RESEARCH ARTICLE

Auditory pitch glides influence time-to-contact judgements of visual stimuli

Carly King[1] · Steven L. Prime[1]

Abstract

A common experimental task used to study the accuracy of estimating when a moving object arrives at a designated location is the time-to-contact (TTC) task. The previous studies have shown evidence that sound motion cues influence TTC estimates of a visual moving object. However, the extent to which sound can influence TTC of visual targets still remains unclear. Some studies on the crossmodal correspondence between pitch and speed suggest that descending pitch sounds are associated with faster speeds compared to ascending pitch sounds due to an internal model of gravity. Other studies have shown an opposite pitch-speed mapping (i.e., ascending pitch associated with faster speeds) and no influence of gravity heuristics. Here, we explored whether auditory pitch glides, a continuous pure tone sound either ascending or descending in pitch, influence TTC estimates of a vertically moving visual target and if any observed effects are consistent with a gravity-centered or gravity-unrelated pitch-speed mapping. Subjects estimated when a disc moving either upward or downward at a constant speed reached a visual landmark after the disc disappeared behind an occluder under three conditions: with an accompanying ascending pitch glide, with a descending pitch glide, or with no sound. Overall, subjects underestimated TTC with ascending pitch glides and overestimated TTC with descending pitch glides, compared to the no-sound condition. These biases in TTC were consistent in both disc motion directions. These results suggest that subjects adopted a gravity-unrelated pitch-speed mapping where ascending pitch is associated with faster speeds and descending pitch associated with slower speeds.

Keywords Crossmodal correspondence · Auditory pitch · Prediction motion · Time-to-contact

Introduction

Predicting the movements of objects is important for many everyday activities such as crossing a busy street, driving in traffic, or catching a ball. This task of predicting the future location of moving objects is commonly called a prediction motion (PM) task (DeLucia and Liddell 1998; Gottsdanker 1955; Tresilian 1995). One specific type of PM task that is often used in the laboratory to study how accurate observers are in making such predictions is the time-to-contact (TTC) task. In a typical TTC task, subjects estimate when a moving object reaches a designated target location some time after the object disappeared (e.g., moving behind an occluding bar). Subjects are usually required to make a button press

response coinciding with the exact time they judge the object has reached the target.

To perform the TTC task accurately, subjects must base their TTC estimates on a variety of motion cues (DeLucia et al. 2003). Most real-world moving objects provide both visual and auditory motion cues. The majority of TTC studies have used only unimodal motion stimuli, either only visual motion stimuli (e.g., Alderson and Whiting 1974; DeLucia et al. 2003; Huber and Krist 2004; Peterken et al. 1991) or auditory motion stimuli (e.g., Gray 2011; Gordon et al. 2013, Rosenblum et al. 1993). A few studies have included combined audiovisual stimuli to investigate the effectiveness of integrating auditory and visual motion information in TTC judgements. In general, TTC judgements of audiovisual motion stimuli have been shown to be more accurate than of unimodal auditory stimuli and as accurate (or nearly so) as that of unimodal visual stimuli (Hofbauer et al. 2004; Keshavarz et al. 2017; Schiff and Oldak 1990; Zhou et al. 2007). This failure to find an advantage for audiovisual conditions over unimodal visual conditions does not

✉ Steven L. Prime
steve.prime@usask.ca

[1] Neurocognition and Psychophysics Laboratory, Department of Psychology, University of Saskatchewan, 9 Campus Drive, 154 Arts, Saskatoon, SK S7N 5A5, Canada

necessarily indicate that audiovisual TTC judgements are driven solely by visual information, while auditory information is completely ignored. Indeed, more recent studies have shown that auditory and visual information is optimally integrated in audiovisual TTC judgements (Wuerger et al. 2010) with visual motion cues weighted more than auditory motion cues (DeLucia et al. 2016; Keshavarz et al. 2017; Zhou et al. 2007). Moreover, other findings demonstrate that auditory information can modulate audiovisual PM performance. Estimating the future position of audiovisual motion stimuli is significantly less accurate when auditory cues are uncorrelated or incongruent with visual cues (Gordon and Rosenblum 2005; Prime and Harris 2010).

The types of sound stimuli used to produce auditory motion cues in audiovisual PM studies have been limited. Of the previously cited audiovisual PM studies, some have used either looming (approaching) stimuli of increasing visual image size and sound intensity (DeLucia et al. 2016; Gordon and Rosenblum 2005; Keshavarz et al. 2017; Schiff and Oldak 1990; Zhou et al. 2007) or translating stimuli of a visual target and a clicking or continuous pure tone sound moving together laterally in front of the subject (Hofbauer et al. 2004; Prime and Harris 2010; Wuerger et al. 2010). In a recent audiovisual PM study, Chotsrisuparat et al. (2017) used novel auditory stimuli where they varied the rhythm of auditory beeps. They found that slower rhythms resulted in longer TTC overestimations of the visual object as though subjects judged the object as moving slower. In addition to providing further evidence that TTC judgements of visual objects can be influenced by auditory stimuli, the researchers postulated that their subjects' TTC judgements were influenced by a learned crossmodal association between auditory rhythm and movement speed. The extent to which TTC judgements might be influenced by other crossmodal associations with other auditory features has yet to be explored.

Here, we sought to extend this previous research by examining the crossmodal effects of continuous pure tones ascending or descending in pitch (i.e., pitch glide) on TTC judgements of visual objects. Several studies have demonstrated crossmodal correspondences between auditory pitch and a variety of visual features, such as spatial location, spatial frequency, size, brightness, and shape (Bernstein and Edelstein 1971; Evans and Treisman 2011; Gallace and Spence 2006; Jamal et al. 2017; Marks 1987; Melara 1989; for review see Spence 2011). More relevant to the present study are the crossmodal correspondences between pitch and both motion direction and speed. The previous studies have shown that ambiguous visual motion stimuli are perceived as moving either upward or downward when paired with ascending or descending pitch glides, respectively (Carnevale and Harris 2016; Jain et al. 2008; Maeda et al. 2004; Sadaghiani et al. 2009). These findings demonstrate an up–down mapping of pitch where observers associate

ascending pitch with upward moving objects and descending pitch with falling objects (for a similar conclusion see Trimble 1934; Walker 1981, 1985). Moreover, the perception of ascending and descending pitch sounds appears to be differentially influenced by an internal model of gravity. Specifically, compared to ascending pitch tones, subjects are more likely to misjudge the pitch of a target tone embedded in a sequence of descending pitch tones as shifted in the direction of the pitch change (i.e., lower than its true pitch), suggesting that descending pitch sequences are perceived as traveling faster through pitch space (Henry et al. 2009; Hubbard 1995; Hubbard and Ruppel 2013). Indeed, in explaining this asymmetrical bias, the researchers of the previously cited studies postulated that, in addition to the mapping of descending (ascending) pitch sounds with downward (upward) motion, subjects applied their implicit knowledge of gravity, and thus, they likely heard descending pitch tones as an object accelerating as it falls and ascending pitch tones as an object decelerating as it rises. This is consistent with Godøy's (2001) argument that sounds evoke a mental image of the object that produced those sounds, including a visual image. In fact, a similar gravity-congruent bias has been demonstrated with spatial judgements of the last seen location of moving visual targets; downward moving visual targets tend to be judged as having been moved farther (more downward) than upward moving visual targets (Hubbard 1995).

