Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-024-02985-5

The role of attention in eliciting a musically induced visual motion aftereffect

Hannah Cormier[1] · **Christine D. Tsang**[1,2] · **Stephen C. Van Hedger**[1,2,3]

Accepted: 23 October 2024

Abstract
Previous studies have reported visual motion aftereffects (MAEs) following prolonged exposure to auditory stimuli depicting motion, such as ascending or descending musical scales. The role of attention in modulating these cross-modal MAEs, however, remains unclear. The present study manipulated the level of attention directed to musical scales depicting motion and assessed subsequent changes in MAE strength. In Experiment 1, participants either responded to an occasional secondary auditory stimulus presented concurrently with the musical scales (diverted-attention condition) or focused on the scales (control condition). In Experiment 2 we increased the attentional load of the task by having participants perform an auditory 1-back task in one ear, while the musical scales were played in the other. Visual motion perception in both experiments was assessed via random dot kinematograms (RDKs) varying in motion coherence. Results from Experiment 1 replicated prior work, in that extended listening to ascending scales resulted in a greater likelihood of judging RDK motion as descending, in line with the MAE. In contrast, the MAE was eliminated in Experiment 2. These results were internally replicated using an in-lab, within-participant design (Experiment 3). These results suggest that attention is necessary in eliciting an auditory-induced visual MAE.

Keywords Attention · Perception · Music · Motion · Vision · Cross-modal perception

Introduction

Musical understanding is indelibly shaped by language. It is difficult to imagine describing music without using directional terms (e.g., "the *rising* melody" or "the *descending* bassline"). Yet, this *pitch-verticality* association is by no means universal. There is nothing about increasing the number of oscillations of a soundwave (i.e., increasing frequency) that inherently maps onto perceived location in vertical space. To further illustrate this point, a speaker of Farsi would likely describe the same musical event in terms of thickness (e.g., "the *thinning* melody" or "the *thickening* bassline"), not verticality. Still other languages would describe pitch changes in terms of size, brightness, or even age (Fernández-Prieto et al., 2015; Marks, 1974; Zbikowski, 2008), and listeners who speak a language that describes pitch in terms of verticality (such as English) can readily understand these various pitch mappings (Eitan & Timmers, 2010).

Based on the observed flexibility of describing pitch changes along different dimensions (Eitan & Timmers, 2010), one might reason that these linguistic labels have no direct consequence on listeners' perceptions of pitch. However, many studies have demonstrated that well-developed, cultural metaphors for understanding pitch cannot be reversed and influence perceptual judgments (Melara & Marks, 1990; Melara & O'Brien, 1987), even under conditions of linguistic interference (Dolscheid et al., 2013). In further support of the idea that conceptual metaphors for understanding pitch exist independently of language, studies have reported the presence of pitch-verticality mappings in prelinguistic infants, where preferential looking was observed towards stimuli corresponding to both height-pitch associations and thickness-pitch associations

✉ Stephen C. Van Hedger
svanhedg@uwo.ca

1 Department of Psychology, Huron University College at Western: London, 1349 Western Road, London, ON N6G 1H3, Canada

2 Department of Psychology, Western University: London, London, ON, Canada

3 Western Institute for Neuroscience, Western University: London, London, ON, Canada

(Dolscheid et al., 2014). These findings imply that language might strengthen some pre-existing understandings of pitch changes, rather than creating entirely novel, arbitrary ones.

Due to the strength of the pitch-verticality metaphor in Western cultures, it is plausible that perceived motion in pitch (e.g., through ascending or descending musical scales) has potential consequences for visual motion perception (e.g., judging whether a visual object is moving up or down). In support of this idea, Maeda et al. (2004) showed that ascending and descending pitch sweeps can influence the perception of ambiguous visual motion in a congruent manner. For example, listening to ascending tone sweeps made participants more likely to judge simultaneously presented ambiguous visual motion as ascending. These findings suggest that auditory stimuli might influence vision at the perceptual level, independent of factors such as eye movements and cueing. However, the demand characteristics in this kind of paradigm are potentially high, given that ascending pitches are associated with judging ambiguous visual motion as ascending. As such, it is unclear whether these findings are driven by perceptual mechanisms or post-perceptual decisions. To more strictly test whether pitch change influences the perception of visual motion at a perceptual level, researchers can take advantage of a non-intuitive perceptual illusion – the visual motion aftereffect (MAE).

MAEs are perceptual illusions resulting from prolonged exposure to a continuous stimulus with unidirectional movement, in which static stimuli observed immediately following motion adaptation are perceived as moving in the opposite direction. For example, sustained fixation on a visual stimulus with continuous, unidirectional leftward motion would evoke a temporary rightward motion aftereffect (i.e., a perceived motion illusion in the opposite direction of the adapted stimulus). This effect has been documented after adaptation to visual stimuli conveying both real motion (e.g., Anstis et al., 1998), implied motion from static images (Winawer et al., 2010), and even linguistic descriptions of motion (Dils & Boroditsky, 2010). The ability of stimuli such as static images implying motion and motion language to elicit MAE-like effects suggests that MAEs might be elicited by a broader set of stimuli that depict motion more conceptually or abstractly.

Previous work has found that auditory stimuli with strong vertical directionality (ascending and descending musical scales) can elicit responses consistent with visual MAEs (Hedger et al., 2013). In this work, listening to several seconds of descending and ascending musical scales made participants more likely to judge visual motion, presented via a random dot kinematogram (RDK; e.g., Newsome & Pare, 1988), as ascending or descending, respectively. These results suggest that the pitch-verticality metaphor has a perceptual basis and can elicit similar perceptual judgments as adapting to real ascending or descending visual motion. However, in this study, participants were simply instructed to listen to the musical scales. As such, the role of attention in inducing this cross-modal MAE is presently unclear.

There are several reasons to predict that an auditorily induced visual MAE might be modulated by attention. Adaptation occurs at multiple levels of the visual system (Webster, 2015), opening a role for the influence of attention at multiple processing levels. Additionally, previous work has found that attention enhances stimulus representations and low levels of the perceptual system (Carrasco et al., 2004; Yeshurun & Carrasco, 1998), which in the present context could potentially make the conveyed motion by ascending or descending musical scales more salient, thereby enhancing the cross-modal MAE. However, empirical work examining the role of attention in motion adaptation more specifically is mixed. One reason for these mixed findings may be that attention does not have a static effect in modulating perceptual processing over time. Specifically, Ling and Carrasco (2006) found that attention enhanced perceptual processing in early temporal windows, but actually inhibited perceptual sensitivity over sustained periods. In an experiment conducted by Morgan and Solomon (2019), attentional distraction had no significant effect on motion adaptation strength, as measured by its duration and asymptote. As motion adaptation is thought to underlie MAEs, a similar result may be inferred in relation to MAEs. Based on this finding, it may be hypothesized that motion adaptation occurs at a preattentive stage of visual processing. Conflicting research findings suggest that attention magnifies the adaptation effects; this attention-adaptation relationship has been referred to as "adaptation gain" (Rezec et al., 2004). Bartlett et al. (2018) identified and controlled potential experimental factors that may have resulted in experimental inconsistencies in relation to the attention-adaptation relationship, but it is clear that the role of attention in MAEs – in particular, cross-modal MAEs – warrants further investigation.

The present study therefore aimed to conceptually replicate the cross-modal MAE reported by Hedger et al. (2013) while additionally exploring the role of attention in inducing this cross-modal MAE. At present, it is unclear whether attention modulates MAEs that are more conceptual in nature. In particular, the audiovisual nature of the MAE outlined by Hedger et al. (2013) might particularly rely on a participant attending to the perceived direction of the sound, as attention may be necessary to for successful integration of information across auditory and visual modalities (Spence & Frings, 2020). In support of this idea, prior research has reported that demands placed on attention can disrupt audiovisual processing in domains such as speech (Alsius et al., 2005). On the other hand, research has found that attention has little-to-no influence on other domains of multisensory integration, such as the ventriloquist effect (Bertelson et al., 2000; Vroomen et al., 2001). These seemingly contradictory findings have been explained in terms of different task demands – for example, considering whether the specific paradigm requires perceptual decisions in both modalities, compared to paradigms in which one modalities can be entirely ignored (Donohue et al., 2015). This suggests that the specific role of attention in audiovisual processing is complex and likely task dependent.

In the current experiments, we present participants with adapting stimuli in the auditory modality (continually ascending or descending musical scales) and measure responses to visual motion stimuli (RDKs) varying in motion coherence. Critically, we manipulate participants' attention to these musical scales to assess the role of attention in eliciting a cross-modal MAE. In Experiment 1, participants either are instructed to focus on the scales (control condition) or on a secondary auditory task, presented occasionally alongside the scales (diverted-attention condition). In Experiment 2, all participants engaged in a more continuously demanding secondary auditory task, providing a stronger attentional manipulation compared to Experiment 1. In Experiment 3, we manipulate attention using a within-participant design, using a controlled in-lab setting. We hypothesized that participants who focused attention on the ascending and descending auditory pitches would experience significantly stronger MAEs than those who are directed to attend to a secondary stimulus presented alongside the ascending and descending auditory pitches (i.e., diverted-attention condition in Experiment 1, all participants in Experiment 2, and trials in which participants divert attention in Experiment 3). We additionally predicted a graded effect of attention, in that the less demanding attention manipulation in Experiment 1 would not attenuate the MAE to the same degree as the more continuously demanding attention manipulations in Experiments 2 and 3.

