frontiers | Frontiers in Cognition

TYPE Original Research
PUBLISHED 18 May 2023
DOI 10.3389/fcogn.2023.1170422

Check for updates

OPEN ACCESS

EDITED BY
Alice Mado Proverbio,
University of Milano-Bicocca, Italy

REVIEWED BY
Julia Föcker,
University of Lincoln, United Kingdom
Jose Pablo Ossandon,
University of Hamburg, Germany

*CORRESPONDENCE
Laura Marie Getz
✉ lgetz@sandiego.edu

RECEIVED 20 February 2023
ACCEPTED 21 April 2023
PUBLISHED 18 May 2023

CITATION
Getz LM (2023) Competition between audiovisual correspondences aids understanding of interactions between auditory and visual perception.
Front. Cognit. 2:1170422.
doi: 10.3389/fcogn.2023.1170422

COPYRIGHT
© 2023 Getz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Competition between audiovisual correspondences aids understanding of interactions between auditory and visual perception

Laura Marie Getz*

Department of Psychological Sciences, University of San Diego, San Diego, CA, United States

An audiovisual correspondence (AVC) refers to an observer's seemingly arbitrary yet consistent matching of sensory features across the two modalities; for example, between auditory pitch height and visual height or visual size. Research on AVCs frequently uses a speeded classification procedure in which participants are asked to rapidly classify the pitch of a sound accompanied either by a congruent or an incongruent visual object (e.g., high pitches are congruent with higher/smaller visual objects and incongruent with lower/larger visual objects). To investigate the strength of these pitch AVCs (height, size, spatial frequency, brightness, sharpness), trials where the height AVC competed with each other AVC in terms of pitch congruency were created. For example, when classifying pitch height, participants were presented with trials where both visual height and size were congruent or incongruent with pitch; additionally, there were trials where height was congruent but size was incongruent (i.e., high pitch matched with large object at high height) and trials where size was congruent but height was incongruent (i.e., high pitch matched with small object at low height). Based on previous work, congruency between pitch and height was expected to be more important than congruency between pitch and spatial frequency, brightness, sharpness, or size. As predicted, in all four studies, RTs when only height was congruent were just as fast as when both dimensions were congruent. In contrast, RTs when only spatial frequency, brightness, sharpness, or size was congruent (and height was incongruent) were just as slow as when both dimensions were incongruent. These results reinforce the superiority of the pitch-height AVC and can be interpreted based on the metaphor used for pitch in English, showing the importance of semantic/linguistic effects to understanding AVCs.

KEYWORDS

cross-modal correspondence, audiovisual correspondence, pitch, speeded classification, competition

1. Introduction

An audiovisual correspondence (AVC) refers to an observer's consistent matching of sensory features across the auditory and visual modalities (see Spence and Sathian, 2020; Spence, 2022; for recent reviews). In this paper, I focus on the auditory dimension of pitch and its association with five different visual stimulus dimensions: elevation/height (e.g., Melara and O'Brien, 1987; Ben-Artzi and Marks, 1995; Patching and Quinlan, 2002; Evans and Treisman, 2010; Chiou and Rich, 2012; Dolscheid et al., 2014; Jamal et al., 2017

McCormick et al., 2018), size (Mondloch and Maurer, 2004; Gallace and Spence, 2006; Parise and Spence, 2009; Spector and Maurer, 2009; Evans and Treisman, 2010; Bien et al., 2012; Fernandez-Prieto et al., 2015), brightness (e.g., Marks, 1987; Martino and Marks, 1999; Mondloch and Maurer, 2004), angularity/sharpness (e.g., O'Boyle and Tarte, 1980; Marks, 1987; Parise and Spence, 2009; Maurer et al., 2012), and spatial frequency (e.g., Evans and Treisman, 2010).

The main paradigm that has been used to study AVCs is the speeded classification task, where participants categorize a multimodal stimulus based on one modality while ignoring values of the other modality. For example, participants would be asked to classify whether the second of two pitches was higher or lower than the first, while at the same time viewing shapes on the screen that vary in height, size, brightness, sharpness, or spatial frequency. For each correspondence, matched endpoints (i.e., "congruent" trials) tend to result in faster and/or more accurate processing than when the endpoints are reversed (i.e., "incongruent" trials). The consensus mappings for each AVC studied here is included in Table 1. It should be noted, however, that not all experiments find congruency effects, often depending on the nature of the task (e.g., Gallace and Spence, 2006; Keetels and Vroomen, 2011; Heron et al., 2012; Klapetek et al., 2012; Chiou and Rich, 2015; Getz and Kubovy, 2018).

In addition to occasional failures to replicate the congruency advantage, a number of studies show that top-down factors can also affect or change the congruency effect, including features of the stimulus or task and individual differences among observers (Welch and Warren, 1980; see Chen and Spence, 2017 for review). For example, Chiou and Rich (2015) argued that cross-modal associations are mediated by volitional rather than automatic attentional mechanisms and Klapetek et al. (2012) showed that participants can use cross-modal correspondences in a strategic manner only when they are informative. Getz and Kubovy (2018) created a modified version of the speeded classification task to quantify the degree of bottom-up and top-down influence on a variety of AVCs. Participants did not know until after each trial whether they would be asked to respond to the pitch of visual dimension, and blocks were created where participants had to respond to the incongruent pairings rather than congruent pairings. As an example, a participant might respond with one response button if the second pitch was higher than the first or the second circle was larger than the first, which are incongruent dimensions according to the consensus mapping. This allowed the researchers to determine whether participants could match the AVCs in the opposite direction without a loss in reaction time in order to show the amount of top-down influence from a change in task instructions.

Getz and Kubovy (2018) found a continuum of how easily participants could reverse their response mappings across the five different AVCs. Participants were just as fast to pair high pitches with small and large circles, but struggled to pair high pitches with circles lower on the screen, and spatial frequency, brightness, and sharpness correspondences fell in between the extremes of size and height. They concluded that the differences in top-down influence may be a function of metaphors used to describe pitch (c.f., Eitan and Timmers, 2010), with the dominant metaphor of pitch height (in English) showing the least top-down influence and largest bottom-up association.

A number of other studies reinforce the importance of the pitch-height correspondence (see Cian, 2017 for review). This correspondence likely stems from a combination of structural, statistical, and semantic/linguistic mechanisms (Spence, 2011; Parise, 2016), with innate biases facilitating adaptive learning of incoming environmental signals that are reinforced by semantic overlap (Spector and Maurer, 2009; Eitan and Timmers, 2010; Maurer et al., 2012; Parkinson et al., 2012; Dolscheid et al., 2013). For example, Fernandez-Prieto et al. (2017) found that language overlap strengthened cross-modal associations, with English speakers showing a stronger congruency advantage for pitch and height than Spanish/Catalan speakers who use different words for pitch height and spatial verticality. Others have shown that a congruency advantage was still elicited when using the verbal labels "high" and "low" rather than sensory stimuli (Melara and Marks, 1990; Gallace and Spence, 2006). Further, developmental investigations shown that the acquisition of the pitch-height correspondence is often dependent on linguistic input (Dolscheid et al., 2022).

Despite the abundance of evidence showing the importance of semantic or linguistic effects, not all researchers agree. For example, the fact that participants can be readily trained to use an unfamiliar pitch mapping (used in a different culture), but not a mapping unused by any culture (Dolscheid et al., 2013) suggests that statistical learning or natural biases may play more of a role than language knowledge in forging the pitch–height mapping. There is also evidence that Westerners can understand pitch metaphors not used in their culture (Eitan and Timmers, 2010) and evidence for pitch–height congruency among individuals who do not use such a metaphor in their native language (Parkinson et al., 2012). Some neural evidence also shows minimal evidence for mediation by semantic processing when investigating cross-modal correspondences using fMRI (McCormick et al., 2018). This lack of agreement shows that more work needs to be done to understand the bottom-up and top-down mechanisms involved in audiovisual correspondence perception in general and the pitch-height correspondence in particular.

