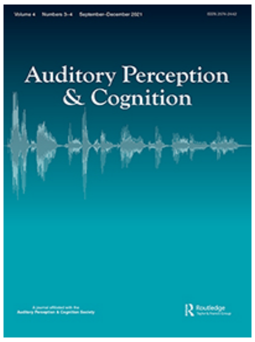

Auditory Perception & Cognition

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rpac20

Does Affective Content of Sounds Affect Auditory Time-to-collision Estimation?

Adam M. Braly, Patricia DeLucia & Daniel Oberfeld

AUDITORY PERCEPTION & COGNITION
2021, VOL. 4, NOS. 3–4, 212–234
https://doi.org/10.1080/25742442.2021.1997064

Does Affective Content of Sounds Affect Auditory Time-to-collision Estimation?

Adam M. Braly [a,b], Patricia DeLucia and Daniel Oberfeld [c]

[a]Department of Psychological Sciences, Texas Tech University, Lubbock, Texas, USA; [b]Department of Psychological Sciences, Rice University, Houston, Texas, USA; [c]Department of Psychology, Johannes Gutenberg-Universität, Mainz, Germany

ABSTRACT

In prior studies, mean estimates of time-to-collision (TTC) of approaching objects were shorter for threatening pictures (e.g., snakes) than non-threatening pictures (e.g., rabbits) but judgments of analogous auditory stimuli were not examined. The aim of the present study was to determine whether the affective content of an approaching object presented in the auditory domain can affect TTC estimates. Using simulations of sound-emitting objects selected from the International Affective Digitized Sound system, we compared TTC judgments of approaching objects that were threatening (hissing rattlesnake, buzzing bee, dentist drill) and control objects that had the same loudness and spectral properties but were unidentifiable. The snake sound, but not the drill or bee sounds, was judged as arriving earlier than the corresponding control sound, providing partial support that previously reported threat-related effects on TTC estimates for visual objects can also occur in the auditory domain.

ARTICLE HISTORY
Received 28 May 2020
Accepted 25 September 2021

KEYWORDS
Time-to-collision; audition; tau; affect; motion in depth

Introduction

Avoiding a collision with a potentially dangerous moving object is paramount for survival. Before an action can be executed to avoid a collision, the observer must estimate the time remaining before a collision will occur, or the time-to-collision (TTC). Most prior studies of TTC estimation focused on situations in which visual information about TTC is available in the optical expansion pattern of an approaching object (e.g., tau; Hoyle, 1957; Lee, 1976). Examples include driving (e.g., Caird & Hancock, 1994; DeLucia & Tharanathan, 2009; Levulis et al., 2015), catching (e.g., Mazyn et al., 2004), table tennis (e.g., Bootsma & van Wieringen, 1990), and visual training (e.g., Braly & DeLucia, 2020). However, approaching objects rarely present solely visual information. Ordinary events, such as a bus approaching passengers at a bus stop, provide both visual and auditory information about arrival time.

CONTACT Patricia DeLucia pat.delucia@rice.edu Department of Psychological Sciences, Rice University, Houston, Texas, USA

Auditory TTC

There are numerous examples of individuals using auditory information to estimate TTC. For example, infants recoiled away from approaching sounds that increased in intensity (Freiberg et al., 2001). Blind individuals used auditory information to estimate TTC with accuracy comparable to sighted individuals who used visual information (Schiff & Oldak, 1990). Perceived urgency and loudness influenced TTC estimates of approaching objects (Gordon et al., 2013), and auditory warnings that increased in intensity were effective in warning drivers about imminent collisions (Gray, 2011).

Prior research also examined whether presenting auditory and visual information concurrently (or alternately; see Gordon & Rosenblum, 2005) can improve TTC estimation compared to either auditory or visual information alone. Results showed that visual information resulted in higher accuracy than auditory information in some cases (Hassan, 2012; Schiff & Oldak, 1990; Zhou et al., 2007), auditory information resulted in higher accuracy in other cases (DeLucia et al., 2016), and simultaneous presentation of both did not result in a multimodal advantage (DeLucia et al., 2016; Zhou et al., 2007).

Affective Visual TTC Estimation

Although prior research investigated the effects of visual and auditory stimuli on TTC estimation, few studies examined whether the *emotional content* of an approaching object can affect TTC estimation. In one such study in the visual domain, participants viewed threatening images of approaching animals (e.g., snakes, spiders) or non-threatening animals (e.g., butterflies, rabbits) (Vagnoni et al., 2012). Participants completed a prediction-motion (PM) task in which the image approached them and then disappeared (Schiff & Detwiler, 1979); participants pressed a button at the exact moment they thought the image would reach them after it disappeared. Mean TTC estimates were shorter for threatening pictures than non-threatening pictures, and ratings of fear for threatening pictures were correlated with shorter TTC estimates. In a follow-up study by Vagnoni et al., TTC estimates for a neutral picture (blue disc) were not shorter when preceded by a threatening picture, suggesting that the prior results could be attributed to the threatening images per se and not a general effect of arousal.

In a related study, participants viewed approaching images of frontal attacks that implied immediate danger (e.g., masked attacker with a knife, biting snake) or images that were neutral (e.g., plant, lamp) (Brendel et al., 2012). The images were selected from the International Affective Picture system (IAPS; Lang et al., 2005). Threatening pictures were chosen based on high arousal and low valence ratings, and neutral pictures were chosen based on intermediate to low arousal ratings and intermediate to high valence ratings. Participants again completed a PM task. Mean TTC estimates were shorter for threating pictures compared to neutral pictures but only when the objects in the pictures could be identified. When the pictures were scrambled to mask the identity of the objects while maintaining the same low-level image features, the difference in estimated TTC between threatening and neutral pictures was not significant, implicating that differences between the original threatening and neutral pictures were due to differences in content and not in low-level image features. Another important result of this study was that the effect of threat on

TTC estimates occurred only for relatively long presentation durations (800 ms vs. 200 ms). These results suggest that the effect of emotion occurs only when there is time for observers to cognitively process the approaching image (Brendel et al., 2012).

Another study examined the effects of threatening pictures and threatening faces on TTC estimates (DeLucia et al., 2014). Threatening pictures were selected from the sample used by Brendel et al. (2012), and threatening faces were selected from a prior study showing a threat advantage in a visual search task (Öhman et al., 2001), as well as the NimStim set of facial expressions from photographs (Tottenham et al., 2009). Two classes of TTC estimation tasks were employed that putatively differed in their reliance on cognitive processes (a) the PM task, as described earlier and (b) the relative judgment (RJ) task, in which observers indicated which of two approaching stimuli would reach them first after they disappeared. The PM task putatively involves cognitive processing; when designed appropriately, the RJ task does not (Tresilian, 1995). In these experiments, the RJ task was designed to preclude cognitive processing by using a 1-s presentation duration and requiring responses shortly after stimulus offset. Across five experiments, results showed that threatening pictures produced shorter TTC estimates than neutral pictures in the PM task, but this effect did not occur with threatening faces. The difference between threatening pictures and threatening faces was not significant in the RJ task. DeLucia et al. proposed that the difference between pictures and faces occurred because the threat was explicit in the pictures (a frontal attack indicates immediate harm) but not in the faces because a facial expression does not unambiguously indicate the harmful actions that might occur. The effect of emotion in the PM task but not the RJ task was attributed to the involvement of cognitive processes in the PM task but not the RJ task. Consistent with Brendel et al. (2012), the effect of emotion on TTC estimation appears to require sufficient time to cognitively process the approaching image.

Affective Auditory TTC Estimation

Although prior research examined auditory and audiovisual TTC estimation (e.g., DeLucia et al., 2016), and effects of emotional stimuli on visual TTC estimation (e.g., Brendel et al., 2012), it is not known whether the emotional content of an auditory stimulus can affect TTC estimation of an approaching object. We considered three reasons that it should. First, if a mechanism exists in the visual domain that results in effects of emotional content on TTC estimates (threatening pictures; Brendel et al., 2012; DeLucia et al., 2014; Vagnoni et al., 2012), it is reasonable to expect that a similar mechanism exists in the auditory domain (people rely on both modalities concurrently) or that there is a supramodal mechanism at a higher processing stage that receives input from both sensory domains. In fact, the auditory system is particularly suited for alerting and thus sensitive to threat (Haas & Edworthy, 2006). Second, unpleasant approaching sounds resulted in more intense emotional responses compared to unpleasant receding sounds (Tajadura-Jiménez et al., 2010). Third, approaching sounds resulted in shorter reaction times and elicited more amygdala activity compared to receding sounds (Bach et al., 2008). The implication is that approaching sounds elicited a fear response resulting in hastened responses.