Experiment 1

Method

Participants

We recruited 100 participants (control: $n = 50$; diverted-attention: $n = 50$) from Amazon Mechanical Turk via the Cloud Research recruitment platform (Litman et al., 2017). Participants had to reside in the USA, have a minimum 90% approval ratings from prior Mechanical Turk tasks, and had to have passed internal attention checks administered by Cloud Research to be eligible to participate in the experiment. Participants of all genders were included. Additional recruitment criteria were developed based on potential factors that may limit one's attentional or perceptual (i.e., auditory or visual) capabilities. Participants had to be between 18 and 60 years old ($M = 37.92$ years old, $SD = 9.68$ years old, range: 20–59 years old), and additionally had to have normal (or corrected-to-normal) vision and hearing. The study excluded 16 participants on the basis of a failure to adequately perform practice or attentional components of the task (see *Data exclusion* for more details), leaving 84 participants in the primary analyses (control: $n = 46$; diverted-attention: $n = 38$). Out of the 84 assessed participants, 39 (46.4%) reported prior musical training.

Materials

A letter of information was provided to potential participants with details about the current experiment, such as confidentiality procedures, inclusion and exclusion criteria, contact information of the researchers, descriptions of the tasks involved in the experiment, and potential benefits, costs, and risks involved in participation. The letter of information did not mention motion perception, nor did it discuss MAEs, in an effort to mitigate demand characteristics.

The experiment was programmed in jsPsych 6 (de Leeuw, 2015). The random dot kinematogram (RDK) stimuli were generated within jsPsych using a customizable plugin (Rajananda et al., 2018). Each RDK displayed 200 dots. Dots on each frame were either designated as coherent (i.e., moving in a consistent up or down direction) or incoherent (i.e., disappearing and reappearing at random positions within the 500-pixel-wide square aperture). Although there are several ways to specify motion within RDKs, the designation of coherent and incoherent dots was assigned randomly on each frame and weighted based on the coherence level for the present experiment. This manipulation is conceptually similar to the approach taken by Hedger et al. (2013). In the RDK Practice Task, dot coherence levels were high (90%, 70%, and 50% ascending/descending motion). In the Main Task, dot coherence levels were considerably more ambiguous (30% and 15% ascending/descending motion, as well as 0% coherence). The inclusion of some RDKs with no coherent motion (0% coherence) permitted the assessment of how musical scales influenced truly ambiguous RDKs with no genuine motion signal. Each RDK stimulus was 1,000 ms in duration.

The musical scales used Shepard tones to elicit the percept of continuously ascending or descending auditory motion (Shepard, 1964). Shepard tones are complex tones constructed via frequencies that are octave relations (i.e., a 2:1 frequency ratio) of one another. Given that tones separated by octaves belong to the same pitch class (e.g., 440 Hz and 880 Hz would both be labeled as the note "A"), each tone by itself has a clear pitch chroma (e.g., A or C#) but is ambiguous with respect to pitch height (e.g., the adjacent octave above or below middle C on a piano). When Shepard tones are used to play musical scales, which typically contain small adjacent changes in auditory frequency, listeners often report a perceptual illusion of a continually rising or falling auditory sequence, similar to the *barber pole illusion* in vision (Wallach, 1935). In this sense, Shepard tones are ideal stimuli for the present experiment, as they strongly evoke ascending or descending auditory motion yet control for several factors that would typically differ for ascending and descending scales (e.g., starting and ending frequencies).

Each Shepard tone was 166.67 ms in duration and contained energy at five octaves – two octaves above and below a specified fundamental frequency. Three octaves of Shepard tones were stacked and arranged to create a chromatic scale with repeating notes, as informal pilot testing suggested that this construction resulted in the most consistent perceptions of continual ascending or descending motion. The Shepard tones were generated in Matlab (Mathworks: Natick, MA). Given that both the ascending and descending chromatic scales contained 24 notes, each scale was 4000 ms in duration. Details regarding the Shepard tones and chromatic scale construction are provided in Fig. 1.

The "noise bursts," which were the focus of attention for participants in the diverted-attention condition, were 50-ms samples of pink noise. These were embedded in the Shepard scales (-10 dB SNR) to make the attentional task more difficult and to also prevent distracting participants in the control condition, given that all participants heard the same sounds. There was a 50% probability of hearing a noise burst embedded within the Shepard scale on each trial in the Main Task. Participants accessed the experiment with their own computer device. Data processing, analyses, and visualization was done in R/RStudio.

Procedure

The experiment was completed in two "runs" separated by condition. The diverted-attention condition was run first, and the control condition was run second. Both conditions were run within one week of each other, and participants who had completed the diverted-attention version of the experiment were not eligible to participate in the control condition. This approach was chosen as opposed to randomly assigning participants to a condition to ensure that both conditions had equal sample sizes.

Participants were first presented with a letter of information, which detailed the terms of the study. Those who decided to participate after reading the letter of information clicked on a checkbox affirming their consent to participate. Participants could not continue the study without checking the consent box on the computer screen. Participants then completed several preliminary assessments, which were implemented to ensure adequate auditory calibration and to familiarize all participants with the RDK stimuli. First, participants heard a 30-second calibration

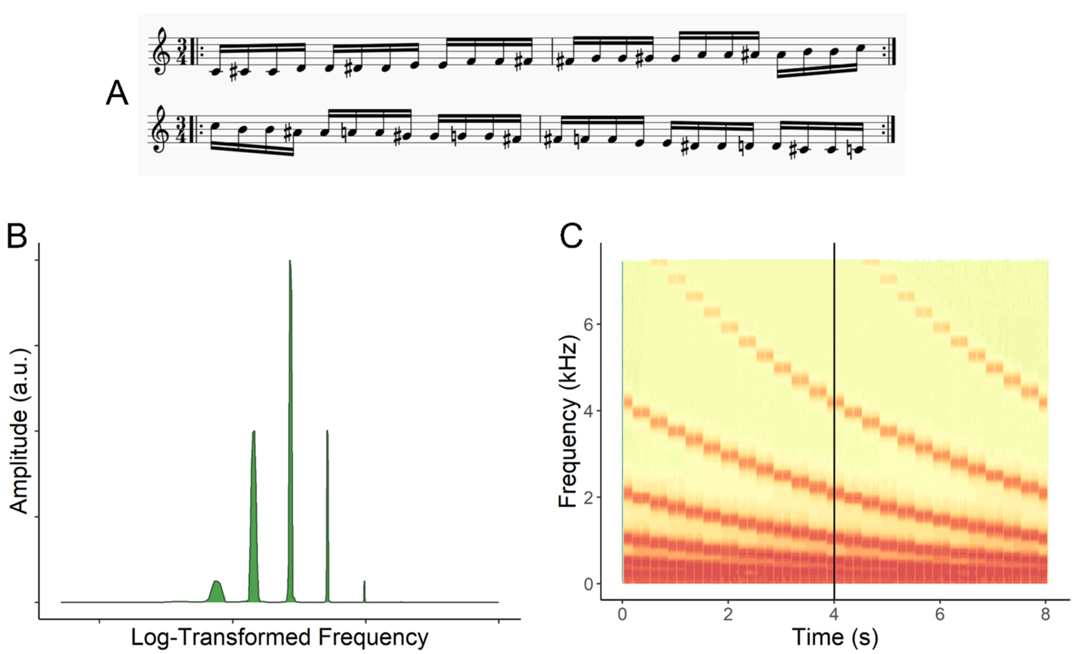

Fig. 1 Depiction of the scales used in the experiment. **Panel A** provides musical notation for both the ascending (top row) and descending (bottom row) chromatic scales. **Panel B** provides the harmonic spectrum of a sample Shepard tone used in the experiment, with each adjacent peak being separated by one octave. Panel C provides a spectrogram of two consecutive descending scales. The vertical line at 4 s represents the point at which the scale repeats

noise, root-mean-square normalized to the same level as the Shepard scales, and adjusted their computer volume to a comfortable listening level. Next, participants engaged in a simple loudness judgment task, which is meant to differentiate participants using versus not using headphones (Woods et al., 2017). Following this auditory calibration, participants in the diverted-attention condition engaged in a "noise burst" practice task, which familiarized participants with the secondary task they would be completing during the presentation of the Shepard scales. This practice task featured short (50-ms) bursts of pink noise. These noise bursts were presented in either the left or right audio channel and participants had to determine whether the bursts were coming from the left or right by pressing designated keys on the keyboard. Feedback was given after each response, for a total of 10 trials. The RDK Practice Task was completed by participants in both conditions to ensure RDK familiarity. The RDK Practice Task consisted of 12 trials and six kinds of RDKs, which varied on motion coherence (i.e., 90%, 70%, 50%) and motion direction (i.e., ascending, descending). Participants judged whether the dots were moving primarily upwards or downwards by pressing the up and down arrows on the keyboard, respectively, and feedback was provided after each response.

After these calibration and practice assessments, participants were then presented with the instructions for the Main Task. In the diverted-attention condition, participants were told to respond with a key press whenever they heard a short noise burst, similar to those heard during the practice assessment. Participants in the diverted-attention condition were alerted to the fact that these noise bursts would be embedded within musical scales but were told to ignore the musical sounds and focus on the noise bursts. Participants in the control condition, in contrast, were told to listen carefully to the musical scales, as the stopping of the scales was a cue that the RDK was about to be presented; the noise bursts were not mentioned. Thus, participants in both conditions were played the same auditory stimuli; the only difference was how participants were instructed to attend to the sounds.

The Main Task consisted of 100 trials (four blocks; 25 trials per block). All trials in the Main Task followed an identical procedure: (1) Shepard scale, (2) RDK presentation, and (3) a forced-choice judgment about the motion of the RDK. Participants in the diverted-attention condition responded to noise bursts that were intermittently played during the Shepard scales (on average, 50% of Shepard scales contained a noise burst). The first trial in each block consisted of a 24-s Shepard scale, followed by an RDK, with subsequent trials in each block (i.e., trials 2–25) containing an 8-s Shepard scale (i.e., trials 2-25). RDK motion coherence and direction were manipulated such that there were five RDK conditions: (1) 30% coherence of descending motion, (2) 15% coherence of descending motion, (3) 0% coherence, (4) 15% coherence of ascending motion, and (5) 30% coherence of ascending motion. These five RDK types were randomly presented five times each within a block. Direction of musical scales was fixed within a block and was interleaved across blocks; participants randomly received one of two orderings (descending-ascending-descending-ascending or ascending-descending-ascending-descending). Figure 2 provides an overview of the experimental paradigm.