In addition to highlighting the superiority of the pitch-height correspondence, Getz and Kubovy (2018) results also showed potential differences in the strength of AVC association mappings, with the greater degree of top-down influence providing evidence for weaker associations. AVC strength has been exhibited in a number of different ways in the literature, including confidence in mapping, consensus among population, vividness of association, consistency over time, and resistance to interference (Spence, 2022).

In the current study, I propose a new method of investigating correspondence strength; namely, competition between correspondences. The majority of research on AVCs has happened one pair at a time using simple tasks (Parise, 2016). However, this is counter to how perception happens in everyday life, "since real-world objects do not have only two sensory dimensions" (Jonas et al., 2017, p. 1104) and instead typically appear "as part of a much more complex and dynamically changing multisensory perceptual environment" (Klapetek et al., 2012, p. 1156). Indeed,

TABLE 1 Consensus mapping for each audiovisual correspondence examined in this study based on previous studies that found a significant congruency effect.

Visual dimension	Low-pitch pairing	High-pitch pairing	Relevant references
Elevation/height	Low	High	Melara and O'Brien, 1987; Ben-Artzi and Marks, 1995; Patching and Quinlan, 2002; Evans and Treisman, 2010; Chiou and Rich, 2012; Dolscheid et al., 2014; Jamal et al., 2017; McCormick et al., 2018
Size	Large	Small	Mondloch and Maurer, 2004; Gallace and Spence, 2006; Parise and Spence, 2009; Spector and Maurer, 2009; Evans and Treisman, 2010; Bien et al., 2012; Fernandez-Prieto et al., 2015
Brightness/contrast	Dark	Bright	Marks, 1987; Martino and Marks, 1999; Mondloch and Maurer, 2004
Angularity/sharpness	Rounded	Sharp	O'Boyle and Tarte, 1980; Marks, 1987; Parise and Spence, 2009; Maurer et al., 2012
Spatial frequency	Low (Wide)	High (Narrow)	Evans and Treisman, 2010

investigating multiple visual components at once in comparison to auditory inputs has biological significance. For example, pitch-size mappings often conflict with pitch-height mappings in the environment: larger individuals producing lower pitches may be taller (spatially higher) than smaller individuals (spatially lower) producing higher pitches. The goal of the present study was therefore to investigate what happens when the congruency of multiple correspondences is manipulated at once. Two alternative hypotheses are shown in Figure 1: an additive effect of congruency or a hierarchy of AVC strength, with greater emphasis on pitch and height than the other visual dimensions. I predicted a hierarchy of correspondence strength, with precedence given to pitch-height over all other correspondences (see Getz and Kubovy, 2018, Figure 5). Therefore, I specifically looked at the competition between height and each of the other four visual components: size, brightness, sharpness, and spatial frequency. For example, on the pitch-height-size task, participants might respond to a high pitch accompanied by a large (incongruent size) circle high on the screen (congruent height) or to a high pitch accompanied by a small (congruent size) circle low on the screen (incongruent height).

Additionally, I focus in this paper on responses to the auditory dimension (with responses to visual dimension in Supplementary material). Even though many prior studies focus on visual-relevant responses, because I put the two visual dimensions in competition with each other here in how they relate to pitch congruency, the effects are most straightforward in the pitch-relevant responding condition. Further, prior studies that have focused on the difference in responding across modalities often show larger effects with auditory responding (Evans and Treisman, 2010) or even sometimes find congruency effects when responding to the auditory dimension but not the visual dimension (Jamal et al., 2017).

2. Method

2.1. Participants

There were a total of 221 participants from the University of San Diego across the four versions of the experiment who participated in exchange for credit in their introductory psychology course. All participants reported normal or corrected-to-normal vision and normal hearing. There were 56 participants in the size experiment (M_{age} = 18.64, SD_{age} = 0.67; 42 women, 12 men, 2 unreported), 51 participants in the sharp experiment (M_{age} = 18.73, SD_{age} = 0.82; 39 women, 7 men, 5 unreported), 57 participants in the spatial frequency experiment (M_{age} = 18.89, SD_{age} = 0.85; 33 women, 16 men, 8 unreported), and 57 participants in the bright experiment (M_{age} = 18.56, SD_{age} = 0.75; 38 women, 10 men, 9 unreported). Although no a priori power calculation was completed, sample sizes here were in line with or larger than previous research using a similar speeded classification task (e.g., Gallace and Spence, 2006; Evans and Treisman, 2010; Getz and Kubovy, 2018).

2.2. Stimuli

2.2.1. Auditory pitches

All sounds were 400 ms sine tones with 10 ms rise and decay times normalized to 76 dB. There was a standard pitch of 261.6 Hz and target pitches ±2, 5, and 10 semitones higher or lower than the standard (lower = 233.1 Hz, 196.0 Hz, 146.8 Hz; higher = 293.7 Hz, 349.2 Hz, 466.2 Hz) when pitch was the relevant dimension. When one of the visual dimensions was relevant, the ±5 semitone target pitches were used. These pitch ranges were chosen to be smaller than previous studies using large pitch differences (e.g., 300 Hz vs. 4,500 Hz in Gallace and Spence, 2006) but still far enough apart to be easily distinguishable (JND value for adults is as low as 0.25 semitones; Heller Murray et al., 2019).

2.2.2. Visual shapes

Examples of stimulus pairs for size, sharpness, spatial frequency, and brightness are shown in Figure 2.

For *size* (Figure 2A), sounds were paired with white circles on a black background; the standard circle had a 5.6° radius and the target circles were adjusted ±10%, ±20%, and ±40% in radius when size was the relevant dimension (smaller = 5.06°, 4.5°, 3.4°; larger = 6.2°, 6.8°, 7.9°). When height and pitch were relevant, the ±20% target circles were used.

For *sharp* (Figure 2B), sounds were paired with white shapes on a black background. Six angular and rounded shape pairs were generated using Matlab (as in Getz and Kubovy, 2018); angular shapes included between 4 and 30 polar coordinates sorted and successively connected on a Cartesian grid. Rounded shapes were

