



Visually induced gains in pitch discrimination: Linking audio-visual processing with auditory abilities

Cecilie Møller^{1,2} · Andreas Højlund^{3,4} · Klaus B. Bærentsen¹ · Niels Chr. Hansen^{2,5} · Joshua C. Skewes⁴ · Peter Vuust²

Published online: 22 February 2018
© The Psychonomic Society, Inc. 2018

Abstract

Perception is fundamentally a multisensory experience. The principle of inverse effectiveness (PoIE) states how the multisensory gain is maximal when responses to the unisensory constituents of the stimuli are weak. It is one of the basic principles underlying multisensory processing of spatiotemporally corresponding crossmodal stimuli that are well established at behavioral as well as neural levels. It is not yet clear, however, how modality-specific stimulus features influence discrimination of subtle changes in a crossmodally corresponding feature belonging to another modality. Here, we tested the hypothesis that reliance on visual cues to pitch discrimination follow the PoIE at the interindividual level (i.e., varies with varying levels of auditory-only pitch discrimination abilities). Using an oddball pitch discrimination task, we measured the effect of varying visually perceived vertical position in participants exhibiting a wide range of pitch discrimination abilities (i.e., musicians and nonmusicians). Visual cues significantly enhanced pitch discrimination as measured by the sensitivity index d' , and more so in the crossmodally congruent than incongruent condition. The magnitude of gain caused by compatible visual cues was associated with individual pitch discrimination thresholds, as predicted by the PoIE. This was not the case for the magnitude of the congruence effect, which was unrelated to individual pitch discrimination thresholds, indicating that the pitch-height association is robust to variations in auditory skills. Our findings shed light on individual differences in multisensory processing by suggesting that relevant multisensory information that crucially aids some perceivers' performance may be of less importance to others, depending on their unisensory abilities.

Keywords Multisensory processing · Hearing

During normal waking consciousness, our perceptual experience is made up by an integrated unity of multiple differentiated sensory qualities (Edelman & Tononi, 2000).

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13414-017-1481-8>) contains supplementary material, which is available to authorized users.

✉ Cecilie Møller
cecilie@clin.au.dk

¹ Department of Psychology, Aarhus University, Aarhus, Denmark

² Center for Music in the Brain, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Nørrebrogade 44, 10G-4-03, DK-8000 Aarhus C, Denmark

³ Center of Functionally Integrative Neuroscience, Aarhus University Hospital, Aarhus, Denmark

⁴ Interacting Minds Centre, Aarhus University, Aarhus, Denmark

⁵ Cognitive and Systematic Musicology Laboratory, School of Music, Ohio State University, Columbus, OH, USA

Perception in everyday activities is oriented toward events and objects that have multisensory qualities and are perceived as integrated entities (e.g., Auvray & Spence, 2008; N. A. Bernstein, 1996). The ability to perceive multimodal entities depends on multisensory processing capabilities of the brain. Multisensory processing facilitates rapid detection and correct identification of objects and events in everyday life (Calvert, Spence, & Stein, 2004). A large body of literature has reported improved detection rates (Frassinetti, Bolognini, & Ladavas, 2002; Gescheider, Kane, Sager, & Ruffolo, 1974; Lovelace, Stein, & Wallace, 2003; Vroomen & Gelder, 2000) as well as faster and more accurate responses to multimodal stimuli than to unimodal stimuli (Alais, Newell, & Mamassian, 2010; Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002; Rowe, 1999). The intent of the present study is to add to this literature by quantifying visually induced gains in pitch discrimination in people with varying degrees of auditory sensitivity.

Several factors influence multisensory processing (Spence, 2007), ranging from low-level structural factors such as the spatiotemporal correspondence of the stimuli to be perceived

(Stein & Meredith, 1993) to high-level cognitive factors such as those involved in audio-visual object recognition (Molholm, Ritter, Javitt, & Foxe, 2004), semantic congruency (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004), and the associations formed through repeated exposure to common perceptual objects such as barking dogs and creaking doors (Chen & Spence, 2010). While most research has focused on the end points of this continuum of complexity, comparatively less is known about intermediate levels i.e., how multisensory processing is influenced by the experimental manipulation of modality-specific stimulus features (Doehrmann & Naumer, 2008; Laurienti et al., 2004) such as color or pitch, the feature of interest in the present study. Such features are often characterized as simple low-level features, yet they are different from amodal properties (duration, rhythm, intensity) in that they carry modal content (Lickliter & Bahrick, 2004) and they are likely influenced by and are themselves influencing higher level cognitive factors (Spence, 2011). Indeed, research on crossmodal correspondences has unequivocally demonstrated tight links between modality-specific features of multisensory objects. Of particular importance in the present context, people consistently map high-frequency sounds to objects positioned high in space (I. H. Bernstein & Edelstein, 1971; Evans & Treisman, 2010; Melara & O'Brien, 1987; Pratt, 1930; Stumpf, 1883), though pitch also maps onto a large number of other domains (Eitan & Timmers, 2010).

Whether intermediate-level and higher-level multisensory processing is influenced by the principles governing low-level multisensory processing is still under debate. One such principle is the principle of inverse effectiveness (henceforth, PoIE), which denotes stronger benefit from multimodal information when responses to the unimodal information are weak. As a case in point, most people recognize the situation where visual lip-reading enhances speech recognition in noisy environments and becomes more important with increasing levels of noise (Erber, 1975; Sumbly & Pollack, 1954). The PoIE was originally derived from single-neuron studies on the superior colliculus in cats (Meredith & Stein, 1986; Stein, Laurienti, Wallace, & Stanford, 2002; Stein & Meredith, 1993), and, interestingly, some support of its application to human behavioral responses has been reported as well, especially within the speech-perception literature (Albouy et al., 2015; Diederich & Colonius, 2004; Laurienti, Burdette, Maldjian, & Wallace, 2006; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007).

Nevertheless, conclusions on the direct applicability of the PoIE from single neurons to the complex level of human behavior must be drawn with due care. The superior colliculus is a structure engaged in simple detection and orientation behaviors, whereas, for example, speech is concerned with several higher-level cognitive processes, including semantic recognition (Ross et al., 2007). Despite the attractiveness of studying naturally occurring perceptual stimuli, results from such studies may be confounded by the added contribution of higher-

order semantic and linguistic features that interact with the basic sensory processes (Laurienti et al., 2004; Van Engen, Phelps, Smiljanic, & Chandrasekaran, 2014). Evidence that the PoIE applies to human behavioral responses is still sparse (Ross et al., 2007), and for psychology to progress as a cumulative science (Meehl, 1978), such evidence has to be built from the bottom up. Hence, investigations using highly controlled and simple yet perceptually relevant stimulus features rather than complex speech stimuli are an important and necessary piece of the puzzle.