Following the Main Task, participants completed a brief questionnaire. The questionnaire recorded sociodemographic information such as age, gender, and highest level of education achieved. Participants' first language and up to two additional languages (including self-rated proficiency) were also collected, along with status of hearing aid use (yes/no) and musical background. Musical experience was assessed in terms of musical training (yes/no; number of years if yes) and self-reported musical skill on a 6-point Likert scale (1 = *not skilled*, 6 = *highly skilled*). Participants were then debriefed and paid for their participation.

Data exclusion

There were two considerations for removing participants from the main analyses. The first consideration was poor performance on the RDK Practice Task. Given that the tested coherence levels in the RDK Practice Task were higher than those used in the Main Task, poor performance on the RDK Practice Task would indicate that participants were unable to reliably perceive motion for *all* of the RDK stimuli in the Main Task. If participants were thus unable to achieve at least 75% accuracy (9 of 12 correct) on the RDK Practice Task, they were removed from further analyses. This consideration removed 15 participants (control: $n = 4$, diverted-attention: $n = 11$).

The second data culling measure only applied to participants in the diverted-attention condition, as it was concerned with "noise burst" detection in the Main Task. Poor performance on this task was taken as evidence that participants did not successfully divert attention to detect the noise bursts, meaning these participants did not comply with the instructions. Performance was assessed in terms of d-prime. A *hit* was defined as a key press between 200 ms and 2,000 ms after a noise. A key press that fell outside of this window was coded as a *false alarm*. The mean performance on the noise burst detection task was quite high ($M = 3.16$, $SD = 1.40$). Of the participants who passed the RDK Practice Task, one participant had a d-prime value below zero (-1.17), suggesting an inability to detect the noise bursts (i.e., logging more false alarms than hits). This participant was thus excluded from subsequent analyses. Thus, the final participant count was 84 (control: $n = 46$, diverted-attention: $n = 38$).

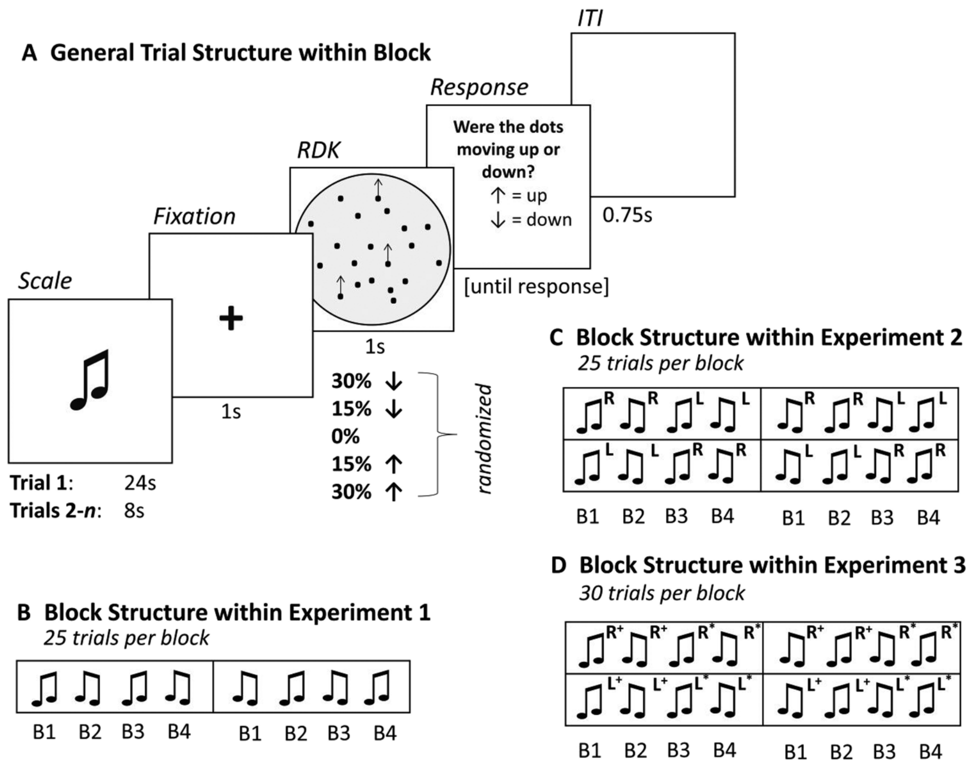

Fig. 2 Overview of experimental paradigm. **Panel A** depicts the trial structure within each block. On each trial, participants heard either a descending or ascending musical scale for either 24 s (Trial 1) or 8 s (Trials 2–25 in Experiments 1 and 2 and Trials 2–30 in Experiment 3). This was followed by a 1-s fixation, which was then followed by a 1-s random dot kinematogram (RDK). There were five coherence levels for the RDKs, and the ordering of RDK coherence levels within a block was randomized. Following the RDK, participants provided a forced choice (up or down). Panels B, C, and D depict the block structure within Experiment 1, 2, and 3 (respectively) for the four blocks of the experiment (B1–B4). In Experiment 1, there were two possible block orderings, relating to whether participants would listen to ascending or descending scales first. In Experiments 2 and 3, there were four possible block orderings, relating to both scale direction and audio channel of the scales (left or right; represented by the superscript L or R). In Experiment 3, participants always attended away from the ear containing the musical scales for B1 and B2 (denoted with +), and always attended to the ear containing the musical scales for B3 and B4 (denoted with *)

The results of the headphone test were not used as formal exclusion criteria, but rather to get a sense of how many participants followed the researcher's recommendation of wearing headphones. Based on a threshold of at least five of six correct responses, taken from Woods et al. (2017), 76% of the participants passed the assessment.

Data analysis

We constructed generalized linear mixed-effects models with probit links using the "lme4" package in R (Bates et al., 2015). The dependent variable was whether a participant responded that the RDK was ascending or descending (arbitrarily coded as a 1 or 0, respectively). Each model included the interaction of condition (diverted-attention, control), scale direction (ascending, descending) and RDK coherence (-2, -1, 0, 1, and 2, with descending motion arbitrarily coded as negative and coherence levels recoded to facilitate model convergence). Participants were modeled with random intercepts. To assess the relative importance of scale direction and condition, we created two nested models from the full model described above - one without scale direction and one without condition. These models were then compared to the full model on the basis of corrected AIC values (Akaike, 1998; Burnham & Anderson, 2004), with negative values of this difference score (ΔAIC) indicating a better fit. The RDK term was initially modeled with a polynomial fit, allowing an examination of linear, quadratic, and cubic fits across RDK coherence. However, the linear fit of the RDK coherence provided the best fit of the data, with the additional higher-order terms resulting in a non-converging model and not providing a better fit ($\Delta AIC = 2.6$). As such, all reported models use linear fits of RDK coherence. We additionally used the "MixedPsy" package in R (Balestrucci et al., 2022) to calculate the point of subjective equality (PSE) from the mixed-effects models. The PSE in the present context refers to the coherence level at which participants are equally likely to judge an RDK as ascending or descending. For ease of interpretation, the PSE values are reported as original coherence values (i.e., percent coherence) as opposed to the re-coded values (-2 to 2) used in the mixed-effects models.

To assess how variables from the questionnaire (specifically age, musical training, self-reported musical skill, and reported bilingualism) related to the present findings, we used the extracted participant slopes for scale direction from the full mixed-effects model described in the previous paragraph. We then assessed via Pearson correlations how this slope was associated with these measures, considered in terms of all participants, as well as split by condition. These correlations thus provide a sense of whether the relative magnitude of the scale direction effect was associated with these measured variables, as well as whether any observed associations differed as a function of condition. These analyses are considered exploratory.

Results

Participants were more likely to report that an RDK was ascending as a function of RDK coherence, which demonstrates that participants could correctly judge the motion in the RDK stimuli given the coding of descending RDKs as negative and ascending RDKs as positive (B = 0.54, SE = 0.08, $p < .001$). Participants were also more likely to report that an RDK was *descending* for trials in the ascending scale blocks relative to trials in the descending scale blocks, as evidenced by a significant main effect of scale direction (B = -0.27, SE = 0.07, $p < .001$). This effect, which is plotted in Fig. 3, conceptually replicates Hedger et al. (2013). No other term was significant in the model, including the main effect, two-way, and three-way interactions including condition. Removing scale direction from the model resulted in a significantly worse fit relative to the full model (ΔAIC = 20.2). In contrast, removing condition from the model resulted in a *better* fitting model (ΔAIC = -6.2), which is perhaps not surprising considering (1) the non-significant effects of condition in the full model, and (2) the fact that AIC penalizes for extra fitted parameters.