However, other studies show that observers do not always adopt gravity-centered heuristics when interpreting sounds ascending or descending in pitch. These other studies have instead shown that when subjects were asked to categorize these sounds with respect to perceived direction or speed, ascending pitch sounds were associated with moving upward and fast or speeding up, while descending pitch sounds were associated with moving downward and slow or slowing down (Collier and Hubbard 2001; Eitan and Granot 2006). These crossmodal associations appear to be acquired through experience with sounds produced by objects moving at different speeds and accelerations either in the real world or on television and in film (e.g., the high-pitch whine of a motorcycle at high speed or the low-pitch rumble of train slowing to a stop). To our knowledge, no study has directly examined whether either of these gravity-centered or gravity-unrelated relationships between pitch and motion direction and speed has a crossmodal influence on visual TTC judgements.

In the present study, subjects estimated the TTC of a visual disc moving at a constant speed either upward or downward along the vertical plane and disappearing behind an occluder (a black bar). Subjects were required to press a button at the exact time they judged the moving disc would have reached a marked target location on the occluder. Subjects performed this task under two audiovisual conditions,

both consisting of a concurrent auditory signal of either an ascending or descending pitch glide. TTC estimations from these audiovisual conditions were compared to a similar no-sound control condition. Our goal was to determine if ascending and descending pitch glides influence visual TTC estimations and if such effects are consistent with either the gravity-centered or gravity-unrelated pitch-speed correspondences described earlier. Based on the findings of these studies, we propose that two different possible hypotheses can be made.

First, if a pitch effect on visual TTC judgements is consistent with the gravity-centered pitch-speed mapping as reported by some of the previously cited studies (Henry et al. 2009; Hubbard 1995; Hubbard and Ruppel 2013), then descending pitch glides would activate its association with an accelerating object as it falls and ascending pitch glides would activate its association with a decelerating object as it rises. We would then predict that, compared to the no-sound condition, descending pitch glides paired with downward moving discs will result in smaller TTC estimations (i.e., the disc will be judged as arriving at the target location sooner than it actually did), whereas ascending pitch paired with upward moving discs will result in larger TTC estimations (i.e., the disc will be judged as arriving later than it actually did). Regarding the other two audiovisual conditions that involve incongruent directions with respect to pitch glide and disc motion (ascending pitch with downward motion and vice versa), predictions under the gravity-centered hypothesis are less clear. Under conditions of incongruency, we think that it is reasonable to hypothesize that subjects will either (1) rely more on the visual motion information, and thus, TTC estimations in these conditions will more resemble the no-sound condition, or (2) be similarly influenced by auditory signals as hypothesized for the congruent conditions as evidenced by a similar bias for shorter TTC estimations with descending pitch glides and longer TTC estimations with ascending pitch glides compared to the no-sound condition.

The second possible hypothesis, on the other hand, is that, if a pitch effect is consistent with the gravity-unrelated pitch-speed mapping as shown by other studies (Collier and Hubbard 2001; Eitan and Granot 2006), then ascending pitch glides would activate its association with fast or accelerating speeds and descending pitch glides would activate its association with slow or decelerating speeds. We would then predict that ascending pitch glides will be associated with smaller TTC estimations and descending pitch glides with larger TTC estimations, compared to the no-sound condition. Since subjects in Collier and Hubbard (2001) and Eitan and Granot (2006) are also associated ascending pitch with upward motion and descending pitch with downward motion, it is possible that such predicted effects of pitch will be strongest when the direction of the pitch glide and

the disc motion are congruent (e.g., ascending pitch glide with upward moving disc). On the other hand, it is possible that this gravity-unrelated pitch-speed correspondence of faster speeds with ascending pitch and slower speeds with descending pitch is independent of perceived direction. In which case, we would predict that we would find these TTC biases regardless of disc motion direction.

Methods

Subjects

Thirty-seven subjects (15 males, mean age 23.8 years) participated in this study. All subjects had normal or corrected-to-normal vision and normal, uncorrected hearing according to self-report. All procedures were approved by the University of Saskatchewan's Behavioural Research Ethics Board. Informed consent was obtained from all individual research participants included in the study.

Apparatus

Stimulus presentation and data recordings were controlled by a specialized experiment software package, PsychoPy v1.83 (Pierce 2007), on a personal computer (3.2 GHz and 8 GB RAM). Subjects recorded their responses on the computer keyboard. Visual stimuli were presented on a 144 Hz 26-inch Benq LCD monitor. Auditory stimuli were presented from a pair of speakers (Logitech Z130) mounted behind and on opposite sides of the monitor. A chin rest used to align subjects eyes to the center of the screen. The distance from the eye to the computer screen was approximately 60 cm. The subjects were tested in a dimly lit room.

Stimuli

The moving visual object was a white disc (subtending 0.5 cm). The disc moved at a constant speed (5.5 cm/s) toward and eventually behind an occluder. The occluder was a black bar 1 cm in width and 19 cm in length. On the occluder was a thin white line drawn across the occluder's width that specified the arrival location. The line was presented at one of three different distances relative to the end of the occluder where the disc disappeared, either at 8.5, 11, or 13.5 cm. The TTC for these different distances was 1545, 2000, and 2455 ms, respectively, relative to the time the disc disappeared behind the occluder.

The auditory pitch glide in the *Ascending Pitch* condition was a continuous tone whose frequency swept linearly upward from 500 to 825 Hz over a duration of 5.5 s. The pitch glide in the *Descending Pitch* condition was similar except it swept linearly downward from 450 to 125 Hz over

the same duration. These frequency ranges used to create our pitch glides were chosen, because they were similar to pitch-changing sound stimuli used in the other studies (Collier and Hubbard 2001; Henry et al. 2009; Hubbard and Ruppel 2013; Maeda et al. 2004; Mossbridge et al. 2011; Walker 1987), and they were judged to be the most comfortable to listen to, while still being well perceived. Sound immediately stopped regardless of what point it was in the pitch glide when subjects made their TTC response. The rationale for the sound presentation to continue after the disc was occluded was to approximate a typical real-world prediction motion scenario where observers might only see a moving object briefly before looking away or the object is obscured, but they can often still hear the object after it is no longer visually available. Unlike visual information, which can be interrupted by looking away, blinking, or the object being obscured, real-world auditory information is often continuously available. For example, crossing a street with traffic. A pedestrian might see a car approaching from one direction and then turn to see if there are cars coming from the other direction. Although the pedestrian has no view of the car anymore, the pedestrian can still hear it and supposedly use sound information. Our paradigm where the auditory presentation continues after occlusion is similar to one of the author's previous studies (Prime and Harris 2010) and other studies (e.g., Chotsrisuparat et al. 2017). Auditory stimuli were presented at ~68 dB SPL, measured at the subjects' head position.