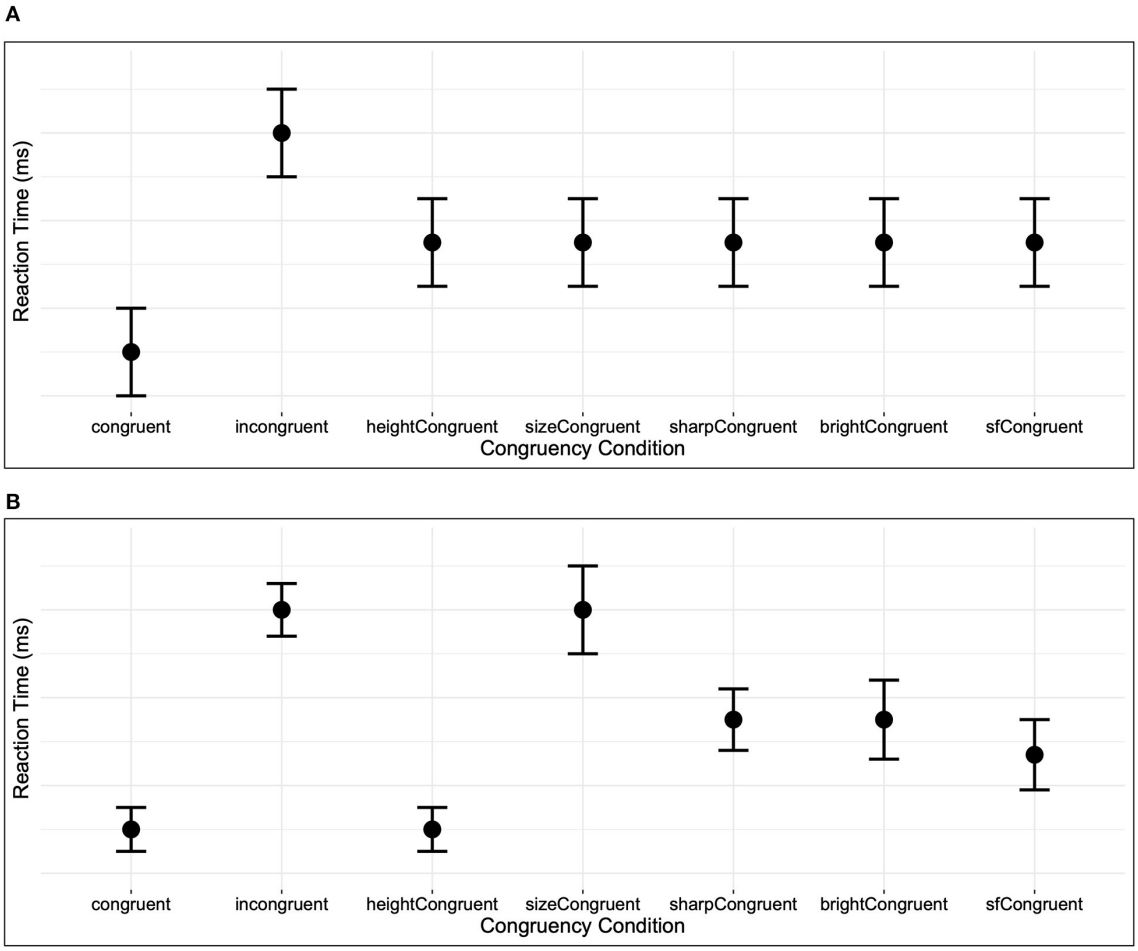

FIGURE 1
Alternative hypotheses. **(A)** Additive effect: both visual dimensions being congruent with pitch would result in faster reaction times than both visual dimensions being incongruent with pitch, and any single dimension being congruent with pitch would result in an intermediate reaction time. **(B)** Hierarchical effect: When pitch and height are congruent (and the other visual dimension is incongruent), reaction times would be just as fast as if height and another visual dimension are both congruent with pitch. When pitch and size are congruent (and height is incongruent), reaction times would be just as slow as if both visual dimensions were incongruent. The sharpness, brightness, and spatial frequency correspondences would fall in between, as in Getz and Kubovy (2018; Figure 5).

created by performing a quadratic spline on the angular shapes, thus controlling for overall size and number of edges.

For *spatial frequency* (Figure 2C), sounds were paired with circles that were 13.1° in diameter and included high-contrast black and white sinusoidal gratings oriented 45° to the left presented on a gray background. The comparison grating was 0.07 cycles/pixel, and the grating was adjusted ±2, 4, and 6 cycles/pixel when spatial frequency was the relevant dimension (narrower: 0.09, 0.11, 0.13 cycles/pixel; wider: 0.05, 0.03, 0.01 cycles/pixel). When height and pitch were relevant, the ±4 cycles/pixel target gratings were used.

For *bright* (Figure 2D), sounds were paired with grayscale circles on a black background; the standard circle had a standard brightness of 150 [on a 0–255 colormap scale] and the target circles were adjusted ±50, ±75, and ±100 colormap units (darker = 100, 75, 50; brighter = 200, 225, 250) when bright was the relevant dimension (as in Getz and Kubovy, 2018). When height and pitch were relevant, the ±75 brightness was used.

Each of these 4 visual dimensions was placed in competition with *height*; the standard height was vertically centered on the screen and target circles were ±10%, ±20%, and ±40% vertically displaced from the screen's center (±2.9°, 5.8°, and 11.5°) when height was the relevant dimension. When pitch and the second visual dimension were relevant, the ±20% target displacements were used.

2.3. Design

Participants were randomly assigned to the size, bright, sharp, or spatial frequency dimension. Each session consisted of three parts (described below), where participants were to focus on the relevant feature of auditory pitch, visual height, or the other visual dimension in a counterbalanced order in one 60-min session. Full details of the three tasks are included in Table 2.

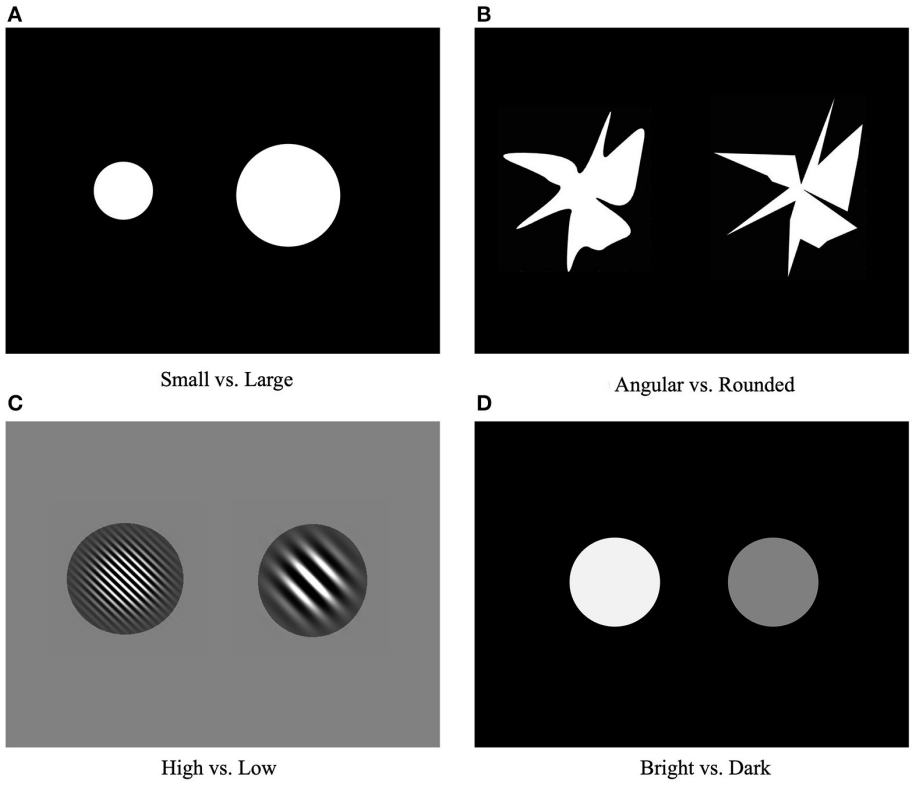

FIGURE 2
Example stimuli for the **(A)** size, **(B)** angularity, **(C)** spatial frequency, and **(D)** brightness experiments.

Within each part of the session, there were five different congruency conditions: (a) *congruent*, where both visual dimensions were congruent with pitch (e.g., high pitch paired with a small circle high on the screen and low pitch paired with a large circle low on the screen); (b) *incongruent*, where both visual dimensions were incongruent with pitch (e.g., high pitch paired with a large circle low on the screen and low pitch paired with a small circle high on the screen); (c) *height-congruent*, where height and pitch were congruent and the other visual dimension and pitch were incongruent (e.g., high pitch paired with a large circle high on the screen and low pitch paired with a small circle low on the screen); (d) *visual dimension-congruent*, where the other visual dimension and pitch were congruent and height and pitch were incongruent (e.g., high pitch paired with a small circle low on the screen and low pitch paired with a large circle high on the screen); and (e) *unimodal*, where pitches were presented with no accompanying shapes.

In the pitch-relevant task, there were six different comparison conditions, with target pitches 2, 5, and 10 semitones higher and lower than the standard pitch. This led to a total of 5 (congruency conditions) × 3 (comparison pitch) × 2 (directions) trials per block. In the height-relevant task, target shapes were ±10, ±20, and ±40% vertically displaced from the screen center. This led to a total of 5 (congruency conditions) × 3 (comparison height) × 2 (directions) trials per block. In the other visual dimension-relevant task, the six comparison conditions depended on visual dimension. For size, the target circles were ±10, ±20, and ±40% different in size from the standard circle. This led to a total of 5 (congruency conditions) × 3 (comparison size) × 2 (directions) trials per block. For brightness, the target circles were ±50, ±75, and ±100 units adjusted from the standard circle. This led to a total of 5 (congruency conditions) × 3 (comparison brightness) × 2 (directions) trials per block. For spatial frequency, the target gratings were ±2, 4, and 6 cycles/pixel adjusted from the standard circle grating. This led to a total of 5 (congruency conditions) × 3 (comparison frequency) × 2 (directions) trials per block. For sharpness, there were six angular and rounded shape pairs. This led to a total of 5 (congruency conditions) × 6 (comparison pairs) trials per block. Each participant completed eight blocks for each relevance task, for a total of 240 pitch-relevant trials, 240 height-relevant trials, and 240 other visual dimension-relevant trials. Before each task, participants also completed 10 practice trials to familiarize them with the new task.