The PSE analyses suggested that scale direction influenced the coherence level at which participants were equally likely to judge an RDK as ascending or descending. When listening to descending musical scales, participants' mean PSE was +3.11% (SE = 2.37%). The 95% confidence interval included zero [-0.20, 7.74], suggesting that participants listening to descending musical scales were not independently shifted from the objective null point (i.e., 0% coherence) in terms of their PSE. However, when participants listened to ascending musical scales, their mean PSE was 11.62% (SE = 3.28%). Unlike the descending musical scale condition, the 95% confidence interval did not include zero [5.17, 18.06]. Thus, the PSE analyses suggest that the effect of scale direction in the mixed-effect models is asymmetrical, with ascending musical scales making participants less sensitive to ascending RDK motion. In contrast, descending

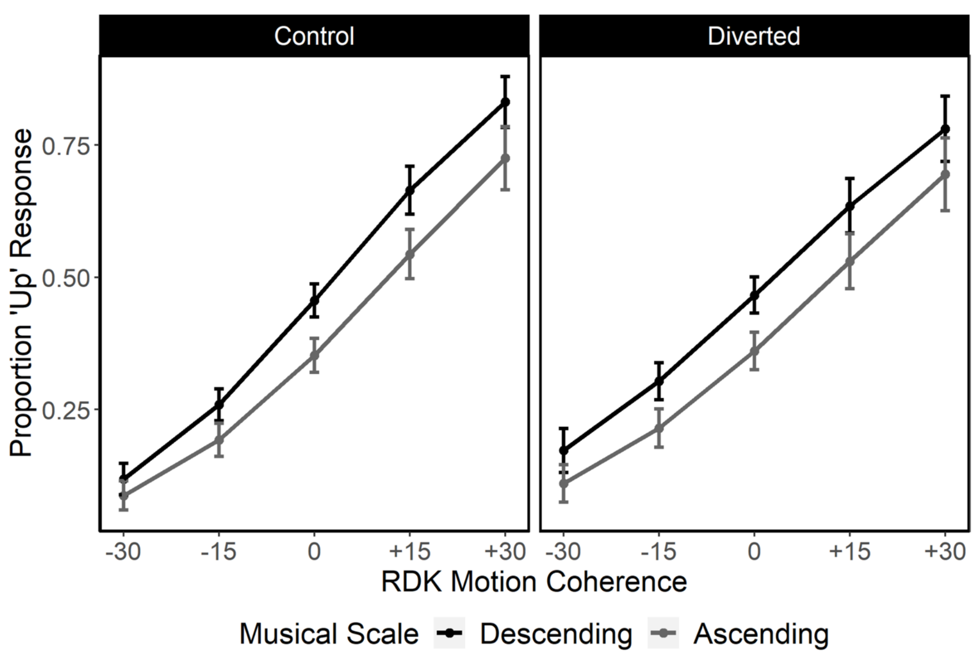

Fig. 3 Influence of scale direction on visual motion judgments from Experiment 1. Error bars represent ± 1 standard error of the mean. Random dot kinematogram (RDK) motion coherence is plotted on the x-axis, with descending motion arbitrarily coded as negative. Control condition participants are plotted in the left panel, and diverted-attention condition participants are plotted in the right panel

musical scales did not significantly shift participants' points of subjective equality below 0% coherence.

The exploratory correlational analyses did not show any associations between the strength of the observed MAE (represented as the extracted participant slope for scale direction) and the analyzed variables (age, musical training, self-reported musical skill, and bilingualism). Table 1 provides a summary of these analyses. In fact, the only significant association observed was between musical training and self-reported musical skill. Thus, the relative degree to which participants were influenced by the musical scales appeared to be independent of age and self-reported music and language factors.

Discussion

In Experiment 1, participants listened to identical auditory stimuli but were instructed to attend to different features of the sound. Participants in the control condition were told to focus on the scales, whereas participants in the diverted-attention condition were given a secondary task – detecting and responding to noise bursts – and were told to focus on responding to these bursts while ignoring the scales. Although participants were overall less likely to report that visual motion was descending following prolonged experience listening to ascending scales, conceptually replicating the cross-modal MAE reported by Hedger et al. (2013), this effect was not modulated by attention in the present experiment. Thus, the present results suggest that the musical scales used in the current study were sufficiently strong to engender motion percepts, that these motion percepts influenced visual motion perception in line with MAEs, and that attention might not modulate the strength of this cross-modal MAE.

One unexpected finding from Experiment 1 was the asymmetry between ascending and descending musical scales in modulating visual motion judgments. Specifically, ascending musical scales significantly shifted participants' PSEs from zero, whereas descending musical scales did not independently shift participants' PSEs from zero. This asymmetry is particularly surprising as previous work has suggested that the descending musical scales' elicitation of (descending) vertical motion is *stronger* than the elicitation of (ascending) vertical motion from ascending musical scales (Eitan & Granot, 2006). One possible explanation is that participants had a general descending bias in judging the motion within the RDKs. Although such a downward bias may be grounded by a *gravity prior* (cf. Jörges & López-Moliner, 2017), to our knowledge, a descending motion bias (relative to ascending motion) is not well documented in the literature, particularly for simple motion stimuli like RDKs. This asymmetrical effect would therefore benefit from future investigation.

Finally, there are several issues to consider before endorsing claims that attention does not influence the strength of the MAE. Most notably, Experiment 1 emphasized keeping the auditory environments identical across conditions, likely at the expense of attentional difficulty. Noise bursts were chosen as it was expected that these relatively quiet and short bursts of sound would go unnoticed in the control condition. Yet, this approach may have ultimately resulted in similar attentional demands regardless of condition. As participants in the diverted-attention condition were specifically instructed to attend to the concurrent auditory stimuli (i.e., the noise bursts) without knowing the frequency of their presentation, this may have caused an attentional focus to *all* auditory stimuli, rather than the intended manipulation of selective attention to the noise bursts at the expense of the scales. Moreover, the control condition may have initially directed their attention to the scales, and implicitly diverted their attention to the brief noise bursts due to their novel, sudden, and unpredictable nature (cf. alerting attention; Fan et al., 2002). In sum, it is possible that both conditions may have had similar attentional demands in terms of focusing on the scales, albeit for different reasons.

As such, Experiment 2 was designed to provide a stronger test of whether attention is necessary in eliciting a cross-modal MAE. By introducing a continuously demanding secondary auditory task (a 1-back task using spoken letters) – presented in the opposite ear of the musical scales and played louder relative to the scales – we predicted that participants would show an attenuated or entirely eliminated

Table 1 Exploratory correlational analysis of scale direction effect and survey measures

	Scale direction	Age	Musician	Musical skill	Bilingual
Scale Direction	–				
Age	.03 / -.10 / .19	–			
Musician	.06 / -.07 / -.05	.07 / .01 / .18	–		
Musical Skill	.07 / .00 / .17	-.12 / -.19 / .01	.44*** / .33* / .56***	–	
Bilingual	.04 / .07 / .00	.09 / .03 / .16	.02 / -.03 / .08	.03 / .03 / .02	–

Note: The first value in each cell represents the overall correlation across all participants, the second value represents the correlation for the control condition participants, and the third value represents the correlation for the divided-attention participants. * $p < .05$ *** $p < .001$

MAE, even if they had explicit awareness of the musical scales in the unattended ear.

Experiment 2

Method

Participants

We recruited 50 participants (M_{age} = 38.22 years, SD = 9.42 years, range: 23–59 years) from Amazon Mechanical Turk via the Cloud Research recruitment platform (Litman et al., 2017). The recruitment criteria were identical to Experiment 1. The study excluded 11 participants on the basis of a failure to adequately perform practice and attentional components of the task (see *Data exclusion* for more details), leaving 39 participants in the primary analyses. Of the 39 participants, 19 (48.7%) reported prior musical training. Participants who completed Experiment 1 were ineligible to participate in Experiment 2.

Fewer participants were recruited compared to Experiment 1 because Experiment 2 did not contain a between-participant condition. Thus, Experiment 2 was essentially treated as an additional between-participant attentional condition, for which we recruited n = 50 per condition in Experiment 1. It should also be noted that a bootstrapped power analysis (Supplemental Material) suggested that the achieved analyzable sample size of 39 could detect cross-modal MAEs on 98.8% of simulations based on the data from Experiment 1.

Materials

The experiment was programmed in jsPsych 6 (de Leeuw, 2015). The letter of information, RDK stimuli, and Shepard scales were all identical to Experiment 1. Unlike Experiment 1, the Shepard scales were combined with strings of spoken letters. The spoken letters had an interstimulus interval of 667 ms, meaning exactly 36 letters were presented for the longer, initial trial of each block and exactly 12 letters were presented for the subsequent trials in each block. One-third of the letters were *targets* (i.e., a repeat of the previous letter), and the remaining two-thirds of the letters were *non-targets* (i.e., a different letter compared to the previous letter). The spoken letters were adapted from prior n-back assessments using spoken letters (Jaeggi et al., 2010). The letters and the Shepard scales were combined dichotically with the letters being presented at a more favorable amplitude (+15 dB). We created two versions of each stimulus – one in which the letters were presented in the right channel, and one in which the letters were presented in the left channel. Similar to Experiment 1, participants accessed the experiment with their own computer device. Data processing, analyses, and visualization was done in R/RStudio.

Procedure

Participants underwent the same consent and auditory calibration procedure as reported in Experiment 1. Following the auditory calibration, participants completed two practice tasks to familiarize themselves with the components of the Main Task. The first practice task introduced the letter-judgment component of the Main Task. In this practice task, participants were told that they would be listening to a rapid stream of spoken letters and should press a designated key (spacebar) as quickly as possible whenever they heard a repeated letter. There were five practice trials, with each trial containing 12 spoken letters (four targets and eight non-targets). Participants received feedback about their performance, represented in terms of percentage correct, after each trial. The second practice task (the RDK Practice Task) was identical to the one described in Experiment 1.

After these calibration and practice assessments, participants were then presented with the instructions for the Main Task. Participants were instructed that they would hear streams of spoken letters, similar to the letter-practice task, and should press the spacebar as quickly as possible whenever the current letter matched the previous letter. Participants were alerted to the fact that the letters would be only presented to one ear and were told to focus solely on the letters.

Similar to Experiment 1, the Main Task consisted of 100 trials (four blocks; 25 trials per block). All trials in the Main Task followed an identical procedure: (1) Shepard scale presented dichotically with spoken letters, (2) RDK presentation, and (3) a forced-choice judgment about the motion of the RDK. The first trial in each block consisted of a 24-s Shepard scale (and 36 spoken letters), with subsequent trials in each block (i.e., trials 2–25) containing an 8-s Shepard scale (and 12 spoken letters). RDK motion coherence and direction were identical to Experiment 1. In addition to interleaving musical scale direction across block (either ascending-descending-ascending-descending or descending-ascending-descending-ascending, between-participant), the channel in which the letters were presented was also counterbalanced across participants (left-left-right-right or right-right-left-left), as depicted in Fig. 2B. Following the Main Task, participants completed an identical questionnaire as the one described in Experiment 1, with the exception of a single additional question. This additional question asked participants to describe what they heard in the study, to assess the extent to which participants reported noticing the Shepard scales in the unattended ear.