Procedure

Figure 1 shows the general experimental design of our TTC task. Each trial began with the presentation of the disc, the occluder, and the line marking the arrival point. In half of the trials, the disc was presented stationary at the bottom of the display (Fig. 1a), and in the other half of the trials, the disc was presented at the top of the display (Fig. 1b). The occluder was always presented as vertically oriented on the opposite side of the display to the disc. The distance between the disc's starting locations and the center of the screen were 13 cm and to the nearest edge of the occluder was 9.5 cm. Stimulus layout (disc starting from the top vs. bottom) and arrival line position were randomly interleaved.

Subjects pressed the spacebar to start the disc's motion. The disc moved vertically, either downward from the top or upward from the bottom of the display, toward and eventually behind the occluder. Subjects were told that the disc would continue to move when behind the occluder. Subjects were instructed to estimate TTC as accurately as possible by pressing the spacebar as soon as they thought the disc would have reached the line. In the *Ascending* and *Descending Pitch* trials, the onset of the sound stimulus began simultaneously with the motion onset of the disc. In the *No-Sound* trials, the disc moved by itself with no-sound stimulus. The *No-Sound* condition acted as the control condition, because we wanted to compare the sound conditions to baseline task performance when relying strictly on visual information, and, thus, be able to assess the extent to which the pitch glides influence TTC estimates relative to visual-only
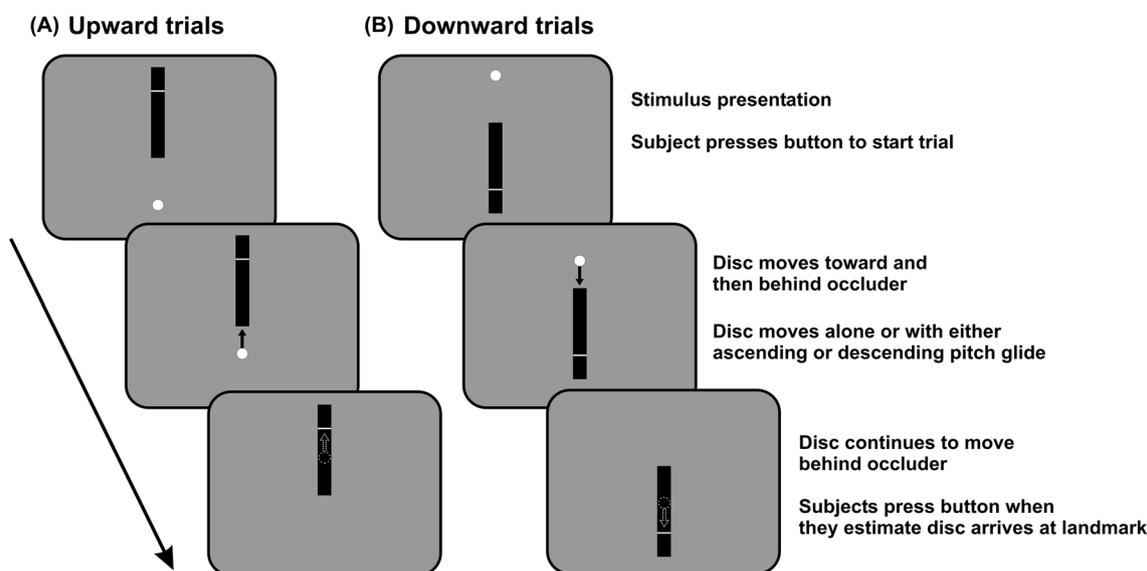

Fig. 1 General experimental paradigm of the time-to-contact (TTC) task. Subjects had to judge when a moving disc arrived at the white line on the black occluder. Upon the subject pressing a computer key, the disc moved either upward from the bottom (**a**) or downward from the top of the screen (**b**). The disc moved toward and then behind the occluder (illustrated by the dashed circle and arrow). Subjects estimated the TTC between the disc and the line by pressing a computer key as soon as they judged the disc reached the line

conditions. The order of the *Ascending Pitch*, *Descending Pitch*, and *No-Sound* trials were randomly interleaved. Upon subjects pressing the spacebar to record their TTC estimation, all visual stimuli disappeared, and in the case of the *Ascending Pitch* and *Descending Pitch* conditions, the sound stimulus stopped. The subjects were not given any instructions about how to follow the disc or how to interpret the sound stimuli. No response feedback was given. Subjects performed 12 practice trials before performing the experiment. The experiment consisted of five blocks of 54 trials for a total of 270 trials. After each block was a self-timed rest break.

Data analysis

Consistent with the previous TTC studies (e.g., Baures et al. 2018; Bennett et al. 2010; Rosenblum 1993; Zheng and Maraj 2018), we analysed two common measurements of TTC performance: constant error and variable error. Constant error (CE) is a measure of the accuracy and bias of the estimated TTC. CE was calculated as the difference between the subjects' estimation of the time the disc arrived at the line and the actual time of arrival (i.e., $TTC_{estimated} - TTC_{actual}$). Negative CE means that subjects responded before the disc arrived at the line, and therefore, the TTC was underestimated. Positive CE means that subjects responded after the disc arrived at the line, and therefore, the TTC was overestimated. Variable error (VE) is a measure of TTC consistency and is the standard deviation of the estimated

TTC responses. We analysed CE and VE using repeated-measures ANOVAs with movement direction (upward and downward), sound condition (*Descending Pitch*, *Ascending Pitch*, and *No-Sound*), and line distance (near, middle, and far) as factors.

Results

The main goal of this study was to compare the TTC estimations between the three sound conditions: *Ascending Pitch*, *Descending Pitch*, and *No-Sound*. Before conducting our analyses, we excluded trials where the subjects pressed the response button before the disc disappeared behind the occluder (anytime before 1.73 s of total travel time) or after the disc remerged from behind the occluder (anytime after 5.18 s of total travel time). These excluded trials accounted for only 1.5% of all trials. For all analyses, a *p* value of 0.05 was adopted for significance. Post hoc comparisons were analysed using the Hochberg step-up method (Hochberg 1988; Keselman 1994). All reported within-subject 95% confidence intervals were calculated using the Cousineau–Morey method (Cousineau 2005; Morey 2008).