Some human behavioral studies investigating responses to low-level stimulus dimensions report results that are consistent with the PoIE. Senkowski, Saint-Amour, Hofle, and Foxe (2011) found stronger audio-visual interactions for low-intensity stimuli, consistent with the PoIE. Caclin et al. (2011) found evidence that concurrent sounds (pink noise bursts) only improved visual (Gabor patches) detection thresholds in a subgroup of subjects exhibiting the poorest performance in the visual-only conditions. Despite its account for effects of sound on vision rather than vice versa, this latter finding is particularly interesting in the context of the present study. This is the case because the reported differences in performance between groups suggest that a given individual's perceptual gain from multimodal information can be predicted by his or her unisensory abilities. In order to further substantiate such a proposition, however, individual-level analyses of the correlations between unisensory abilities and multimodal gain rather than group comparisons are necessary.

Only limited attention has been given to visually induced enhancement of pitch discrimination despite the relevance of pitch perception to everyday tasks such as speech perception, music, and auditory scene analysis (Oxenham, 2012). A recent study claimed some support to behavioral level inverse effectiveness by showing larger visual facilitation of subtle pitch change detection in participants with amusia who have very poor auditory-only abilities, than in matched controls (Albouy et al., 2015). Analyses of reaction time data also revealed that the audio-visual benefit is related to task difficulty as it varied as a function of pitch interval size (50 cents, 25 cents, and 12.5 cents) and group, with no gains in conditions where the task was too simple for controls (50 cents) or too difficult for participants with amusia (12.5 cents). Importantly, however, the visual components of the audio-visual stimuli used in the experiment conveyed information about onsets and offsets of the tones but were uninformative with respect to changes in pitch. Thus, while a small number of studies have used inverse effectiveness to explain their behavioral data at the group level, none of these have investigated whether visually induced enhancements in pitch discrimination at the individual level are directly related to pitch discrimination thresholds.

Here, we aimed to characterize visually induced enhancements of subtle pitch discrimination in people with varying levels of pitch sensitivity. We recruited a subgroup of

professional musicians because musicians generally show superior sensitivity to pitch changes at the behavioral as well as neural level (Tervaniemi, Just, Koelsch, Widmann, & Schroger, 2005; Vuust, Brattico, Seppanen, Naatanen, & Tervaniemi, 2012). Vertical position was used as the visual feature, as its correspondence with auditory pitch, also known as the pitch-height association, is well established (Parise, Knorre, & Ernst, 2014; Parise, Spence, & Deroy, 2016). As such, the visual stimuli enabled us to mimic naturally occurring crossmodal correspondences—however, in a controlled setting—for the purpose of quantifying the beneficial contribution of relevant visual cues to pitch discrimination. By comparing responses to crossmodally corresponding trials against responses to trials with no visual cues and incongruent visual cues, respectively, two kinds of facilitatory effects constituting visually induced enhancements were quantified: (1) the gain associated with crossmodally congruent pairings of audio-visual stimuli compared to performance in a condition without visual cues, henceforth denoted bimodal compatibility gain (BCG), and (2) the difference in performance between conditions with crossmodally congruent and incongruent cues, henceforth denoted congruence effect (CE). Motivated by the findings that semantic (Laurienti et al., 2004) as well as crossmodal (Spence, 2011) congruence plays a key role in multisensory processing of *salient* changes in audio-visual stimuli, we hypothesized that crossmodally matching visual cues would also facilitate detection of the *subtle* pitch changes in our experiment, as seen in enhanced performance not only compared to trials without visual cues, the BCG, but also compared to trials with crossmodally mismatching cues, the CE. We also measured participants' pitch discrimination thresholds, as this allowed us to perform individual-level analyses in order to test the main hypothesis that higher (poorer) pitch discrimination thresholds are associated with stronger BCGs, as predicted by the PoIE (Stein & Meredith, 1993).

Method

This behavioral experiment was a separate part of a larger study that also included separate sessions of data acquisition using magnetoencephalography (MEG), magnetic resonance imaging (MRI), a test of pitch direction sensitivity, and two questionnaires. For clarity, only measures relevant to the hypotheses described above are included and analyzed here, that is, data from the behavioral experiment, the individual pitch threshold estimations, and the Musical Ear Test (Wallentin, Nielsen, Friis-Olivarius, Vuust, & Vuust, 2010), which measures musical skills.

Considering the novelty of the scope of the study, an estimate of optimal sample size through a priori power analyses based on any specific previous report would potentially be inaccurate and possibly misleading. Instead, the criterion used

to determine sample size was based on a comparison with studies highlighted in the introduction (Albouy et al., 2015; Caclin et al., 2011; Laurienti et al., 2004; Senkowski et al., 2011), where sample sizes between 11 and 34 yielded reasonable effect sizes. Hence, recruitment of ~50 participants was considered adequate for testing the main hypotheses while still allowing for exclusion of participants performing at chance and ceiling levels.

Participants

Forty-nine participants (32 nonmusicians, mean age = 23.9 years, $SD = 3.2$, 18 female, and 17 musicians, mean age = 24.1 years, $SD = 4.1$, eight female) volunteered to participate in the study. Thirteen of these participants (three nonmusicians, 10 musicians) performed at ceiling level at the largest (easiest) pitch level in the behavioral experiment. Ceiling performance may limit the size of the visually induced gain. This, in turn, could potentially bias the results in favor of the hypothesis of inverse effectiveness. Data from these 13 participants were therefore excluded. Hence, 36 participants (seven musicians) were included in the main statistical analyses reported here. As a consequence of the high exclusion rate within the group of musicians, the statistical power dropped to the extent that it was not possible to draw solid conclusions on between-group comparisons. Therefore, the group factor was omitted from the main analyses. A subsequent exploratory analysis focusing on responses to the smallest (most difficult) pitch level allowed reinclusion of six participants (all musicians) for this particular analysis, as these participants only performed at ceiling at the largest pitch level. This analysis is reported in the Supplementary Material and should still be interpreted with due care due to the limited statistical power.

All participants were right-handed, all reported normal or corrected-to-normal visual acuity and no hearing impairments. Musicians were full-time conservatory students or professional musicians. Nonmusicians had never received any formal music training other than mandatory primary school music lessons and had never played any kind of musical instrument, including singing on a regular basis. Participants gave their written consent before participation and received a taxable compensation of DKK 400 for participating in the full study, which took place on 2 separate days. The study protocol was approved by The Central Denmark Regional Committee on Health Research Ethics (Project ID: M-2014-52-14).