Data exclusion

There were two considerations for removing participants from the main analyses. The first consideration was poor performance on the RDK Practice Task, using the same 75% accuracy threshold as Experiment 1. This consideration removed ten participants. The second data culling measure was related to performance on the letter-judgment task in the Main Task. Poor performance on this task was taken as evidence that participants did not successfully divert attention to the letter-judgment task, essentially meaning these participants did not comply with the instructions. Performance was assessed in terms of d-prime. A "hit" was defined as a key press between 200 ms and 1,200 ms after the onset of the repeated letter. The upper limit of the response window was lowered relative to Experiment 1 given the continuous nature of the task, and was selected based on prior work using response times in the context of rapid and continuous presentations of auditory tokens (e.g., Batterink & Paller, 2017). A key-press that fell outside of this window was coded as a "false alarm." Of the participants who passed the RDK Practice Task, one participant had a d-prime value near zero (0.28), which was 2.82 standard deviations lower than the sample excluding this participant ($M = 3.68$, $SD = 1.20$) and was the only d-prime below one (the next lowest value was 1.15). As such, this participant was excluded from subsequent analyses, leaving 39 participants in the sample.

The results of the headphone test were not used as formal exclusion criteria, but rather to get a sense of how many participants followed the researcher's recommendation of wearing headphones. Based on a threshold of at least five of six correct responses, taken from Woods et al. (2017), 62% of the participants passed the assessment.

Data analysis

Similar to Experiment 1, we constructed generalized linear mixed-effects models with probit links using the "lme4" package in R (Bates et al., 2015). To assess the relative importance of scale direction, we created a nested model without scale condition and then compared this model to the full model on the basis of corrected AIC values. The PSE analyses were conducted in an identical manner to Experiment 1, as were the exploratory analyses associating the relative effect of scale direction with the questionnaire variables.

We included two additional analyses in Experiment 2. First, we combined the present data with Experiment 1 to assess whether any differences in the relative magnitude of scale direction significantly differed from those observed in Experiment 1. Second, we used the additional question in which participants described their auditory experience to categorize participants based on their awareness of the Shepard scales. Specifically, participants were binned into three groups: (1) those who made no mention of the Shepard scales, (2) those who mentioned the Shepard scales but did not explicitly mention directionality (e.g., one participant described hearing faint "carnival music"), and (3) those who mentioned the Shepard scales and described them in terms of motion (e.g., using terms such as *up, down, ascending, descending,* or *scales*). This variable was then correlated with the extracted scale direction slope for each participant, similar to the other questionnaire variables to assess whether explicit awareness of the scale direction was associated with MAE strength.

Results

Mixed-effect modeling of performance from Experiment 2

Participants were more likely to report that an RDK was ascending as a function of RDK coherence, which demonstrates that participants could correctly judge the motion in the RDK stimuli given the coding of descending RDKs as negative and ascending RDKs as positive (B = 0.50, $SE = 0.09$, $p < .001$). Unlike Experiment 1, there was no main effect of scale direction (B = 0.11, $SE = 0.06$, $p = .061$; Fig. 4). Indeed, the effect of scale direction was marginal and trending in the *opposite* direction (i.e., more likely to classify RDKs as ascending during ascending music blocks). There was also no interaction of RDK coherence level and scale direction (B = -0.01, $SE = 0.04$, $p = .720$). Removing scale direction from the model resulted in a nominally *better* fit relative to the full model (ΔAIC = -2.0).

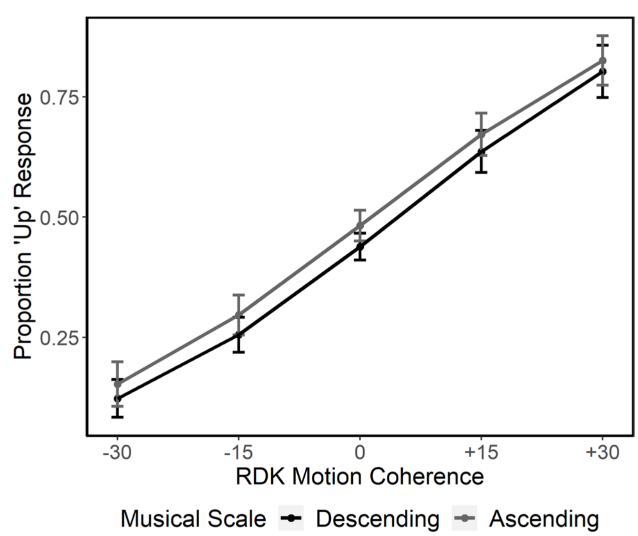

Fig. 4 Influence of scale direction on visual motion judgments from Experiment 2. Error bars represent ± 1 standard error of the mean. Random dot kinematogram (RDK) motion coherence is plotted on the x-axis, with descending motion arbitrarily coded as negative

Point of subjective equality analyses

The PSE analyses suggested that scale direction did not meaningfully influence the coherence level at which participants were equally likely to judge an RDK as ascending or descending. When listening to descending musical scales, participants' mean PSE was +4.60% ($SE = 2.29\%$). The 95% confidence interval technically did not contain zero [0.12, 9.09]; however, the confidence interval functionally included zero on the low end and additionally was in the *opposite* direction as predicted with the MAE (as descending musical scales made participants' PSE shift toward weakly ascending RDKs). When participants listened to ascending musical scales, their mean PSE was 1.35% ($SE = 2.48\%$), and the 95% confidence interval included zero [-3.51, 6.21]. Thus, the PSE analyses suggest that the musical scales did not meaningfully shift participants' PSE values and, if anything, descending scales made participants' PSE shift in an opposite direction as what would be predicted under the MAE.

Exploratory correlations among questionnaire variables

The exploratory correlational analyses did not show any associations between the influence of scale direction on RDK responses and the analyzed variables (age, musical training, self-reported musical skill, bilingualism, and explicit awareness of the musical stimuli). Table 2 provides a summary of these analyses. Similar to Experiment 1, the only significant association observed was between musical training and self-reported musical skill. There was a marginally significant association in which musicians were more likely to report greater detail about the unattended scales. However, no measure was associated with the relative influence of scale direction, which is perhaps not surprising as we did not observe an effect of scale direction as in Experiment 1.

Comparison of Experiment 2 with Experiment 1

In the model incorporating the data from both experiments, there was a significant interaction between scale direction and experiment ($B = 0.31$, $SE = 0.07$, $p < .001$), characterized by an attenuation of scale direction in Experiment 2 relative to Experiment 1. The main effects of RDK coherence ($B = 0.37$, $SE = 0.01$, $p < .001$) and scale direction ($B = -0.21$, $SE = 0.04$, $p < .001$) were also significant. Removing the term coding for experiment resulted in a significantly worse fitting model ($\Delta AIC = 10.1$). Comparing the data from Experiment 2 to just the diverted-attention condition of Experiment 1 elicited the same effect - there was a significant interaction of experiment and scale direction ($B = 0.32$, $SE = 0.08$, $p < .001$). Thus, the effects of the present experiment represent a significant attenuation of the MAE compared to Experiment 1, even when limiting the analyses to participants who were given a secondary auditory task to divert attention away from the scales.

Discussion

Experiment 2 was designed to provide a stronger assessment of whether attention is necessary in eliciting a cross-modal MAE. Relative to Experiment 1, this was accomplished in three ways. First, the secondary auditory task was presented *continuously* alongside the musical scales, requiring constant vigilance to the spoken letters. Second, the two sources of auditory information – spoken letters (to-be-attended) and musical scales (to-be-ignored) were spatially separated. Third, the relative amplitude for the two auditory streams was set to heavily favor the spoken letters (+15 dB relative to the scales).

The results of Experiment 2 suggest that attention is necessary in eliciting a cross-modal MAE. The overall effect of scale direction was not significant and was nominally in the opposite direction of an aftereffect – i.e., a slight bias to respond that RDKs were ascending following exposure to ascending musical scales. This null effect, moreover, was contextualized by directly comparing the

Table 2 Exploratory correlational analysis of scale direction effect and survey measures in Experiment 2

	Scale direction	Age	Musician	Musical skill	Bilingual	Describe scales
Scale Direction	–					
Age	.06	–				
Musician	.18	.19	–			
Musical Skill	.15	.10	.59***	–		
Bilingual	.02	-.14	.01	.16	–	
Describe Scales	.24	.20	.29+	.22	.17	

Note: "Describe scales" is a coded value representing the level of detail participants used in describing the unattended Shepard scales. + $p < .10$
*** $p < .001$

present results to those of Experiment 1. This analysis demonstrated that the more demanding attention manipulation indeed resulted in a significant attenuation of the MAE, as observed in Experiment 1. Critically, the results from Experiment 2 cannot be explained by a simple failure of participants noticing the musical scales in the unattended audio channel. A majority of participants (25 of 39) explicitly mentioned the music when asked about their auditory experiences in the experiment, with over one quarter of participants (11 of 39) providing descriptions of the *motion* of this music without explicit prompting to do so. This reported level of awareness of the musical scales did not relate to the strength of the observed MAE (and if anything, as demonstrated by the nominally positive correlation in Table 2, greater awareness of the musical scales was associated with a pattern of results opposite of an MAE).

The results of Experiment 2 can also help contextualize the unexpected asymmetry between ascending and descending musical scales from Experiment 1. Despite the efficacy of the attentional manipulation in Experiment 2, suggesting that the musical scales did not have an influence on visual motion judgments, we still observed a downward motion bias, as seen in Fig. 4 and partially supported by the PSE analyses (in which participants needed ascending motion before being equally likely to judge an RDK as ascending or descending). These findings tentatively suggest that the asymmetry in ascending and descending musical scales in Experiment 1 may be best explained by a general bias to report that RDKs were descending in the present paradigm.

