



OPEN

On the generalization of tones: A detailed exploration of non-speech auditory perception stimuli

Michael Schutz^{1,2}  & Jessica Gillard²

The dynamic changes in natural sounds' temporal structures convey important event-relevant information. However, prominent researchers have previously expressed concern that non-speech auditory perception research disproportionately uses simplistic stimuli lacking the temporal variation found in natural sounds. A growing body of work now demonstrates that some conclusions and models derived from experiments using simplistic tones fail to generalize, raising important questions about the types of stimuli used to assess the auditory system. To explore the issue empirically, we conducted a novel, large-scale survey of non-speech auditory perception research from four prominent journals. A detailed analysis of 1017 experiments from 443 articles reveals that 89% of stimuli employ amplitude envelopes lacking the dynamic variations characteristic of non-speech sounds heard outside the laboratory. Given differences in task outcomes and even the underlying perceptual strategies evoked by dynamic vs. invariant amplitude envelopes, this raises important questions of broad relevance to psychologists and neuroscientists alike. This lack of exploration of a property increasingly recognized as playing a crucial role in perception suggests future research using stimuli with time-varying amplitude envelopes holds significant potential for furthering our understanding of the auditory system's basic processing capabilities.

When designing research studies, scientists strive to minimize confounds potentially confusing experimental outcomes. The most famous cautionary tale of failing to control for extraneous variables can be found in Hans the counting horse, who delighted early 20th century audiences by appearing to answer basic arithmetic questions through sequential taps of his hoof. Subsequent investigation revealed the true source of his seemingly remarkable talent—rather than calculating, 'Clever Hans' merely recognized the reactions of humans who moved with excitement after seeing the correct number of taps¹. Although disappointing for his fans, it provided such an invaluable lesson in experimental control that it is still routinely discussed in introductory psychology textbooks^{2,3}—a century after Hans's debut.

Today, researchers take great pains to avoid confounding factors through carefully designed paradigms employing tightly controlled stimuli. Although this approach has undoubtedly contributed to psychology's success in explaining many complex phenomena, overuse of simplified tones in experiments can lead to inaccurate perspectives on perceptual processing. Here we examine this issue of broad importance through an in-depth study of the stimuli used to assess non-speech auditory perception, an exploration holding important implications for interpreting a wide body of perceptual research.

Controlled auditory stimuli

Sounds synthesized with temporal shapes ("amplitude envelopes") consisting of rapid onsets followed by sustain periods and rapid offsets afford precise quantification and description—qualities of obvious methodological value. However as William Gaver argued in a different context, fixating on simplistic sounds can lead researchers astray when attempting to explore the processes used in everyday listening^{4,5}. For example, a sound's amplitude envelope is rich in information, allowing listeners to discern the materials involved in an event^{6,7}, or even an event's outcome—such as whether a dropped bottle bounced or broke⁸. However, this cue is largely absent in synthesized tones with abrupt offsets, as their short decays provide no information about sound-producing events and materials. Therefore the simplistic structures of tone beeps, buzzes, and clicks do not necessarily trigger the

¹School of the Arts, McMaster University, Hamilton, Canada. ²Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Canada. ✉e-mail: schutz@mcmaster.ca

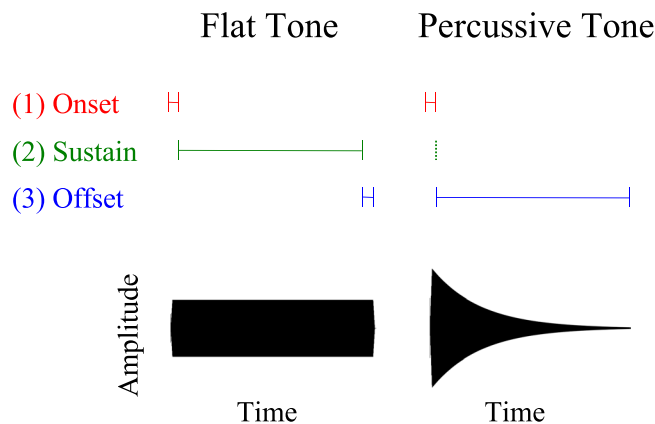


Figure 1. Flat and percussive amplitude envelopes. The rapid onset segment (1) is often similar in flat and percussive tones. The sustain segment (2) constitutes a large percentage of a flat tones, but is non-existent in percussive tones. Conversely the offset segment (3) is typically brief in flat tones whereas it constitutes the majority of percussive tones.

same perceptual processes as natural sounds—potentially complicating attempts to generalize from experimental outcomes to our processing of sounds outside the laboratory.

The ecological relevance of auditory stimuli outside of speech has ironically grown more problematic as the field evolves. Early experiments employed natural sounds such as balls dropping on hard surfaces and hammers striking plates⁹. However with the invention of the vacuum tube and then modern computers, many researchers eagerly traded natural sounds for precisely controlled tones¹⁰. Concern with this decision is hardly novel, as colleagues have previously expressed worry that much of auditory psychophysics “lack[s] any semblance of ecological validity”¹⁰ given the dearth of amplitude invariant (i.e. “flat”) tones in the natural world¹¹. Although some have articulated the merits of using stimuli with more varied amplitude envelopes¹², to the best of our knowledge there has been no large-scale formal exploration of non-speech auditory perception stimuli—a useful step in understanding the current state of the field so as to improve future approaches.

Amplitude Envelope’s Crucial Role in Perceptual Organization. Although amplitude envelope’s importance in timbre is widely recognized^{13–15}, its role in other perceptual constructs and processes has often received less attention. Consequently many experiments are conducted with a single type of amplitude envelope—the temporally simplistic flat tone. Their artificial characteristics embody the concern clearly articulated by Gaver^{4,5} and others¹⁰ warning of a divide between the auditory system’s use in everyday listening and its assessment in the laboratory. The following series of experiments on audio-visual integration illustrates one specific example of problems endemic with over-using a single type of stimulus to pursue a generalized understanding of psychological processes.

Videos of a renowned musician using long and short striking movements illustrate that vision can strongly affect judgments of musical note duration¹⁶. This illusion persists when using impact (but not sustained) sounds from other events¹⁷, point-light simplifications of the movements¹⁸, and even a single moving dot¹⁹. Curiously however, it breaks with widely-accepted thinking that vision exerts little influence on auditory judgments of event duration^{20–22}. This conflict has its roots in the dynamically decaying amplitude envelope (i.e. “sound shape”) of sounds created by natural impacts such as those produced by the marimba (Fig. 1). Further explorations demonstrate that pure tones shaped with the amplitude envelopes characteristic of impacts integrate with visual information, whereas the same pure tones shaped with flat amplitude envelopes (i.e., traditional “beeps”) do not²³. This illustrates that conclusions derived from experiments with flat tones do not necessarily generalize to real-world tasks, as their simplified temporal structures fail to trigger the same perceptual processes as natural sounds.

Amplitude envelope’s effect on audio-visual integration can be seen in other tasks. For example, a click simultaneous with two disks overlapping after moving across a screen increases the probability of perceiving a ‘bounce’ rather than the circles passing through one another²⁴. However, damped tones (i.e. decreasing in intensity over time) elicit stronger bounce percepts than ramped tones (i.e. increasing in intensity over time), presumably as they are event-consistent²⁵. These two studies illustrate that in addition to amplitude envelope affecting vision’s influence on audition^{16,17} it can affect audition’s influence on vision²⁵.

Repeated findings of amplitude envelope’s role in audio-visual integration^{17,23,25,26} complement a growing body of work on differences in the processing of tones with rapid increases vs. decreases in intensity (i.e., “ramped” or “looming” vs. “damped” or “receding”) in auditory processing. Although merely time-reversed and therefore spectrally matched, these sounds are perceived as differing in duration^{27–31}, loudness^{32–34}, and loudness change^{35,36}. These observations of differences in the perception of tones distinguished only by amplitude envelope shape raise questions about whether the disproportionate use of flat tones as experimental stimuli could lead to broader problems with generalization. For example, the durations of amplitude invariant tones can be evaluated using a ‘marker strategy’—marking tone onset and offset. This approach is consistent with Scalar Expectancy

| Journal | Database | Search Terms |
|------------------|----------------|--|
| APP (1966–2017) | PsycInfo | Attention, Perception & Psychophysics (Publication Name) AND Auditory (Identifier) NOT Speech (Identifier) NOT Language (Identifier) NOT Phonetic (Identifier) NOT Dialect (Identifier) |
| JEP (1975–2017) | PsycInfo | Journal of Experimental Psychology: Human Perception & Performance (Publication Name) AND Auditory (Identifier) NOT Speech (Identifier) NOT Language (Identifier) NOT Phonetic (Identifier) NOT Dialect (Identifier) NOT Word (Identifier) |
| HR (1978–2017) | Web of Science | Hearing Research (Publication Name) AND Auditory (Topic) NOT Speech (Topic) NOT Language (Topic) NOT Phonetic (Topic) NOT Dialect (Topic) Article (Document Type) Social Sciences & Arts Humanities (Categories) EXCLUDE Zoology (Subject Area) EXCLUDE Veterinary Sciences (Subject Area) EXCLUDE Plant Sciences (Subject Area) EXCLUDE Agriculture (Subject Area) EXCLUDE Food Science Technology (Subject Area) |
| JASA (1950–1975) | JASA Portal | Psychophysic* (full bibliographic record) NOT Speech (abstract/title/keyword) NOT Animal (abstract/title/keyword) |
| JASA (1976–2017) | Web of Science | Journal of the Acoustical Society of America (Publication Name) AND Psychophysic* (topic) NOT Animal (topic) NOT Speech (topic) |

Table 1. Summary of Search Terms.