Stimuli and paradigm

Auditory stimuli were delivered via Sennheiser HDA 200 headphones at approximately 70 dB SPL. Visual stimuli were presented on a desktop computer screen with a refresh rate of 60 Hz. Using Presentation software (Neurobehavioral Systems Inc., Albany, CA, USA) the audio-visual stimuli were

presented with a stimulus onset asynchrony (SOA) of 800 ms in an oddball paradigm with 80% standards. To reduce predictability, the deviants were pseudorandomly presented with a minimum of three and a maximum of seven standards in between two deviants, that is, the distribution of the number of standards was centered on four, with a right skew, giving relatively fewer instances of six and seven standards. The standard stimulus consisted of a 523.25 Hz sinusoidal tone of 100 ms duration, including 5 ms fade in/out, followed by 700 ms interstimulus interval (ISI). The tone was coupled with an image of a light gray disc behind a cross in a static rectangle (see Fig. 1) centrally positioned on a computer screen. The duration of the visual stimuli was 800 ms (i.e., with no ISI). This method was preferred because pilot tests indicated that the perceptual salience of a flickering visual stimulus (i.e., one with a 700-ms ISI) much exceeded the perceptual salience of the excursion of the disc and thus added noise as well as discomfort for the participants.

Target deviants were two levels of pitch change, that is, 20 and 30 cents deviating in both directions (high and low). These four tones were coupled in all possible combinations with the image of the disc in three vertical positions: centrally, above, or below the center. Viewed from a distance of 60 cm, the approximate location of the participant, the disc subtended 3 degrees visual angle, and the displacement of the disc from the center was 0.5 degrees visual angle. This displacement was small enough to be perceived as an excursion of the disc, rather than as the sudden pop-up of a new disc, but large enough to be clearly visible. Following the pattern where higher/lower pitch corresponds to higher/lower vertical position, the audio-visual target deviants were two pitch levels of three categories: crossmodally matching visual cue (MC), where the auditory and visual components of the stimulus deviate in the same direction; crossmodally mismatching visual cue (MmC), where they deviate in opposite directions; and no visual cue (NVC), where only the auditory subcomponent deviates.

In addition to the target deviants, nontarget visual-only deviants were included in the paradigm to ensure that a change in visual position was not always associated with a pitch change. Based on our pilot experiment, we expected that participants with the lowest (best) pitch discrimination thresholds would perform at or near ceiling level on 30 cents deviants, whereas participants with the highest (worst) thresholds would perform at chance level on 20 cents deviants. Hence, to increase the sensitivity of the paradigm to both high and low performance levels, we included both levels of pitch change.

Procedure

Participants were seated in front of a computer screen in a sound-attenuated experimental lab with ambient lighting. All necessary instructions were presented in writing on the screen. Their task was to focus on the cross at the center of the screen and press the space bar as fast as possible without making mistakes whenever they detected a tone that deviated from the train of standard tones. Before each block, they were reminded to focus on the cross at the center of the screen. The experiment consisted of five blocks with 1-minute breaks in between and included four experimental blocks of 4 minutes 40-s duration and one auditory only (AUD) control block of 5 minutes 20-s duration. The AUD block was randomly presented in Position 2, 3, or 4 of the five blocks, and in this block the images on the screen were replaced by a fixation cross. Each of the four experimental blocks contained five instances of each of the 14 deviants (12 targets, i.e., with sound deviance; and two nontargets, i.e., visual only) as well as 280 (80%) standards. The AUD block contained 20 instances of each pitch level deviating in each direction as well as 320 standards. In this way, the total number of deviant trials (20 per trial type) was kept constant throughout the experiment. The duration of the experiment was approximately 30 minutes. A 1-minute training block preceding the experiment served to familiarize participants with the task. This was

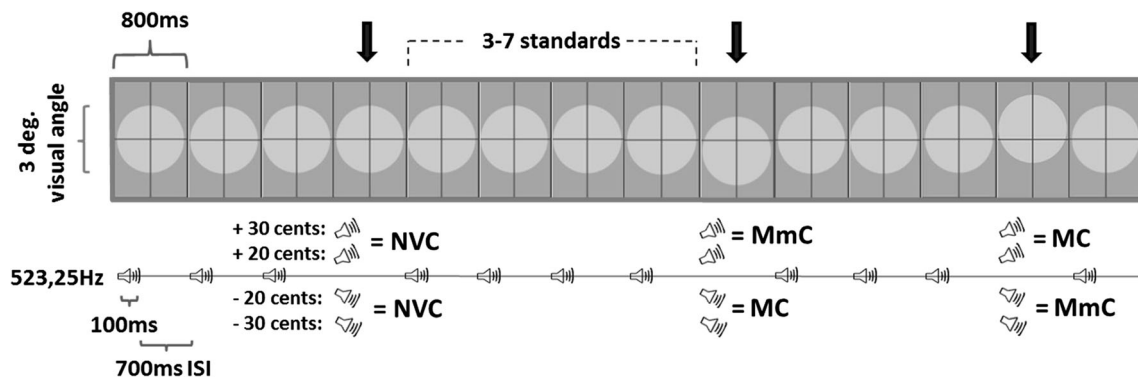


Fig. 1 Stimuli. Examples of audio-visual standards (no arrows) and target deviants (black arrows). A target deviant is any one of four possible changes in pitch: either 20 or 30 cents in both directions (i.e., either high or low). A pitch change is presented simultaneously with either no

visual cue (NVC), a crossmodally matching cue (MC) where the auditory and visual stimuli deviate in the same direction, or a crossmodally mismatching cue (MmC) where they deviate in opposite directions. ISI = interstimulus interval

identical to the four experimental blocks except that it included only one instance of each of the 14 deviants.

Pitch threshold estimation

On a separate day, 2 to 4 weeks after the experimental session, individual pitch discrimination thresholds (PDT) were estimated using a two-down, one-up adaptive staircase procedure that converges on the 70.7% performance level on the psychometric function (Levitt, 1971). The staircase was adapted from Williamson, Liu, Peryer, Grierson, and Stewart (2012) to match the stimuli and participants of the present study, and it employed a criterion-free AXB forced-choice task. The reference (X) was always a 523.25-Hz sinusoidal tone, and the task was to state whether the first (A) or the last (B) tone differed from the two other tones. The tones were 100 ms and the SOA was 400 ms. The staircase terminated after 14 reversals, and the threshold was calculated on the basis of an average of the last six reversals. The duration was approximately 3 to 5 minutes, depending on participants' responses.