Although Experiment 2 suggests that attention is necessary in eliciting a cross-modal MAE, there are some notable factors that limit this interpretation. First, Experiment 2 changed the attention manipulation relative to Experiment 1 by adopting a dichotic listening paradigm. This change was implemented in an effort to make the secondary task more demanding; however, it is unclear whether such a paradigm could *ever* elicit a cross-modal MAE (i.e., even when participants were instructed to attend to the musical scales and ignore the spoken letters), given the steady presentation of spoken letters at a more favorable signal-to-noise ratio relative to the musical scales. Second, the continued use of a web-based sample in Experiment 2 limits the experimental control over the audiovisual setup of the experiment. This is particularly important in the context of sound calibration, as participants might have lowered their volume following the calibration (e.g., to quiet the volume of the spoken letters), which could have made the musical scales not audible. These limitations are addressed in Experiment 3, which continues to use the dichotic listening paradigm to manipulate attention but (1) does so in a within-participant manner, and (2) uses an in-lab setup where audiovisual calibration can be monitored and controlled.

Experiment 3

Method

Participants

We recruited 25 participants (M_{age} = 23.52 years, SD = 6.79 years, range: 18–44 years), and 22 participants were included in the primary analyses (see Data exclusion for details). Unlike Experiments 1 and 2, participants completed the study in-person, in a designated laboratory space. The lower sample size in the present experiment relative to Experiments 1 and 2 was determined by two primary factors. First, the shift to a within-participant, in-person design was expected to increase statistical power to detect effects. Second, using a bootstrapping procedure (see Online Supplemental Material), we determined that the minimum sample size to find significant effects of scale direction from Experiment 1 on at least 80% of bootstrapped simulations was 14. The present sample size of 22, assuming comparable effects of scale direction as what was observed in Experiment 1, detected significant effects of scale direction on 90% of bootstrapped simulations. Participants were recruited from the larger Huron and Western University communities. If participants were enrolled in select courses, they could opt to receive course credit for participating; otherwise, participants received $15 CAD for completing the experiment.

Materials

The experiment materials were largely the same as those used in Experiment 2 with a couple of exceptions. First, the stream of spoken letters was embedded within multi-talker babble (Wilson et al., 2012) at a favorable signal-to-noise ratio (-15 dB). This was done to increase perceptual load, which has been argued to be necessary for effective manipulations of selective attention (e.g., Lavie, 1995, 2005), in part by discouraging dip listening during periods between the spoken letters (cf. Brungart, 2001). Second, given the within-participant design, we reimplemented the noise bursts task for blocks in which participants were instructed to listen to the ear containing the musical scales. The noise bursts were identical to those presented in Experiment 1 and were presented within the audio channel containing the musical scales.

Participants listened to the auditory stimuli through Sony MDR-7506 studio monitor headphones, connected to the testing computer via a Steinberg UR-12 USB audio interface. Moreover, we used a REED R8050 sound level meter to calibrate the volume of the sounds. The channel containing the spoken letters and multitalker babble was calibrated to 70 dB SPL (A-weighting). At this level, it was confirmed through informal pilot testing (n = 5) that the musical scales were perceptible, including directionality (ascending or descending).

Procedure

After providing informed consent, participants completed the letter repeat practice task practice and the RDK practice task in this order. The auditory calibration was removed from the present experiment given the in-lab setting, in which headphone use and volume calibration was done prior to participants arriving for the session.

Following these practice tasks, participants were introduced to the main task, in a similar manner as Experiment 2. Participants were instructed that they would hear streams of spoken letters, similar to the letter repeat practice task, and should press the spacebar as quickly as possible whenever the current letter matched the previous letter. Participants were alerted to the fact that the letters would be only presented to one ear and that the letters would be embedded within a faint background babble and were told to focus solely on the ear containing the letters.

Participants then completed two blocks (30 trials each; six repetitions of each of the five RDK coherence levels in a randomized order) in which they responded to letter repeats. Participants always completed the letter repeat blocks first, as there was a concern that if participants attended to the musical scales first, this might lessen the effectiveness of the attentional manipulation (as participants would know that there were ascending and descending scales). Although this design choice meant that attentional manipulation was confounded with time a reanalysis of the results from Experiment 2 suggested that the relative strength of the MAE did not change over the course of the experiment (see Online Supplemental Material). One block contained ascending scales, and the other block contained descending scales in the unattended ear (ordering counterbalanced across participants). Additionally, the ear of the letter repeat task (left or right) was counterbalanced across participants but fixed throughout the experiment (e.g., letters would be heard in the right or left ear for the entirety of the experiment).

After completing the first two blocks of the experiment, participants were then introduced to the noise burst detection task. Participants first completed four practice trials, in which they were instructed to press the spacebar whenever they heard a noise burst embedded within a musical scale. No letters or multitalker babble were presented during the practice trials. Following practice, participants were given updated instructions for the main task, in which they were told to ignore the spoken letters, attend to the ear containing the musical scales, and press the spacebar as quickly as possible whenever they detected a burst. Participants completed two more blocks (30 trials each) of the main task while performing the noise burst detection task, attending to the same ear as the musical scales. Scale order (ascending first or descending first) was counterbalanced across participants. For the initial trial of each block (in which a longer musical scale was presented), there were four noise bursts. The timing of the bursts was pseudo-randomly determined (the first occurred between 1.2 and 4.8 s, the second occurred between 7.2 and 10.8 s, the third occurred between 13.2 and 16.8 s, and the fourth occurred between 19.2 and 22.8 s). For subsequent trials within each block, there were two noise bursts, also pseudo-randomly determined (the first occurred between 1.2 and 3.9 s, and the second occurred between 5.1 and 6.8 s). This pseudo-randomization was done (1) to ensure that the noise burst did not occur too close to the start or end of a musical scale, and (2) to prevent participants from learning an exact pattern of when the bursts would occur.

Following the main task, participants completed a questionnaire identical to Experiment 2. Musical training was not analyzed in the present experiment as Experiments 1 and 2 did not find any evidence of an effect of musical training. Due to a programming error, the questionnaire was only presented to 21 of 25 participants. Following the questionnaire, participants were debriefed and compensated for their participation.

Data exclusion

Participants were excluded if (1) they did not achieve at least 75% accuracy on the RDK practice task, (2) if their d-prime was lower than one for the main letter repeat judgment task, and (3) if their d-prime was lower than one for the main noise burst detection task. Three participants failed to achieve the 75% accuracy threshold in the RDK practice task. All remaining 22 participants passed the 1-back letter repeat and noise burst performance thresholds, suggesting that they engaged with each task and effectively allocated attention to the appropriate ear. Mean performance (d') on the letter repeat task was 3.43 (SD = 0.86), and mean performance (d') on the noise burst detection task was 4.17 (SD = 0.83).

Data analysis

Similar to Experiments 1 and 2, we constructed generalized linear mixed-effects models with probit links using the "lme4" package in R (Bates et al., 2015). To assess the relative importance of scale direction, we created a nested model without scale condition and then compared this model to the full model on the basis of corrected AIC values. Unlike Experiment 2, the models contained an additional term (within-participant) for the attention manipulation, specifically coding whether participants were attending to the ear containing the musical scales (yes, no). The PSE analyses were conducted in an identical manner to Experiments 1 and 2.

Results

Mixed-effect modeling of performance from Experiment 3

Participants were more likely to report that an RDK was ascending as a function of RDK coherence, which demonstrates that participants could correctly judge the motion in the RDK stimuli given the coding of descending RDKs as negative and ascending RDKs as positive (B = 0.90, *SE* = 0.13, $p < .001$). There were additionally main effects of scale direction (B = -0.58, *SE* = 0.11, $p < .001$) and attention manipulation (B = -0.50, *SE* = 0.09, $p < .001$), with both ascending musical scales and attending to speech associated with a lower likelihood of classifying RDKs as ascending. These main effects, however, cannot be meaningfully interpreted, as there was also an interaction of scale direction and attention manipulation (B = 0.77, *SE* = 0.13, $p < .001$). This interaction (Fig. 5) is best characterized as the predicted MAE pattern of results for the blocks in which participants attended to the channel containing the musical scales (i.e., contrastive influence of musical scale direction on RDK motion judgment), as opposed to no discernible influence of musical scale direction on RDK motion judgment for the blocks in which participants attended to the channel containing the spoken letters. There was additionally an interaction of attention manipulation and RDK coherence level (B = 0.17, *SE* = 0.08, $p = .030$), characterized by a stronger linear fit for the trials in which participants attended to the channel containing the musical scales. No other term was significant in the model. Removing scale direction from the model resulted in a significantly worse fit relative to the full model (ΔAIC = 38.6).

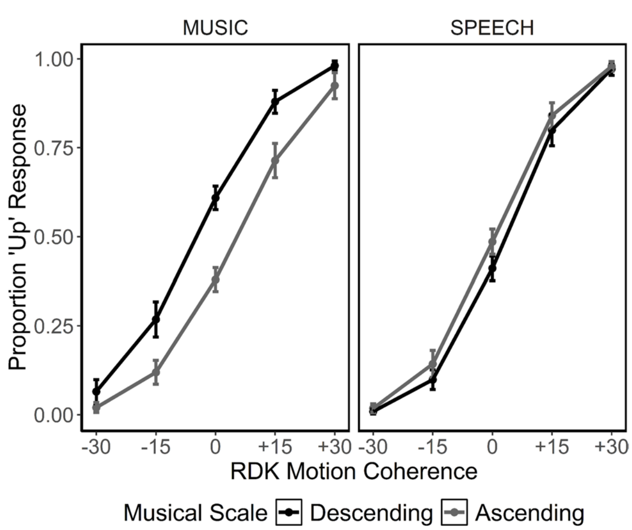

Fig. 5 Influence of scale direction on visual motion judgments from Experiment 3. Error bars represent ± 1 standard error of the mean. Random dot kinematogram (RDK) motion coherence is plotted on the x-axis, with descending motion arbitrarily coded as negative

Point of subjective equality analyses

The PSE analyses suggested that scale direction influenced the coherence level at which participants were equally likely to judge an RDK as ascending or descending, but only for the trials in which participants were attending to the channel containing the musical scales. For the trials in which participants attended to the channel containing the musical scales, participants' mean PSE was -5.70% (*SE* = 1.91%) when listening to descending musical scales. The 95% confidence interval did not contain zero [-9.44, -1.96]. When participants listened to ascending musical scales, their mean PSE was +6.65% (*SE* = 2.18%), and the 95% confidence interval did not include zero [2.38, 10.92].