Several previous studies examined effects of sounds on judgments of movement in depth or TTC or time-to-arrival (TTA). For example, when a negatively valenced sound (stationary crying baby) was played concurrently with a looming sound (an object approaching from the side), judgments of the sound's TTA was shorter compared to a positively valenced auxiliary sound (stationary laughing baby) or no auxiliary sound, and removing spectral features that define the emotion eliminated such effects (Neuhoff et al., 2014). However, the emotional content of the moving stimulus for which TTC judgments were required was not varied. Ferri et al. (2015) showed that the emotional content (e.g., positive or negative valence) of a stimulus that changed in intensity and seemed to approach resulted in a change in the perception of peripersonal space, but TTC judgments were not measured. Similarly, Tajadura-Jiménez et al. (2010) reported that approaching sounds (rising intensity) resulted in more intense emotional responses than receding sounds (decreasing intensity). In addition, unpleasant approaching sounds resulted in more unpleasant and arousing affect compared to corresponding receding sounds. However, TTC estimation was not measured. Using a method most closely related to ours (prediction motion task), Wilkie and Stockman (2020) measured TTC judgments of sounds that increased in intensity and were terminated 300 ms before actual TTC. The nature of the sound (real vs artificial) did not affect TTC estimates perhaps because the "occlusion time" was too short, preventing observers from mentally extrapolating the perceived motion (DeLucia & Liddell, 1998).

The purpose of the current study was to determine whether the affective content of a sound-emitting approaching object affects TTC estimates. We expected that sounds rated as low in valence and high in arousal (i.e., threatening; Brendel et al., 2012) would result in shorter TTC estimates compared to control sounds that could not be identified (e.g., as a snake) but had the same long-term frequency spectrum and the same loudness. We used the conventional PM task to measure TTC judgments of sounds that simulated head-on approach motion and that varied with respect to characteristics that are associated with threat. We compared our results to those of our prior study (Brendel et al., 2012), which also measured effects of threat on TTC judgments but in the visual domain (pictures).

Looming Vs TTC

It is important to distinguish between looming and TTC. Looming refers to the increasing optical size (of a visually presented object) or sound intensity (of an aurally presented object) that occurs when an object approaches an individual. TTC refers to the time remaining until an approaching object would contact an individual. For approaching objects that are experienced visually, TTC information is provided by the ratio of the object's optical size to its instantaneous rate of expansion, known as tau (Lee, 1976). For approaching objects that are experienced aurally, there is an analogous ratio based on sound intensity (Shaw et al., 1991). Although looming signifies approach (Schiff et al., 1962), it does not specify when an object would arrive because objects of different sizes, speeds, and distances can result in the same looming pattern. Tau is independent of size, speed and distance.

In studies of TTC estimation, observers typically report the exact time at which an approaching object would reach them. Such reports typically reflect underestimations; observers report that the object would arrive sooner than it actually would (Schiff &

Detwiler, 1979). Judgments of looming are not typically included in studies of TTC perception. However, there is a separate and related literature on auditory looming bias. In such studies, observers listen to a sound that increases or decreases in intensity and report movement direction (approach vs recede), speed, distance, and other parameters. Observers typically report that their distance to an approaching object is closer than it really is, known as the looming bias. Observers also report that looming sounds are perceived as faster than receding sounds (Neuhoff, 2016). This is consistent with underestimations of TTC, but as noted earlier, looming per se does not provide unambiguous information about TTC. In the current study, we measured TTC judgments using established methods in the TTC literature. We did not examine looming bias.

In short, studies of looming judgments and TTC judgments differ in the nature of the task and the information available for the task, and each has distinct theories and literatures. We focused on TTC estimations of auditory stimuli, which is essential because we wanted to compare our results to those in our prior study of TTC estimates of visual stimuli (Brendel et al., 2012).

Method

Participants

Twenty university students (10 females) between the ages of 18 and 44 years ($M = 22.55$, $SD = 5.86$) participated in the study; sample size was selected to be the same as our prior study of effects of visual threat to which we wanted to compare results (Brendel et al., 2012). All participants reported normal or corrected visual acuity, normal hearing, and normal motor control. Participants received $20 for completing the study. Written informed consent was obtained from each participant. This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Boards at Texas Tech University (IRB2017-403) and Rice University (IRB-FY2018-447).

Apparatus and Stimuli

The stimuli simulated an approaching sound-emitting object, with the acoustic intensity increasing as the distance between the object and the observer decreased (DeLucia et al., 2016; Shaw et al., 1991). The relation between distance and intensity followed the inverse law: The sound pressure of a point source in the acoustic-free field is inversely proportional to its distance from the receiver (Hartmann, 2005). The object approached for 2.0 seconds on a straight path along the midsagittal plane and at a constant velocity and then became inaudible. The object's actual TTC at the time of its disappearance was 0.75, 1.5, or 3.0 seconds.

Stimuli were created with a Dell OptiPlex 775 Intel Core 2.33 GHz computer with 5.0 GB of RAM and integrated Intel Q35 graphics. Instructions were presented on a 17" color monitor in 1024 × 768 resolution. An Audient iD14 High Performance USB audio interface was connected to the computer for audio playback. To ensure that the source of the approaching sound was perceived as being in front of the listener, auditory stimuli were presented through a mono speaker that was located on top of the computer monitor right in front of the participant. We chose not to employ the use of headphones because this often results in sounds being perceived as localized within the head (Begault et al., 2001; Best et al., 2020).

Auditory stimuli were created with MATLAB (The MathWorks, 1993) and presented using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Three sounds were selected from the International Affective Digitized Sound system (IADS; Bradley & Lang, 1999), based on (high) arousal and (low) valence ratings. These three sounds were selected because they were comparable to the ratings of the threatening images used by Brendel et al. (2012). These criteria were important because we wanted to determine whether these effects of threat on TTC estimations we found in the visual domain (Brendel et al., 2012) also occur in the auditory domain. The threatening sounds were a buzzing bee (#115), a hissing and rattling rattlesnake (#134), and a dentist drill (#719). Because these sounds contain considerable sound-level fluctuations that could interfere with TTC estimation based on intensity changes, level-smoothed versions of the original IADS sounds were used in the experiment. In the first step, a temporal portion of each of the IADS sounds was selected manually that showed a relatively small-level variation. This selected part of the sounds was then smoothed in level by dividing the amplitude of each digital sample by the root-mean-square amplitude in a moving temporal window of 20 ms (bee, dentist drill) or 50 ms (snake) centered around the given sample. The resulting three level-smoothed versions ($115_{smoothed}$, $134_{smoothed}$, and $719_{smoothed}$) served as the experimental test stimuli. Table 1 shows the duration of the selected part for each IADS sound as well as a measure of the amount of level fluctuations.

Control stimuli consisted of sounds with the same spectral properties as the level-smoothed sounds but which were not threatening because the object could not be identified. For these threat-absent control stimuli, we used loudness-matched noise stimuli with the same long-term frequency spectrum (LTS) as the level-smoothed sounds, but without the envelope information or temporal fine structure information contained in the level-smoothed sounds. Presenting control stimuli with the same frequency spectrum as the test stimuli is important because spectral differences can result in differences in loudness (for an overview see Jesteadt & Leibold, 2011) and in auditory intensity processing (Plack & Carlyon, 1995). In our experiment, auditory TTC estimation was based on dynamic changes in acoustic intensity. In addition, shorter auditory TTC estimates were reported for loud (compared to soft) sound sources

Table 1. Sound statistics for the three versions of each of the three sounds from the International Affective Digitized Sound system (IADS; Bradley & Lang, 1999) used in the current study.

		Level SD		
IADS sound file	Selected time range	original	smoothed	LTS
Buzzing bee (#115)	0.6–5.6	4.1 dB	1.0 dB	1.1 dB
Rattlesnake (#134)	0.2–3.0	8.2 dB	4.6 dB	0.7 dB
Dentist drill (#719)	2.0–6.0	3.4 dB	0.7 dB	0.9 dB

Note. The second column shows the time range (in seconds) selected from the IADS sound (start seconds – end seconds). Level SD: standard deviation of the RMS-envelope (computed with a window size of 10 ms). The columns "original," "smoothed," and "LTS" refer to the original IADS sound, level-smoothed version, and randomly scrambled version with the same long-term spectrum (LTS), respectively.

(DeLucia et al., 2016; Keshavarz et al., 2017). Consequently, using an emotionally "neutral" sound from IADS (i.e., with neutral valence and low arousal) as a control stimulus might confound effects of potential differences in loudness and sound spectrum with the effects of threatening (compared to non-threatening) sounds. To create control sounds with the same LTS as the test sounds, we used a temporal scrambling-and-averaging approach. A 200-ms random temporal portion of the level-smoothed sound was selected. Then 50-ms $\cos^2$ on- and off-ramps were applied in order to prevent spectral splatter, and the ramped sound snippet was positioned at a random temporal position of the target waveform, which was identical in duration to the level-smoothed sound. This process was repeated 40,000 times for each of the level-smoothed sounds, resulting in three control sounds with the same LTS as their level-smoothed counterparts (115_{LTS}, 134_{LTS}, and 719_{LTS}). To avoid effects of loudness differences, each control sound was presented at the same loudness as the corresponding threatening sound, based on individual loudness matches. In the TTC-estimation task, a randomly selected 2.0-s temporal portion of a given sound was presented on each trial. Figure 1 shows the spectrograms of the two sound types (level-smoothed versus LTS), for each of the three sounds.