Theory (SET), a widely accepted timing framework^{37,38}. However such a strategy would be problematic for sounds with decaying offsets, as their moment of acoustic completion is ambiguous (Fig. 1).

What sounds are used in auditory perception research? In order to explore the types of stimuli used to study non-speech auditory perception, we analyzed a representative sample of experiments drawn from several decades of four well-respected journals (two focused on general psychological research, and two with a specific auditory focus). This approach builds on our team's previous survey of *Music Perception*, which revealed surprisingly that over one-third of its studies omitted definition of amplitude envelope³⁹. That survey focused heavily on musical stimuli and examined only experiments using single tones or isolated series of tones. Furthermore, it drew unequally from different time periods, making it difficult to discern trends. In order to broaden our approach, here we conducted a survey (a) exploring a variety of non-speech auditory perception tasks, (b) incorporating diverse paradigms, (c) assessing multiple stimulus properties (i.e. spectral structure, duration), and (d) involving multiple journals widely recognized for their rigor and prestige. Consequently, this project offers useful insight into sounds used to explore the auditory system—the stimuli upon which numerous theories of perceptual processing are built.

Methods

In order to obtain a representative sample of experiments we used databases indexing articles in four highly regarded journals regularly publishing auditory perception research on human subjects. We initially began with two journals focused on general psychological processing: *Attention, Perception & Psychophysics* (henceforth referred to as APP) and *Journal of Experimental Psychology: Human Perception & Performance* (JEP)—both of which are indexed by *PsycInfo*. Later when expanding the survey to include the auditory-focused *Hearing Research* (HR) we turned to *Web of Science*, as HR is not indexed by *PsycInfo*. Although adequate for HR, *Web of Science* only indexes *Journal of the Acoustical Society of America* (JASA) back to 1976. Therefore we used *Web of Science* for articles published in or after 1976 to align as much as possible with our approaches to HR, and used *JASA Portal* for earlier articles.

Selection of articles to classify. Differences in each journal's scope necessitated slightly different search terms in order to obtain a consistent focus. For example, although our searches of APP and JEP naturally resulted in papers focused on human participants, an equivalent focus in HR required filtering out non-human animal studies. Similarly, whereas the wide range of psychophysical studies in APP and JEP necessitated use of the search term “audition”, this was unnecessary for JASA. However, JASA's broad acoustical focus, including issues such as underwater sound transmission^{40,41} instead compelled use of “psychophysic*” —a term obviously unnecessary for APP. Complete terms used are displayed in Table 1.

| Journal | Article Pool (N) | Selection | Sample (N) | Experiments (N) |
|------------------|------------------|--|------------------|-----------------|
| APP (1966–2017) | 466 | First 2 articles from each year | 104 | 228 |
| JEP (1975–2017) | 210 | First 4 articles from each year | 113 | 392 |
| HR (1978–2017) | 1820 | First 3 articles from each year involving human subjects | 114 | 180 |
| JASA (1950–2017) | 2126 | First 2 articles from each year | 112 ^a | 217 |
| Total | | | 443 | 1017 |

Table 2. Summary of Article Selection and Number of Experiments. ^aWeb of Science returned 2005 articles (1975–2017), and the JASA Portal 121 (1950–1975).

| Article | Exp. # | Sound Categories | Functional Category | Point Weight | Envelope Category |
|---|--------|------------------|---------------------|--------------|-------------------|
| Pfordresher, 2008 ¹³⁵ <i>Three experiments, each using a single sound</i> | 1 | 1 | auditory feedback | 1.00 | percussive |
| | 2 | 1 | auditory feedback | 1.00 | percussive |
| | 3 | 1 | auditory feedback | 1.00 | percussive |
| Kirby, Browning, Brennan, Spratford, & McCreery, 2015 ¹³⁶ <i>One experiment using two sounds</i> | 1 | 2 | reference | 0.50 | undefined |
| | | | target | 0.50 | undefined |
| Stilp, Alexander, Kiefe, & Kluender, 2010 ⁸² <i>Two experiments, each using three types of sounds</i> | 1 | 3 | target A | 0.333 | other |
| | 1 | | target B | 0.333 | other |
| | 1 | | precursor | 0.333 | other |
| | 2 | 3 | target A | 0.333 | other |
| | 2 | | target B | 0.333 | other |
| | 2 | | precursor | 0.333 | other |
| Møller & Jho, 1989 ¹³⁷ <i>One experiment, using three types of sounds</i> | 1 | 3 | signal A | 0.111 | flat |
| | | | signal A | 0.111 | flat |
| | | | signal A | 0.111 | flat |
| | | | signal B | 0.167 | click train |
| | | | signal B | 0.167 | click train |
| | | | masker | 0.333 | undefined |

Table 3. Examples of Point Weighting Distributions. Note: Each experiment received a single point, which was distributed equally amongst the functional categories of the sounds used.

This process resulted in a pool of 4622 potential articles. In order to select a manageable number we used a stratified quota sampling technique⁴², taking the first two to four articles per journal per year. This balanced competing desires for a sample representative of that journal's history and rough equivalence in the number of articles per journal. For example, we selected a maximum of two articles per year from JASA (dating back to 1950), but up to four per year for JEP (established in 1975). Adapted for our purposes based on best practices for accurate sampling in public opinion polls and market research⁴³, this approach yielded a final corpus of 443 papers split relatively evenly amongst the four journals (see Table 2).

Analysis and classification of individual experiments. We coded all experiments ($n = 1017$) individually within the 443 articles, classifying only the auditory components of multisensory stimuli. Due to the diversity of designs encountered, we fractionally distributed one point amongst all sound categories within each experiment—refining our team's earlier approaches. For example, if an experiment used two sound categories (i.e. a target and distractor), each sound category received a half point. In an experiment with four types of targets and two types of distractors, each target and distractor received 0.125 and 0.25 points respectively (sample point weightings appear in Table 3). This avoided over-emphasizing individual experiments using a large number of stimuli—such as the 64 different sounds employed by Gygi and Shafiro (2011).

Classification of Amplitude Envelope. We initially grouped sounds into one of five categories based on the descriptions given in the article and online links: (i) *flat*, (ii) *percussive*, (iii) *click train*, (iv) *other*, and (v) *undefined*. Our “flat” category included sounds with a period of invariant sustain and defined rise/fall times, such as “a 500-Hz sinusoid, 150 msec in duration...gated with a rise-decay time of 25 msec”⁴⁴. Similarly, we classified sounds described as “rectangularly gated”⁴⁵, having a “rectangular envelope”⁴⁶, “trapezoidal envelope”^{47,48}, “square-gate”⁴⁹, “fade-ins and fade-outs to avoid clicks”⁵⁰ or “abrupt onsets and offsets”⁵¹ as flat. Samples of sounds falling into this category appear in the top row of Fig. 2.

Our second category, “percussive,” encompassed sounds with sharp onsets followed by gradual decays with no sustain period (i.e. impact sounds). This included sounds from cowbells⁵², bongos⁵³, drums⁵⁴, chimes and bells⁵⁵, marimbas⁵⁶, vibraphones⁵⁷, and pianos (in which hammers impact strings)—both natural⁵⁸ and synthesized^{52,59}. Environmental impact sounds such as hand claps⁵⁵, footsteps⁶⁰, dropped⁶¹ and struck objects^{62,63} also fell into this

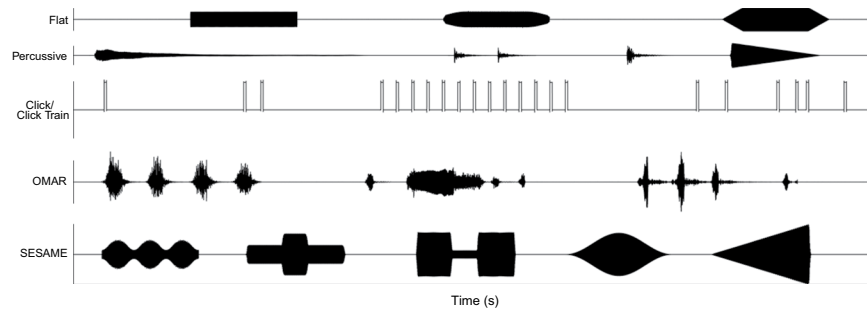


Figure 2. Examples of defined amplitude envelope categories: (a) various *Flat* tones, (b) *Percussive* sounds including a bell, hand claps, and bongo as well as a pure tone synthesized with a linear offset, (c) *Clicks* (left) and *Click trains* (right), (d) *OMAR* stimuli such as a dog barking, chicken clucking and bird chirping, and (e) *SESAME* stimuli including an amplitude modulated tone, two pedestal tones, a speedbump tone, rising tone.

category. In addition to natural sounds, this category included synthesized tones with ‘damped’ envelopes^{64–69}. For example, we considered a “target tone (5-ms rise time) ... [that] terminated with a 95-ms linear offset ramp”⁶⁸ to be a percussive ‘damped’ tone. Waveforms of stimuli categorized as percussive are shown in the second row of Fig. 2 and are summarized in detail in Supplemental Table 1.