For the trials in which participants attended to the channel not containing the musical scales, there was a small influence of scale direction on participants' PSE values in the *opposite* direction of a MAE, similar to Experiment 2. When listening to descending musical scales, participants' mean PSE was +4.27% (*SE* = 1.62%). The 95% confidence interval did not contain zero [1.09, 7.45]. When participants listened to ascending musical scales, their mean PSE was -0.53% (*SE* = 1.83%), and the 95% confidence interval included zero [-4.12, 3.06]. Thus, the PSE analyses replicated what was observed in Experiment 1 for the blocks in which participants attended to the channel containing the musical scales and replicated Experiment 2 for the blocks in which participants did not attend to the channel containing musical scales.

Discussion

The findings from Experiment 3 suggest that attention is necessary in eliciting a cross-modal MAE from musical scales. When participants were attending to spoken letters, presented in the opposite ear of the ascending or descending musical scales, there was no clear effect of scale direction on visual motion judgments, similar to Experiment 2. In contrast, when these same participants shifted their attention to detecting intermittent noise bursts, presented in the same ear as the ascending or descending musical scales, we observed evidence for a cross-modal MAE, similar to Experiment 1. Additionally, the in-lab setting of Experiment 3 provided assurances related to auditory volume calibration, headphone use, and consistent audiovisual hardware, that could only be inferred or could not be confirmed with the web-based samples of Experiments 1 and 2.

The results from Experiment 3 can also inform the unexpected asymmetry of musical scales observed in Experiment 1. Unlike Experiment 1, we did not observe an asymmetry in musical scale direction influencing visual motion judgments for blocks in which participants attended to the scales. Rather, ascending musical scales significantly shifted participants' PSEs below zero, whereas descending musical scales significantly shifted participants' PSEs above zero. These findings suggest that the asymmetry from Experiment 1 and the general downward bias observed in Experiment 2 might have been in part due to the web-based nature of Experiments 1 and 2, in which there was definitionally greater variability in the presentation hardware of the visual stimuli. Thus, while the finding of a cross-modal MAE in Experiment 1 is encouraging for future web-based psychophysical research, having data from both online and in-lab samples may be advisable to resolve potential discrepancies or unexpected findings within the online sample.

General discussion

The present set of experiments assessed whether attention to musical scales is necessary in eliciting a visual MAE. In Experiment 1, we conceptually replicated prior work (Hedger et al., 2013), finding that participants were more likely to report that a visual RDK was ascending following descending musical scales. In Experiment 2, however, we found that this effect could be entirely attenuated when presenting participants with a secondary auditory task requiring sustained attention. Importantly, the majority of participants reported hearing the unattended music in Experiment 2 and a sizable minority even described its motion properties (e.g., describing it as ascending or descending, without being explicitly prompted to provide information about the music's direction). Yet, this level of awareness of the musical stimuli did not relate to the strength of the MAE, suggesting that the complete attenuation of the MAE in Experiment 2 cannot be explained by participants not processing the musical scales – for example, due to energetic masking (Brungart, 2001). In Experiment 3, we conceptually replicated the findings from Experiments 1 and 2 using a within-participant, in-lab design, further strengthening the role of attention in eliciting a cross-modal MAE. Specifically, in Experiment 3, when participants attended to the ear containing the musical scales, we observed an MAE similar to Experiment 1. In contrast, when these same participants attended to the ear not containing musical scales, there was no discernible influence of musical scale direction on RDK motion judgments. Taken together, the present findings replicate the music-induced MAE in an online (Experiments 1 and 2) and in-lab (Experiment 3) sample and further suggest that attention to the music is necessary to elicit the MAE.

These findings have implications for the understanding of musical pitch change as vertical motion. Although the pitch-verticality mapping is well documented, existing in some capacity among prelinguistic infants (Dolscheid et al., 2014) and influencing perceptual judgments in both audition (e.g., Shu et al., 1993) and vision (e.g., Maeda et al., 2004), the present results suggest that pitch changes may not automatically influence perceptual processing in line with this mapping. Thus, attention may be required to process pitch changes in a manner that measurably influences cross-modal perceptual decisions. In this sense, the present results may be aligned with recent work (Antović et al., 2020), suggesting that cross-modal understandings of pitch are best understood using a conceptual framework, based on abstract schemata rather than lower-level perceptual features. The present findings can also be integrated with neural investigations of how auditory signals influence visual motion processing (Sadaghiani et al., 2009). Specifically, Sadaghiani et al. (2009) found that metaphoric auditory motion (ascending and descending pitch) fell between natural motion and linguistically descriptions of motion in terms of representations in both lower-level perceptual areas and higher-level convergence areas. Thus, this positioning of pitch changes in terms of both lower- and higher-level cortical activity suggests that attentional modulations may have varied effects in terms of influencing visual judgments.

The present results can also be interpreted in the broader context of the role of attention on perceptual processing. Attention has been shown to enhance perceptual processes across a variety of visual and auditory domains. In vision, attention has been shown to modulate visual contrast sensitivity (Carrasco et al., 2004), perceived size of visual motion displays (Anton-Erxleben et al., 2007) and, most importantly for the present study, the perception of visual motion coherence (Liu et al., 2006). In audition, attention has been shown to modulate aspects of auditory stream segregation (see Snyder et al., 2012 for a review), influencing the tracking of a conversational partner in background noise (Price & Bidelman, 2021), and even help differentiate neural responses to sounds with highly overlapping auditory feature spaces (Allen et al., 2019). Although work in the visuospatial domain has convincingly demonstrated that these attentional effects operate at a perceptual (rather than a post-perceptual, response bias) level (e.g., Carrasco et al., 2004; Liu et al., 2006; Yeshurun & Carrasco, 1998), the metaphoric depiction of motion in the present set of experiments makes it less clear whether attention is changing perception directly as opposed to altering post-perceptual decision processes (cf. Prinzmetal et al., 2008; Schneider & Komlos, 2008). Gallagher et al. (2021) explored this question in the context of MAEs resulting from implied motion in static images, concluding that the MAE is better explained by biasing decision making rather than changing perceptual processing directly. Future work is needed to determine whether the present cross-modal MAE is best explained in terms of biasing decision making as opposed to altering perceptual processing more directly.

There are several limitations to consider in the present set of experiments. First, the online administration of Experiments 1 and 2 necessarily increased variability in auditory and visual experiences relative to prior in-lab work (e.g., Hedger et al., 2013). Given this limitation, it is particularly notable that we observed results consistent with the MAE in Experiment 1. Additionally, this limitation was addressed in Experiment 3 using an in-lab sample, which suggests that future research in this area may benefit from online recruitment, particularly if recruiting specialized populations (e.g., speakers of languages that do not use verticality as the primary metaphor for understanding auditory pitch). Second, the present experiments used fixed coherence levels for the RDKs across all participants, as opposed to individually calibrated RDK coherence values based on initial testing (cf. Hedger et al., 2013; Winawer et al., 2008). This likely resulted in increased variability in RDK responses, particularly for participants that might have had higher coherence thresholds and would have experienced all coherence levels as essentially ambiguous. Despite this limitation, all analyzed participants passed an initial RDK practice and every model demonstrated clear separation of responses as a function of coherence level, demonstrating reliable motion percepts at least considered at the aggregate level.

Although the present results highlight the importance of attention in eliciting a cross-modal MAE, there are several ways in which future research could help contextualize the present results. First, the current experiments cannot speak to whether attentional demands can modulate the MAE in a more graded fashion. Approaches that systematically vary the degree to which attention is engaged (e.g., perhaps through increasing the length of time between spoken letters in the case of an *n*-back) might allow participants increasingly longer periods of time to process the musical scales, akin to the "glimpsing" model of speech perception in noise (Cooke, 2006) and possibly leading to an attenuated but statistically present MAE. Second, the particular nature of the attentional manipulation in Experiments 2 and 3 (consistent monitoring of spoken letters) makes it unclear whether the MAE was disrupted due to the linguistic nature of the secondary task. Research in other perceptual domains (e.g., color perception) has shown that language-based perceptual differences – such as the differences in processing dark and light blue by Russian speakers compared to English speakers – are eliminated under conditions of linguistic interference (Winawer et al., 2007). In contrast, linguistic interference does not appear to have the same influences of attenuating or eliminating perceptual interpretations of pitch in terms of verticality (Dolscheid et al., 2013). As such, replicating the present findings with using non-linguistic tokens could help contextualize these prior results. Third, the successful replication of the music-induced MAE in an online format opens up possibilities for testing a more diverse participant pool varying in conceptualizations of auditory pitch (e.g., Farsi speakers who primarily conceptualize pitch along a thickness dimension).

Conclusion

The present experiments were designed to assess whether explicit attention to ascending and descending musical scales was necessary for eliciting a music-induced visual MAE. Across three experiments, using both web-based and in-lab samples, as well as using both between-participant and within-participant designs, we found evidence consistent with the claim that attention is necessary to elicit a visual MAE from repeated listening to ascending or descending musical scales. Overall, these results suggest that pitch-verticality associations might not be automatic and appear to require some degree of attention. As such, the dominant association between pitch and vertical change found in Western music may require sufficient attention to process and understand to influence motion perception.

Supplementary information The online version contains supplementary material available at https://doi.org/10.3758/s13414-024-02985-5.

Open practices statement The data, analysis script, and materials for the study are available on the Open Science Framework (https://doi.org/10.17605/OSF.IO/QW4VE).