2.4. Procedure

The experiment was completed using OpenSesame 3.2.8 (Mathôt et al., 2012) running on an Intel Core i5 CPU. The visual stimuli were presented on a 21.5" Dell 1080p LED monitor (screen resolution: 1,024 × 768 pixels) and sounds were played through Sennheiser HD 559 headphones. All procedures were in accordance with the University of San Diego Institutional Review Board and all participants provided informed consent before the experiment began.

TABLE 2 Stimulus characteristics for each correspondence based on responding dimension.

Correspondence	Responding dimension	Stimulus characteristics	
Size	**Pitch relevant**	Pitch compare	**261.6 Hz +/– 2, 5, 10 semitones**
		Height compare	0 +/– 5.8° up/down
		Size compare	5.6° +/– 1.1° radius
	Height relevant	Pitch compare	261.6 Hz +/– 5 semitones
		Height compare	**0** +/– 2.9°, 5.8°, 11.5° **up/down**
		Size compare	5.6° +/– 1.1° radius
	Size relevant	Pitch compare	261.6 Hz +/– 5 semitones
		Height compare	0 +/– 5.8° up/down
		Size compare	5.6° +/– 0.6 °, 1.1°, 2.2° radius
Bright	**Pitch relevant**	Pitch compare	**261.6 Hz +/– 2, 5, 10 semitones**
		Height compare	0 +/– 5.8° up/down
		Bright compare	150 +/– 75 RGB
	Height relevant	Pitch compare	261.6 Hz +/– 5 semitones
		Height compare	**0 +/–** 2.9°, 5.8°, 11.5° **up/down**
		Bright compare	150 +/– 75 RGB
	Bright relevant	Pitch compare	261.6 Hz +/– 5 semitones
		Height compare	0 +/– 5.8° up/down
		Bright compare	**150 +/– 50, 75, 100 RGB**
Sharp	**Pitch relevant**	Pitch compare	**261.6 Hz +/– 2, 5, 10 semitones**
		Height compare	0 +/– 5.8° up/down
		Sharp compare	6 sharp/round pairs
	Height relevant	Pitch compare	261.6 Hz +/– 5 semitones
		Height compare	**0 +/–** 2.9°, 5.8°, 11.5° **up/down**
		Sharp compare	6 sharp/round pairs
	Sharp relevant	Pitch compare	261.6 Hz +/– 5 semitones
		Height compare	0 +/– 5.8° up/down
		Sharp compare	**6 sharp/round pairs**
Spatial Frequency	**Pitch relevant**	Pitch compare	**261.6 Hz +/– 2, 5, 10 semitones**
		Height compare	0 +/– 5.8° up/down
		SF compare	0.07 +/– 4 cycles/pixel
	Height relevant	Pitch compare	261.6 Hz +/– 5 semitones
		Height compare	**0 +/–** 2.9°, 5.8°, 11.5° **up/down**
		SF compare	0.07 +/– 4 cycles/pixel
	Spatial frequency relevant	Pitch compare	261.6 Hz +/– 5 semitones
		Height compare	0 +/– 5.8° up/down
		SF compare	**0.07 +/– 2, 4, 6 cycles/pixel**

Participants were instructed to respond as quickly and as accurately as possible. Figure 3 illustrates the sequence of events during each trial. Each trial began with a 500 ms fixation cross, followed by the presentation of the standard shape left of center (always centered vertically) accompanied by the 400 ms standard pitch (261.6 Hz). After a 500 ms blank screen, the target shape appeared right of center (adjusted in radius/spatial frequency/brightness/sharpness and displaced vertically based on congruency condition) accompanied by the 400 ms target pitch (±2, 5, or 10 semitones up/down from standard depending on task). After both pitch/shape pairings, a response screen displayed the options for participants to indicate how the second target shape/pitch compared to the first standard shape/pitch. In the pitch-relevant task, participants indicated whether the second pitch

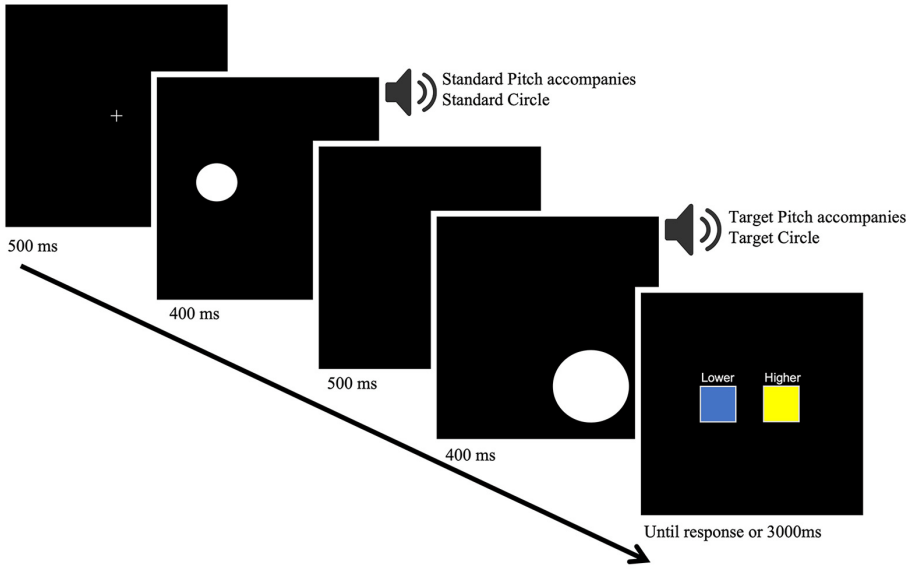

FIGURE 3
Schematic illustration of the trial sequence used in the experiment. Illustration shows a pitch-relevant trial for the height/size competition.

was lower or higher than the first pitch; in the height-relevant task, participants indicated whether the second shape was below or above the first shape; in the size-relevant task, participants indicated whether the second circle was smaller or larger than the first circle; in the bright-relevant task, participants indicated whether the second circle was darker or brighter than the first circle; in the sharp-relevant task, participants indicated whether the second shape was rounder or sharper than the first shape; in the spatial frequency-relevant task, participants indicated whether the grating on the second circle was wider or narrower than the second circle. Participants responded using a MilliKey LH-8 r1 button-box and their response times were recorded. The response screen remained until the participant responded, at which point there was a 200 ms inter-trial interval before the next fixation screen appeared. Participants could take a break every 30 trials for as long as they liked, and pressed the SPACEBAR to continue to the next block of trials.

3. Results

All analyses were performed in R (R Development Core Team, 2022) using the packages lme4 (Bates et al., 2015) and multcomp (Hothorn et al., 2008). Reaction time (RT) analyses only included correct responses. RTs <50 ms were discarded, assuming that the participant had started their response before the response cue appeared on the screen; RTs > 3,000 ms were also discarded because participants were told to respond within 3 seconds.

In the main text, I focus on the pitch-relevant condition; details about the height-relevant and visual dimension-relevant conditions are included in the Supplementary material. For each correspondence competition the RT data were modeled with linear mixed-effects models (LMMs) with congruency (5 levels) and pitch difference (treated as continuous) as fixed effects and subject-by-subject variations as a random effect. Because of the interest in how each of the five congruency levels (congruent, incongruent, height congruent, size congruent, unimodal) compare in terms of reaction time, a subsequent Tukey test for multiple comparisons designed for linear mixed effects models was performed.