Our third category of “click/click train” contained sounds described as clicks or a series of repeated stimuli over a short duration (refining our earlier approaches³⁹). This included sounds explicitly identified as “clicks”^{70,71} or “transients”⁷², as well as as “click trains”⁷³, “pulse trains”^{74,75}, “pulses in a train”⁷⁶, or stimuli “presented in rapid, successive bursts”⁷⁷. We also included click trains of variable rates⁷⁸ within this category (see third line of Fig. 2 for examples).

Our fourth category of “other” initially contained all sounds with defined amplitude envelopes other than those previously described. We subsequently split this category based upon referentiality—whether or not the sounds originated from real world events. Referential sounds included environmental sounds^{79–81}, recordings of animals such as dogs and/or chickens^{54,55}, and collections of sounds such as those heard at bowling alleys, beaches, and construction sites⁵⁴. This also included a variety of non-percussive musical sounds such as brass^{55,82,83}, string^{81,82}, and woodwind instruments^{57,84}, including instrument sounds later shortened^{85,86} or filtered⁸². Additionally, excerpts of popular music⁸⁷ as well as choral singing⁸⁸ fell into this category. We named this new group *OMAR* as it encompassed Other Musical And Referential sounds (i.e. referential sounds other than those included in the percussive category). Despite its broad nature this category ultimately contained the smallest percentage of sounds (fourth row of Fig. 2).

The *other* category also included non-referential sounds, i.e. those lacking a real-world referent. This includes amplitude modulated tones⁸⁹, pedestal tones^{90,91}, tones with defined rise/fall times and no sustain; both symmetric (e.g. 50 ms rise/fall time)^{92–95} and asymmetric (e.g. 15 ms rise 45 ms fall)⁹⁶, as well as reversed-damped or ‘ramped’ tones^{64,66,68,69,97}. We named this subcategory *SESAME*—Sounds Exhibiting Simple Amplitude Modulating Envelopes. These sounds include some amplitude variation beyond onset/offset, yet lack real world referents (note that although rising tones are often regarded as mimicking approaching sounds^{35,36}, this only holds if the approaching sounds are flat⁹⁸). Although this category’s definition is somewhat broad, it ultimately contained the second fewest number of stimuli (after *OMAR*). Depictions of these stimuli appear in the final line of Fig. 2, and Supplemental Table 1 provides a detailed breakdown of sounds classified under this category.

Finally, we used a fifth category of “undefined” for sounds whose amplitude envelopes could not be discerned from the information provided. For example, we classified the amplitude envelope of sounds described as ‘a 500 ms, 1000 Hz tone’ as undefined. We treated this as a category of last resort, using it only when unable to discern any information regarding temporal structure. For example, when authors stated they used stimuli defined in other papers^{99–102} or included links to online repositories^{55,103}, we obtained and analyzed the supplementary information. This avoided labeling stimuli as undefined when authors had merely been judicious with space.

Definition of six crucial properties. We also coded stimulus duration, as well as the presence or absence of information on additional characteristics such as spectral structure and intensity, and technical equipment details such as delivery device (i.e. headphone/speaker) and tone generator make/model. This expanded our team’s previous approach³⁹ of classifying these properties only for stimuli with undefined amplitude envelopes.

We created three categories for coding these properties: Specific, Approximate or Undefined (see Table 4 for examples). For example, we coded the intensity of stimuli described at “70 dB” as *Specific*, those “at a comfortable level” as *Approximate*, and those lacking any information on intensity as *Undefined*. Similarly, we coded delivery device information of “Sennheiser HD265 headphones” as *Specific*, general mention “headphones” as *Approximate*, and the lack of any information about sound delivery as *Undefined*. This helps contextualize our exploration of amplitude envelope by providing useful comparators for levels of definition of five other properties.

| Property | Undefined | Approximate | Specific |
|------------------------|------------------------------|---|---|
| Amplitude Envelope | No information provided | Trapezoidal envelope, MIDI Cowbell, Recording of bird song, Click train | 25 ms rise/fall time and 200 ms sustain, Train of five rectangular pulses (0.2 ms wide, 10 ms interval) presented bilaterally every 2 s |
| Spectral Structure | Tone, Signal, Stimulus | 500 Hz Tone (i.e. tone with specified fundamental) | Sinusoid/Pure Tone, Broadband White Noise, Piano, Bird Chirp |
| Duration | No information provided | Continuous, Variable, 60 ms-1000 ms (range) | 400 ms, 1.5 s, 20 μ s |
| Intensity | No information provided | Comfortable listening level, 30 dB above absolute threshold | 70 dB SPL, 88 dB |
| Delivery Device | No information provided | Headphones Loudspeaker Computer monitor speaker | TDH-49 headphones, Sennheiser HD265 headphones, Sony Speakers (Model SRS-A91) |
| Sound Generator/Source | No information provided | Digital recordings of Instruments, Pure-tone oscillator, 8-bit digital-to-analog converter | Grason-Stadler 455 C noise generator, IBM Thinkpad 560X with an audiocard (TDK DMC9000), Hewlett-Packard Model 200 ABR oscillator |

Table 4. Examples of Undefined, Approximate and Specific Descriptions of Properties.

Results and discussion

Our analysis illustrates a surprising lack of attention to the reporting of amplitude envelope, with 37.6% of stimuli from 1017 experiments omitting any information about their temporal structure (Fig. 3). This varied somewhat by journal: 53.1% (APP), 35.7% (JEP), 35.1% (HR), 26.9% (JASA), providing useful perspective on our team's previous survey of the journal *Music Perception*, which fell within this range³⁹. As the lack of definition is fairly consistent across duration categories (Fig. 4), it is not driven by the use of extremely short sounds in which amplitude changes would be imperceptible.

To contextualize the under-reporting of amplitude envelope, we compared its definition to that of other stimulus properties (spectral structure, duration, and intensity), as well as technical equipment information—such as the exact make and model of delivery device (e.g., Sennheiser HD265 headphones, Sony SRS-A91 Speakers) and sound generating equipment (e.g., Grason-Stadler 455 C noise generator, Hewlett-Packard Model 200 ABR oscillator) used. As shown in Table 5, we observed significantly less detail about amplitude envelope than most surveyed properties. Authors omitted duration information for only 16.7% of stimuli, and spectral structure for a mere 4.1%. This contrasts with amplitude envelope's lack of definition for 37.6% of stimuli—the highest of all properties surveyed. Curiously, we found authors significantly more likely to include the *exact model* of delivery device than *any information* about amplitude envelope ($\chi^2 = 5.87$, $p = 0.015$).

Interpreting the undefined tones (and illuminating the larger problem). Although the lack of definition regarding amplitude envelope is surprising, we believe the more important issue illuminated by this survey is the heavy focus on flat tones in non-speech auditory research. As shown in the grand summary of all four journals (pie chart in Fig. 3), flat tones formed the largest group in the survey—39.2% of sounds encountered. Clicks/Click trains formed the second largest group of defined stimuli (6.85%). Percussive sounds formed the third largest group (6.64%), followed by SESAME tones (5.63%) and OMAR sounds (4.08%). The use of flat tones outnumbered that of all other classifications combined—62.8% of defined stimuli. Furthermore, we strongly suspect that the vast majority of undefined stimuli are in fact flat.

Given the prominence of both the authors and journals surveyed, we find it unlikely that researchers neglected to disclose amplitude changes in their synthesized sounds. Additionally, based on feedback from conferences flat tones appear to serve as a go-to stimulus for assessing hearing, and we have often encountered surprise from colleagues when realizing that descriptions of a “short tone” could refer to anything else. Furthermore although their prevalence ranged considerably amongst journals, Fig. 3 shows remarkable consistency in “presumed flat” tones—a combination of the flat and undefined categories: 82.4% (APP), 74.2% (JEP), 73.9% (HR), 77.9% (JASA). For these reasons we strongly suspect that undefined tones are in fact flat. Therefore presumed flat tones constitute over three quarters (76.8%) of surveyed stimuli, with the majority of the remaining non-flat tones either Clicks/Click Trains or SESAME sounds.

The role of temporal complexity and referential sounds. In the process of defining stimulus categories for this project, we realized the utility of grouping sounds based on their referentiality—whether they refer to physical events. Both Percussive and OMAR sounds (Fig. 2) originate from real-world events outside the laboratory. Percussive sounds are created by musical instruments (drums, pianos) or natural impacts such as footsteps⁶⁰, as well as synthesized tones mimicking receding⁶⁹, departing⁶⁶ damped⁶⁴ or “dull”⁶⁸ sounds. OMAR sounds include musical tones produced by blowing or bowing (including synthesized versions), as well as soundscape recordings of the beach and/or forest⁵⁴ and specific events such as animal vocalizations^{80,83}, and water poured into a glass⁵⁴. We also consider sounds produced by helicopters⁷⁹ trains⁵⁵ and car engines¹⁰⁴ to be referential, as they are derived from physical events.