Constant error (CE) analyses

Figure 2 shows the mean CE results for all three sound conditions as a function of line distance in the downward disc condition (Fig. 2a) and upward disc condition (Fig. 2b).
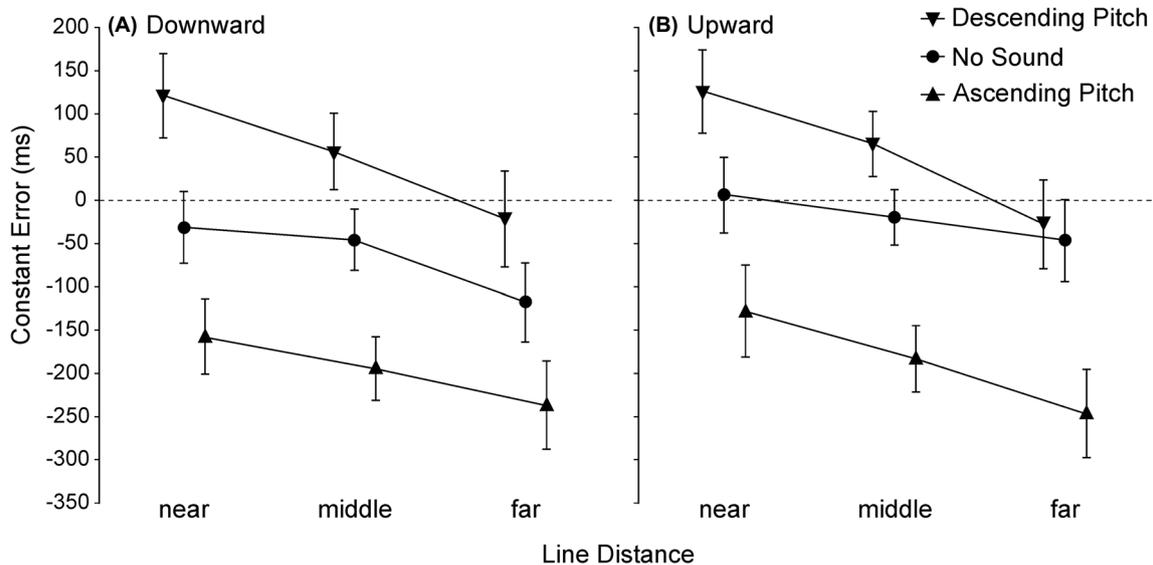

Fig. 2 TTC constant error (CE) shown separately for each sound condition, line location, and direction. Negative and positive CE indicates under- and overestimations, respectively. Zero CE indicated perfect TTC accuracy. Data curves are slightly jittered to aid visibility of the data points and error bars. Overall, subjects overestimated

TTC in the *Descending Pitch* condition and underestimated TTC in the *Ascending Pitch* condition relative to the *No-Sound* condition in both downward (**a**) and upward (**b**) disc motion directions. Error bars represent 95% Cousineau–Morey confidence intervals for within-subjects designs (Cousineau 2005; Morey 2008)

Analysing the CE allowed us make a simple assessment of the overall accuracy of the TTC estimations among the three sound conditions for both disc motion directions. Our main interest here was to compare and assess the direction of any potential differences between both the *Ascending* and *Descending Pitch* conditions and the *No-Sound* condition (i.e., overestimations or underestimations), and characterize these patterns of differences for each disc direction. A visual inspection of these CE data in Fig. 2 shows that, in general, TTC was more underestimated in the *Ascending Pitch* condition and overestimated in the *Descending Pitch* condition relative to the *No-Sound* condition, regardless of disc direction. In addition, TTC estimates overall decreased with increasing line distance (i.e., time disc is occluded), which is a typical finding in TTC experiments (e.g., Hancock and Manser 1997; Schiff and Oldak 1990). These observations were confirmed by an initial three-way repeated-measures ANOVA (Direction × Sound Condition × Line Distance), which revealed significant results for the main effects of Sound Condition [$F_{(2,72)} = 58.38$, $p < 0.001$, $\eta_p^2 = 0.62$] and Line Distance [$F_{(2,72)} = 13.5$, $p < 0.001$, $\eta_p^2 = 0.27$], but not for Direction [$F_{(1,36)} = 1.45$, $p = 0.24$]. However, the Sound Condition × Direction interaction was significant [$F_{(2,72)} = 4.62$, $p = 0.013$, $\eta_p^2 = 0.11$] and the Direction × Sound Condition × Line Distance interaction approached significance [$F_{(4,144)} = 2.07$, $p = 0.08$, $\eta_p^2 = 0.05$].

To further explore the nature of these effects more closely, we conducted separate two-way repeated-measures ANOVAs (Sound Condition × Line Distance) for each disc direction. For the downward condition (Fig. 2a), significant results were found for both main effects of Sound Condition [$F_{(2,72)} = 54.23$, $p < 0.001$, $\eta_p^2 = 0.6$] and Line Distance [$F_{(2,72)} = 12.02$, $p < 0.001$, $\eta_p^2 = 0.25$], while the Sound Condition × Line Distance interaction approached significance [$F_{(4,144)} = 2.34$, $p = 0.058$, $\eta_p^2 = 0.06$]. Hochberg post hoc analyses of the Sound Condition main effect showed significant differences between the *No-Sound* condition ($M = -65$ ms; 95% CI − 90 to − 41) and both the *Ascending Pitch* ($M = -196$ ms; 95% CI − 228 to − 165) and *Descending Pitch* ($M = 52$ ms; 95% CI 17–86) conditions, as well as a significant difference between *Ascending Pitch* and *Descending Pitch* (all comparisons $p < 0.01$). These results are consistent with our main observation that TTC was more underestimated in trials with an ascending pitch compared to no-sound trials and overestimated in trials with a descending pitch, indicating that subjects judged the disc as faster to arrive at the line with ascending pitch and slower with descending pitch. Post hoc analyses of the Line Distance main effect showed significant differences between the far distance ($M = -126$ ms; 95% CI − 158 to − 94) and both the near ($M = -23$ ms; 95% CI − 54 to 8) and middle ($M = -62$ ms; 95% CI − 75 to − 48) distances (comparisons

$p < 0.01$), but not between the near and middle distances ($p = 0.07$). These results show that overall greater line distances resulted in larger underestimations of TTC when collapsing across sound conditions.