3.1. Size vs. height

Analysis began with 12,330 observations from 56 participants; 274 trials were removed for short/long responses and 982 inaccurate responses were removed, leaving 11,074 trials. Accuracy was high overall, though slightly lower in the size-congruent (88.7%, $SE = 0.78\%$) and both incongruent (89.8%, $SE = 0.78\%$) conditions than the both congruent (93.8%, $SE = 0.80\%$), height-congruent (94.3%, $SE = 0.78\%$), and unimodal (92.7%, $SE = 0.78\%$) conditions. A LMM on accuracy with congruency as a fixed effect and subject-by-subject variations as a random effect confirmed that the size-congruent and both incongruent conditions were significantly less accurate than the both-congruent, height-congruent, and unimodal conditions (all linear Tukey contrasts comparisons $p < 0.002$).

Table 3 presents all possible congruency condition comparisons resulting from the LMM analysis for the size vs. height competition. As shown in Figure 4, the typical congruency advantage was displayed ($b = 39.94$, $p = 0.004$), with faster responses in the both congruent condition than the both incongruent condition. Further, the height-congruent/size-incongruent condition was just as fast as the both congruent condition ($b = -14.54$, $p = 0.689$), whereas the size-congruent/height-incongruent condition was just as slow as the both incongruent condition ($b = 5.65$, $p = 0.988$). There was also a significant effect of pitch difference ($b = -112.75$, $SE = 4.42$, $z = -25.50$, $p < 0.001$), meaning participants were

slower to respond to target pitches that were closer in semitones to the comparison pitch and faster to respond to target pitches that were more semitones apart from the comparison pitch.

As detailed in Supplementary material Section 2, although trending in the expected direction, there was no congruency advantage when height was the relevant dimension and no difference between the pitch-congruent and size-congruent conditions. When size was the relevant dimension, there was a marginal congruency advantage, but no difference between the pitch-congruent and height-congruent conditions.

3.2. Sharp vs. height

Analysis began with 12,480 observations from 51 participants; 459 trials were removed for short/long responses. Accuracy was high overall, though slightly lower in the sharp congruent (89.4%, $SE = 0.77\%$) and both incongruent (90.0%, $SE = 0.77\%$) conditions than the both congruent (94.1%, $SE = 0.99\%$), height-congruent (93.1%, $SE = 0.77\%$), and unimodal (92.3%, $SE = 0.77\%$) conditions. A LMM on accuracy with congruency as a fixed effect and subject-by-subject variations as a random effect confirmed that the sharp-congruent and both incongruent conditions were significantly less accurate than the both-congruent, height-congruent, and unimodal conditions (all linear Tukey contrasts comparisons $p < 0.027$). Inaccurate responses were removed from 986 trials, leaving 11,035 trials.

Table 4 presents all possible congruency condition comparisons resulting from the LMM analysis for the sharpness vs. height competition. As shown in Figure 5, the

TABLE 3 Size experiment; pitch-relevant condition.

Comparison	Estimate	SE	*z*	*p*
Height congruent–congruent	−14.54	11.16	−1.30	0.689
Incongruent–congruent	39.94	11.30	3.54	0.004
Size congruent–congruent	45.59	11.33	4.02	<0.001
Unimodal–congruent	42.35	11.21	3.78	0.002
Incongruent–height congruent	54.48	11.27	4.83	<0.001
Size congruent–height congruent	60.13	11.31	5.32	<0.001
Unimodal–height congruent	56.89	11.18	5.09	<0.001
Size congruent–incongruent	5.65	11.44	0.49	0.988
Unimodal–incongruent	2.41	11.32	0.21	0.999
Unimodal–size congruent	−3.24	11.35	−0.29	0.998

Tukey's *post-hoc* test for multiple comparisons of RT values across congruency conditions.

TABLE 4 Sharp experiment; pitch-relevant condition.

Comparison	Estimate	SE	*z*	*p*
Height congruent–congruent	7.93	13.12	0.61	0.974
Incongruent–congruent	56.32	13.22	4.26	<0.001
Sharp congruent–congruent	71.89	13.25	5.43	<0.001
Unimodal–congruent	43.89	13.15	3.34	0.008
Incongruent–height congruent	48.38	13.25	3.65	0.002
Sharp congruent–height congruent	63.95	13.28	4.82	<0.001
Unimodal–height congruent	35.95	13.19	2.73	0.050
Sharp congruent–incongruent	15.57	13.38	1.16	0.772
Unimodal–incongruent	−12.43	13.29	−0.94	0.883
Unimodal–sharp congruent	−28.00	13.32	−2.10	0.219

Tukey's *post-hoc* test for multiple comparisons of RT values across congruency conditions.

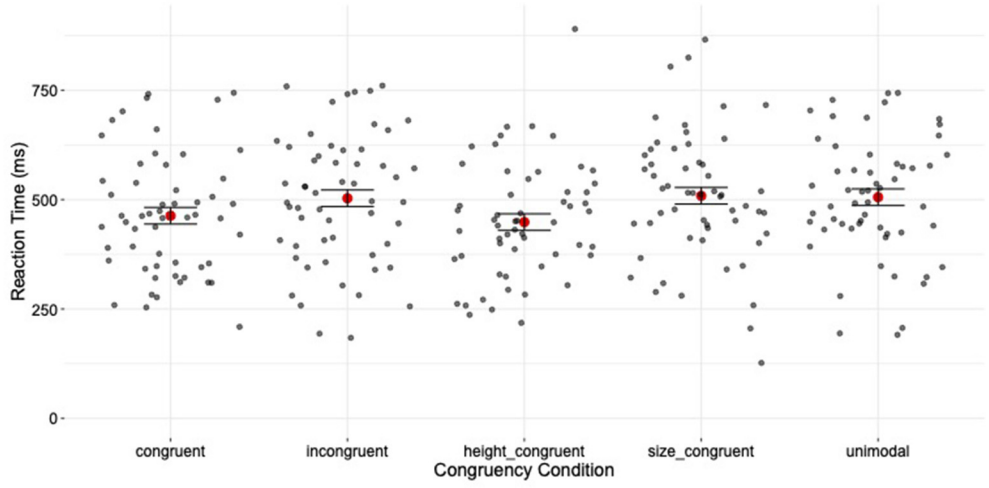

FIGURE 4
Size experiment; Pitch-relevant task. Reaction times for the five congruency conditions. Participants responded significantly faster in the congruent and height congruent conditions than in the size congruent, incongruent, and unimodal conditions. Error bars represent standard error. Individual points show each participant's average reaction time per condition.

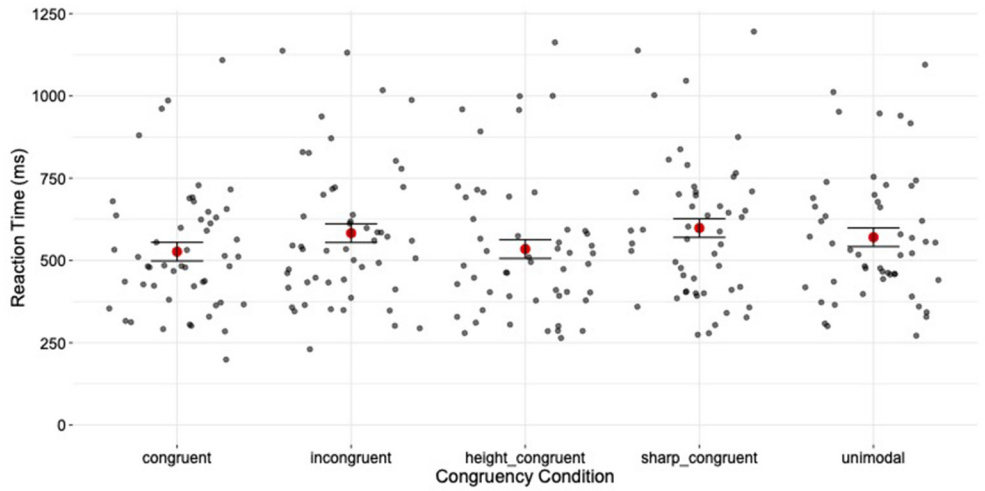

FIGURE 5
Sharp experiment; Pitch-relevant task. Reaction times for the five congruency conditions. Participants responded significantly faster in the congruent and height congruent conditions than in the sharp congruent, incongruent, and unimodal conditions. Error bars represent standard error. Individual points show each participant's average reaction time per condition.

typical congruency advantage was replicated ($b = 56.32$, $p < 0.001$), with faster responses in the both congruent condition than the both incongruent condition. Further, the height-congruent/sharp-incongruent condition was just as fast as the both congruent condition ($b = 7.93$, $p = 0.974$), whereas the sharp-congruent/height-incongruent condition was just as slow as the both incongruent condition ($b = 15.57$, $p = 0.772$). A significant effect of pitch difference was also found ($b = -125.72$, $SE = 5.19$, $t = -24.21$, $p < 0.001$), meaning participants were slower to respond to target pitches that were closer in semitones to the comparison pitch.