None of the stimuli sounded like either white noise at one extreme or pure tones at the other extreme. The bee stimulus sounded like a low-pitched buzzing of multiple bees. The dentist drill stimulus sounded like a high-pitched whirring of a dentist's drill, with a clear tonal component. The snake stimulus sounded like a medium pitched rattle and hissing of a rattlesnake. The smoothed versions sounded very similar to the original sounds but with less modulation in amplitude. The LTS versions contained nearly no audible modulation. It is critical to emphasize that the smoothed and LTS versions of each of the three sounds were identical in spectrum so that we could make valid comparisons between them.

Procedure and Design

Participants listened to auditory stimuli from a 45.72-cm (18 in) distance while seated. Head movements were not restricted, but the experimenter monitored participants to make sure their head remained at the correct distance for the simulation. There were 4 sessions that lasted a total of approximately 3 hours.

Loudness Matching

In the first two sessions, the six different sounds (level-smoothed versions of bee, snake, and drill, and corresponding control sounds with identical long-term frequency spectrum) were equalized in loudness for each participant individually to ensure that any differences between threatening and control sounds were due to the sound's content rather than to differences in perceived loudness, which is known to have an effect on auditory TTC estimates (DeLucia et al., 2016; Keshavarz et al., 2017). The loudness of each smoothed IADS sound ($115_{smoothed}$, $134_{smoothed}$, $719_{smoothed}$) was matched to the loudness of the other two sounds separately using an adaptive two-interval, two-alternative forced-choice procedure with a one-up, one-down rule (Levitt, 1971), tracking the 50%-point on the psychometric function. Within each of the three pairs of smoothed IADS sounds (e.g, $115_{smoothed}$ vs. $134_{smoothed}$), both of the sounds once served

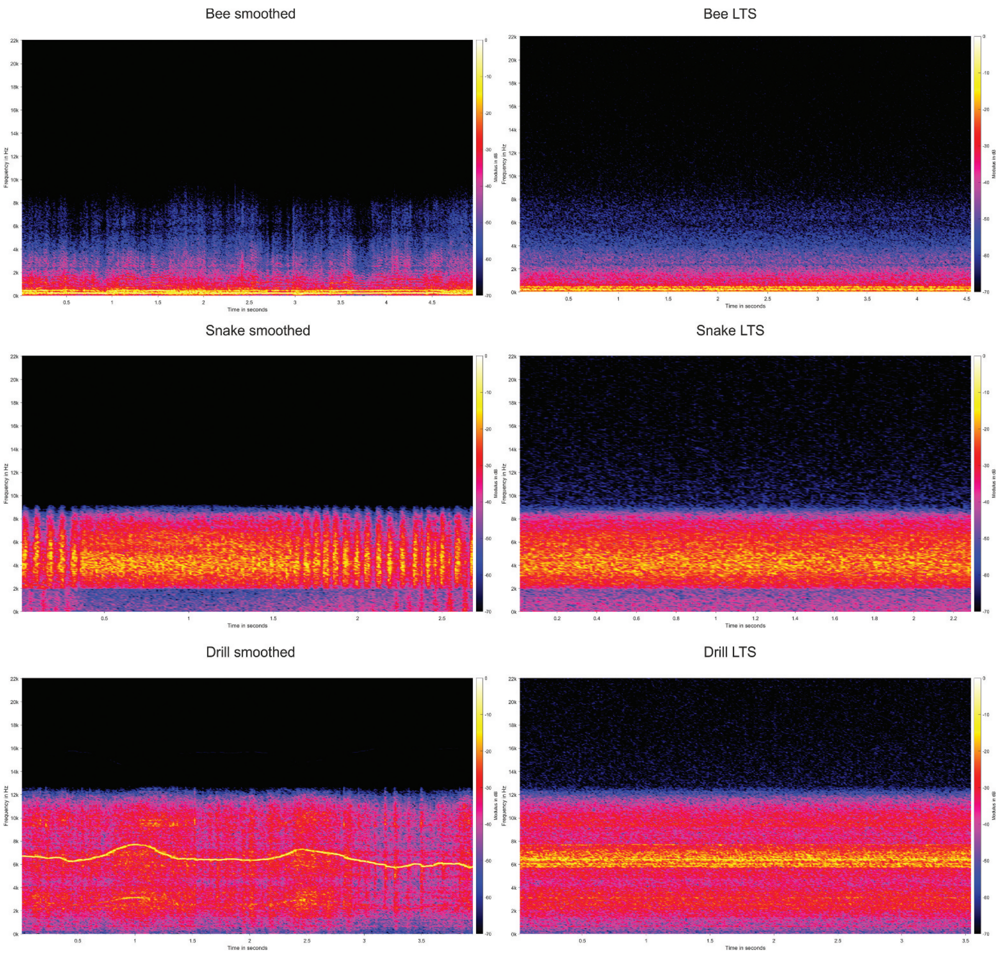

Figure 1. Spectrograms of the three sound types. Left panels: level-smoothed version of the original IADS sounds. Right panels: static sounds with the same long-term spectrum (LTS) as the corresponding sounds on the left. Upper row: bee sound. Middle row: rattlesnake sound. Lower row: dentist drill sound.

as the reference stimulus (or standard stimulus; presented at fixed level of 70 dB SPL) and once as the test stimulus (or comparison stimulus; level varied by an adaptive procedure, discussed next), in separate adaptive tracks. The test stimulus was presented either in the first or in the second interval, to reduce order effects. Reference sound was crossed with reference sound interval and resulted in four adaptive tracks for each of the three sound pairs. To further reduce effects of biases, these four adaptive tracks were randomly interleaved within an experimental block (cf. Buus et al., 1997; Oberfeld et al., 2012; Verhey, 1999). Each experimental block contained only one pair of sounds, and two blocks were presented for each of the six sound pairs.

On each trial, the participant heard two sounds and indicated which one was louder by pressing the corresponding button on a computer keyboard. The sound duration was 1000 ms (including 10-ms $\cos^2$ on- and off-ramps), and the two sounds were separated by

800 ms of silence. The adaptive track started with the level of the test stimulus equal to the level of the reference stimulus. If the participant responded that the test stimulus was louder than the reference stimulus, the level of the test stimulus was decreased; otherwise, it was increased. This initial increment or decrement in level was 5 dB. After four reversals (i.e., "peaks" and "valleys" in the adaptive track; Levitt, 1971), the track continued with a step size of 2 dB, until another eight reversals had occurred or 60 trials had been presented, whichever happened first. For each track, the arithmetic mean of the test sound levels at these final eight reversals was used to calculate the level difference between the test stimulus and the equally loud reference stimulus, that is, the loudness match. Using the same method, loudness matches were obtained between each of the smoothed IADS sounds and its corresponding control sound with the same long-term spectrum (e.g., 115_{smoothed} versus 115_{LTS}). The data from the loudness-matching task were used to apply level corrections for each participant so that all sounds were equally loud for each participant.

TTC Estimation: Prediction-motion Task

In the third session, participants completed a prediction-motion (PM) task (Schiff & Detwiler, 1979). Participants were instructed to press a button on the keyboard when they thought that the approaching object would reach them, had the object continued to move with the same velocity after it was no longer audible. The duration of the approach (sound) was 2 s, which matched the stimulus duration used by Brendel et al. (2012). TTC judgments were measured as the time between the last audio sample of the stimulus and the time when the participant pressed the button. To provide a variety of stimuli and be consistent with prior studies (e.g., DeLucia et al., 2014; Schiff & Detwiler, 1979), within each of the three presented TTCs, the simulated auditory object started and ended at either a near distance or a far distance from the virtual listener (i.e., the object in the near condition ended with a higher sound level compared to the far condition). The far distance from the listener in the final sample was two times the near distance in the final sample. Because in the simulation the sound intensity was varied dynamically according to the inverse law, the level was thus $20 \log_{10}(D_{\text{final far}}/D_{\text{final near}}) = 20 \log_{10}(2) = 6.02$ dB higher in the near distance condition than in the far distance condition. The two final distances and thus also the final levels were identical for each presented TTC value. For the sound 115_{LTS}, which served as the reference sound in the loudness matches (discussed previously), the final levels were 74.44 and 80.46 dBA for the far and near distance, respectively. For the other sounds, the final levels were adjusted on the basis of the loudness matches so that the final loudness was identical among all of the six sound types. The sound of the object (snake, bee, or drill), nature of the sound (threatening or control), TTC (0.75, 1.5, or 3.0 s), and distance (near or far) were factorially crossed which resulted in 36 unique conditions. For each condition, a total of 10 trials was presented per participant.

