAJIMA, Y., HOOPEN, G. T., & VAN DER WILK, R. (1991). A new illusion of time perception. *Music Perception, 8*, 431-448.

NAKAJIMA, Y., TSUMURA, T., MATSUURA, S., & MINAMI, H. (1988). Dynamic pitch perception for complex tones derived from major triads. *Music Perception, 6*, 1-20.

NEUHOFF, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology, 13*(2), 87–110.

NIELZÉN, S., & OLSSON, O. (1993). The influence of duration on verbal-attribute ratings of complex protomusical sounds. *Music Perception, 11*, 73-85.

NUSSECK, M., & WANDERLEY, M. M. (2009). Music and motion – How music-related ancillary body movements contribute to the experience of music. *Music Perception, 26*, 335-353.

PATEL, A., WONG, M., FOXTON, J., LOCHY, A., & PERETZ, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception, 25*, 357-368.

PLATZ, F., & KOPIEZ, R. (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception, 30*, 71-83.

Sandhu, R., Schutz, M., & Dyson, B. (2013, April). An ERP investigation of duration estimations of flat and percussive tones. *Poster presentation at the 20th Annual Meeting of the Cognitive Neuroscience Society*. San Francisco, CA.

Schneider, S., Münte, T., Rodriguez-Fornells, A., Sailer, M., & Altenmüller, E. (2010). Music-supported training is more efficient than functional motor training for recovery of fine motor skills in stroke patients. *Music Perception*, *27*, 271-280.

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.

Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review, 3*, 83-108.

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

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

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

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

Schutz, M., & Manning, F. (2012). Looking beyond the score: The musical role of percussionists' ancillary gestures. *Music Theory Online*, *18*, 1-14.

Schutz, M., & Stefanucci, J. (2010). Amplitude envelope and auditory alarms. In S. M. Demorest, S. J. Morrison, & P. S. Campbell (Eds.), *Proceedings of the International Conference on Music Perception and Cognition*. Seattle, WA: University of Washington.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Cross-modal dynamic capture: Congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 330-345.

Strong, W., & Clark, M. (1967). Perturbations of synthetic orchestral wind-instrument tones. *Journal of the Acoustical Society of America*, *41*, 277-285.

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

Vallet, G., Shore, D., & Schutz, M. (2014). *Exploring the role of amplitude envelope in duration estimation*. Manuscript submitted for publication.

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.