As detailed in Supplementary material Section 3, when height was the relevant dimension, there was a significant congruency advantage and participants were faster to respond in the pitch congruent condition than in the both incongruent condition. However, there were no differences between conditions when sharp was the relevant dimension.

3.3. Spatial frequency vs. height

Analysis began with 13,920 observations from 57 participants; 455 trials were removed for short/long responses. Accuracy was high overall, though slightly lower in the spatial frequency congruent (89.8%, $SE = 0.75\%$) and both incongruent (88.4%, $SE = 0.76\%$) conditions than the both congruent (93.4%, $SE = 0.85\%$), height-congruent (93.2%, $SE = 0.76\%$), and unimodal (91.5%, $SE = 0.76\%$) conditions. A LMM on accuracy with congruency as a fixed effect and subject-by-subject variations as a random effect confirmed that the spatial frequency-congruent and both incongruent conditions were significantly less accurate than the both-congruent and height-congruent (all linear Tukey contrasts comparisons $p < 0.001$). Inaccurate responses were removed from 1,178 trials, leaving 12,287 trials.

TABLE 5 Spatial frequency experiment; pitch-relevant condition.

Comparison	*Estimate*	*SE*	*z*	*p*
Height-congruent–congruent	19.26	10.76	1.79	0.380
Incongruent–congruent	49.01	10.92	4.49	<0.001
Spatial frequency congruent–congruent	48.92	10.85	4.51	<0.001
Unimodal–congruent	43.34	10.82	4.01	<0.001
Incongruent–height congruent	29.75	10.94	2.72	0.051
Spatial frequency congruent–height congruent	29.66	10.87	2.73	0.050
Unimodal–height congruent	24.08	10.84	2.22	0.171
Spatial frequency congruent–incongruent	−0.09	11.02	−0.01	0.999
Unimodal–incongruent	−5.67	10.99	−0.52	0.986
Unimodal–Spatial frequency congruent	−5.58	10.92	−0.51	0.986

Tukey's *post-hoc* test for multiple comparisons of RT values across congruency conditions.

Table 5 presents all possible congruency condition comparisons resulting from the LMM analysis for the spatial frequency vs. height competition. As shown in Figure 6, the typical congruency advantage was again replicated ($b = 49.01$, $p < 0.001$), with faster responses in the both congruent condition than the both incongruent condition. Further, the height-congruent/spatial frequency-incongruent condition was just as fast as the both congruent condition ($b = 19.26$, $p = 0.380$), whereas the spatial frequency-congruent/height-incongruent condition was just as slow as the both incongruent condition ($b = -0.09$, $p = 0.999$). There was also a significant effect of pitch difference ($b = -107.65$, $SE = 4.27$, $t = -25.24$, $p < 0.001$), meaning participants were

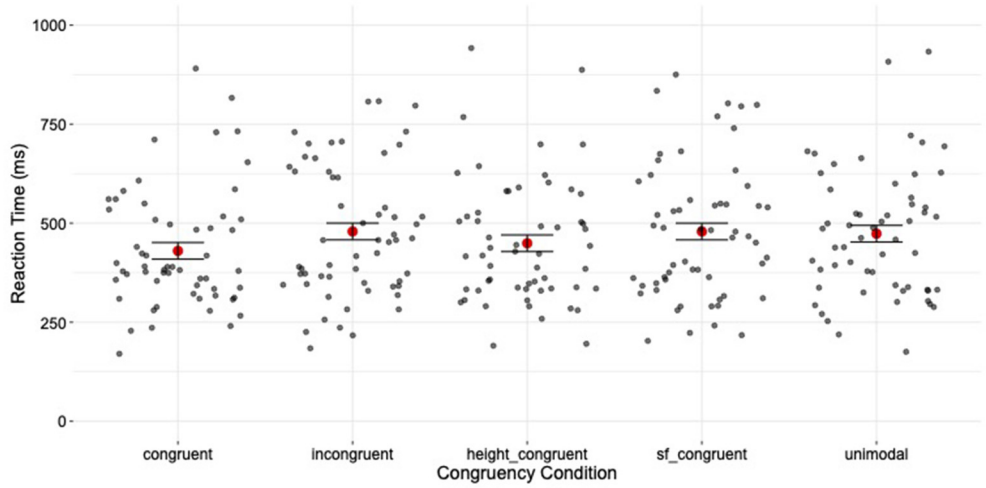

FIGURE 6
Spatial frequency experiment; Pitch-relevant task. Reaction times for the five congruency conditions. Participants responded significantly faster in the congruent and height congruent conditions than in the bright congruent, incongruent, and unimodal conditions. Error bars represent standard error. Individual points show each participant's average reaction time per condition.

slower to respond to target pitches that were closer in semitones to the comparison pitch.

As detailed in Supplementary material Section 4, there was a marginal congruency advantage when height was the relevant dimension, but no differences between pitch congruent and spatial frequency congruent conditions. There were no significant differences between audiovisual conditions when spatial frequency was the relevant condition.

3.4. Brightness vs. height

Analysis began with 14,096 observations from 58 participants; 460 trials were removed for short/long responses. Accuracy was high overall, though slightly lower in the bright congruent (87.6%, $SE = 1.28\%$) and both incongruent (85.7%, $SE = 0.79\%$) conditions than the both congruent (93.1%, $SE = 0.79\%$), height-congruent (91.9%, $SE = 0.79\%$), and unimodal (90.1%, $SE = 0.79\%$) conditions. A LMM on accuracy with congruency as a fixed effect and subject-by-subject variations as a random effect confirmed that the bright-congruent and both incongruent conditions were significantly less accurate than the both-congruent, height-congruent, and unimodal conditions (all linear Tukey contrasts comparisons $p < 0.011$). Inaccurate responses were removed from 1,409 trials, leaving 12,227 trials.

Table 6 presents all possible congruency condition comparisons resulting from the LMM analysis for the brightness vs. height competition. As shown in Figure 7, the typical congruency advantage was again replicated ($b = 75.29$, $p < 0.001$), with faster responses in the both congruent condition than the both incongruent condition. Further, although the height-congruent/sharp-incongruent condition was significantly slower than the both congruent condition ($b = 36.14$, $p = 0.014$), it was also significantly faster than the both incongruent condition ($b = 39.15$, $p = 0.007$), whereas bright-congruent/height-incongruent condition was just as slow the both incongruent condition ($b = 9.58$, $p = 0.926$). There was also a significant effect of pitch difference ($b = -109.97$, $SE = 4.56$, $t = -24.14$, $p < 0.001$), meaning participants were slower to respond to target pitches that were closer in semitones to the comparison pitch.

As detailed in Supplementary material Section 5, there was a marginal congruency advantage when height was the relevant dimension, but no differences between pitch congruent and bright congruent conditions. When brightness was the relevant dimension, there was a significant congruency advantage, but there were no noticeable differences between the pitch congruent and height congruent conditions.

TABLE 6 Bright experiment; pitch-relevant condition.

Comparison	*Estimate*	*SE*	*z*	*p*
Congruent–bright congruent	−65.71	11.55	−5.69	<0.001
Height congruent–bright congruent	−29.57	11.59	−2.55	0.080
Incongruent–bright congruent	9.58	11.74	0.82	0.926
Unimodal–bright congruent	−2.59	11.62	−0.22	0.999
Height congruent–congruent	36.14	11.45	3.16	0.014
Incongruent–congruent	75.29	11.62	6.48	<0.001
Unimodal–congruent	63.12	11.47	5.50	<0.001
Incongruent–height congruent	39.15	11.65	3.36	0.007
Unimodal–height congruent	26.98	11.51	2.34	0.131
Unimodal–incongruent	−12.17	11.69	−1.04	0.836

Tukey's *post-hoc* test for multiple comparisons of RT values across congruency conditions.