Musical Ear Test

The melodic part of the Musical Ear Test (MET; Wallentin et al., 2010) was administered to assess the musical abilities of the participants in an auditory-only setting. This subtest employs a same–different task (i.e., participants are asked to judge whether two musical phrases are identical or not). There are 52 trials consisting of pairs of melodic piano phrases, each containing 3–8 tones. MET scores have been shown to correlate with results of musical imitation tests typically used in auditions for music conservatories (Wallentin et al., 2010).

Analyses and results

Preprocessing

IBM SPSS Statistics for Windows, Version 24, was used for preprocessing and for all analyses. Measures of hits and reaction time were collected from all trials in the experiment, and only responses between 200 ms and 1,000 ms after stimulus onset were included in the analysis (see justification in the Supplementary Materials, Fig. S1). Initial paired-samples *t* tests showed no significant differences between numbers of correct responses to high and low deviants within each pitch level of each category. Hence, the factor of pitch direction was omitted from the analysis, yielding 40 trials per condition (two pitch levels of MC, MmC, and NVC).

The sensitivity index d' (pronounced 'd-prime') was calculated for each participant, each condition, using the formula $d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$ (Green & Swets, 1966). Because hit rates of zero or 100% present a problem to the calculations of

Z , we first applied a standard correction that entailed adding 0.5 to the number of hits and false alarms and 1 to the total number of trials within each condition (Hautus, 1995). Ceiling performance was assessed by hit rates, that is, not corrected for false alarms. Data from participants who performed at ceiling level in conditions containing the largest deviant were excluded from the main analyses. As already stated, this was the case for three nonmusicians (all males) and 10 musicians (six males), who responded correctly to 100% of trials in the easiest condition (i.e., matching visual cue to 30 cents deviants). In the case of conditions containing the smallest pitch level (20 cents), only four ceiling performers (musicians, three males; all also performing at ceiling in the easiest condition) were identified and hence excluded from the exploratory individual level analyses focusing on 20 cents deviants only. These analyses are reported in Supplementary Materials. All analyses were performed using d' as the dependent measure.

Preliminary plotting of the data showed that increases in d' were associated with decreases in reaction time, indicating that there was no speed–accuracy trade-off. This was also the case in the control block. Control block analyses also revealed significant correlations between performance in the auditory only block and the PDT, and between performance in the auditory-only block and performance in the NVC trials in the experimental blocks (see Supplementary Materials).

Statistics

The statistical analysis was a two-step process: in the first step, the global experimental effect was assessed with a two-way repeated-measures ANOVA, with condition (matching visual cue [MC], mismatching visual cue [MmC], no visual cue [NVC]) and pitch level (20 cents, 30 cents) as within-subjects factors.

From these factors, we extracted the two variables necessary for further individual-level analysis: (1) bimodal compatibility gain (BCG) was quantified for each participant by subtracting the d' measured in the NVC condition from the d' measured in the MC condition, and (2) the congruence effect (CE) was quantified for each participant by subtracting the MmC from the MC condition. Both of these variables were calculated on the basis of the simple main effect of condition (i.e., the mean d' of the 20 and 30 cents deviants within each condition). In the second step, we ran two Pearson product-moment correlation analyses across all participants to determine whether individual PDTs were correlated with (1) the magnitude of the BCG in accordance with the PoIE and the main hypothesis, and (2) the magnitude of the CE.

Statistical significance was determined by the conventional alpha level of .05 (two-tailed). Uncorrected *p* values are reported. When applicable, Bonferroni-adjusted alpha levels were set to correct for multiple comparisons. For transparency, a three-way mixed ANOVA, which also included

musicianship as a between-subjects factor, is reported in the Supplementary Materials. Note, though, that because of the substantial group size differences (seven musicians and 29 nonmusicians) and associated differences in statistical power, caution should be taken when interpreting these results.

Two-way ANOVA results

Thirty-six participants (mean age = 23.9 years, $SD = 3.5$, 14 male), seven of which were professional musicians, were included in the analysis. Following Maxwell and Delaney (2004), Greenhouse–Geisser-corrected F values are reported and interpreted for all within-subjects effects whether or not the assumption of sphericity was met, according to Mauchly's test. Table 1 summarizes the descriptive statistics derived from each of the trial types.

The ANOVA revealed statistically significant main effects of both factors (i.e., pitch level and condition) and of the interaction between them. The main effect of pitch level, $F(1, 35) = 264.347$, $p < .001$, $\eta_p^2 = .883$, indicated that performance in the 30 cents conditions exceeded performance in the 20 cents conditions. The main effect of condition, $F(1.259, 44.070) = 97.259$, $p < .001$, $\eta_p^2 = .735$, was broken down by pairwise comparisons, that showed statistically significant differences in all comparisons with a Bonferroni-corrected significance level of $\alpha = .0167$: Relative to the trials that contained no visual cues, visual cues facilitated pitch discrimination significantly, whether the cue was crossmodally matching ($p < .001$) (this difference is henceforth denoted bimodal compatibility gain, or BCG) or crossmodally mismatching ($p < .001$). Furthermore, participants performed significantly better in the MC condition than in the MmC condition ($p = .008$) (this difference is henceforth denoted congruence effect, or CE).

Importantly, the main effect of condition should be interpreted in light of the statistically significant two-way

interaction that was found between condition and pitch level, $F(1.855, 64.916) = 8.298$, $p = .001$, $\eta_p^2 = .192$. Simple pairwise comparisons showed that although performance was better in trials with matching than with mismatching cues, this comparison constituting the congruence effect did not reach statistical significance at the largest pitch level (20 cents: $p = .001$; 30 cents: $p = .068$; see interaction plot in Supplementary Materials, Fig. S3). Statistically significant differences between conditions were found in all remaining comparisons at both pitch levels ($p < .001$).

Individual-level results

Two Pearson product-moment correlation analyses were run to determine whether individual PDTs were correlated with the CE and the BCG, respectively. PDT data were not obtained from two participants (both nonmusicians) who did not attend the final session of the study; thus, 34 (seven musicians) participants' data were included in these two correlation analyses. Problems with nonnormally distributed data points were solved by log-transforming the PDT values using the natural logarithm before running the parametric correlation analyses.

The correlation analyses revealed no statistically significant correlation between mean CE and PDT, $r = .044$, $n = 34$, $p = .805$. However, a statistically significant positive correlation was found between mean BCG and PDT, indicating that the larger (poorer) the thresholds, the larger the visually induced gain, $r = .602$, $n = 34$, $p < .001$. This is in accordance with the principle of inverse effectiveness (PoIE). Figure 2 shows the mean CE (top) and the mean BCG (middle) plotted against PDT.