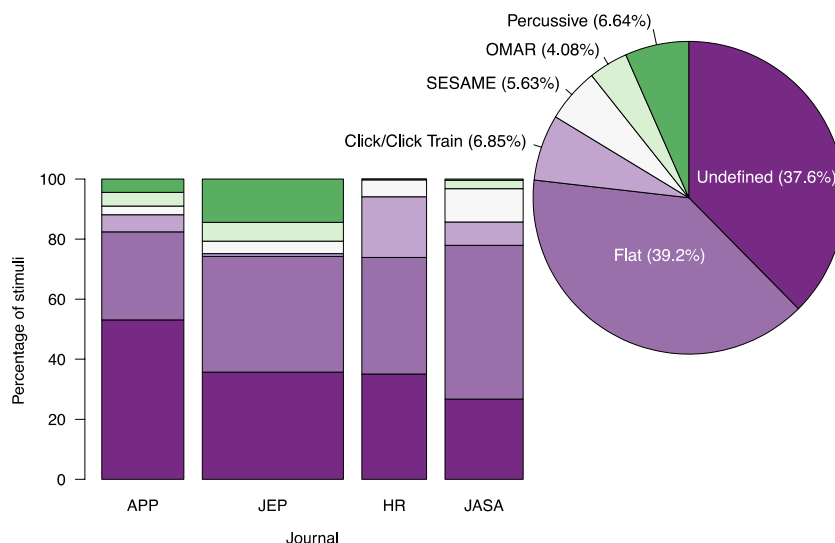


Figure 3. Amplitude envelope distribution. Bars indicate distribution within each journal, with width indicating the journal's relative points. JEP contained more multi experiment papers and therefore contributed the most points (see Table 2 for a detailed breakdown). Pie chart shows the grand summary across all stimuli.

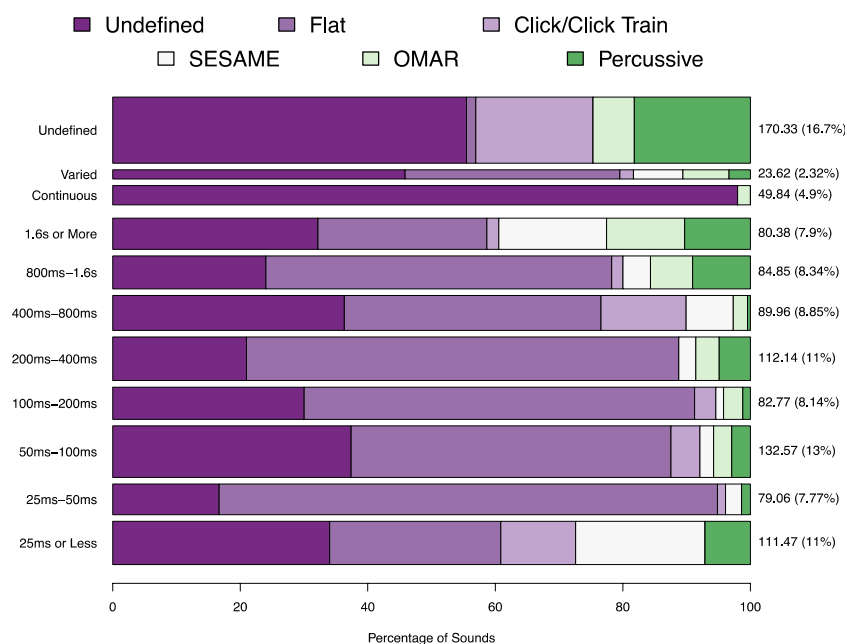


Figure 4. Distribution of stimuli by duration. The lack of definition is not confined to short sounds. The lowest row groups stimuli less than 25 ms, with each row doubling in duration. The top three rows indicate envelope distribution for stimuli with undefined durations (~17% of observed stimuli), as well as those with defined durations that varied, or sounded continuously (i.e., background noise). Bar width reflects relative number of points, with specific points (and percentages of total points) to the right of each bin.

Despite its broad definition, only 10.7% of the total stimuli encountered are referential (20.7% JEP; 9.0% APP; 3.2% JASA; 0.3% HR). Therefore 89.3% of these auditory stimuli have no connection to real-world events. As this sample is likely representative of non-speech auditory perception research as a whole, we consider this an important insight, given that everyday listening is so grounded in its utility for understanding the environment—such as using sound to inform our understanding of objects and events^{4,5}.

How have stimulus selections changed historically? In order to examine changes in stimulus selection over time, we grouped our data into five-year bins starting in 1970 and going back to 1950 (Fig. 5). This illustrates growth in the use of referential sounds, particularly in the last two decades. Although encouraging, it indicates less an embrace of complex sounds than a *broadening of research questions*. For example, this includes a

| Comparisons of defining amplitude envelope vs. other properties | | | |
|---|----------------------|----------------------|------------------------------|
| Property | Undefined Raw Points | Undefined Percentage | Chi Squared Test |
| Spectral Structure | 42.15 | 4.1% | $\chi^2 = 272.91, p < 0.001$ |
| Duration | 170.33 | 16.7% | $\chi^2 = 81.51, p < 0.001$ |
| Delivery Device | 200.17 | 19.7% | $\chi^2 = 57.12, p < 0.001$ |
| Intensity | 204.38 | 20.1% | $\chi^2 = 54.13, p < 0.001$ |
| Sound Generator/Source | 353.00 | 34.7% | $\chi^2 = 1.19, p = 0.275$ |
| Amplitude envelope | 382.63 | 37.6% | N/A |

Table 5. Definition levels of six properties. All other properties of sound coded were defined at significantly higher rates than amplitude envelope.

2013 study of how music affects tinnitus⁸⁷, a 2015 exploration of how airplane sound affects the taste of food¹⁰⁵, and a 2015 study of how street noise affects perception of naturalistic street scenes¹⁰⁶. Other tasks with referential sounds include a 2009 study of animal identification⁵⁵, and a 2008 study of identifying a walker's posture⁶⁰. Therefore this increased use of referential sounds appears to indicate an expansion of the types of questions investigated, rather than a reassessment of basic theories and models derived, tested, and refined with an overwhelming focus on temporally constrained stimuli.

Conclusions and Implications. Amplitude envelope's significance²³ in explaining why a novel audio-visual illusion breaks with accepted theory¹⁶ sparked our interest in understanding its importance in other aspects of auditory processing. Our team's findings regarding its role in audio-visual integration^{16,17,19,107} duration assessment²⁶, musical timbre¹⁰⁸, associative memory¹⁰⁹, and even perceived product value¹¹⁰ complement a growing literature with others documenting its importance in perceptual organization^{24,25,111}, as well as evaluations of event duration^{27–31}, loudness^{32–34}, and loudness change^{35,36}. Together, these studies suggest that research focused heavily on flat tones might overlook and/or misrepresent the capabilities and capacities of the auditory system. In several instances their disproportionate use has demonstrably led to faulty conclusions—for example misunderstanding the role of vision in duration estimation^{16,17,19,107}.

Despite long-standing speculation amongst leading figures in auditory perception^{5,10} and explicit notes of concern in the literature^{11,12,112}, to the best of our knowledge there has not previously been a detailed survey of this nature. Consequently our examination of over one thousand auditory experiments from four highly regarded journals offers three insights of broad relevance: (1) under-reporting of amplitude envelope, (2) defaulting to the use of flat tones for non-speech research, and (3) relatively little attention to the importance of referential aspects of sounds. We will now discuss each point in turn, placing them in the context of ongoing areas of inquiry.

Lack of attention to the reporting of amplitude envelope. The lack of attention to the reporting of amplitude envelope is our most surprising outcome. Well-respected authors publishing in highly regarded journals neglected to define amplitude envelope for 37.6% of stimuli. It is one thing to find a particular property to be under-researched; it is quite another to realize its importance has been so underappreciated that manuscripts fail to convey information about it in over one third of prominent auditory experiments. Although some may argue that descriptions such as “a 500 ms tone” imply flat tones, this ambiguous description fits a wide range of sounds. For example, all of the SESAME and flat stimuli shown in Fig. 2 are in fact 500 ms tones.

This lack of definition does not result from mere technicalities such as the prominence of very short tones (Fig. 4), or general inattention to methodological detail (Table 5). Curiously, our data suggest authors, reviewers and editors gave more emphasis to definition of the exact model of headphones used to deliver tone beeps, clicks, and bursts than any information regarding amplitude envelope. As every article included in this survey passed peer review in highly regarded journals, we see this oversight less as a failing of individual papers than as a cautionary note for the discipline as a whole. Among other concerns, this observation raises important questions regarding best scientific practice as researchers replicating these studies would in theory lack information needed to definitively recreate the sounds used. Our goal in clearly articulating this oversight is not to dismiss previous insights into the the auditory system, but merely to draw attention to the fact that this is an area in which we can improve as a discipline. Science progresses through critical reflection leading to refinement of best practices, and we are hopeful this survey will spark useful discussions about documentation in future research studies.