Authors' contributions **HC:** conceptualization, methodology, formal analysis, writing – original draft **CDT**: supervision, writing – review and editing **SVH**: conceptualization, methodology, software, formal analysis, data curation, writing – review and editing, supervision, funding acquisition. methodology, investigation, writing – original draft, writing – review and editing, project administration.

Funding This study was supported through internal start-up funds from Huron University awarded to SVH.

Availability of data and materials All data and study materials are available through the Open Science Framework (https://doi.org/10.17605/OSF.IO/QW4VE)

Code availability The study code (both the code to run the study as well as the code to analyze the data) is available through the Open Science Framework (https://doi.org/10.17605/OSF.IO/QW4VE)

Declarations

Conflicts of interest/Competing interests The authors declare no conflicts of interests.

Ethics approval This research was approved by the Huron University Research Ethics Boards (#14-202109 and #15S-202011).

Consent to participate All participants were provided with a Letter of Information, providing the details of the study, at the beginning and were provided with a Debriefing Letter, explaining the purpose of the study, at the end. All participants provided consent to participate in the study.

Consent for publication Not applicable.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer. https://doi.org/10.1007/978-1-4612-1694-0_15

Allen, E. J., Burton, P. C., Mesik, J., Olman, C. A., & Oxenham, A. J. (2019). Cortical correlates of attention to auditory features. *The Journal of Neuroscience, 39*(17), 3292–3300. https://doi.org/10.1523/JNEUROSCI.0588-18.2019

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15*(9), 839–843. https://doi.org/10.1016/j.cub.2005.03.046

Anstis, S., Verstraten, F. A. J., & Mather, G. (1998). The motion aftereffect. *Trends in Cognitive Sciences, 2*(3), 111–117. https://doi.org/10.1016/S1364-6613(98)01142-5

Anton-Erxleben, K., Henrich, C., & Treue, S. (2007). Attention changes perceived size of moving visual patterns. *Journal of Vision, 7*(11), 5. https://doi.org/10.1167/7.11.5

Antović, M., Mitić, J., & Benecasa, N. (2020). Conceptual rather than perceptual: Cross-modal binding of pitch sequencing is based on an underlying schematic structure. *Psychology of Music, 48*(1), 84–104. https://doi.org/10.1177/0305735618785242

Balestrucci, P., Ernst, M. O., & Moscatelli, A. (2022). *Psychophysics with R: The R Package MixedPsy* [Preprint]. Neuroscience. https://doi.org/10.1101/2022.06.20.496855

Bartlett, L. K., Graf, E. W., & Adams, W. J. (2018). The effects of attention and adaptation duration on the motion aftereffect. *Journal of Experimental Psychology: Human Perception and Performance, 44*(11), 1805–1814. https://doi.org/10.1037/xhp0000572

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex, 90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics, 62*(2), 321–332. https://doi.org/10.3758/BF03205552

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101–1109. https://doi.org/10.1121/1.1345696

Burnham, K. P., & Anderson, D. R. (Eds.). (2004). *Model Selection and Multimodel Inference*. New York: Springer. https://doi.org/10.1007/b97636

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7*(3), 308–313. https://doi.org/10.1038/nn1194

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America, 119*(3), 1562–1573. https://doi.org/10.1121/1.2166600

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dils, A. T., & Boroditsky, L. (2010). Visual motion aftereffect from understanding motion language. *Proceedings of the National Academy of Sciences, 107*(37), 16396–16400. https://doi.org/10.1073/pnas.1009438107

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science, 24*(5), 613–621. https://doi.org/10.1177/0956797612457374

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science, 25*(6), 1256–1261. https://doi.org/10.1177/0956797614528521

Donohue, S. E., Green, J. J., & Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli. *Frontiers in Integrative Neuroscience*, *9*. https://doi.org/10.3389/fnint.2015.00032

Eitan, Z., & Granot, R. Y. (2006). How music moves: Musical parameters and listeners' images of motion. *Music Perception, 23*(3), 221–247. https://doi.org/10.1525/mp.2006.23.3.221

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition, 114*(3), 405–422. https://doi.org/10.1016/j.cognition.2009.10.013

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*(3), 340–347. https://doi.org/10.1162/089892902317361886

Fernández-Prieto, I., Navarra, J., & Pons, F. (2015). How big is this sound? Crossmodal association between pitch and size in infants. *Infant Behavior and Development, 38*, 77–81. https://doi.org/10.1016/j.infbeh.2014.12.008

Gallagher, R. M., Suddendorf, T., & Arnold, D. H. (2021). The implied motion aftereffect changes decisions, but not confidence. *Attention, Perception, & Psychophysics, 83*, 3047–3055. https://doi.org/10.3758/s13414-021-02331-z

Hedger, S. C., Nusbaum, H. C., Lescop, O., Wallisch, P., & Hoeckner, B. (2013). Music can elicit a visual motion aftereffect. *Attention, Perception, & Psychophysics, 75*(5), 1039–1047. https://doi.org/10.3758/s13414-013-0443-z

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning—Implications for training and transfer. *Intelligence, 38*(6), 625–635. https://doi.org/10.1016/j.intell.2010.09.001

Jörges, B., & López-Moliner, J. (2017). Gravity as a Strong Prior: Implications for Perception and Action. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00203

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology Human Perception and Performance, 21*(3), 451–468. https://doi.org/10.1037/0096-1523.21.3.451

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9*(2), 75–82. https://doi.org/10.1016/j.tics.2004.12.004

Ling, S., & Carrasco, M. (2006). When sustained attention impairs perception. *Nature Neuroscience, 9*(10), 1243–1245. https://doi.org/10.1038/nn1761

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Liu, T., Fuller, S., & Carrasco, M. (2006). Attention alters the appearance of motion coherence. *Psychonomic Bulletin & Review, 13*(6), 1091–1096. https://doi.org/10.3758/BF03213931

Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology, 14*(23), R990–R991. https://doi.org/10.1016/j.cub.2004.11.018

Marks, L. E. (1974). On Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness. *The American Journal of Psychology, 87*(1/2), 173. https://doi.org/10.2307/1422011

Melara, R. D., & Marks, L. E. (1990). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance, 16*(2), 398–414. https://doi.org/10.1037/0096-1523.16.2.398

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General, 116*(4), 323–336. https://doi.org/10.1037/0096-3445.116.4.323

Morgan, M. J., & Solomon, J. A. (2019). Attention and the motion aftereffect. *Quarterly Journal of Experimental Psychology, 72*(12), 2848–2864. https://doi.org/10.1177/1747021819864552

Newsome, W., & Pare, E. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *The Journal of Neuroscience, 8*(6), 2201–2211. https://doi.org/10.1523/JNEUROSCI.08-06-02201.1988

Price, C. N., & Bidelman, G. M. (2021). Attention reinforces human corticofugal system to aid speech perception in noise. *NeuroImage, 235*, 118014. https://doi.org/10.1016/j.neuroimage.2021.118014

Prinzmetal, W., Long, V., & Leonhardt, J. (2008). Involuntary attention and brightness contrast. *Perception & Psychophysics, 70*(7), 1139–1150. https://doi.org/10.3758/PP.70.7.1139

Rajananda, S., Lau, H., & Odegaard, B. (2018). A random-dot kinematogram for web-based vision research. *Journal of Open Research Software, 6*(1), 6. https://doi.org/10.5334/jors.194

Rezec, A., Krekelberg, B., & Dobkins, K. R. (2004). Attention enhances adaptability: Evidence from motion adaptation experiments. *Vision Research, 44*(26), 3035–3044. https://doi.org/10.1016/j.visres.2004.07.020

Sadaghiani, S., Maier, J. X., & Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *Journal of Neuroscience, 29*(20), 6490–6499. https://doi.org/10.1523/JNEUROSCI.5437-08.2009

Schneider, K. A., & Komlos, M. (2008). Attention biases decisions but does not alter appearance. *Journal of Vision, 8*(15), 3–3. https://doi.org/10.1167/8.15.3

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America, 36*(12), 2346–2353. https://doi.org/10.1121/1.1919362

Shu, Z. J., Swindale, N. V., & Cynader, M. S. (1993). Spectral motion produces an auditory after-effect. *Nature, 364*(6439), 721–723. https://doi.org/10.1038/364721a0

Snyder, J. S., Gregg, M. K., Weintraub, D. M., & Alain, C. (2012). Attention, Awareness, and the Perception of Auditory Scenes. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00015

Spence, C., & Frings, C. (2020). Multisensory feature integration in (and out) of the focus of spatial attention. *Attention, Perception, & Psychophysics, 82*(1), 363–376. https://doi.org/10.3758/s13414-019-01813-5

Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics, 63*(4), 651–659. https://doi.org/10.3758/BF03194427

Wallach, H. (1935). Über visuell wahrgenommene Bewegungsrichtung. *Psychologische Forschung, 20*(1), 325–380. https://doi.org/10.1007/BF02409790

Webster, M. A. (2015). Visual adaptation. *Annual Review of Vision Science, 1*(1), 547–567. https://doi.org/10.1146/annurev-vision-082114-035509

Wilson, R. H., McArdle, R., Watts, K. L., & Smith, S. L. (2012). The revised speech perception in noise test (R-SPIN) in a multiple signal-to-noise ratio paradigm. *Journal of the American Academy of Audiology, 23*(08), 590–605. https://doi.org/10.3766/jaaa.23.7.9

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, 104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Winawer, J., Huk, A. C., & Boroditsky, L. (2008). A motion aftereffect from still photographs depicting motion. *Psychological Science, 19*(3), 276–283. https://doi.org/10.1111/j.1467-9280.2008.02080.x

Winawer, J., Huk, A. C., & Boroditsky, L. (2010). A motion aftereffect from visual imagery of motion. *Cognition, 114*(2), 276–284. https://doi.org/10.1016/j.cognition.2009.09.010

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics, 79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature, 396*(6706), 72–75. https://doi.org/10.1038/23936

Zbikowski, L. M. (2008). Metaphor and music. In R. W. Gibbs Jr. (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 502–524). Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.030