To assess effects of musical skills, two Pearson product-moment correlation analyses were run to determine whether individual absolute scores (correct responses) on the melodic part of the Musical Ear Test

Table 1 Descriptive statistics for data included in the two-way ANOVA ($n = 36$)

Pitch level	No Visual Cue (NVC)	Mismatching Cue (MmC)	Matching Cue (MC)	BCG (MC – NVC)	CE (MC – MmC)
	d' , M (SD)	d' , M (SD)	d' , M (SD)	d' , M (SD)	d' , M (SD)
20 cents	1.74 (0.95)	2.45 (0.77)	2.60 (0.84)	0.86 (0.41)	0.15 (0.25)
30 cents	2.47 (1.07)	2.98 (0.80)	3.05 (0.77)	0.59 (0.51)	0.07 (0.23)
meanBCG / meanCE				0.72 (0.40)	0.11 (0.16)

Note. Descriptive statistics—mean d' (M) with standard deviations (SD) in parentheses—for responses to deviants within each condition, including the variables derived from them: the means across pitch levels of the bimodal compatibility gain (meanBCG) (mean MC minus mean NVC) and congruence effect (meanCE) (mean MC minus mean MmC) were fed into further analyses. The maximum attainable d' in this experiment was 5.57

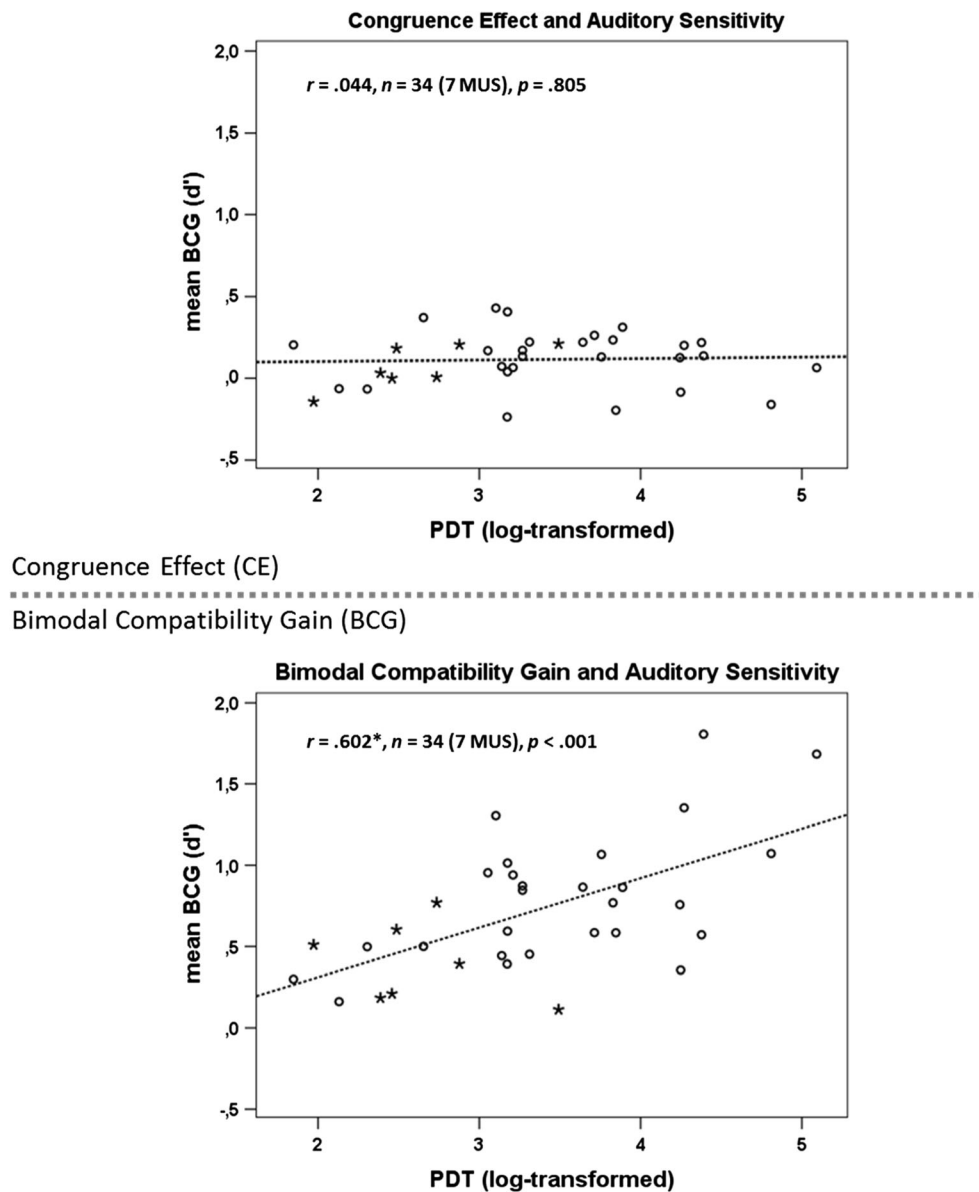


Fig. 2 Scatterplots show the congruence effect (CE, top) and bimodal compatibility gain (BCG, bottom) as a function of the log-transformed pitch discrimination thresholds (PDT). Pearson correlation analyses

showed that BCG and PDT were significantly correlated. Stars = musicians, open circles = nonmusicians, lines are fitted to all data points, ignoring musicianship

(MET) were correlated with the CE and the BCG, respectively. This analysis was run using the same 34 participants as the previous analysis. The correlation analyses revealed no statistically significant correlation between mean CE and MET, $r = .109, n = 34, p = .541$. However, a statistically significant negative correlation was found between mean BCG and MET (see Fig. 3), indicating that less advanced musical skills is associated with more benefit from visual cues in pitch discrimination, $r = .431, n = 34, p = .011$. There was a statistically significant negative correlation between PDT and MET, $r = -.563, n = 34, p = .001$, reflecting the association between auditory sensitivity and musical skills.