Encouragingly, we note a slight increase in the amount of specification of amplitude envelope over time, with fewer undefined stimuli in more recent years (Fig. 5). We are hopeful this trend will continue, as definition of this property can only help to further clarify our understanding of its important role.

Challenges with the use of flat tones as a default stimulus. More important than the lack of definition is the fact that flat tones account for over three quarters (77%) of stimuli encountered (when treating undefined tones as flat). As the survey drew upon on a representative selection of auditory research from four major journals, we believe this is indicative of standard approaches to auditory perception research. Flat tones hold certain methodological benefits such as avoiding potential confounds from associations with referential sounds, offering tight control, and/or minimizing variation between research teams. However, as they are processed differently than temporally varying sounds in a variety of contexts^{24–36,107,109–111} they should not be assumed to fully

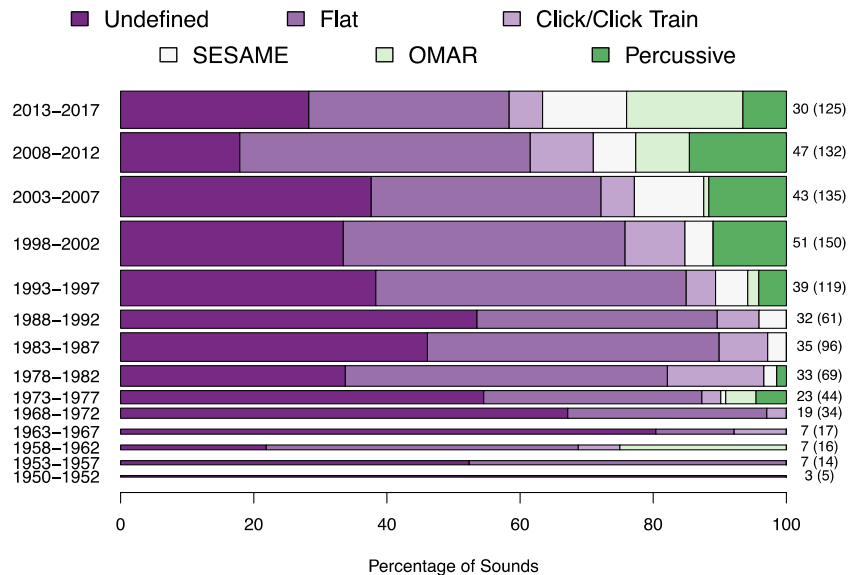


Figure 5. Changes in stimulus distribution over time. Researchers have used more diverse sounds in recent decades. However, note that even in the latest time bin, over half of stimuli surveyed are either flat or presumed flat, and less than 25% use referential sounds. Bar width indicates number of points associated with a given bin. Specific information on the number of papers appears to the right, with the number of points derived from these papers (i.e. the total number of experiments) in parenthesis. The earliest years are more sparsely sampled in part as they contain only JASA prior to 1966.

assess the limits or even the basic capabilities of the auditory system. Consequently, an over-reliance on flat tones poses serious problems for building a generalized picture of the auditory system's capabilities.

To draw a lesson from other areas of perceptual inquiry, visual researchers have long recognized that we cannot fully appreciate object recognition by assessing vision using only static, 2D images¹¹³. Although unmoving stimuli are methodologically convenient (simple to generate and easier to equate than moving images), overreliance on them overlooks the crucial importance of movement¹¹⁴. Consequently, a full understanding of the visual system requires stimuli exhibiting cues posing challenges for experimental control. In many ways temporal variation in amplitude is “auditory movement,” and previous research documents that amplitude envelope plays an important role in signalling both the materials involved in an event^{6,7} as well as the event's outcome. For example, amplitude envelope is helpful in understanding whether a dropped bottle bounced or broke⁸, as well as in determining an object's hollowness¹¹⁵. Research focused disproportionately on sounds lacking the kinds of complex dynamic properties found in natural sounds may overlook crucial aspects of auditory processing—much as visual research using only static images can overlook motion's role in visual processing.

The literature on duration assessment provides a useful example of potential problems arising from the overuse of flat tones (beyond numerous previously discussed examples in audio-visual integration). As mentioned in the Introduction, research on SET (Scalar Expectancy Theory)^{37,38} explores the perceptual processing of duration, positing in essence the use of a marker strategy—marking tone onset and offset and calculating the difference. However this strategy would be ill-suited for sounds with decaying offsets, which might instead be processed with a prediction strategy estimating tone completion from decay rate. A direct experimental test of duration assessment strategies found evidence consistent with the idea that different underlying strategies are used for sounds with flat tones and sounds with natural decays²⁶, which might help explain why flat tones elicit different experimental outcomes than sounds with time varying amplitude envelopes in various perceptual organization tasks^{23,25}. Although further research is needed to fully explore the issue, a bias towards the use of flat tones in assessing SET could lead to problematic situations where numerous experiments converge on and confirm one particular theoretical perspective for duration processing—which then fails to explain how duration is actually processed in natural sounds which often lack abrupt offsets.

Problems with the pervasive nature of non-referential sounds. In many ways we see the most important outcome of this survey to be that so few non-speech auditory stimuli—just over 10%—emerge from real world events. Intriguingly, closer exploration of these referential sounds reveals that the vast majority are used in experiments *requiring real-world referents*. For example, studies exploring the recognition of animal vocalizations⁵⁵, how street noise affects perception of street scenes¹⁰⁶, and whether a walker's posture can be identified by their footsteps⁶⁰ simply could not be conducted without animal vocalizations, street sounds, and walkers' footsteps respectively. Studies using referential sounds for traditional tasks such as sound localization^{85,86} and auditory-haptic interactions⁵⁸ constitute only a small fraction of the 10.7% of referential sounds encountered.

It appears that non-referential (and in particular flat) tones serve as the default auditory stimuli for non-speech research. Tone beeps, clicks and SESAME tones are used for the vast majority of research on core theoretical issues, such as the perception of loudness^{32–34,116} and duration¹¹⁷ as well as sound-in-noise detection^{48,89,118}

localization¹¹⁹, and stream segregation^{120,121}. This raises important questions about the stimuli best suited for exploring auditory processing—for although beeps and clicks offer precise control, the lack of real-world referents presents the perceptual system with sounds that differ in crucial ways from those encountered outside the lab¹⁰⁸.

Given that the perceptual system evolved in an environment where sounds emanate from events (i.e. rocks falling) and actors (i.e. animal vocalizations), the disproportionate use of non-referential sounds in its assessment can lead to problematic conclusions regarding fundamental processes. For example, research on the ‘unity assumption’¹²² and/or ‘identity decision’¹²³ explores the degree to which the kinds of supra-modal congruence cues pervasive in natural events affect cross-modal binding, a process essential for our ability to function in a multi-sensory world. This includes but is not limited to semantic congruencies^{124,125}, synesthetic correspondences¹²⁶, and learned associations between arbitrarily-paired stimuli¹²⁷. Understanding binding in this context requires the use of co-occurring sights and sounds (which are by definition referential). As this makes the tight control desirable for experiments challenging, research on the unity assumption serves as a useful domain for illustrating problems with the relative paucity of naturalistic sounds used in psychophysical experiments.

To apply controlled methodology to a domain that has long been studied with less rigorous methods, Vatakis and Spence documented stronger integration of gender-matched (vs. mis-matched) faces and voices, providing important evidence for the unity assumption in a tightly controlled psychophysical context¹²⁴. Subsequent expansions assessed whether non-speech events could trigger the unity assumption—such as notes played on the piano vs. classical guitar¹²⁸. They found videos of a piano key being depressed integrated similarly with the sound of a piano as well as a guitar (and that the guitar plucking gesture also integrated similarly with both sounds). Vatakis and Spence interpreted these data as indicating that event unity (i.e., the pairing of gestures and sounds emanating from the same event) had no meaningful effect on multi-modal binding. These outcomes along with others using non-musical impact sounds such as noises from objects being struck vs. dropped¹²⁸ and vocalizations by humans vs. monkeys¹²⁹ led to their conclusion that the unity assumption did not extend beyond speech.

Curiously, Vatakis and Spence’s experiments overlooked the crucial role of amplitude envelope. Notes produced by the piano and guitar share similar temporal structures, with a sharp attack and immediate decay resulting from either a hammer striking a string (piano) or the plucking of a string (guitar). Our team replicated their paradigm using notes from instruments with different amplitude envelopes—either percussive (marimba) or sustained (cello). In doing so, we found clear evidence for the unity assumption in a non-speech task¹⁰⁷, in contrast to its absence in a similar task involving piano/guitar pairs¹²⁸. This discrepancy is consistent with a broader literature on the importance of cross-modal congruency in the binding of impact sounds—particularly with respect to the role of amplitude envelope^{17,19,25,111}.