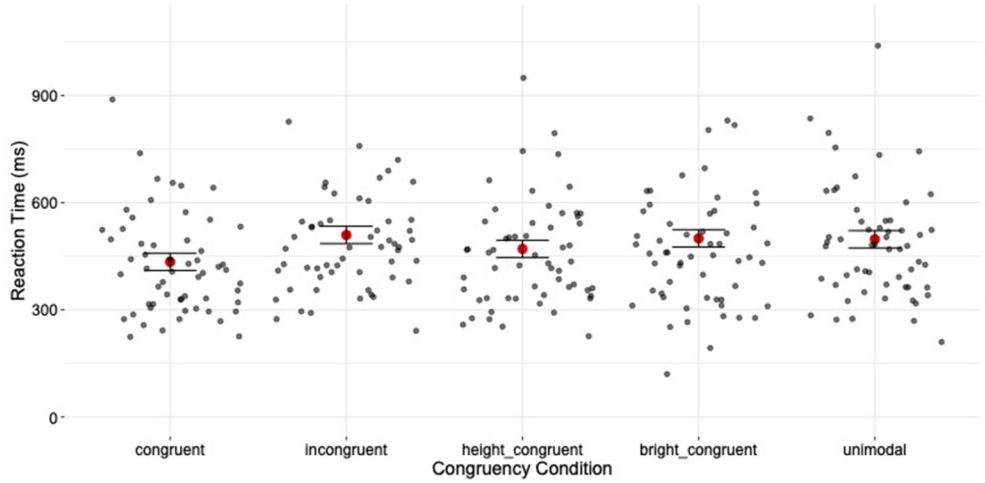

FIGURE 7
Bright experiment; Pitch-relevant task. Reaction times for the five congruency conditions. Participants responded significantly faster in the congruent and height congruent conditions than in the bright congruent, incongruent, and unimodal conditions. Error bars represent standard error. Individual points show each participant's average reaction time per condition.

3.5. Combined analysis

The RT data were modeled with a combined LMM with congruency (8 levels: congruent, incongruent, unimodal, height-congruent, size-congruent, sf-congruent, sharp-congruent, bright-congruent), dimension (4 levels: size, sharp, spatial frequency, bright), and their interaction as fixed effects as subject-by-subject variations as a random effect. There was a main effect of congruency, χ^2 (7, N = 221) = 139.02, $p < 0.001$. There was also a main effect of dimension, χ^2 (3, N = 221) = 11.22, p = 0.011, with slower responses in the sharp dimension than the spatial frequency (-97.38, SE = 33.22, p = 0.018) and bright (-101.20, SE = 31.95, p = 0.008) dimensions. However, there was no significant interaction, χ^2 (9, N = 221) = 11,21, p = 0.262, meaning the difference between congruency conditions was stable across visual dimensions. *Post-hoc* analyses showed that the congruent condition and the height congruent condition were both significantly faster than each of the visual (size, sharp, bright, spatial frequency) congruent conditions, which did not differ from the incongruent condition or from each other. These results mirror those of the individual competition results in sections 3.1–3.4 and highlight the superiority of the pitch-height correspondence over all other audiovisual correspondences tested.

4. Discussion

The current study created a competition between different audiovisual correspondences in order to investigate correspondence strength; namely, looking at the competition between height and each of four other visual components—size, brightness, sharpness, and spatial frequency—compared to the auditory dimension of pitch. Given prior research showing the importance of the pitch-height correspondence (c.f., Getz and Kubovy, 2018), I predicted a hierarchy of strength with congruency between pitch and height mattering more than congruency between pitch and any other visual dimension. As predicted on the pitch-relevant trials, participants were faster to respond when pitch was congruent with height (and incongruent with size, sharpness, spatial frequency, or brightness) and slower to respond when pitch was congruent with size, sharpness, spatial frequency, or brightness (and incongruent with height). For size, sharpness, and spatial frequency, the height congruent condition was also just as fast as when both dimensions were congruent. Although the height congruent condition was slower than the both congruent condition in the brightness experiment, in all four cases the height incongruent condition was also just as slow as the both incongruent condition. As shown in the Supplementary material, there were fewer congruency effects when responding to the visual dimensions, which is in line with previous work showing larger congruency effects with auditory-relevant than visual-relevant responding (Evans and Treisman, 2010; Jamal et al., 2017), but the effects present did match the hypothesis that congruency between pitch and height is more important than congruency between pitch and the other visual dimensions.

Together these results reinforce the superiority of the pitch-height correspondence (Spence, 2011; Parise, 2016; Cian, 2017) over other pitch-visual dimension correspondences. Although not the case in all languages (Eitan and Timmers, 2010), the metaphor used for pitch in English incorporates the same words "high" and "low" to describe visual elevations and auditory pitch height. Thus the pitch-height congruency effect shows the importance of semantic or linguistic factors in understanding this audiovisual correspondence, which is in line with an abundance of previous research (Melara and Marks, 1990; Gallace and Spence, 2006; Spence, 2011; Fernandez-Prieto et al., 2017).

However, it will be important for future research to continue investigating developmental comparisons in AVC congruency, as prior work has shown that the number and strength of such

associations increases over the lifespan (Speed et al., 2021). As a specific example, Dolscheid et al., 2022 investigated the emergence of the pitch-height and pitch-thickness associations in Dutch and Turkish five-, seven-, nine-, and 11-year-old children. They found that the pitch-thickness association was robust across ages and languages, whereas the pitch-height association was not reliable until age 11, and only then in Dutch speakers. Further, during a conflict task where children had to choose which association made more sense to describe high vs. low pitches (e.g., a thick line high in space vs. a thin line low in space paired with a high pitch would be congruent for height and incongruent for thickness), "children opted for thickness over height, regardless of language background" (Dolscheid et al., 2022, p. 9). Because this result shows the opposite effect of our results with adults, with pitch-height being a weaker correspondence in children and stronger correspondence in adults, it highlights the need for more developmental research to investigate the role of experience in understanding audiovisual correspondences.

Continuing to address the competition between correspondences is another important avenue for future research. Given that real-world objects often differ on multiple dimensions rather than just changing one visual or auditory feature at a time, it is interesting that the majority of AVC research has focused on changing single correspondence dimensions (Parise, 2016). One study that did investigate how correspondences interact was Jonas et al. (2017), who looked at how visual dimensions of luminance, saturation, size, and height jointly influenced pitch judgments. In their unspeeded classification task, they varied two, three, or four visual characteristics per trial in order to determine whether the correspondences had a summative or hierarchical effect on pitch judgments. They found more evidence for a summation model than the hierarchical model, though both models were significant with two and three features changing. Our current results seem to suggest a hierarchical model, with height accounting for more of the congruency advantage than size, sharpness, spatial frequency, or brightness. One difference between the two studies was the task, with Jonas et al. (2017) using an unspeeded classification and the current study using a speeded classification. Future research could more directly compare speeded to unspeeded judgments while changing different numbers of visual dimensions in order to better understand this complex multisensory environment. It may also be interesting for future studies to use more realistic stimuli than circles and sine tones, such as studies investigating how the pitch of a speaker's voice or the background music might affect the perception of an advertised product's size (e.g., Tran and Getz, submitted; Lowe and Haws, 2017) and to extend the competition paradigm beyond audiovisual correspondences.

In summary, the current study created a novel way of investigating the strength of cross-modal correspondences by putting them in competition with one another. Results show that pitch-height congruency is more important than the congruency between pitch and visual dimensions of size, sharpness, spatial frequency, and brightness, but more work is needed to look at the developmental origins of these audiovisual correspondences and how correspondences hold up in more naturalistic settings.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by University of San Diego Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LG designed and programmed the experiments, analyzed the data, and wrote the manuscript. Data collection was completed with the help of undergraduate research assistants at USD.