The two-way ANOVA in the upward condition (Fig. 2b) yielded significant results for both main effects of Sound Condition [$F_{(2,72)} = 52.27$, $p < 0.001$, $\eta_p^2 = 0.6$] and Line Distance [$F_{(2,72)} = 12.72$, $p < 0.001$, $\eta_p^2 = 0.26$], and significance for the Sound Condition × Line Distance interaction [$F_{(4,144)} = 4.56$, $p < 0.01$, $\eta_p^2 = 0.11$]. Post hoc analyses of the Sound Condition main effect showed significant differences between the *No-Sound* condition ($M = -20$ ms; 95% CI − 45 to 4) and both the *Ascending Pitch* ($M = -187$ ms; 95% CI − 217 to − 157) and *Descending Pitch* ($M = 55$ ms; 95% CI 18–92) conditions, as well as a significant difference between *Ascending Pitch* and *Descending Pitch* (all comparisons $p < 0.01$). As in the downward condition, TTC was more underestimated in trials with an ascending pitch and overestimated in trials with a descending pitch compared to no-sound trials, and like before, indicates that subjects judged the disc as faster to arrive at the line with ascending pitch and slower with descending pitch. Post hoc analyses of the Line Distance main effect showed significant differences between the far distance ($M = -108$ ms; 95% CI − 141 to − 74) and both the near ($M = 13$ ms; 95% CI − 17 to 44) and middle ($M = -46$ ms; 95% CI − 61 to − 32) distances (comparisons $p < 0.01$), and between the near and middle distances ($p = 0.032$). Again, as before, these results show that overall greater line distances resulted in larger underestimations of TTC when collapsing across sound conditions. Finally, the interaction effect of Sound Condition and Line Distance was due to significant comparisons between all the different sound conditions at each line distance (all comparisons: $p < 0.01$), but not between *Descending Pitch* and *No-Sound* at the far line ($p = 0.5$).

Variability error (VE) analyses

Figure 3 shows the mean VE results for all three sound conditions as a function of line distance in the downward disc condition (Fig. 3a) and upward disc condition (Fig. 3b). As with the CE analyses, we conducted an initial three-way repeated-measures ANOVA (Direction × Sound Condition × Line Distance), which revealed a significant result for the main effect of Line Distance [$F_{(2,72)} = 31.28$, $p < 0.001$, $\eta_p^2 = 0.47$], but not for the main effects of Sound Condition [$F_{(2,72)} = 2.62$, $p < 0.08$] or Direction [$F_{(1,36)} = 0.52$, $p = 0.47$]. The Sound Condition × Direction interaction was also significant [$F_{(2,72)} = 4.44$, $p = 0.015$, $\eta_p^2 = 0.11$].

As before, we studied more closely the nature of these effects by conducting separate two-way repeated-measures ANOVAs (Sound Condition × Line Distance) for each disc
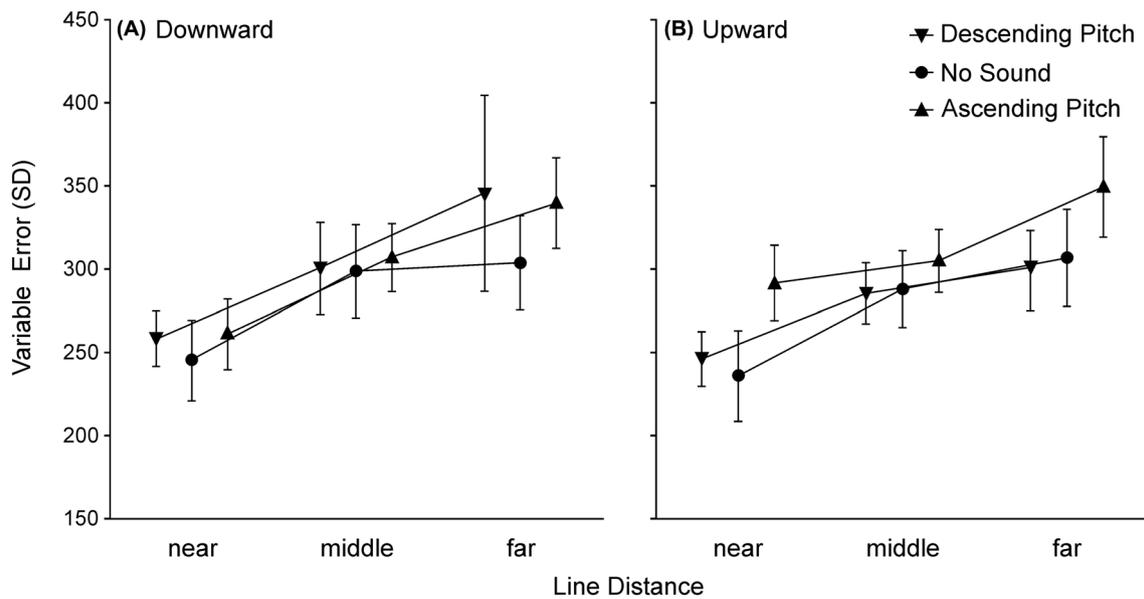
Fig. 3 TTC variable error (VE) shown for each sound condition, line location, and distance. Overall, VE increased with line distance across all sound condition in both the downward (**a**) and upward (**b**) disc motion directions. Data curves are slightly jittered to aid visibil-ity of the data points and error bars. Error bars represent 95% Cous-ineau–Morey confidence intervals for within-subjects designs (Cous-ineau 2005; Morey 2008)

direction. For the downward condition (Fig. 3a), the main effect of Line Distance was significant [$F_{(2,72)} = 19.21$, $p < 0.001$, $\eta_p^2 = 0.35$], indicating that variability of TTC esti-mates increased progressively with line distance, which is consistent with the TTC research literature (e.g., Rosenblum 1993; Schiff and Oldak 1990; Tresilian 1995). Post hoc tests confirmed this observation, showing that VE was signifi-cantly greater for both the far ($M = 329$; 95% CI 313–346) and middle ($M = 302$; 95% CI 289–315) line distances com-pared to the near ($M = 255$; 95% CI 243–266) line distance (all comparisons $p < 0.001$) and for the comparison between the middle and far line distances ($p = 0.027$). No significance was found for the main effect of Sound Condition [$F_{(2,72)} = 0.99$, $p < 0.38$] and the Sound Condition × Line Distance interaction [$F_{(4,144)} = 0.56$, $p = 0.69$].

For the upward condition (Fig. 3b), the main effect of Line Distance was significant [$F_{(2,72)} = 19.4$, $p < 0.001$, $\eta_p^2 = 0.35$], similar to the VE results in the downward condi-tion. Post hoc comparisons showed significant results between the near ($M = 257$; 95% CI 247–268) and both the middle ($M = 293$; 95% CI 283–303) and far ($M = 318$; 95% CI 306–331) line distances (both comparisons $p < 0.01$) and between the middle and far line distances ($p = 0.01$). The Sound Condition × Line Distance interaction also was not significant [$F_{(4,144)} = 1.27$, $p = 0.28$]. The main effect of Sound Condition was significant [$F_{(2,72)} = 5.99$, $p = 0.004$, $\eta_p^2 = 0.14$]. Post hoc analyses of the Sound Condition main effect showed that this significant result was driven by the

greater VE in the *Ascending Pitch* condition ($M = 315$; 95% CI 302–329) compared to both the *Descending Pitch* ($M = 276$; 95% CI 265–289) and *No-Sound* ($M = 277$; 95% CI 259–294) conditions (both comparisons $p < 0.01$), while the comparison between *Descending Pitch* and *No-Sound* was not significant ($p = 0.99$). To further investigate the sig-nificant difference between the *Ascending Pitch* and *No-Sound* conditions, we conducted additional post hoc com-parisons between these two sound conditions separately at each line distance. These comparisons yielded only one significant difference at the near line distance ($p < 0.01$); the non-significant comparisons between the *Ascending* and *No-Sound* conditions at the middle and far line distances were $p = 0.33$ and $p = 0.14$, respectively.