Although oversight of amplitude envelope’s crucial role in the unity assumption by an internationally renowned research team is surprising, it is consistent with the relative lack of attention to natural sounds. If only ~10% of stimuli have real-world referents, it is understandable that important distinctions within this category have gone overlooked. This illustrates one challenge with disproportionately using non-referential stimuli such as beeps, buzzes, and clicks. Sounds with temporal variations constitute the majority of our everyday listening—as well as the entirety of our evolutionary history. Yet they appear to be avoided whenever possible in basic non-speech auditory perception research. Although their complexity comes with obvious challenges, avoiding them risks overlooking the ways in which this same complexity is routinely and effectively used by the auditory system in basic processing—similar to problems using only static stimuli to understand object recognition¹¹⁴ which are gaining increasing attention in visual research¹¹³.

Final thoughts. Although most relevant to those working in audio-visual integration, there are at least three reasons why this survey holds important messages for the field of auditory perception as a whole. First, amplitude envelope is recognized as playing a role in shaping perception of musical timbre^{13–15,108} as well as duration^{26–31} loudness^{32–34}, loudness change^{35,36} and even associative memory¹⁰⁹. Consequently there is good reason to believe its importance could extend widely beyond the context in which it has been most clearly shown to play a role—audio-visual integration^{16,17,19,24,25,107,111}. Second, further evidence of amplitude envelope’s effects on key theories and models can only be discovered by recognizing the value of broadening our stimulus toolset. As contemporary sound synthesis programs can easily facilitate the precise generation of tones with more amplitude variation¹³⁰, the primary barrier to their use is no longer technical but historical—choosing flat tones by default. Consequently this survey illustrates trends difficult to observe from any single experiment, and provides unique insight into challenges with current approaches. Third, the use of time varying envelopes holds tremendous immediate potential for use in applied work. For example the International Electrotechnic Commission mandates the use of flat tones in many auditory alarms¹³¹, which is one (of several) well documented problems^{132,133}. Alternative amplitude envelope shapes can improve their suitability for wide-spread use¹³⁴ yet have been rarely explored to date. Therefore efforts to raise awareness of this issue are pertinent for the auditory community as a whole, and for projects both theoretical and applied. To aid with this issue we have also created an online tool offering interactive visualizations of our survey data at www.maplelab.net/survey.

In conclusion, we strongly encourage both (a) the greater specification of amplitude information and (b) the use of a more diverse stimulus set in future studies. To be clear, we do not think flat tones should be avoided entirely, nor should non-referential tones be eliminated from our repertoire. Both offer certain benefits, and in some situations are adequate or even ideal—particularly when a lack of previous associations is desirable. Our concern is not that such sounds are used in auditory research, but rather that they are *used so disproportionately*. Greater consideration of how experimental outcomes might vary with sounds exhibiting natural amounts of temporal complexity would help address concerns from leading researchers that the world is “[not] replete with examples of naturally occurring auditory pedestals [i.e., flat amplitude envelopes]”¹¹ and that more attention is needed to sounds with amplitude envelopes “closer to real-world tasks faced by the auditory system”¹².

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 7 February 2019; Accepted: 13 March 2020;

Published online: 12 June 2020

References

- Pfungst, O. *Clever Hans: (the horse of Mr. Von Osten.) A contribution to experimental animal and human psychology.* (Holt, Rinehart and Winston, 1911).
- Dewey, R. A. *Clever Hans. Psychology: An Introduction* (2007). Available at, https://www.intropsych.com/ch08_animals/clever_hans.html. (Accessed: 7th September 2018).
- Kalat, J. W. *Introduction to psychology.* (Brooks/Cole Publ., 1996).
- Gaver, W. How do we hear in the world?: Explorations in ecological acoustics. *Ecol. Psychol.* **5**, 285–313 (1993).
- Gaver, W. What in the world do we hear?: An ecological approach to auditory event perception. *Ecol. Psychol.* **5**, 1–29 (1993).
- Klatzky, R. L., Pai, D. K. & Krotkov, E. P. Perception of material from contact sounds. *Presence Teleoperators Virtual Environ.* **9**, 399–410 (2000).
- Lutfi, R. A. Human Sound Source Identification. in *Auditory Perception of Sound Sources* (eds. Yost, W. A., Fay, R. R. & Popper, A. N.) 13–42 (Springer, 2007).
- Warren, W. H. & Verbrugge, R. R. Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 704–712 (1984).
- Fechner, G. *Elements of psychophysics. Vol. I. Elements of psychophysics. Vol. I.* (New York, 1966).
- Neuhoff, J. G. *Ecological psychoacoustics.* (Elsevier Academic Press, 2004).
- Phillips, D. P., Hall, S. E. & Boehnke, S. E. Central auditory onset responses, and temporal asymmetries in auditory perception. *Hear. Res.* **167**, 192–205 (2002).
- Joris, P. X., Schreiner, C. E. & Rees, A. Neural processing of amplitude-modulated sounds. *Physiol. Rev.* **84**, 541–577 (2004).
- Grey, J. M. Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.* **61**, 1270–1277 (1977).
- McAdams, S., Winsberg, S., Donnadieu, S., de Soete, G. & Krimphoff, J. Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychol. Res.* **58**, 177–192 (1995).
- Rossing, T. D., Moore, R. F. & Wheeler, P. A. *The science of sound.* (Pearson Education Limited, 2013).
- Schutz, M. & Lipscomb, S. Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, <https://doi.org/10.1068/p5635> (2007).
- Schutz, M. & Kubovy, M. Causality and cross-modal integration. *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 1791–1810 (2009).
- Schutz, M. & Kubovy, M. Deconstructing a musical illusion: Point-light representations capture salient properties of impact motions. *Can. Acoust.* **37**, 23–28 (2009).
- Armontrout, J. A., Schutz, M. & Kubovy, M. Visual determinants of a cross-modal illusion. *Atten. Percept. Psychophys.*, <https://doi.org/10.3758/APP.71.7.1618> (2009).
- Guttman, S. E., Gilroy, L. A. & Blake, R. Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychol. Sci.* **16**, 228–235 (2005).
- Walker, J. T. & Scott, K. J. Auditory-visual conflicts in the perceived duration of lights, tones and gaps. *J. Exp. Psychol. Hum. Percept. Perform.* **7**, 1327–1339 (1981).
- Welch, R. B. & Warren, D. H. Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* **88**, 638–667 (1980).
- Schutz, M. *Crossmodal integration: The search for unity.* (University of Virginia, 2009).
- Sekuler, R., Sekuler, A. B. & Lau, R. Sound alters visual motion perception. *Nature* **385**, 308 (1997).
- Grassi, M. & Casco, C. Audiovisual bounce-inducing effect: Attention alone does not explain why the discs are bouncing. *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 235–243 (2009).
- Vallet, G., Shore, D. I. & Schutz, M. Exploring the role of amplitude envelope in duration estimation. *Perception* **43**, 616–630 (2014).
- Schlauch, R. S., Ries, D. T. & DiGiovanni, J. J. Duration discrimination and subjective duration for ramped and damped sounds. *J. Acoust. Soc. Am.* **109**, 2880–2887 (2001).
- Grassi, M. & Pavan, A. The subjective duration of audiovisual looming and receding stimuli. *Atten. Percept. Psychophys.* **74**, 1321–1333 (2012).
- Grassi, M. & Darwin, C. J. The subjective duration of ramped and damped sounds. *Percept. Psychophys.* **68**, 1382–1392 (2006).
- DiGiovanni, J. J. & Schlauch, R. S. Mechanisms responsible for differences in perceived duration for rising-intensity and falling-intensity sounds. *Ecol. Psychol.* **19**, 239–264 (2007).
- Grassi, M. Sex difference in subjective duration of looming and receding sounds. *Perception* **39**, 1424–1426 (2010).
- Ries, D. T., Schlauch, R. S. & DiGiovanni, J. J. The role of temporal-masking patterns in the determination of subjective duration and loudness for ramped and damped sounds. *J. Acoust. Soc. Am.* **124**, 3772–3783 (2008).
- Stecker, G. C. & Hafter, E. R. An effect of temporal asymmetry on loudness. *J. Acoust. Soc. Am.* **107**, 3358–3368 (2000).
- Teghtsoonian, R., Teghtsoonian, M. & Canévet, G. Sweep-induced acceleration in loudness change and the ‘bias for rising intensities’. *Percept. Psychophys.* **67**, 699–712 (2005).
- Neuhoff, J. G. An Adaptive Bias in the Perception of Looming Auditory Motion. *Ecol. Psychol.* **13**, 87–110 (2001).
- Neuhoff, J. G. Perceptual bias for rising tones. *Nature* **395**, 123–124 (1998).
- Machado, A. & Keen, R. Learning to time (LET) or scalar expectancy theory (SET)? A critical test of two models of timing. *Psychol. Sci.* **10**, 285–290 (1999).
- Gibbon, J. Scalar expectancy theory and Weber’s law in animal timing. *Psychol. Rev.* **84**, 279–325 (1977).
- Schutz, M. & Vaisberg, J. M. Surveying the temporal structure of sounds used in Music Perception. *Music Percept. An Interdiscip. J.* **31**, 288–296 (2014).
- Root, J. A. & Rogers, P. H. Performance of an underwater acoustic volume array using time-reversal focusing. *J. Acoust. Soc. Am.* **112**, 1869–1878 (2002).
- Yang, L. & Chen, K. Performance and strategy comparisons of human listeners and logistic regression in discriminating underwater targets. *J. Acoust. Soc. Am.* **138**, 3138–3147 (2015).
- Kothari, C. *Research methodology: methods and techniques.* Vasa (New Age International, 2004).
- Smith, T. M. F. On the validity of inferences from non-random sample. *J. R. Stat. Soc. Ser. A* **146**, 394–403 (1983).
- Watson, C. S. & Clopton, B. M. Motivated changes of auditory sensitivity in a simple detection task. *Percept. Psychophys.* **5**, 281–287 (1969).
- Robinson, C. E. Reaction time to the offset of brief auditory stimuli. *Percept. Psychophys.* **13**, 281–283 (1973).
- Franěk, M., Mates, J., Radil, T., Beck, K. & Pöppel, E. Sensorimotor synchronization: Motor responses to regular auditory patterns. *Percept. Psychophys.* **49**, 509–516 (1991).
- McAnally, K. I. & Calford, M. B. A psychophysical study of spectral hyperacuity. *Hear. Res.* **44**, 93–96 (1990).