We used the prediction-motion task which is one of the most frequently used paradigms for studying TTC estimation and was the method used by Brendel et al. (2012). The object is occluded during its approach because if it is presented until it reaches the observer, TTC estimation is not necessary; the participant just needs to press the button when the object reaches him or her.

Affective ratings. In the fourth session, participants rated the perceived valence, arousal, and dominance of each sound according to the Self-Assessment Manikin (SAM) procedure (Bradley & Lang, 1994). Dominance was included for completion and was not analyzed. SAM is a non-verbal picture-oriented measure that contains five images for each of the three affective dimensions, rated on a 9-point scale. The purpose of this procedure was to compare ratings of the sounds in the current study to the original values reported in IADS, and to compare the emotion ratings between the level-smoothed and the LTS versions. Specifically, the SAM ratings were obtained for stimuli presented at a constant level (i.e., without a "looming" intensity profile). This corresponds to the procedure also used in studies on effects of visual emotional stimuli on TTC estimation, in which emotion ratings are typically obtained for static (not moving) pictures (e.g., Brendel et al., 2014). Participants heard a sound three times and then made ratings on all of the three SAM scales, using the keyboard while viewing the picture-oriented scales on the computer monitor. This process was repeated until ratings for all six sounds were collected. Finally, we included a posttest questionnaire to determine whether participants could correctly identify the sounds.

Results

Our objective was to determine whether the affective content of a sound-emitting approaching object affects TTC estimates. We expected that the smoothed IADS sounds rated as low in valence and high in arousal (i.e., threatening; Brendel et al., 2012) would result in shorter TTC estimates compared to control (LTS) sounds that could not be identified (and thus were not perceived as threatening) but had the same spectral properties. This would be analogous to our previously reported effects of threatening pictures on TTC estimates of approaching objects (e.g., Brendel et al., 2012).

TTC Estimates

Effects of Long-term Frequency Spectrum

In a first analysis, the mean TTC estimates for the three LTS sounds (assumed to represent no threat; shown by the hatched bars in Figure 2) were analyzed with a 2 (distance: near, far) × 3 (sound: Bee, Drill, Snake) × 3 (TTC: 0.75, 1.5, 3.0 s) repeated-measures analysis of variance (rmANOVA). Here and in the subsequent analyses, a univariate approach with the Huynh-Feldt correction for the degrees of freedom was used where applicable, and the value of $\tilde{\varepsilon}$ is reported. The effect of sound was not significant, $F(2, 38) = 0.54$, $p = 0.53$, $\tilde{\varepsilon} = 0.69$, $\eta_p^2 = 0.02$, and none of the interactions involving sound was significant [sound x distance: $F(2, 38) = 1.83$, $p = 0.18$, $\tilde{\varepsilon} = 0.80$, $\eta_p^2 = 0.09$; sound x TTC: $F(4, 76) = 2.01$, $p = 0.10$, $\tilde{\varepsilon} = 1.0$, $\eta_p^2 = 0.10$; sound x distance x TTC: $F(4, 76) = 0.49$, $p = 0.69$, $\tilde{\varepsilon} = 0.74$, $\eta_p^2 = 0.03$]. This indicates that the different spectral properties of the three types of sounds did not have a systematic effect on the TTC estimates.

Effects of Threat

Next, mean TTC estimates were analyzed with a 2 (distance: near, far) × 2 (threat: present, absent) × 3 (sound: Bee, Drill, Snake) × 3 (TTC: 0.75, 1.5, 3.0 s) rmANOVA. The level-smoothed sounds were assumed to represent threat, and the corresponding control (LTS) versions of those sounds were assumed to represent absence of threat.

The main effect of threat was not significant, $F(1, 19) = 0.49$, $p < .49$, $\eta_p^2 = 0.03$. However, there was a significant interaction between sound and threat, $F(2, 38) = 13.49$, $p < .001$, $\tilde{\varepsilon} = 0.85$, $\eta_p^2 = 0.42$. To determine the effect of threat at each level of sound, follow-up tests were conducted with separate one-way rmANOVAs at each level of sound. For the snake sound, there was a significant effect of threat, $F(1, 19) = 12.31$, $p = .002$, $\eta_p^2 = 0.39$. As Figure 2 shows, mean TTC estimates were significantly shorter for the threatening snake sound compared to the control snake sound. However, the opposite occurred for the drill sound, $F(1, 19) = 8.92$, $p = .008$, $\eta_p^2 = 0.32$, and for the bee sound, $F(1, 19) = 5.63$, $p = .03$, $\eta_p^2 = 0.23$. In the latter two cases, the mean TTC estimate was significantly longer for the threatening than the control sound, which was unexpected.

Effects of Sound, Distance and TTC

Averaged across the level-smoothed and LTS versions of the sounds, mean TTC judgments were significantly shorter for the snake sound ($M = 1.19$ s, $SD = 0.60$ s) than for the bee sound ($M = 1.47$ s, $SD = 0.67$ s) and the drill sound ($M = 1.53$ s, $SD = 0.70$ s), as indicated by a main effect of sound, $F(2, 38) = 9.71$, $p = .0015$, $\tilde{\varepsilon} = 0.75$, $\eta_p^2 = 0.34$. A potential explanation is that the (level-smoothed) snake sound was perceived as more threatening than the drill or bee sounds. We examined this further with analyses of the SAM emotion ratings, reported subsequently.

Mean TTC judgments were shorter for near sounds ($M = 1.10$ s, $SD = 0.47$ s) compared to far sounds ($M = 1.69$ s, $SD = 0.81$ s) as indicated by a main effect of distance, $F(1, 19) = 32.01$, $p < .001$, $\eta_p^2 = 0.63$. Because the level at the ear of a listener is higher when the sound source is closer to the position of the listener (in our experiment, the level difference between the near and far condition was 6.02 dB; see Methods), the data confirm the intensity-arrival effect reported by DeLucia et al. (2016) and Keshavarz et al. (2017). The two latter studies consistently showed that if two auditory objects are presented with the same TTC but different final sound levels, a shorter TTC is estimated for the object with the higher final sound level. The intensity-arrival effect is an auditory analog of the effects of final size on TTC estimates in the visual domain (size-arrival effect; DeLucia, 1991, 2013; DeLucia & Warren, 1994).

The main and interactive effects of TTC were not statistically significant, possibly due to the needed level smoothing (*F*s < 3.34, *p*s > 0.061). However, generally, participants overestimated the short TTC and underestimated the longer TTC, consistent with prior studies (e.g., ., Caird & Hancock, 1994; Schiff & Detwiler, 1979): On average, the mean estimated TTCs were longer than the actually presented TTC of .75 s, and shorter than the actually presented TTC of 3 s.

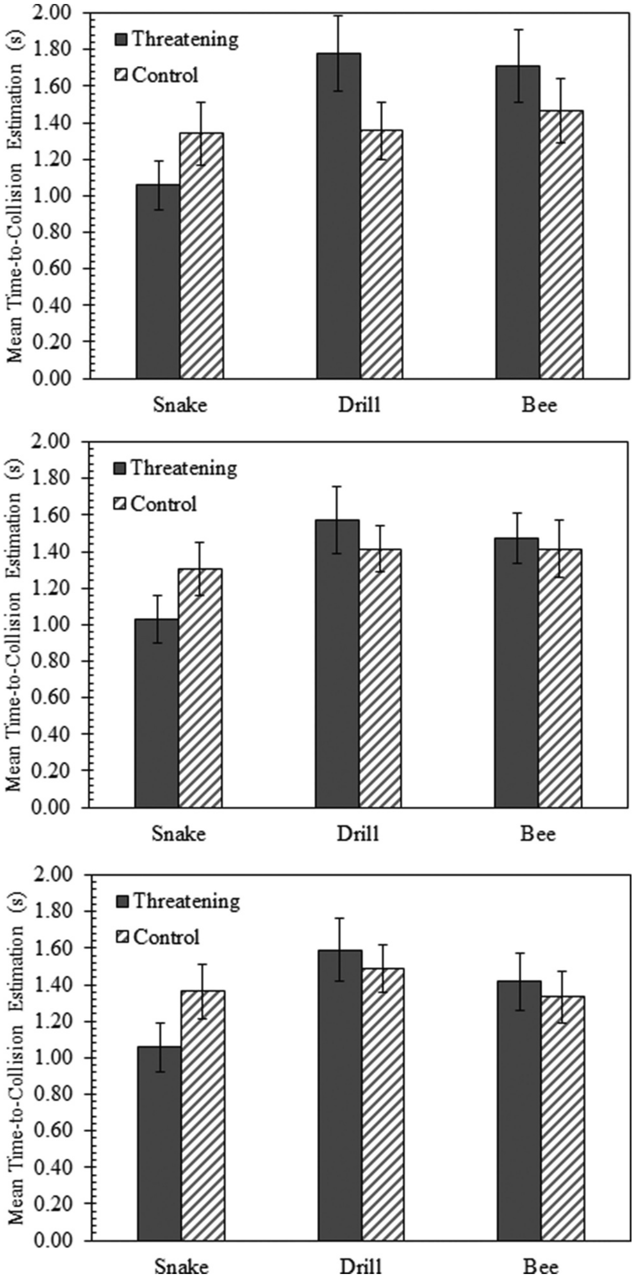

Figure 2. Mean time-to-collision estimation as a function of sound (snake, drill, bee) and threat (present: gray bars vs. absent: hatched bars). Top: TTC is .75 s. Middle: TTC is 1.5 s. Bottom: TTC is 3 s. Error bars represent ± 1 standard error of the mean.