Discussion

The aim of the present study was to explore whether, and if so how, ascending and descending auditory pitch glides influence TTC estimates of a visual stimulus. Our investiga-tion focused on whether any such influence would be more consistent with a gravity-centered or a gravity-unrelated pitch-speed correspondence. According to the gravity-cen-tered pitch-speed mapping hypothesis, observers employ gravity heuristics when making perceptual judgements of the pitch glides, whereby descending pitch glides are perceived as an object falling and moving faster or accelerating and ascending pitch glides are perceived as an object rising and

possibly decelerating. A prediction based on this hypothesis is that TTC would be underestimated (i.e., subjects judge the disc as arriving sooner) when the disc was moving downward and paired with a descending pitch glide. Moreover, we might predict that TTC would be overestimated (i.e., subjects judge the disc as arriving later) when the disc was moving upward and paired with an ascending pitch glide. Conversely, according to the gravity-unrelated pitch-speed mapping hypothesis, descending pitch glides are associated with slower or decelerating motion, whereas ascending pitch glides are associated with faster or accelerating motion. A prediction based on this hypothesis is that, in general, TTC would be underestimated with ascending pitch glides and overestimated with descending pitch glides.

Our main results show that auditory pitch glides influenced TTC judgments consistent with the gravity-unrelated pitch-speed mapping hypothesis. Specifically, subjects showed a greater underestimation of the TTC with ascending pitch glides and a greater overestimation of the TTC with descending pitch glides compared to the no-sound condition. These biases in TTC judgments were consistent across both upward and downward dot motion directions. Furthermore, in general, presenting visual and auditory stimuli simultaneously did not affect the variability of the TTC estimates. Overall, variable errors of the TTC estimates were not statistically different across the three conditions. Altogether, these findings provide further evidence that sound can influence prediction motion judgments under audiovisual stimulus conditions, consistent with the previous crossmodal prediction motion studies (Chotsrisuparat et al. 2017; Gordon and Rosenblum 2005; Prime and Harris 2010; Wuerger et al. 2010). Moreover, we extend this previous research with novel results that demonstrate that ascending and descending pitch glides systematically bias TTC judgments of a visual target in a manner that suggest an influence on the perceived speed of the target unrelated to motion direction and without adopting gravity heuristics.

Our results suggest that our subjects attributed a faster speed to the ascending pitch glides and a slower speed to the descending pitch glides, which is consistent with the pitch-speed mapping reported in Collier and Hubbard (2001) and Eitan and Granot (2006). However, our results do not challenge the findings of those previously cited studies that show an up–down mapping of ascending and descending pitch sounds (Carnevale and Harris 2016; Jain et al. 2008; Maeda et al. 2004; Sadaghiani et al. 2009) and evidence that acceleration is attributed to descending pitch consistent with an internal representation of gravity (Henry et al. 2009; Hubbard 1995; Hubbard and Ruppel 2013). Indeed, taken together with these previous findings, our results suggest that observers use different mapping strategies in a context-dependent manner based on the nature of the task and the availability and reliability of other information (e.g., from

vision). In the studies showing gravity-congruent effects (Henry et al. 2009; Hubbard 1995; Hubbard and Ruppel 2013), subjects had to judge the pitch height of a target tone embedded in rising or falling pitch tone sequences in the absence of other stimuli. In the studies showing the up–down mapping of pitch (Carnevale and Harris 2016; Jain et al. 2008; Maeda et al. 2004; Sadaghiani et al. 2009), subjects were required to judge the direction of ambiguous visual motion stimuli paired with pitch glides. In both of these sets of studies, subjects were not presented with any additional reliable information about motion direction or speed beyond the auditory pitch-changing stimulus itself. In these cases, it appears as though the ascending and descending pitch sounds alone led subjects to adopt up–down and gravity-centered mapping strategies. However, in the case of the present study, additional unambiguous motion information was provided; the disc was clearly seen moving either up or down before it disappeared behind the occluder. Moreover, our subjects were able to see that the direction of the pitch glides was uncorrelated with the direction of the disc in that congruent and incongruent pitch and disc directions were equally probable. Thus, subjects could see that overall the ascending and descending pitch glides did not imply rising or falling objects, respectively. This context appears to have guided the subjects' mapping strategy in a way that preferred the gravity-unrelated mapping of pitch and speed as per Collier and Hubbard (2001) and Eitan and Granot (2006).

Although we find that this explanation of adopting different mapping strategies based on the availability and nature of additional motion information is appealing, it is not entirely satisfactory. In both Collier and Hubbard (2001) and Eitan and Granot (2006), the sound stimuli were also presented alone. Why, then, did the subjects in these studies not perceive these sounds also according to a gravity-centered mapping of pitch and speed (e.g., subjects ascribing faster speeds or acceleration to descending pitch sounds compared to ascending pitch sounds)?

The tasks in those studies that show up–down or gravity-centered mapping were specifically designed to test vertical judgements of either the direction of the coinciding ambiguous visual motion stimulus (Carnevale and Harris 2016; Jain et al. 2008; Maeda et al. 2004; Sadaghiani et al. 2009) or position of the target tone in pitch space (Henry et al. 2009; Hubbard 1995; Hubbard and Ruppel 2013), respectively. In all of these studies, subjects responded by making a two-alternative forced choice (2AFC) keypress, e.g., up/down or same/different. Moreover, the durations of the sounds in these studies were relatively very short (the sound duration in most of these studies were as brief as a couple hundred milliseconds). Thus, in these tasks where subjects had a short time to process and make ascriptions of the meaning of the sounds and their responses were limited to two discrete choices about vertical direction/position, it appears

up–down and gravity-centered mappings were activated rapidly and preferentially. Conversely, in Collier and Hubbard (2001) and Eitan and Granot (2006), subjects were presented with pitch glides of substantially longer durations (e.g., as much as 6 s in Collier and Hubbard 2001) and completed questionnaires where they rated and categorized the sounds with respect to several perceptual dimensions, including speed, direction, tempo, and brightness. Thus, we speculate that the longer time subjects had to process and ascribe different meanings to the pitch glides with the task of judging the sounds among multiple dimensions among multiple response alternatives (e.g., selecting its rate of speed and acceleration along a Likert scale) activated additional crossmodal mappings. This idea is consistent with the previous arguments that the nature of the response options themselves affects response choice (Einhorn and Hogarth 1981; Payne, Bettman, and Johnson 1988; Tversky and Sattath 1979). Under these conditions, the crossmodal mapping between the pitch glide direction and both speed and acceleration became dissociated with an internal gravity representation; although subjects in Collier and Hubbard (2001) and Eitan and Granot (2006) judged ascending pitch glides as faster or accelerating and descending pitch glides as slower or decelerating, they also judged the ascending pitch glides as upward motion and descending pitch glides as downward motion. Similarly, in the present study, subjects' responses were not limited to two discrete choices about vertical direction/position and they also had more time to process the sounds compared to studies showing up–down and gravity-centered mappings (the sounds in the present study played throughout the entire trial in the audiovisual conditions until the subjects made a keypress). This, coupled with the fact that, as mentioned before, the disc provided unambiguous motion information uncorrelated with pitch glide direction suggests that, in addition to taking into account the contextual information of the visual motion, subjects also take into account the task demands as a guide to which mapping of pitch and speed to adopt.