48. Treisman, M. & Faulkner, A. The setting and maintenance of criteria representing levels of confidence. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 119–139 (1984).
49. Mott, J. B., Norton, S. J., Neely, S. T. & Warr, W. B. Changes in spontaneous otoacoustic emissions produced by acoustic stimulation of the contralateral ear. *Hear. Res.* **38**, 229–242 (1989).
50. Bertelson, P., Vroomen, J., de Gelder, B. & Driver, J. The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept. Psychophys.* **62**, 321–332 (2000).
51. Hübner, R. & Hafter, E. R. Cuing mechanisms in auditory signal detection. *Percept. Psychophys.* **57**, 197–202 (1995).
52. Pfordresher, P. Q. & Palmer, C. Effects of hearing the past, present, or future during music performance. *Percept. Psychophys.* **68**, 362–376 (2006).
53. Radeau, M. & Bertelson, P. Cognitive factors and adaptation to auditory–visual discordance. *Percept. Psychophys.* **23**, 341–343 (1978).
54. Gygi, B. & Shafiro, V. The incongruity advantage for environmental sounds presented in natural auditory scenes. *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 551–565 (2011).
55. Gregg, M. K. & Samuel, A. G. The importance of semantics in auditory representations. *Attention, Perception, Psychophys.* **71**, 607–619 (2009).
56. Keller, P. E., Dalla Bella, S. & Koch, I. Auditory imagery shapes movement timing and kinematics: Evidence from a musical task. *J. Exp. Psychol. Hum. Percept. Perform.* **36**, 508–513 (2010).
57. Bey, C. & McAdams, S. Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *J. Exp. Psychol. Hum. Percept. Perform.* **29**, 267–279 (2003).
58. Rinaldi, L., Lega, C., Cattaneo, Z., Girelli, L. & Bernardi, N. F. Grasping the sound: Auditory pitch influences size processing in motor planning. *J. Exp. Psychol. Hum. Percept. Perform.* **42**, 11–22 (2016).
59. Repp, B. H. Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 600–621 (2001).
60. Pastore, R. E., Flint, J., Gaston, J. R. & Solomon, M. J. Auditory event perception: The source–perception loop for posture in human gait. *Percept. Psychophys.* **70**, 13–29 (2008).
61. Grassi, M. Do we hear size or sound? Balls dropped on plates. *Percept. Psychophys.* **67**, 274–284 (2005).
62. Wagman, J. B. & Abney, D. H. Transfer of recalibration from audition to touch: Modality independence as a special case of anatomical independence. *J. Exp. Psychol. Hum. Percept. Perform.* **38**, 589–602 (2012).
63. Kunkler-Peck, A. J. & Turvey, M. T. Hearing shape. *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 279–294 (2000).
64. Carlyon, R. P. Spread of excitation produced by maskers with damped and ramped envelopes. *J. Acoust. Soc. Am.* **99**, 3647–3655 (1996).
65. Golubock, J. L. & Janata, P. Keeping timbre in mind: Working memory for complex sounds that can't be verbalized. *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 399–412 (2013).
66. Cusack, R., Deeks, J., Aikman, G. & Carlyon, R. P. Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 643–656 (2004).
67. Lewkowicz, D. J. Perception of auditory–visual temporal synchrony in human infants. *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 1094–1106 (1996).
68. Mondor, T. A., Zatorre, R. J. & Terrio, N. A. Constraints on the selection of auditory information. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 66–79 (1998).
69. McGuire, A. B., Gillath, O. & Vitevitch, M. S. Effects of mental resource availability on looming task performance. *Attention, Perception, Psychophys.* **78**, 107–113 (2016).
70. Berg, K. M. Temporal masking level differences for transients: Further evidence for a short-term integrator. *Percept. Psychophys.* **37**, 397–406 (1985).
71. Ikeda, K. Binaural interaction in human auditory brainstem response compared for tone-pips and rectangular clicks under conditions of auditory and visual attention. *Hear. Res.* **325**, 27–34 (2015).
72. Wit, H. P. & Ritsma, R. J. Evoked acoustical responses from the human ear: Some experimental results. *Hear. Res.* **2**, 253–261 (1980).
73. Shinn-Cunningham, B. Adapting to remapped auditory localization cues: A decision-theory model. *Percept. Psychophys.* **61**, 33–47 (2000).
74. Pollack, I. Discrimination of restrictions in sequentially blocked auditory displays: Shifting block designs. *Percept. Psychophys.* **9**, 335–338 (1971).
75. Zhu, Z., Tang, Q., Zeng, F.-G., Guan, T. & Ye, D. Cochlear-implant spatial selectivity with monopolar, bipolar and tripolar stimulation. *Hear. Res.* **283**, 45–58 (2012).
76. Richardson, B. L. & Frost, B. J. Tactile localization of the direction and distance of sounds. *Percept. Psychophys.* **25**, 336–344 (1979).
77. Soto-Faraco, S., Spence, C. & Kingstone, A. Cross-modal dynamic capture: Congruency effects in the perception of motion across sensory modalities. *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 330–345 (2004).
78. Riedel, H. & Kollmeier, B. Auditory brain stem responses evoked by lateralized clicks: Is lateralization extracted in the human brain stem? *Hear. Res.* **163**, 12–26 (2002).
79. Gregg, M. K. & Samuel, A. G. Change deafness and the organizational properties of sounds. *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 974–991 (2008).
80. Fairnie, J., Moore, B. C. J. & Remington, A. Missing a trick: Auditory load modulates conscious awareness in audition. *J. Exp. Psychol. Hum. Percept. Perform.* (2016).
81. Mayr, S. & Buchner, A. Evidence for episodic retrieval of inadequate prime responses in auditory negative priming. *J. Exp. Psychol. Hum. Percept. Perform.* **32**, 932–943 (2006).
82. Stilp, C. E., Alexander, J. M., Kiefe, M. & Kluender, K. R. Auditory color constancy: Calibration to reliable spectral properties across nonspeech context and targets. *Attention, Perception, Psychophys.* **72**, 470–480 (2010).
83. McAnally, K. I. *et al.* A dual-process account of auditory change detection. *J. Exp. Psychol. Hum. Percept. Perform.* **36**, 994–1004 (2010).
84. Gates, A., Bradshaw, J. L. & Nettleton, N. C. Effect of different delayed auditory feedback intervals on a music performance task. *Percept. Psychophys.* **15**, 21–25 (1974).
85. Möller, M., Mayr, S. & Buchner, A. Target localization among concurrent sound sources: No evidence for the inhibition of previous distractor responses. *Attention, Perception, Psychophys.* **75**, 132–144 (2013).
86. Möller, M., Mayr, S. & Buchner, A. Effects of spatial response coding on distractor processing: Evidence from auditory spatial negative priming tasks with keypress, joystick, and head movement responses. *Attention, Perception, Psychophys.* **77**, 293–310 (2015).
87. Vanneste, S. *et al.* Does enriched acoustic environment in humans abolish chronic tinnitus clinically and electrophysiologically? A double blind placebo controlled study. *Hear. Res.* **296**, 141–148 (2013).
88. Valente, D. L., Braasch, J. & Myrbeck, S. A. Comparing perceived auditory width to the visual image of a performing ensemble in contrasting bi-modal environments. *J. Acoust. Soc. Am.* **131**, 205–217 (2012).
89. Riecke, L., van Opstal, A. J. & Formisano, E. The auditory continuity illusion: A parametric investigation and filter model. *Percept. Psychophys.* **70**, 1–12 (2008).