Emotion Ratings

Effects of Threat

Because we observed the expected significantly shorter TTC estimates for the level-smoothed (supposedly threatening) compared to the LTS (supposedly non-threatening) versions of the snake sound, but not for the two other sounds (see Figure 2), we considered whether our participants perceived only the snake sound as threatening but did not perceive the drill and bees sounds as threatening. Threat is defined as low valence combined with high arousal (e.g., Ho et al., 2015; Knutson et al., 2014). Thus, we analyzed whether participants did or did not perceive a difference in arousal and valence between the level-smoothed and LTS versions of the IADS sounds. Table 2 shows the mean emotion ratings on the 9-point SAM scales for each of the six sounds presented in the experiment.

We analyzed the SAMs ratings with 2 (threat: present vs. absent) × 3 (sound: bees, drill, snake) rmANOVAs separately for valence and arousal. The two-way interaction was significant for valence, $F(2, 38) = 8.58$, $p = .0013$, $\tilde{\varepsilon} = 0.91$, $\eta_p^2 = .31$, but not arousal. Post-hoc pairwise comparisons were conducted with paired-samples t-tests for the difference in mean ratings of valence and arousal between presence and absence of threat (i.e., between level-smoothed and LTS versions of sounds), using correction for multiple testing with the Hochberg (1988) procedure.

As shown by the means in Table 2, for the snake sound, the arousal rating was descriptively but not significantly higher and the valence rating was significantly lower for the level-smoothed version compared to the LTS version of the sound, suggesting that the level-smoothed version was perceived as more threatening than the control sound. For the drill sound, the arousal rating was virtually identical for the two versions, and the valence rating was significantly lower for the level-smoothed compared to the LTS version.

Table 2. Mean and standard deviation of Self-Assessment Manikin (SAM) ratings for each of the six sounds presented in the study, and results of two-tailed paired-samples t-tests on mean difference between threatening (level-smoothed) and control (LTS) versions of the sounds (df = 19).

	Arousal Rating		Valence Rating		Dominance Rating			
	1 = calm 9 = excited		1 = unhappy 9 = happy		1 = dominated 9 = dominant		Negative Arousal Score	
Sound	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Threatening Snake	5.55 (6.98)	2.06 (1.67)	3.65 (3.55)	1.42 (1.99)	4.6 (3.50)	1.70 (1.82)	1.90	2.53
Control Snake	4.70	2.03	4.60	1.35	5.15	1.81	.10	2.55
Difference	0.85		**−0.95**				**−1.80**	
t, p	1.95	.067	**−2.55**	**.020***			**−3.21**	**.005***
Threatening Drill	4.90 (6.91)	2.43 (2.02)	3.40 (2.89)	2.06 (1.67)	4.30 (2.92)	2.13 (2.03)	1.50	3.05
Control Drill	4.85	2.06	4.70	0.98	5.5	1.99	.15	2.62
Difference	0.05		**−1.30**				**−1.35**	
t, p	0.14	.89	**−3.16**	**.005***			**−2.97**	**.008***
Threatening Bee	5.25 (7.03)	1.77 (1.91)	4.30 (2.16)	1.78 (1.33)	4.05 (2.67)	1.82 (1.71)	.95	2.61
Control Bee	4.75	2.31	3.60	1.47	5.00	1.86	1.15	2.91
Difference	0.50		0.70				.20	
t, p	0.83	.42	1.65	.12			.302	.766

Note. On the nine-point SAM rating scales, the value 5 represents the neutral category. Numbers in parentheses represent mean ratings for the original sounds according to the IADS manual. Asterisks and bold font indicate that a pairwise difference was significant at the .05 level according to the Hochberg (1988) procedure.

To analyze the difference in perceived threat more directly, we used the concept of threat being defined as low valence combined with high arousal (e.g., Knutson et al., 2014). As suggested by the latter authors, we computed a ”negative arousal” score, defined as the difference between the arousal rating and the valence rating, for each sound and each participant separately. The mean negative arousal scores are displayed in Table 2.

The negative arousal scores for the level-smoothed and the LTS version of the bee sound were very similar, while for the snake and the dentist drill sounds, the negative arousal score was higher for the smoothed than for the LTS version. We analyzed the negative arousal scores with a 2 (threat: present vs. absent) × 3 (sound: bees, drill, snake) rmANOVA. The effect of threat was significant, $F(1, 19) = 7.29$, $p = .014$, $\eta_p^2 = .28$. The threat × sound interaction was also significant, $F(2, 38) = 3.92$, $p = .035$, $\tilde{\varepsilon} = 0.86$, $\eta_p^2 = .17$. Post-hoc two-tailed paired-samples *t*-tests with the Hochberg (1988) procedure showed that the negative arousal score for the smoothed version was significantly higher than for the LTS version in case of snake and drill, but not for the bee sounds. Also, because the negative arousal scores are computed as the difference between two categorical rating scales ranging from 1 to 9, the negative arousal score can obtain values between −8 (minimally possibly negative arousal) to +8 (maximally possibly negative arousal). Considering this scale range, the maximum mean value of 1.9 observed for the negative arousal score for the snake sound can be considered as representing relatively low negative arousal. In the same line of reasoning, even for the two sounds for which the two versions differed significantly in terms of perceived negative arousal, the maximal observed difference in mean negative arousal was 1.8 (snake sound) and can thus be considered to be small.

How does the pattern of emotion ratings align with the pattern of TTC estimates? Shorter TTC estimates for the level-smoothed sound compared to the LTS version for the snake sound are compatible with the negative arousal scores, which suggested significantly higher perceived threat for the level-smoothed sound compared to the LTS version for the snake sound. Longer TTC estimates for the level-smoothed bee sound compared to the LTS version are also in principle compatible with the slightly lower negative arousal score for the level-smoothed version than the LTS version, although this difference in perceived threat was descriptively very small and non-significant. However, for the drill sound, longer TTC estimates for the level-smoothed version compared to the LTS version were not compatible with the significantly *higher* negative arousal score (i.e., more threatening) for the level-smoothed version than the LTS version of the drill sound.

Relationship between TTC Estimates and Emotion Ratings

To gain more insight into the relationship between TTC estimates and ratings of perceived arousal and valence, we calculated the difference in mean TTC estimates between the smoothed and LTS versions of each sound and for each participant. We also calculated the difference in the SAM ratings of arousal and valence between the smoothed and LTS versions of each sound and for each participant. Using multiple linear regression, we then regressed the difference in TTC estimates on the difference in rated arousal and rated valence, separately for each sound. As shown in Table 3, results indicated that the correlation between the affective ratings and the TTC estimates at

Table 3. Multiple linear regression results. Criterion: difference in time-to-collision estimates and differences in mean arousal (threatening minus control). Predictors: difference in arousal ratings and difference in valence ratings.

Sound	Predictor	Beta (β)	*t*	*p*
Snake	Arousal Difference	.022	0.094	.926
	Valence Difference	−.273	−1.168	.259
Drill	Arousal Difference	.170	0.672	.511
	Valence Difference	−.022	−0.085	.933
Bee	Arousal Difference	−0.299	−1.269	.221
	Valence Difference	0.14	0.593	.561

the level of individual subjects was not significant. That is, a large difference in valence between the smoothed and LTS sounds was not associated with a large difference in the TTC estimates between the smoothed and LTS sounds.

Next, we regressed the difference in TTC estimates between level-smoothed and LTS versions of each sound on the difference in negative arousal scores between the two sound versions. Due to the repeated-measures structure of the data, we used a mixed-model approach (SAS PROC MIXED) with random intercept and slope, fitting an unstructured covariance matrix. The fixed effect of the negative arousal score was not significant, B = −0.013, $t(1) = 0.34$, $p = .79$. Thus, differences in negative arousal did not predict differences in TTC estimates for any of the sounds.

Effects of Sound Identification on Emotion Ratings

Sounds high in affective content were reported to become more neutral when individuals were unable to access the associations that lead to strong emotional responses (i.e., knowing what the sound is; Asutay et al., 2012), which converges on our prior finding that effects of affective content for approaching pictures required cognitive processing (Brendel et al., 2012; DeLucia et al., 2014). Thus, a potential explanation for the relatively small differences in perceived threat between the level-smoothed and the LTS versions of the IADS sounds (see Table 2) is that some of our participants were not able to identify the threatening sounds, resulting in low perceived threat, and consequently no pronounced effect on the TTC estimates. To examine this possibility, we compared the emotion ratings between 9 participants who successfully identified the level-smoothed snake and drill sounds, and the 11 participants who did not identify the sounds during the posttest questionnaire. The snake subset included only data for the level-smoothed and LTS versions of the snake sound, and the drill subset included only data for the level-smoothed and LTS versions of the drill sound. We did not create a subset for the bees sound because 17 out of 20 participants had identified the sound correctly. Means are shown in Table 4.