Funding

Completion of this manuscript was partially funded by a University of San Diego Faculty Research Grant to LG.

Acknowledgments

Special thanks to all of my Language and Music Perception (LAMP) Lab research assistants who helped with data collection for this project: Samantha Eason, Alexandra Griffin, Sarah Mann, Grace Masino, Hannah McIntosh, Victoria Nguyen, Kunal Patel, Makena Spencer, Kailey Taylor, and Jordaine Tran.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcogn.2023.1170422/full#supplementary-material

References

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Ben-Artzi, E., and Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Percept. Psychophys.* 57, 1151–1162. doi: 10.3758/BF03208371

Bien, N., ten Oever, S., Goebel, R., and Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG, and psychophysics study. *NeuroImage* 59, 663–672. doi: 10.1016/j.neuroimage.2011.06.095

Chen, Y. C., and Spence, C. (2017). Assessing the role of the 'unity assumption' on multisensory integration: A review. *Front. Psychol.* 8, 1–22. doi: 10.3389/fpsyg.2017.00445

Chiou, R., and Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception* 41, 339–353. doi: 10.1068/p7161

Chiou, R., and Rich, A. N. (2015). Volitional mechanisms mediate the cuing effect of pitch on attention orienting: The influences of perceptual difficulty and response pressure. *Perception* 44, 169–182. doi: 10.1068/p7699

Cian, L. (2017). Verticality and conceptual metaphors: A systematic review. *J. Assoc. Consumer Res.* 2, 444–459. doi: 10.1086/694082

Dolscheid, S., Çelik, S., Erkan, H., Küntay, A., and Majid, A. (2022). Children's associations between space and pitch are differentially shaped by language. *Dev. Sci.* e13341. doi: 10.1111/desc.13341 [Epub ahead of print].

Dolscheid, S., Hunnius, S., Casasanto, D., and Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychol. Sci.* 25, 1256–1261. doi: 10.1177/0956797614528521

Dolscheid, S., Shayan, S., Majid, A., and Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychol. Sci.* 24, 613–621. doi: 10.1177/0956797612457374

Eitan, Z., and 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, 405–422. doi: 10.1016/j.cognition.2009.10.013

Evans, K. K., and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *J. Vision* 10, 1–12. doi: 10.1167/10.1.6

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

Fernandez-Prieto, I., Spence, C., Pons, F., and Navarra, J. (2017). Does language influence the vertical representation of auditory pitch and loudness? *i-Perception,* 8, 2041669517716183. doi: 10.1177/2041669517716183

Gallace, A., and Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Percept. Psychophys.* 68, 1191–1203. doi: 10.3758/BF03193720

Getz, L. M., and Kubovy, M. (2018). Questioning the automaticity of audiovisual correspondences. *Cognition* 175, 101–108. doi: 10.1016/j.cognition.2018.02.015

Heller Murray, E. S., Hseu, A. F., Nuss, R. C., Harvey Woodnorth, G., and Stepp, C. E. (2019). Vocal pitch discrimination in children with and without vocal fold nodules. *Appl. Sci.* 9, 3042. doi: 10.3390/app9153042

Heron, J., Roach, N. W., Hanson, J. V., McGraw, P. V., and Whitaker, D. (2012). Audiovisual time perception is spatially specific. *Exper. Brain Res.* 218, 477–485. doi: 10.1007/s00221-012-3038-3

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometr. J.* 50, 346–363. doi: 10.1002/bimj.200810425

Jamal, Y., Lacey, S., Nygaard, L., and Sathian, K. (2017). Interactions between auditory elevation, auditory pitch and visual elevation during multisensory perception. *Multisens. Res.* 30, 287–306. doi: 10.1163/22134808-00002553

Jonas, C., Spiller, M. J., and Hibbard, P. (2017). Summation of visual attributes in auditory–visual crossmodal correspondences. *Psychon. Bull. Rev.* 24, 1104–1112. doi: 10.3758/s13423-016-1215-2

Keetels, M., and Vroomen, J. (2011). No effect of synesthetic congruency on temporal ventriloquism. *Attent. Percept. Psychophys.* 73, v209e218. doi: 10.3758/s13414-010-0019-0

Klapetek, A., Ngo, M. K., and Spence, C. (2012). Do crossmodal correspondences enhance the facilitatory effect of auditory cues on visual search? *Attent. Percept. Psychophys.* 74, 1154–1167. doi: 10.3758/s13414-012-0317-9

Lowe, M. L., and Haws, K. L. (2017). Sounds big: The effects of acoustic pitch on product perceptions. *J. Market. Res.* 54, 331–346. doi: 10.1509/jmr.14.0300

Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *J. Exper. Psychol.* 13, 384–394. doi: 10.1037/0096-1523.13.3.384

Martino, G., and Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception* 28, 903–923. doi: 10.1068/p2866

Mathôt, S., Schreij, D., and Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods* 44, 314–324. doi: 10.3758/s13428-011-0168-7

Maurer, D., Gibson, L. C., and Spector, F. (2012). "Infant synaesthesia: New insights into the development of multisensory perception," in *Multisensory development.* eds. A. J. Bremner, D. J. Lewkowicz, and C. Spence (Oxford, UK: Oxford University Press) 229–250. doi: 10.1093/acprof:oso/9780199586059.003.0010

McCormick, K., Lacey, S., Stilla, R., Nygaard, L. C., and Sathian, K. (2018). Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. *Neuropsychologia* 112, 19–30. doi: 10.1016/j.neuropsychologia.2018.02.029

Melara, R. D., and Marks, L. E. (1990). Processes underlying dimensional interactions: Correspondences between linguistic and nonlinguistic dimensions. *Memory Cognit.* 18, 477–495. doi: 10.3758/BF03198481

Melara, R. D., and O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *J. Exper. Psychol.* 116, 323–336. doi: 10.1037/0096-3445.116.4.323

Mondloch, C., and Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cogn. Affect. Behav. Neurosci.* 4, 133–136. doi: 10.3758/CABN.4.2.133

O'Boyle, M. W., and Tarte, R. D. (1980). Implications for phonetic symbolism: The relationship between pure tones and geometric figures. *J. Psycholinguistic Res.* 9, 535–544. doi: 10.1007/BF01068115

Parise, C., and Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE* 4, e5664. doi: 10.1371/journal.pone.0005664

Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisens. Res.* 29, 7–28. doi: 10.1163/22134808-00002502

Parkinson, C., Kohler, P. J., Sievers, B., and Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception* 41, 854–861. doi: 10.1068/p7225

Patching, G. R., and Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *J. Exper. Psychol.* 28, 755–775. doi: 10.1037/0096-1523.28.4.755

R Development Core Team (2022*). R: A language and environment for statistical computing [Computer software manual].* Vienna, Austria. Available online at: http://www.R-project.org/ (accessed March 10, 2022).

Spector, F., and Maurer, D. (2009). Synesthesia: A new approach to understanding the development of perception. *Develop. Psychol.* 45, 175–189. doi: 10.1037/a0014171

Speed, L. J., Croijmans, I., Dolscheid, S., and Majid, A. (2021). Crossmodal associations with olfactory, auditory, and tactile stimuli in children and adults. *i-Perception* 12, 20416695211048513. doi: 10.1177/20416695211048513

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attent. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7

Spence, C. (2022). Exploring group differences in the crossmodal correspondences. *Multisens. Res.* 35, 495–536. doi: 10.1163/22134808-bja10079

Spence, C., and Sathian, K. (2020). "Audiovisual crossmodal correspondences: Behavioral consequences and neural underpinnings," in *Multisensory Perception: From Laboratory to Clinic,* eds. K. Sathian and V. S. Ramachandran (London: Academic Press) 239–258. doi: 10.1016/B978-0-12-812492-5.00011-5

Tran, J., and Getz, L. (submitted). Attention required for advertisement pitch, tempo, and timbre to influence the perception of product features.

Welch, R. B., and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* 3, 638–667. doi: 10.1037/0033-2909.88.3.638