Discussion

This is the first study to investigate the effects of perceptually informative visual cues on subtle pitch change detection in participants with varying levels of pitch discrimination thresholds. Visual cues caused significant improvements in performance, whether the cue was crossmodally matching or mismatching. A correlation analysis revealed larger bimodal compatibility gains (BCG) in participants with poorer pitch discrimination thresholds. This is in accordance with the principle of inverse effectiveness (PoIE) and implies that the realm of this principle may be extended from the single neuron

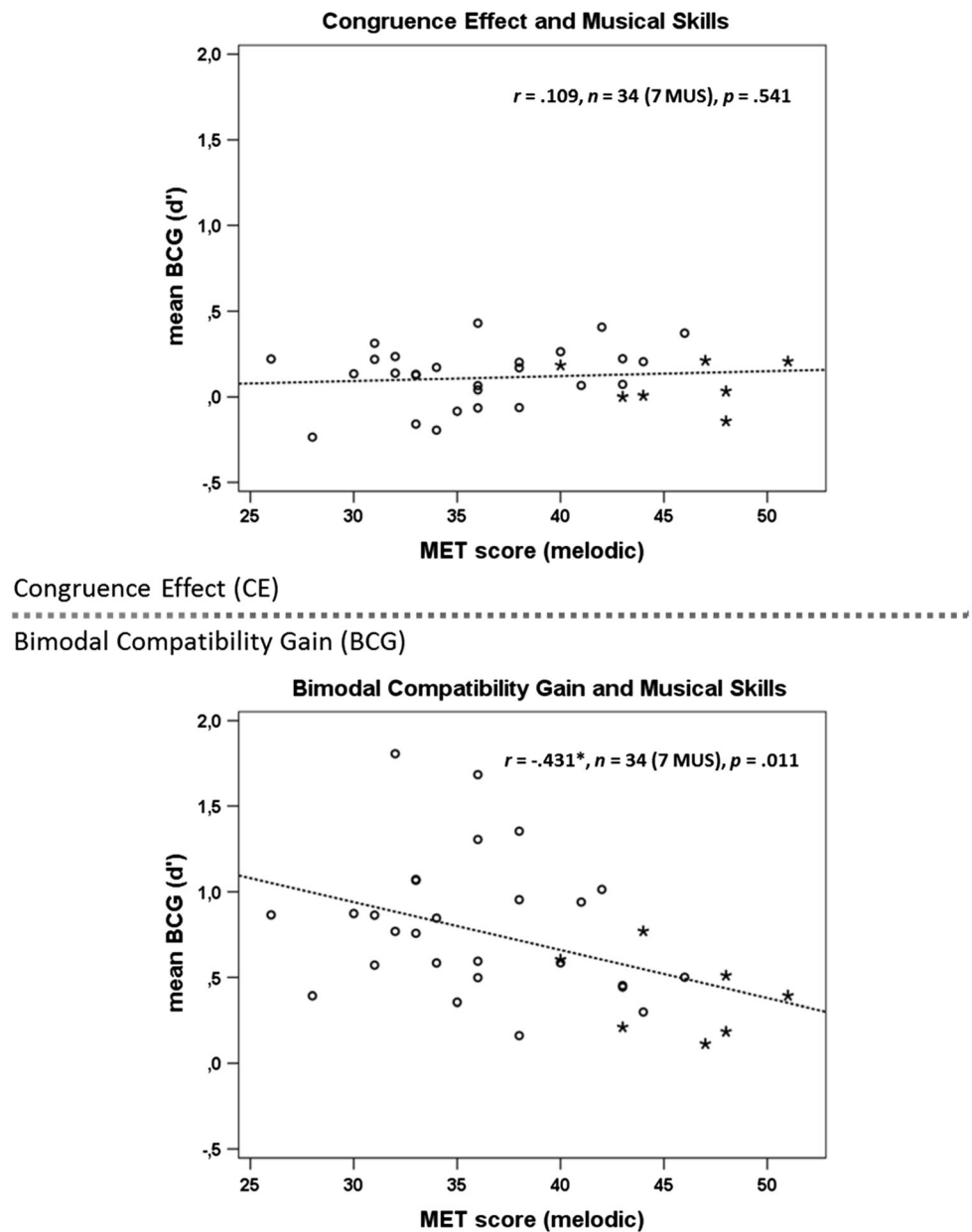


Fig. 3 Scatterplots show the congruence effect (CE, top) and bimodal compatibility gain (BCG, bottom) as a function of the absolute score (correct responses) on the melodic part of the Musical Ear Test (MET).

Pearson correlation analyses showed that BCG and MET were statistically significantly correlated. Stars = musicians, open circles = nonmusicians, lines are fitted to all data points, ignoring musicianship

within-subject scale to also include a behavioral interindividual scale. Similarly, larger gains were associated with poorer performance on a measure of musical abilities, the melodic part of the Musical Ear Test (MET) (Wallentin et al., 2010), indicating more reliance on visual cues with less advanced musical skills. We also found a significant congruence effect (CE), that is, better performance in the matching than in the mismatching condition. This indicates that the association between a high-pitch note and a visually perceived object positioned high in space contributes to visually induced gains in pitch discrimination, and more so than when the mapping is reversed. The BCG is therefore not only a result of increased

vigilance due to the change of the stimulus (orienting reflex) but also a directed influence related to the signal relationship of the crossmodally corresponding stimuli. Correlation analyses revealed that the CE is robust to variations in pitch discrimination thresholds as well as to variations in scores on the MET.

The functional relevance of overlapping sensory systems is obvious not only in clinical cases of sensory substitution (Proulx, Brown, Pasqualotto, & Meijer, 2014), but also in everyday life of the general population. This becomes particularly clear whenever information in one sensory modality is compromised (e.g., in darkness, noisy conditions, or when hearing declines at older age; Laurienti et al., 2006). The

PoIE describes one of the basic mechanisms underlying multisensory integration that is well established at the neurophysiological level in animals (Stein & Meredith, 1993). There is also convincing evidence that simple spatial and temporal relationships have a major influence on whether crossmodal features are combined to produce behavioral benefits at the detection threshold in humans (Bolognini, Frassinetti, Serino, & Ládavas, 2005; Frassinetti et al., 2002). In contrast, much less is known about the benefits of a close correspondence between stimulus contents (Doehrmann & Naumer, 2008; Laurienti et al., 2004), and this gap is even more pronounced with respect to simple features such as pitch, despite a longstanding and now rapidly growing interest in feature-based crossmodal correspondences (Parise et al., 2016). Furthermore, the field is still in its infancy when it comes to assessing the applicability of the PoIE to human behavior (Albouy et al., 2015; Caclin et al., 2011; Ross et al., 2007; Senkowski et al., 2011).

Our study contributes to a closing of this gap by showing how perception of even a low-level feature, such as pitch, is modulated by crossmodally corresponding visual cues. This modulation cannot solely be attributed to spatiotemporal correspondence of the auditory and visual stimulus streams, since we furthermore provide evidence that crossmodal congruence affects the magnitude of the visually induced gain (larger gain in the condition with matching than with mismatching stimulus pairs). Furthermore, our study highlights the link between unisensory abilities and multisensory processing by showing that individual pitch discrimination abilities are associated with the magnitude of the BCG (larger gains with higher [worse] pitch discrimination thresholds) in accordance with the PoIE.