To summarize results shown in Table 2 and , Table 2 shows the mean SAM ratings (for all study participants) for the smoothed and LTS versions of the snake, drill, and bee sounds in the current study. It also shows the mean negative arousal scores as defined by Knutson et al. (2014). The values show lower valence ratings (and higher negative arousal scores) for the smoothed sounds compared to the LTS versions for the snake and the drill but not the bee.

Table 4. Mean valence and arousal ratings, and standard deviations (SD) for groups who did and did not identify the sounds correctly, for threatening and non-threatening snake and drill sounds.

	Group: Identified Sound Correctly		Group: Did Not Identify Sound Correctly	
	M	*SD*	*M*	*SD*
VALENCE				
Threatening Snake	3.00	1.41	4.18	1.25
Control Snake	3.50	.71	4.72	1.36
Threatening Drill	3.44	2.46	3.36	1.80
Control Drill	5.00	0.0	4.65	1.06
AROUSAL				
Threatening Snake	4.67	2.74	6.27	.91
Control Snake	1.50	.71	5.06	1.80
Threatening Drill	5.11	2.57	4.73	2.41
Control Drill	2.33	2.31	5.29	1.72

Note. Means for the bees sound were not included because 17 out of 20 participants identified the sound correctly.

Table 4 shows the mean valence and arousal ratings separately for groups who did and did not identify the snake and drill sounds correctly. Because there was a moderate imbalance in the group sizes (9 versus 11 participants) and traditional procedures for rmANOVAs can be affected by unequal group sizes (cf. Keselman et al., 2001), the analysis of the emotion ratings and the TTC estimates were performed using the function *hrm_test* from the R package HRM (Happ et al., 2018). This analysis procedure shows sufficient control of the Type I error rate even when the design is unbalanced and the variance-covariance matrices differ between groups (Happ et al., 2016, 2017).

Specifically, for each sound, we conducted a 2 (between-subjects factor group: identified sound correctly vs not) × 2 (within-subjects factor threat: present vs absent) hrmANOVA on valence and arousal ratings separately. For the snake sound, for valence ratings, there was a significant effect of group, $F(1, 19.52) = 8.68$, $p < .008$, $\eta_p^2 = 0.31$, and threat $F(1, 14.34) = 5.504$, $p = 0.034$, $\eta_p^2 = 0.28$, but not their interaction $F(1, 14.34) = .02$, $p = .892$, $\eta_p^2 = 0.001$. For arousal ratings, there was no significant effect of group, $F(1, 11.08) = 1.801$, $p = 0.21$, $\eta_p^2 = 0.14$, threat $F(1, 16.86) = 3.22$, $p = 0.10$, $\eta_p^2 = 0.16$, or their interaction $F(1, 16.86) = 1.10$, $p = 0.31$, $\eta_p^2 = 0.06$. For the drill sound, for valence ratings, there was a significant effect of threat, $F(1, 16.45) = 9.37$, $p = 0.007$, $\eta_p^2 = 0.36$, but not of group, $F(1, 18.44) = 0.26$, $p = 0.62$ $\eta_p^2 = 0.01$, or their interaction, $F(1, 16.45) = 0.29$, $p = 0.60$, $\eta_p^2 = 0.02$. For arousal, there was no significant effect of group, $F(1, 18.96) = 0.001$, $p = 0.98$, $\eta_p^2 < 0.001$, threat, $F(1, 12.31) = 0.05$, $p = 0.83$, $\eta_p^2 = 0.004$, or their interaction, $F(1, 12.31) = 0.84$, $p = 0.38$, $\eta_p^2 = 0.063$. The absence of a significant interaction between group and threat shows that emotion ratings did not depend on the ability to identify the snake and drill sounds. Analyses of negative arousal also indicated a significant effect of only threat but not group or interactions with group.

Effects of Sound Identification on TTC Estimates

When we analyzed TTC estimates of the snake sound with a 2 (distance: near, far) × 2 (threat: present, absent) × 2 (group: identified sound correctly vs not) × 3 (TTC: 0.75, 1.5, 3.0 s) hrmANOVA, results were significant for only threat, $F(1, 19.65) = 12.26$, $p = 0.002$, $\eta_p^2 = .38$ and distance, $F(1, 19.75) = 29.77$, $p = 0.001$, $\eta_p^2 = .60$. For the drill sound, results were significant for only threat, $F(1, 11.50) = 7.45$, $p = 0.02$, $\eta_p^2 = .39$, and distance, F(1, 19.26) = 30.23, p = 0.001, η_p^2 = .61, and the interaction between TTC and threat, F(1.75, 30.39) = 4.66, p = 0.021, η_p^2 = .21. Thus, the sound identification had no significant effect on the pattern of TTC estimates.

Discussion

The purpose of the current study was to determine whether threatening sounds would result in shorter auditory TTC estimates compared to control (scrambled) sounds that had the same long-term spectrum (LTS) but were not identifiable (and thus were not perceived as threatening) because the envelope and temporal fine structure of the original sound were removed by the random scrambling. This would indicate that effects of threat on TTC estimations in the visual domain (Brendel et al., 2012) also occur in the auditory domain. Using level-smoothed versions of snake, bee, and dentist drill sounds from the IADS, that we supposed to be perceived as threatening, together with their scrambled control versions, which we supposed to be emotionally neutral, we found that the level-smoothed (and supposedly threatening) snake sounds were judged as arriving sooner than the control snake sounds with the same LTS. However, we did not observe this effect of sound type for drill or bee sounds. The pattern of mean ratings of valence and arousal suggested that a significant difference in threat (defined as low valence combined with high arousal) between the level-smoothed and the LTS version was perceived for the snake and the drill sounds, but not for the bee sounds. The observed shorter TTC estimates for the smoothed compared to the LTS version of the snake sound are compatible with the difference in perceived threat, but the finding of significantly longer TTC estimates for the smoothed drill sound (perceived as more threatening than the LTS version) and for the smoothed bee sound (perceived as equally threatening as the LTS version) cannot be explained on these grounds. Also, regression analyses indicated that the difference in arousal ratings, valence ratings, or negative arousal scores between the level-smoothed and LTS versions of the sounds were not significant predictors of the mean difference in TTC estimates between the two sound versions.

One potential explanation for the latter result is that the differences in perceived threat between the two sound versions were simply too small for consistently causing differences in TTC estimates. Compatible with this view, the analysis of the emotion ratings indicated that the differences in perceived threat were relatively small in magnitude even for the two sounds for which we found a significant difference in negative arousal between the level-smoothed and the LTS version. Also, the difference in emotion ratings between supposedly threatening and neutral control sounds was somewhat smaller in our study than in a study reporting threat-related differences in TTC estimates for visual stimuli (Brendel et al., 2012).

Particularly for the case of the snake and drill sounds, where we observed similarly higher perceived threat for the smoothed compared to the LTS sound version, but opposite TTC differences between the two sound versions, it is interesting to ask whether non-emotional differences between the sounds could have played a role in the observed effects on TTC estimation. First, all presented sounds were loudness-matched so that an intensity-arrival effect (DeLucia et al., 2016) can be ruled out. Second, although the frequency content of the three sounds was different, the TTC estimates for the randomly scrambled versions (LTS) were very similar among the three sounds, suggesting that the sound spectrum does not have a pronounced effect on the TTC estimates. We are aware of only one previous study on spectral effects on auditory TTC estimates (Gordon et al., 2013). This study reported longer estimated TTC for low-frequency compared to high-frequency octave-bands. However, the bands were not loudness matched, and in addition the rate of change in sound level as the (simulated) sound source approached was varied as a function of center frequency because the experiment simulated the frequency-dependent air absorption (see Blauert, 1996). Third, and probably most important, as shown in Table 1, the level-smoothed version of the sound of a rattling and hissing rattle snake contained a considerably higher amount of amplitude modulation than the two other sounds. To date, no data on the potential effect of amplitude modulation on auditory TTC estimation are available. Such research is needed in order to answer the question to which extent this difference in the acoustic characteristics between the sounds might have contributed to our observation of shorter TTC estimates for only the level-smoothed snake sound (compared to the LTS version), but not for the level-smoothed drill sound, despite the fact that we found differences in perceived threat between the two sound versions for both the snake and the drill sounds. Interestingly, rattle snakes adjust the frequency of their rattle when a potentially dangerous object approaches them, although it is at present unclear whether the rattle frequency is distance or TTC dependent (Schutte et al., 2019). The same study reported that humans stop at a larger distance from a (simulated) snake when the presented rattling simulated the behavior of a real snake.