90. Bonnel, A.-M. & Hafter, E. R. Divided attention between simultaneous auditory and visual signals. *Percept. Psychophys.* **60**, 179–190 (1998).
91. Green, D. M. & Nguyen, Q. T. Profile analysis: detecting dynamic spectral changes. *Hear. Res.* **32**, 147–163 (1988).
92. Wright, B. A. & Fitzgerald, M. B. The time course of attention in a simple auditory detection task. *Percept. Psychophys.* **66**, 508–516 (2004).
93. Bacon, S. P. & Healy, E. W. Effects of ipsilateral and contralateral precursors on the temporal effect in simultaneous masking with pure tones. *J. Acoust. Soc. Am.* **107**, 1589–1597 (2000).
94. Killan, E. C. & Kapadia, S. Simultaneous suppression of tone burst-evoked otoacoustic emissions—effect of level and presentation paradigm. *Hear. Res.* **212**, 65–73 (2006).
95. Moore, B. C. J., Glasberg, B. R. & Roberts, B. Refining the measurement of psychophysical tuning curves. *J. Acoust. Soc. Am.* **76**, 1057–1066 (1984).
96. Hasuo, E., Nakajima, Y., Osawa, S. & Fujishima, H. Effects of temporal shapes of sound markers on the perception of interonset time intervals. *Attention, Perception, Psychophys.* **74**, 430–445 (2012).
97. Mondor, T. A. & Terrio, N. A. Mechanisms of perceptual organization and auditory selective attention: The role of pattern structure. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 1628–1641 (1998).
98. Schutz, M. Clarifying amplitude envelope's crucial role in auditory perception. *Can. Acoust.* **44**, 42–43 (2016).
99. Henning, G. B. & Ashton, J. The effect of carrier and modulation frequency on lateralization based on interaural phase and interaural group delay. *Hear. Res.* **4**, 185–194 (1981).
100. Roberts, R. A., Koehnke, J. & Besing, J. Effects of reverberation on fusion of lead and lag noise burst stimuli. *Hear. Res.* **187**, 73–84 (2004).
101. Zwicker, E. & Henning, G. B. The four factors leading to binaural masking-level differences. *Hear. Res.* **19**, 29–47 (1985).
102. Gaskell, H. & Henning, G. B. Forward and backward masking with brief impulsive stimuli. *Hear. Res.* **129**, 92–100 (1999).
103. Visscher, K. M., Kahana, M. J. & Sekuler, R. Trial-to-trial carryover in auditory short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **35**, 46–56 (2009).
104. Keshavarz, B., Campos, J. L., DeLucia, P. R. & Oberfeld, D. Estimating the relative weights of visual and auditory tau versus heuristic-based cues for time-to-contact judgments in realistic, familiar scenes by older and younger adults. *Attention, Perception, Psychophys.* **79**, 929–944 (2017).
105. Yan, K. S. & Dando, R. A crossmodal role for audition in taste perception. *J. Exp. Psychol. Hum. Percept. Perform.* **41**, 590–596 (2015).
106. Tan, J. & Yeh, S. Audiovisual integration facilitates unconscious visual scene processing. *J. Exp. Psychol. Hum. Percept. Perform.* **41**, 1325–1335 (2015).
107. Chuen, L. & Schutz, M. The unity assumption facilitates cross-modal binding of musical, non-speech stimuli: The role of spectral and amplitude cues. *Attention, Perception, Psychophys.* **78**, 1512–1528 (2016).
108. Schutz, M. Acoustic structure and musical function: Musical notes informing auditory research. in *The Oxford Handbook on Music and the Brain* (eds. Thaut, M. H. & Hodges, D. A.) (Oxford University Press).
109. Schutz, M., Stefanucci, J., Baum, S. H. & Roth, A. Name that percussive tune: Associative memory and amplitude envelope. *Q. J. Exp. Psychol.* **70**, 1323–1343 (2017).
110. Schutz, M. & Stefanucci, J. Hearing value: Exploring the effects of amplitude envelope on consumer preference. *Ergon. Des. Q. Hum. Factors Appl.*
111. Grassi, M. & Casco, C. Audiovisual bounce—inducing effect: When sound congruence affects grouping in vision. *Attention, Perception, Psychophys.* **72**, 378–386 (2010).
112. Li, X., Logan, R. J. & Pastore, R. E. Perception of acoustic source characteristics: walking sounds. *J. Acoust. Soc. Am.* **90**, 3036–3049 (1991).
113. Snow, J. C. *et al.* Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects. *Sci. Rep.* **1**, 1–10 (2011).
114. Gibson, J. J. The visual perception of objective motion and subjective movement. *Psychol. Rev.* **61**, 304–314 (1954).
115. Lutfi, R. A. Auditory detection of hollowness. *J. Acoust. Soc. Am.* **110**, 1010–1019 (2001).
116. Humes, L. E. A psychophysical evaluation of the dependence of hearing protector attenuation on noise level. *J. Acoust. Soc. Am.* **73**, 297–311 (1983).
117. Schwent, V. L., Snyder, E. & Hillyard, S. A. Auditory evoked potentials during multichannel selective listening: Role of pitch and localization cues. *J. Exp. Psychol. Hum. Percept. Perform.* **2**, 313–325 (1976).
118. Hicks, M. L. & Bacon, S. P. Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing. *J. Acoust. Soc. Am.* **105**, 326–338 (1999).
119. McCarthy, L. & Olsen, K. N. A. “looming bias” in spatial hearing? Effects of acoustic intensity and spectrum on categorical sound source localization. *Attention, Perception, Psychophys.* **79**, 352–362 (2017).
120. Carlyon, R. P., Cusack, R., Foxtton, J. & Robertson, I. H. Effects of attention and unilateral neglect on auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 115–127 (2001).
121. Handel, S., Weaver, M. S. & Lawson, G. Effect of rhythmic grouping on stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* **9**, 637–651 (1983).
122. Welch, R. B. Meaning, attention, and the “unity assumption” in the intersensory bias of spatial and temporal perceptions. In *Advances in Psychology* **129**, 371–387 (Elsevier, 1999).
123. Bedford, F. L. Analysis of a constraint on perception, cognition, and development: One object, one place, one time. *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 907–912 (2004).
124. Vatakis, A. & Spence, C. Crossmodal binding: Evaluating the ‘unity assumption’ using audiovisual speech stimuli. *Percept. Psychophys.* **69**, 744–756 (2007).
125. Margiotoudi, K., Kelly, S. & Vatakis, A. Audiovisual temporal integration of speech and gesture. *Procedia - Soc. Behav. Sci.* **126**, 154–155 (2014).
126. Parise, C. V. & Spence, C. ‘When birds of a feather flock together’: Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One* **4**, 1–7 (2009).
127. Ernst, M. O. Learning to integrate arbitrary signals from vision and touch. *J. Vis.* **7**, 1–14 (2007).
128. Vatakis, A. & Spence, C. Evaluating the influence of the ‘unity assumption’ on the temporal perception of realistic audiovisual stimuli. *Acta Psychol. (Amst.)* **127**, 12–23 (2008).
129. Vatakis, A., Ghazanfar, A. A. & Spence, C. Facilitation of multisensory integration by the ‘unity effect’ reveals that speech is special. *J. Vis.* **8**, 1–11 (2008).
130. Ng, M. & Schutz, M. Seeing sound: A new tool for teaching music perception principles. *Can. Acoust.* **45** (2017).
131. Commission, I. E. International Standard IEC 60601: Medical electrical equipment. *Part 1-8 Gen. Requir. safety. Collat. Stand. Gen. Requir. tests Guid. Alarm Syst. Med. Electr. Equip. Med. Electr. Syst.* (2006).
132. Edworthy, J. Medical audible alarms: A review. *J. Am. Med. Informatics Assoc.* **20**, 584–589 (2013).
133. Edworthy, J. *et al.* The Recognizability and Localizability of Auditory Alarms: Setting Global Medical Device Standards. *Hum. Factors J. Hum. Factors Ergon. Soc.* **59**, 1108–1127 (2017).

134. Sreetharan, S. & Schutz, M. Improving Human–Computer Interface Design through Application of Basic Research on Audiovisual Integration and Amplitude Envelope. *Multimodal Technol. Interact.* **3**, 4 (2019).
135. Pfordresher, P. Q. Auditory feedback in music performance: The role of transition-based similarity. *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 708–725 (2008).
136. Kirby, B. J., Browning, J. M., Brennan, M. A., Spratford, M. & McCreery, R. W. Spectro-temporal modulation detection in children. *J. Acoust. Soc. Am.* **138**, EL465–EL468 (2015).
137. Møller, A. R. & Jho, H. D. Response from the exposed intracranial human auditory nerve to low-frequency tones: Basic characteristics. *Hear. Res.* **38**, 163–176 (1989).

Acknowledgements

We would like to thank Jonathan Vaisberg, Jennifer Harris, Fiona Manning, Aimee Battcock, and Lorraine Chuen, for their assistance in earlier versions of this work, as well as Cam Anderson for assistance in creating Fig. 2. We would also like to acknowledge Ian Bruce for clarify the structure of complex sounds used in Hearing Research, and Peter Pfordersher for helpful feedback on an earlier version of this manuscript. Finally we found specific suggestions from Dr. Nina Krauss to include journals with an auditory focus, and Dr. Ben Dyson for conversations leading us to broaden our exploration of under-reported properties. We are grateful for funding from NSERC (Natural Science and Engineering Council of Canada), CFI-LOF (Canadian Foundation for Innovation Leaders Opportunity Fund), the Ontario Earlier Research award, and the McMaster Arts Research Board for this project.

Author contributions

J.G. completed coding of stimulus characteristics, most figure preparation, and writing of the methods section. M.S. took the lead in design, directing the data analysis, funding the project, and writing. Both authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-63132-2>.

Correspondence and requests for materials should be addressed to M.S.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020