Whether subjects adopt a gravity-centered mapping could also depend on the nature of the visual scene. Miller et al. (2008) showed that subjects take gravity-congruent acceleration into account to make arrival timing judgements of a falling object when the object was presented within a pictorial scene (i.e., a ball falling along the side of a building's exterior wall), but not when the moving object was presented in a non-pictorial scene (i.e., against a uniform coloured background). These findings suggest that subjects used an internal gravity model when the visual scene provides pictorial references to gravity (e.g., an object falling toward the ground in a real-world scene vs. an object simply moving downward toward another object in an otherwise empty scene). It is possible that, in the present study, our subjects failed to show evidence of adopting a gravity-centered mapping, because we also used a simple visual display similar to the non-pictorial scene presentation in Miller et al. (2008).

Of course, we acknowledge that these different explanations are highly speculative, but we assert that it provides new interpretations that can be tested in future studies. We also acknowledge that our findings might not generalize well to other sound presentations (e.g., using constant-pitch tones or the sound ending when the disc is occluded). Future research can elucidate the potential effects of different stimulus durations, background scenes, pitch stimuli (e.g., constant vs. gliding pitch), perceptual judgement tasks, and response types (e.g., 2AFC vs. ratings) might have on evoking different mappings between pitch and either direction or speed and whether they are gravity-centered or not.

An alternative interpretation suggested by an anonymous reviewer is that our results are due to demand characteristics (for a discussion of how demand characteristics can impact perception research, see Firestone and Scholl 2016). According to this view, subjects responded earlier with ascending pitch and later with descending pitch sounds consistent with the association between these pitch glides and speed, because they believed it was expected of them. Thus, subjects would not be making their button responses to record their actual estimation of when the disc arrived at the line (i.e., estimating TTC). We acknowledge that we cannot completely discount this possibility. However, we argue that this interpretation is unlikely to account fully for our results. First, we only instructed subjects to estimate the disc's TTC by giving them clear instructions to press the computer button at the moment they thought the disc arrived at the line, which they practiced in a block of practice trials under our guidance, and they were not given any instruction about how to interpret the sounds. Second, if subjects were responding according to demand characteristics, we believe that it is reasonable to expect that subjects would have different ideas about how early or late to respond. In ascending pitch trials, for example, some subjects might press the computer button shortly after the disc is occluded, other subjects are just as likely to wait longer and press it when the occluded disc is somewhere in the middle of the time it takes to reach the line, while some others are also likely to wait even longer until the disc is closer to the line. Even if subjects were to still estimate the TTC but adjust their response to be earlier or later, because they think they are expected to, some subjects might make small adjustments (just a little earlier/later), while others are just as likely to make bigger adjustments (moderately or a lot earlier/later). It logically follows, then, that response times in the sound conditions would be highly variable and show greater variability relative to the no-sound condition. Instead, we found no significant differences in the variable errors between the

sound conditions and the no-sound condition. Third, TTC estimations in the no-sound condition were fairly accurate and showed two distinct systematic trends as a function of line distance that are consistent with other TTC studies—specifically, variable error increases and overall response times relative to the actual TTC decrease with increasing line distance. Both of these systematic trends are typical findings in TTC studies (e.g., Hancock and Manser 1997; Rosenblum 1993; Schiff and Oldak 1990; Tresilian 1995). The overall accuracy and these systematic trends consistent with the previous TTC studies clearly indicate that subjects understood the task and were complying with our instructions in the no-sound condition. Finally, although the TTC estimations in the sound conditions were biased, they were still relatively accurate, and, more importantly, show the same two systematic trends in variability and overall response times as the no-sound condition. Again, these trends are also consistent with the previously cited TTC studies. These results suggest to us that task performance in the sound conditions was still driven by estimating TTC, albeit biased by the sound stimuli. Although we cannot entirely discount the possibility that some subjects might have performed according to demand characteristics, if there were any, it appears that they had a minimal impact on the overall results.

To summarize, we have shown that estimating when a moving visual target arrives at a designated location can be influenced by auditory stimuli of ascending and descending pitch glides. When the visual target was paired with an ascending pitch glide, TTC was underestimated. Conversely, when the visual target was paired with a descending pitch glide, TTC was overestimated. These biases in the audiovisual conditions were relative to the baseline no-sound condition and, in general, relative to the actual TTC. Moreover, these biases were consistent in both upward and downward disc directions. These findings suggest that observers interpreted the pitch glides in accordance to a specific mapping strategy of pitch and speed whereby ascending pitch corresponded to faster speed or acceleration and descending pitch corresponded to slower speed or deceleration.

Compliance with ethical standards

References

Alderson GJK, Whiting HTA (1974) Prediction of linear motion. Hum Factors 16:495–502

Baures R, Maquestiaux F, DeLucia PR, Defer A, Prigent E (2018) Availability of attention affects time-to-contact estimation. Exp Brain Res 236:1971–1984

Bennett SJ, Baures R, Hecht H, Benguigui N (2010) Eye movements influence estimation of time-to-contact in prediction motion. Exp Brain Res 206:399–407

Bernstein IH, Edelstein BA (1971) Effects of some variations in auditory input upon visual choice reaction time. J Exp Psychol 87:241–247

Carnevale MJ, Harris LR (2016) Which direction is up for a high pitch? Multisensory Res 29:113–132

Chotsrisuparat C, Koning A, Jacobs R, van Lier R (2017) Auditory rhythms influence judged time to contact of an occluded moving object. Multisensory Res 30:717–738

Collier WG, Hubbard TL (2001) Judgments of happiness, brightness, speed, and tempo change of auditory stimuli varying in pitch and tempo. Psychomusicology 17:36–55

Cousineau D (2005) Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. Tutorials Quant Methods Psychol 1:42–45

DeLucia PR, Liddell GW (1998) Cognitive motion extrapolation and cognitive clocking in prediction motion tasks. J Exp Psychol Hum Percept Perform 24:901–914