There were also individual differences. Only 11 of 20 participants reported that the smoothed snake sound was more arousing than the LTS snake sound, and for the bee and drill sounds this proportion was even lower. Also, 11 of 20 participants reported that the smoothed snake sound had lower valence than the LTS snake sound; 14 of 20 did so for the drill sound. The implication is that our participants may not have perceived the level-smoothed sounds as threatening as the participants who rated the original IADS sounds; this is consistent with the pattern of means in: Mean arousal was lower (less arousing) and mean valence was higher (more positive) in our study than in the original IADS study (Bradley & Lang, 1999). However, for the snake sound, the mean arousal and valence ratings were at least 0.5 higher and lower, respectively, than the neutral rating of 5.0.

Similarly, the difference between threatening and control sounds in the current study may not have been as large as the difference between threatening and control pictures in our prior study (Brendel et al., 2012). Table 5 shows mean emotion ratings reported by Brendel et al. (2012) for the threatening (an average of 3 different threatening pictures) and corresponding scrambled (control) pictures. It also shows mean ratings for original IAPS pictures of snake, bees, and dental drilling (Lang et al., 2008); there were no corresponding control pictures. The difference in arousal between threatening and control sounds in the current study are lower than the difference between threatening and control pictures of Brendel et al. (2012). The difference in valence in the current study are higher for the snake and drill sounds but in the opposite direction for the bee sounds. Similar to Brendel's pictures, the difference between threatening and control sounds in the current study was nearly 1 scale unit for both arousal and valence, which

Table 5. Means and standard deviations of threatening and scrambled **Pictures** used in Experiments 1 and 2 of Brendel et al. (2012), and IAPS Pictures of Snake (#1120), bees (#1390), and dentist drilling tooth of boy (#3280); Lang et al., 2008).

	Arousal Rating 1 = calm 9 = excited		Valence Rating 1 = unhappy 9 = happy		Dominance Rating 1 = dominated 9 = dominant	
Picture	*M*	*SD*	*M*	*SD*	*M*	*SD*
*Threatening	7.16	1.39	2.89	.93	3.86	1.86
Scrambled	6.08	1.24	3.67	1.37	4.13	1.72
IAPS Snake	6.93	1.68	3.79	1.93	3.87	2.31
IAPS Bees	5.29	1.97	4.50	1.56	4.75	1.84
IAPS Dentist	3.22	1.77	5.17	2.18	4.83	1.38

Note. * Average of Snarling Pit Bull, Masked attacker with a knife, biting snake

may account for our findings that the snake sound best fit the pattern of a threatening sound (relatively high arousal and low valence) and the shorter TTC estimates than the bee and drill sounds.

Taken together, our results were partially compatible with the hypothesis that auditory TTC estimates are shortened when an approaching sound is perceived as threatening, similar to threat-related effects on TTC estimates for visual objects. However, a difference in the TTC estimates compatible with a threat-related effect was observed only for one of the presented sounds, and regression analyses showed no systematic relation between differences in perceived threat and differences in auditory TTC estimates. Additional research is needed in order to answer the question to which extent this pattern of results can be attributed to too weak differences in perceived threat in our stimuli, or to non-emotional factors like amplitude modulations or other auditory characteristics of the sounds.

Finally, it is important to consider a limitation of the current study. To make a direct comparison between the results of the current study and the results of our prior study that used visual stimuli (Brendel et al., 2012), we selected sounds from IADS that had ratings of valence and arousal that were as close as possible to the corresponding ratings of the IAPS stimuli used by Brendel et al. (2012). Selecting sounds with different ratings would have invalidated the direct comparison. In that study, the threating pictures (snarling Pit Bull, masked attacker with knife, biting snake) had high arousal and low valence ratings that constitutes the concept of threat in prior literature (Knutson et al., 2014). Consequently, we looked for sounds with similar ratings in the IADS and used them in the current study. However, one limitation of the study is that our participants' Manikin ratings of the sounds did not match the ratings of the original IADS sounds or the pictures used by Brendel et al. (2012). Therefore, it is possible that the perceived threat of the sounds in the current study was not the same as the perceived threat of the pictures in Brendel. This could account for differences in TTC estimates between the current and prior study. The implication is that the ratings of arousal and valence that result in effects of threat on TTC estimation differ between visual and auditory stimuli. Greater degrees of arousal and negative valence may be needed for auditory stimuli to result in effects of threat. We collected ratings of valence and arousal for the stimuli, following the traditional approach in emotion research, in order to be able to differentiate between effects of valence and arousal, and to be consistent with previous studies on effects of emotion on visual TTC estimation. In hindsight, it would have been advantageous to collect additional direct ratings of threat, in addition to the threat scores computed from the valence and arousal ratings based on the approach suggested by Knutson. This could have contributed to another limitation in our comparison of the current results with those of Brendel et al., 2012), that participants did not identify the sounds well (and even for the snake sound only 9 out of 20 participants correctly identified the sound source), but identification was much better in the previous study.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Adam M. Braly http://orcid.org/0000-0001-9395-5057
Patricia DeLucia http://orcid.org/0000-0002-1735-9154
Daniel Oberfeld http://orcid.org/0000-0002-6710-3309

References

Asutay, E., Västfjäll, D., Tajadura-Jimenez, A., Genell, A., Bergman, P., & Kleiner, M. (2012). Emoacoustics: A study of the psychoacoustical and psychological dimensions of emotional sound design. *Journal of the Audio Engineering Society*, *60*, 21–28.

Bach, D. R., Schächinger, H., Neuhoff, J. G., Esposito, F., Salle, F. D., Lehmann, C., Herdener, M., Scheffler, K., & Seifritz, E. (2008). Rising sound intensity: An intrinsic warning cue activating the Amygdala. *Cerebral Cortex*, *18*(1), 145–150. https://doi.org/10.1093/cercor/bhm040

Begault, D. R., Wenzel, E. M., & Anderson, M. R. (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society*, *49*(10), 904–916.

Best, V., Baumgartner, R., Lavandier, M., Majdak, P., & Kopco, N. (2020). Sound externalization: A review of recent research. *Trends in Hearing*, *24 (10)*, https://doi.org/10.1177/2331216520948390

Blauert, J. (1996). *Spatial hearing. The psychophysics of human sound localization* (Revised ed.). MIT Press.

Bootsma, R. J., & van Wieringen, P. C. (1990). Timing an attacking forehand drive in table tennis. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(1), 21. .

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavioral Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Bradley, M. M., & Lang, P. J. (1999). *International Affective Digitized Sounds (IAD): Instruction manual and affective ratings (Tech. Rep. No. B-1)*. University of Florida, The Center for Research in Psychophysiology.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Braly, A., & DeLucia, P. R. (2020). Can stroboscopic training improve judgments of time-to-collision? *Human Factors*, *62*(1), 152–165. https://doi.org/10.1177/0018720819841938

Brendel, E., DeLucia, P. R., Hecht, H., Stacy, R. L., & Larsen, J. T. (2012). Threatening pictures induce shortened time-to-contact estimates. *Attention, Perception & Psychophysics*, *74*(5), 979–987. https://doi.org/10.3758/s13414-012-0285-0

Brendel, E., Hecht, H., DeLucia, P. R., & Gamer, M. (2014). Emotional effects on time-to-contact judgments: Arousal, threat, and fear of spiders modulate the effect of pictorial content. *Experimental Brain Research*, *232*(7), 2337–2347. https://doi.org/10.1007/s00221-014-3930-0

Buus, S., Florentine, M., & Poulsen, T. (1997). Temporal integration of loudness, loudness discrimination and the form of the loudness function. *Journal of the Acoustical Society of America*, *101*(2), 669–680. https://doi.org/10.1121/1.417959

Caird, J. K., & Hancock, P. A. (1994). The perception of arrival time for different oncoming vehicles at an intersection. *Ecological Psychology*, *6*(2), 83–109. https://doi.org/10.1207/s15326969eco0602_1

DeLucia, P. R. (1991). Pictorial and motion-based information for depth perception. *Journal of Experimental Psychology. Human Perception and Performance*, *17*(3), 738–748. 10.1037/0096-1523.17.3.738.

DeLucia, P. R. (2013). Effects of size on collision perception and implications for perceptual theory and transportation safety. *Current Directions in Psychological Science*, *22*(3), 199–204. https://doi.org/10.1177/0963721412471679

DeLucia, P. R., Brendel, E., Hecht, H., Stacy, R. L., & Larsen, J. T. (2014). Threatening scenes but not threatening faces shorten time-to-contact estimates. *Attention, Perception & Psychophysics*, *76*(6), 1698–1708. https://doi.org/10.3758/s13414-014-0681-8

DeLucia, P. R., & Liddell, G. W. (1998). Cognitive motion extrapolation and cognitive clocking in prediction-motion tasks. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 901–914. 10.1037/0096-1523.24.3.901.