The place to look for a more formalized model of this finding could be in Bayesian theories of optimal multisensory integration (see, e.g., Angelaki, Gu, & DeAngelis, 2009; Deneve & Pouget, 2004; Ernst, 2012). The reason for this is that we can assume that people with lower PDTs have a more precise representation of pitch, and when they have that they are less influenced by information from other modalities. This is consistent with behavioral results using computational modeling to show that musicians demonstrate more certain pitch expectations (governed by lower degrees of entropy) than nonmusicians, both when assessed explicitly and when assessed with more indirect, implicit methods (Hansen & Pearce, 2014; Hansen, Vuust, & Pearce, 2016). An influential model of Bayesian optimal integration was provided by Ernst and Banks (2002) and describes how evidence from the two modalities are weighted according to their reliability in order to minimize the variance of the final percept. Indeed, this is consistent with the PoIE (Rowland, 2012). However, while it may be conceptually useful to interpret the present results within a Bayesian framework, the paradigm used in the present study does not lend itself to strict interpretation within such a formalized model. Most importantly, our participants' task was not to estimate an amodal but a *modality-specific* feature (i.e., pitch)

using two sensory modalities (i.e., audition and vision). In this case only one of the sensory modalities (audition) is capable of responding to sound waves and hence to specify the target feature (pitch). The model by Ernst and Banks (2002) was developed with the purpose of determining sensory dominance, specifically, assessing the degree to which vision or haptics dominates in estimating height, the target feature in their task. This is meaningful in the context of amodal feature detection, where both modalities have direct access to the target feature. However, as explained above, this is not the case here.

We interpret the observed congruence effect as a confirmation that the correspondence between auditory pitch and visually perceived vertical position not only modulates responses to very salient pitch changes, as has most often been investigated experimentally (I. H. Bernstein & Edelstein, 1971; Evans & Treisman, 2010; Lu, Ho, Sun, Johnson, & Thompson, 2016; Orchard-Mills, Van der Burg, & Alais, 2015; Patching & Quinlan, 2002) but is also relevant in the context of subtle pitch change detection. Our results show that in the case of subtle pitch changes, congruence is equally effective in modulating pitch discrimination performance irrespective of participants' pitch discrimination and musical abilities in separate auditory-only tasks. The finding that the congruence effect in the present study is unrelated to individual pitch discrimination thresholds and musical skills is noteworthy considering previous studies that demonstrate musicians' increased sensitivity to pitch-height congruency (Eitan & Granot, 2006; Paraskevopoulos, Kraneburg, Herholz, Bamidis, & Pantev, 2015).

This discrepancy could be attributed to the relative contribution of semantic influences in the present and previous studies and may as such be seen as a valuable addition to the discussion regarding the underlying mechanisms and developmental trajectories of crossmodal correspondences (Parise, 2016). Whether and how preattentive, perceptual, semantic, and/or decisional mechanisms contribute to the perceptual consequences of crossmodal correspondences is still debated, and it is not unlikely that several stages of processing influence the crossmodal mapping in question here (Spence, 2011). Though a pitch-verticality mapping has been found in preverbal 3–4-month-old infants (Walker et al., 2010), postperceptual semantic influences undoubtedly play a role as well (Eitan & Timmers, 2010). Within the behavioral paradigm reported here, it is not possible to determine at which level of processing the congruence relations assert their influence. However, a number of steps were taken in an effort to specifically avoid enhancing semantic influences on participants' responses. In contrast to previous studies (Paraskevopoulos et al., 2015; Paraskevopoulos, Kuchenbuch, Herholz, & Pantev, 2012), our participants were naïve to the purpose of the study (i.e., to our interest in investigating the putatively beneficial effect of relevant visual information on pitch change detection). This was achieved through the use of a task unrelated to audio-visual congruency and one which did not require focus on

the direction of pitch changes. Furthermore, we referred to “changes in the tones” rather than “high/low tones” in all encounters with the participants before and during the experiment.

Semantic influences may well explain, for example, why musicians in previous studies have shown increased sensitivity to pitch-height congruency (Eitan & Granot, 2006; Paraskevopoulos et al., 2015), possibly owing to their increased familiarity with musical notation. Indeed, many previous studies have used musical notation or closely resembling visual stimuli; and, not surprisingly, in their familiar domain, musicians have shown increased behavioral benefits compared to nonmusicians (Abel, Li, Russo, Schlaug, & Loui, 2016; Nichols & Grahn, 2016; Paraskevopoulos et al., 2015; Paraskevopoulos et al., 2012). It is indeed plausible that considerable experience with musical notation and terminology results in an association between pitch and vertical position that is quite explicit in musicians, and that this familiarity effect becomes apparent whenever the pitch intervals can actually be represented by musical notation, resulting in musicians’ stronger associations between vertical position and pitch as found in previous studies. However, owing to the focus on the PoIE, the present study targeted visual influences on responses to near-threshold auditory stimuli (20 and 30 cents deviants) as opposed to musical scale intervals (1+ semitones), where musicians may benefit from their explicit knowledge about pitch intervals and their labels (Abel et al., 2016; Nichols & Grahn, 2016). Therefore, the pitch intervals could not be represented by musical notation in the present study. In combination with the explicit steps taken to reduce the semantic bias, the subtle nature of the pitch deviants to be detected here may have produced results that reflect processing at a low-level perceptual stage rather than semantic and/or decisional level responses.

One could raise the potential concern that the visual stimuli in the present experiment acted as a mere warning signal, or that the participants did not follow the instructions given but simply either closed their eyes or responded to the changes in visual rather than auditory components of the audio-visual stimuli. Because we did not video-record participants during the experiment, we cannot rule out that the results were affected by nonconforming participants. However, the noise that this may have induced in the data was not sufficiently strong to eliminate the presence of a significant congruence effect, which indicates that performance was modulated by the specific content of the visual stimulus. Therefore, our findings cannot simply be explained by the increased physical salience of the crossmodally matching stimuli compared to the stimuli containing no visual cue, nor by an increased level of participants’ arousal in response to those stimuli.

Despite the inclusion of a subgroup of professional musicians, this study was not designed specifically for assessing group differences with respect to the PoIE, and such

differences were not part of the initial hypotheses. However, exploratory analyses reported in the Supplementary Materials hint to the notion that a positive correlation between BCG and PDT may be found only in nonmusicians. In other words, it may be the case that musicians are not prone to behavioral level inverse effectiveness when detecting subtle pitch changes. We wish to emphasize, though, that the results of these exploratory analyses should be interpreted with caution. This is the case not only because absence of evidence is not evidence of absence but also because the substantial group size differences cause reasonable concerns about the reliability of the reported group difference. Our further exploratory analysis on *n*-matched and PDT-matched groups (see Supplementary Materials, Fig. S5) showed complementary results, suggesting that power differences alone may not account for the exploratory findings reported here. However, while this analysis may have tackled the statistical issues, preselecting participants within the groups based on their PDTs may come at the expense of representativeness. This may even be questioned in the present study because data from the best performing participants within each group had to be excluded due to ceiling performance, making it unclear to what extent the participants included in the analyses fully represent the nonmusician and musician population, respectively. Therefore, the present data do not support further discussions of potential between-group differences. A wider range of pitch deviance levels are advised in future studies aimed at assessing whether the interaction between the PoIE and musicianship is genuine.