DeLucia PR, Kaiser MK, Bush JM, Meyer LE, Sweet BT (2003) Information integration in judgements of time to contact. Q J Exp Psychol 56A:1165–1189

DeLucia PR, Preddy D, Oberfeld D (2016) Audiovisual integration of time-to-contact information for approaching objects. Multisensory Res 29:365–395

Einhorn HJ, Hogarth RM (1981) Behavioral decision theory: processes of judgment and choice. Annu Rev Psychol 32:53–88

Eitan Z, Granot RY (2006) How music moves: musical parameters and listeners' images of motion. Music Percept 23:221–247

Evans KK, Treisman A (2011) Natural cross-modal mappings between visual and auditory features. J Vis 10:6.1–6.12

Firestone C, Scholl BJ (2016) Cognition does not affect perception: evaluating the evidence for "top-down" effects. Behav Brain Sci 39:E229

Gallace A, Spence C (2006) Multisensory synesthetic interactions in the speeded classification of visual size. Percept Psychophys 68:1191–1203

Godøy RI (2001) Imagined action, excitation, and resonance. In: Godøy RI, Jørgensen H (eds) Musical imagery. Taylor & Francis, New York, pp 237–250

Gordon MS, Rosenblum LD (2005) Effects of intrastimulus modality change on audiovisual time-to-arrival judgments. Percept Psychophys 67:580–594

Gordon MS, Russo FA, MacDonald E (2013) Spectral information for detection of acoustic time of arrival. Atten Percept Psychophys 75:738–750

Gottsdanker RM (1955) A further study of prediction motion. Am J Psychol 68:432–437

Gray R (2011) Looming auditory collision warnings for driving. Hum Factors 53:63–74

Hancock PA, Manser MP (1997) Time-to-contact: more than tau alone. Ecol Psychol 9:265–297

Henry MJ, McAuley JD, Zaleha M (2009) Evaluation of an imputed pitch velocity model of auditory tau effect. Atten Percept Psychophys 71:1399–1413

Hochberg Y (1988) A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75:800–802

Hofbauer M, Wuerger SM, Meyer GF, Roehrbein F, Schill K, Zetzsche C (2004) Catching audiovisual mice: predicting the arrival time of auditory-visual motion signals. Cogn Affect Behav Neurosci 4:241–250

Hubbard TL (1995) Auditory representational momentum: surface form, direction, and velocity effects. Am J Psychol 108:255–274

Hubbard TL, Ruppel SE (2013) A Frohlich effect and representational gravity in memory for auditory pitch. J Exp Psychol Hum Percept Perform 39:1153–1164

Huber S, Krist H (2004) When is the ball going to hit the ground? Duration estimates, eye movements, and mental imagery of object motion. J Exp Psychol Hum Percept Perform 30:431–444

Jain A, Sally SL, Papathomas TV (2008) Audiovisual short-term influences and aftereffects in motion: examination across three sets of directional pairings. J Vision 8:1–13

Jamal Y, Lacey S, Nygaard L, Sathian K (2017) Interactions between auditory elevation, auditory pitch and visual elevation during multisensory perception. Multisensory Res 30:287–306

Keselman HJ (1994) Stepwise and simultaneous multiple comparison procedures of repeated-measures means. J Educ Stat 19:127–162

Keshavarz B, Campos JL, DeLucia PR, 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. Atten Percept Psychophys 79:929–944

Maeda F, Kanai R, Shimojo S (2004) Changing pitch induced visual motion illusion. Curr Biol 14:R990–R991

Marks LE (1987) On cross-modal similarity: auditory–visual interactions in speeded discrimination. J Exp Psychol Hum Percept Perform 13:384–394

Melara RD (1989) Similarity relations among synesthetic stimuli and their attributes. J Exp Psychol Hum Percept Perform 15:212–231

Miller WL, Maffei V, Bosco G, Iosa M, Zago M, Macaluso E, Lacquaniti F (2008) Vestibular nuclei and cerebellum put visual gravitational motion in context. J Neurophysiol 99:1969–1982

Morey RD (2008) Confidence intervals from normalized data: a correction to Cousineau (2005). Tutorials Quant Methods Psychol 4:61–64

Mossbridge JA, Grabowecky M, Suzuki S (2011) Changes in auditory frequency guide visual-spatial attention. Cognition 121:133–139

Payne JW, Bettman JR, Johnson EJ (1988) Adaptive strategy selection in decision making. J Exp Psychol Learn Mem Cogn 14:534–552

Peterken C, Brown B, Bowman K (1991) Predicting the future position of a moving target. Perception 20:5–16

Prime SL, Harris LR (2010) Predicting the position of moving audio-visual stimuli. Exp Brain Res 203:249–260

Rosenblum LD (1993) Acoustical information for controlled collisions. In: Schick A (ed) Contributions to psychological acoustics. Bibliotheks, Oldenburg, pp 303–322

Rosenblum LD, Wuestefeld AP, Saldana HM (1993) Auditory looming perception: influences on anticipatory judgments. Perception 22:1467–1482

Sadaghiani S, Maier JX, Noppeney U (2009) Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. J Neurosci 29:6490–6499

Schiff W, Oldak R (1990) Accuracy of judging time to arrival: effects of modality, trajectory, and gender. J Exp Psychol Hum Percept Perform 16:303–316

Spence C (2011) Crossmodal correspondences: a tutorial review. Atten Percept Psychophys 73:971–995

Tresilian JR (1995) Perceptual and cognitive processes in time-to-contact estimation: analysis of prediction-motion and relative judgment tasks. Percept Psychophys 57:231–245

Trimble OC (1934) Localization of sound in the anterior posterior and vertical dimensions of auditory space. Br J Psychol 24:320–334

Tversky A, Sattath S (1979) Preference trees. Psychol Rev 86:542–573

Walker AR (1981) The presence of internalized images of musical sounds and their relevance to music education. Bull Counc Res Music Educ 66(67):107–111

Walker AR (1985) Mental imagery and musical concepts: some evidence from the congenitally blind. Bull Counc Res Music Educ 85:229–238

Walker R (1987) The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. Percept Psychophys 42:491–502

Wuerger S, Meyer G, Hofbauer M, Zetzsche C, Schill K (2010) Motion extrapolation of auditory-visual targets. Inf Fusion 11:45–50

Zheng R, Maraj BKV (2018) The effect of concurrent hand movement on estimated time to contact in a prediction motion task. Exp Brain Res 236:1953–1962

Zhou L, Yan J, Liu Q, Li H, Xie C, Wang Y, Campos JL, Sun H-J (2007) Visual and auditory information specifying an impending collision of an approaching object. In: Jacko JA (ed) Human–computer interaction, Pt 2, vol 4551. Lecture notes in computer science. Springer, Berlin, pp 720–729