DeLucia, P. R., Preddy, D., & Oberfeld, D. (2016). Audiovisual integration of time-to-contact information for approaching objects. *Multisensory Research*, *29*(4–5), 365–395. https://doi.org/10.1163/22134808-00002520

DeLucia, P. R., & Tharanathan, A. (2009). Responses to deceleration during car following: Roles of optic flow, warnings, expectations and interruptions. *Journal of Experimental Psychology. Applied*, *15*(4), 334–350. 10.1037/a0017877.

DeLucia, P. R., & Warren, R. (1994). Pictorial and motion-based depth information during active control of self motion: Size-arrival effects on collision avoidance. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(4), 783–798. 10.1037/0096-1523.20.4.783.

Ferri, F., Tajadura-Jiménez, A., Väljamäe, A., Vastano, R., & Costantini, M. (2015). Emotion-inducing approaching sounds shape the boundaries of multisensory peripersonal space. Neuropsychologia, 70, 468–475.http://dx.doi.org/10.1016/j.neuropsychologia.2015.03.001

Freiberg, K., Tually, K., & Crassini, B. (2001). Use of an auditory looming task to test infants' sensitivity to sound pressure level as an auditory distance cue. *British Journal of Developmental Psychology*, *19*(1), 1–10. https://doi.org/10.1348/026151001165903

Gordon,M. S., & Rosenblum, L. D. (2005). Effects of intrastimulus modality change on audiovisual time-to-arrival judgments. *Perception & Psychophysics*, *67*(4), 580–594. https://doi.org/10.3758/BF03193516

Gordon, M. S., Russo, F. A., & MacDonald, E. (2013). Spectral information for detection of acoustic time to arrival. *Attention, Perception, & Psychophysics*, *75*(4), 738–750. https://doi.org/10.3758/s13414-013-0424-2

Gray, R. (2011). Looming auditory collision warnings for driving. *Human Factors*, *53*(1), 63–74. https://doi.org/10.1177/0018720810397833

Haas, E., & Edworthy, J. (2006). An introduction to auditory warnings and alarms. In M. S. Wogalter (Ed.), *Human factors and ergonomics. Handbook of warnings* (pp. 189–198). Lawrence Erlbaum Associates Publishers.

Happ, M., Harrar, S. W., & Bathke, A. C. (2016). Inference for low- and high-dimensional multigroup repeated measures designs with unequal covariance matrices. *Biometrical Journal*, *58*(4), 810–830. https://doi.org/10.1002/bimj.201500064

Happ, M., Harrar, S. W., & Bathke, A. C. (2017). High-dimensional repeated measures. *Journal of Statistical Theory and Practice*, *11*(3), 468–477. https://doi.org/10.1080/15598608.2017.1307792

Happ, M., Harrar, S. W., & Bathke, A. C. (2018). HRM: An R package for analyzing high-dimensional multi-factor repeated measures. *The R Journal*, *10*(1), 534–548. https://doi.org/10.32614/RJ-2018-032

Hartmann, W. M. (2005). *Signals, sound, and sensation* (5th ed.). Springer.

Hassan, S. E. (2012). Are normally sighted, visually impaired, and blind pedestrians accurate and reliable at making street crossing decisions? Pedestrian street crossing decisions. *Investigative Ophthalmology & Visual Science*, *53*(6), 2593–2600. https://doi.org/10.1167/iovs.11-9340

Ho, S. M. Y., Mak, C. W. Y., Yeung, D. N., Duan, W. J., Tang, S., Yeung, J. C., & Ching, R. T. (2015). Emotional valence, arousal, and threat ratings of 160 Chinese words among adolescents. *Plos One*, *10*(7), 7. https://doi.org/10.1371/journal.pone.0132294

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, *75*(4), 800–802. https://doi.org/10.1093/biomet/75.4.800

Hoyle, F. (1957). *The black cloud*. Heinemann.

Jesteadt, W., & Leibold, L. (2011). Loudness in the laboratory, Part I: Steady-state sounds. In M. Florentine, A. N. Popper, & R. R. Fay (Eds.), *Loudness* (pp. 109–144). Springer.

Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: A review. *British Journal of Mathematical and Statistical Psychology*, *54*(1), 1–20. https://doi.org/10.1348/000711001159357

Keshavarz, B., Campos, J. L., DeLucia, P. R., & Oberfeld, D. (2017). 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, & Psychophysics*, *79*(3), 929–944. https://doi.org/10.3758/s13414-016-1270-9

Knutson, B., Katovich, K., & Suri, G. (2014). Inferring affect from fMRI data. *Trends in Cognitive Sciences*, *18*(8), 422–428. https://doi.org/10.1016/j.tics.2014.04.006

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual. Technical report A-6*. University of Florida.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Technical manual and affective ratings. Retrieved from https://www2.unifesp.br/dpsicobio/adap/instructions. pdf on October 26, 2021

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception*, *5*(4), 437–459. https://doi.org/10.1068/p050437

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*(2B), 467–477. https://doi.org/10.1121/1.1912375

Levulis, S. J., DeLucia, P. R., & Jupe, J. (2015). Effects of oncoming vehicle size on overtaking judgments. *Accident Analysis & Prevention*, *82*, 163–170. https://doi.org/10.1016/j.aap.2015.05.024

Mazyn, L. I., Lenoir, M., Montagne, G., & Savelsbergh, G. J. (2004). The contribution of stereo vision to one-handed catching. *Experimental Brain Research*, *157*(3), 383–390. https://doi.org/10.1007/s00221-004-1926-x

Neuhoff, J. G. (2016). Looming sounds are perceived as faster than receding sounds. *Cognitive Research: Principles and Implications*, *1*(15), 1–9. 10.1186/s41235-016-0017-4.

Neuhoff, J. G., Hamilton, G. R., Gittleson,A. L., & Mejia, A. (2014). Babies in traffic: Infant vocalizations and listener sex modulate auditory motion perception. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(2), 775–783. 10.1037/a0035071.

Oberfeld, D., Heeren, W., Rennies, J., & Verhey, J. (2012). Spectro-temporal weighting of loudness. *PLOS One*, *7*(11), e50184. https://doi.org/10.1371/journal.pone.0050184

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, *80*(3), 381. https://doi.org/10.1037/0022-3514.80.3.381

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Plack, C. J., & Carlyon, R. P. (1995). Loudness perception and intensity coding. In B. C. J. Moore (Ed.), *Hearing* (pp. 123–160). Academic Press.

Schiff, W., Caviness, J. A., & Gibson, J. J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of looming." *Science*, *136*(3520), 982–983. https://doi.org/10.1126/science.136.3520.982

Schiff, W., & Detwiler, M. L. (1979). Information used in judging impending collision. *Perception*, *8*(6), 647–658. https://doi.org/10.1068/p080647

Schiff, W., & Oldak, R. (1990). Accuracy of judging time to arrival: Effects of modality, trajectory, and gender. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(2), 303. 10.1037/0096-1523.16.2.303.

Schutte, M., Forsthofer, M., Chagnaud, B., & Wiegrebe, L. (2019). Auditory perception of distance to rattlesnakes in an audio-visual virtual environment. *23rd international congress on acoustics integrating 4th EAA Euroregio 2019*. Aachen, Germany.

Shaw, B. K., McGowan, R. S., & Turvey, M. T. (1991). An acoustic variable specifying time-to-contact. *Ecological Psychology*, *3*(3), 253–261. https://doi.org/10.1207/s15326969eco0303_4

Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., & Västfjäll, D. (2010). Embodied auditory perception: The emotional impact of approaching and receding sound sources. *Emotion*, *10*(2), 216. https://doi.org/10.1037/a0018422

The MathWorks. (1993). *MATLAB user's guide*.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Tresilian, J. R. (1995). Perceptual and cognitive processes in time-to-contact estimation: Analysis of prediction-motion and relative judgment tasks. *Attention, Perception & Psychophysics*, *57*(2), 231–245. https://doi.org/10.3758/BF03206510

Vagnoni, E., Lourenco, S. F., & Longo, M. R. (2012). Threat modulates perception of looming visual stimuli. *Current Biology*, *22*(19), R826–R827. https://doi.org/10.1016/j.cub.2012.07.053

Verhey, J. L. (1999). *Psychoacoustics of spectro-temporal effects in masking and loudness perception*. BIS-Verlag.

Wilkie, S., & Stockman, T. (2020). The effect of audio cues and sound source stimuli on the perception of approaching objects. *Applied Acoustics*, *167*, 1–15. https://doi.org/10.1016/j.apacoust.2020.107388

Zhou, L., Yan, J., Liu, Q., Li, H., Xie, C., Wang, Y., Campos, J. L., & Sun, H. J. (2007). Visual and auditory information specifying an impending collision of an approaching object. In J. Jacko (Ed.), *Human-computer interaction: Interaction platforms and techniques. Part II of the 12th international conference, HCI International 2007, Beijing, China, July 2007* (pp. 720–729), Springer, Berlin, Heidelberg.