Conclusion

Perception in everyday activities is aided by a wealth of structured multisensory information, some of which is used by the perceiver and some of which is of less importance or even ignored. This study shows that the magnitude of gain in pitch discrimination caused by compatible visual cues is directly associated with pitch discrimination thresholds, in accordance with the principle of inverse effectiveness, and that the pitch-height association modulates the size of visually induced gains. The idea that perception may depend in systematic ways on unisensory abilities may inspire future research to focus more evenly on controlling the properties of the stimuli presented to participants as controlling the characteristics of the participants themselves.

Acknowledgements We thank Zohar Eitan for much valued inputs in the design stage of the experiment, Sukhbinder Kumar and Victoria Williamson for sharing the original pitch threshold estimation scripts, and Signe Hagner for very competent help with data collection. This project has been supported by seed funding from the Interacting Minds Centre, AU, DK. Center for Music in the Brain is funded by the Danish National Research Foundation (DNRF117).

References

- Abel, M. K., Li, H. C., Russo, F. A., Schlaug, G., & Loui, P. (2016). Audiovisual interval size estimation is associated with early musical training. *PLOS ONE*, *11*(10), e0163589. <https://doi.org/10.1371/journal.pone.0163589>
- Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving*, *23*(1), 3–38. <https://doi.org/10.1163/187847510X488603>
- Albouy, P., Leveque, Y., Hyde, K. L., Bouchet, P., Tillmann, B., & Caclin, A. (2015). Boosting pitch encoding with audiovisual interactions in congenital amusia. *Neuropsychologia*, *67*, 111–120. <https://doi.org/10.1016/j.neuropsychologia.2014.12.006>
- Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: Psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology*, *19*(4), 452–458. <https://doi.org/10.1016/j.conb.2009.06.008>
- Auvray, M., & Spence, C. (2008). The multisensory perception of flavor. *Consciousness and Cognition*, *17*(3), 1016–1031. <https://doi.org/10.1016/j.concog.2007.06.005>
- Bernstein, I. H., & Edelman, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, *87*(2), 241–247.
- Bernstein, N. A. (1996). On dexterity and its development. In M. L. Latash & M. T. Turvey (Eds.), *Dexterity and its development* (pp. 3–237). Mahwah, NJ: Erlbaum.
- Bolognini, N., Frassinetti, F., Serino, A., & Ladavas, E. (2005). “Acoustical vision” of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, *160*(3), 273–282. <https://doi.org/10.1007/s00221-004-2005-z>
- Caclin, A., Bouchet, P., Djoulah, F., Pirat, E., Pernier, J., & Giard, M. H. (2011). Auditory enhancement of visual perception at threshold depends on visual abilities. *Brain Research*, *1396*, 35–44. <https://doi.org/10.1016/j.brainres.2011.04.016>
- Calvert, G., Spence, C., & Stein, B. E. (Eds.). (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.
- Chen, Y., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389–404. <https://doi.org/10.1016/j.cognition.2009.10.012>
- Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology—Paris*, *98*(1), 249–258. <https://doi.org/10.1016/j.jphysparis.2004.03.011>
- Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*(8), 1388–1404.
- Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*, *1242*, 136–150. <https://doi.org/10.1016/j.brainres.2008.03.071>
- Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York, NY: Basic Books.
- Eitan, Z., & Granot, R. Y. (2006). How music moves: Musical parameters and listeners images of motion. *Music Perception: An Interdisciplinary Journal*, *23*(3), 221–248. <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>
- Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders*, *40*(4), 481. <https://doi.org/10.1044/jshd.4004.481>
- Ernst, M. O. (2012). Optimal multisensory integration: Assumptions and limits. In B. E. Stein (Ed.), *The new handbook of multisensory processing* (pp. 527–543). Cambridge, MA: MIT Press.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. <https://doi.org/10.1038/415429a>
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 6.1–6. <https://doi.org/10.1167/10.1.6>
- Forster, B., Cavina-Pratesi, C., Aglioti, S. M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research*, *143*(4), 480–487. <https://doi.org/10.1007/s00221-002-1017-9>
- Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*(3), 332–343. <https://doi.org/10.1007/s00221-002-1262-y>
- Gescheider, G. A., Kane, M. J., Sager, L. C., & Ruffolo, L. J. (1974). The effect of auditory stimulation on responses to tactile stimuli. *Bulletin of the Psychonomic Society*, *3*(3), 204–206.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.
- Hansen, N. C., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, *5*, 1052. <https://doi.org/10.3389/fpsyg.2014.01052>
- Hansen, N. C., Vuust, P., & Pearce, M. (2016). “If you have to ask, you’ll never know”: Effects of specialised stylistic expertise on predictive processing of music. *PLOS ONE*, *11*(10), e0163584. <https://doi.org/10.1371/journal.pone.0163584>
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. <https://doi.org/10.3758/BF03203619>
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, *158*(4), 405–414. <https://doi.org/10.1007/s00221-004-1913-2>
- Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging*, *27*(8), 1155–1163. <https://doi.org/10.1016/j.neurobiolaging.2005.05.024>
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2), 467–477.
- Lickliter, R., & Bahrick, L. E. (2004). Perceptual development and the origins of multisensory responsiveness. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 643–654). Cambridge, MA: MIT Press.
- Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: A psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research*, *17*(2), 447–453. [https://doi.org/10.1016/S0926-6410\(03\)00160-5](https://doi.org/10.1016/S0926-6410(03)00160-5)
- Lu, X., Ho, H. T., Sun, Y., Johnson, B. W., & Thompson, W. F. (2016). The influence of visual information on auditory processing in individuals with congenital amusia: An ERP study. *NeuroImage*, *135*, 142–151. <https://doi.org/10.1016/j.neuroimage.2016.04.043>
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective*. Mahwah, NJ: Erlbaum.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*(4), 806–834. <https://doi.org/10.1037/0022-006X.46.4.806>
- 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>
- Meredith, A. M., & Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, 365(2), 350–354. [https://doi.org/10.1016/0006-8993\(86\)91648-3](https://doi.org/10.1016/0006-8993(86)91648-3)
- Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex (New York, NY: 1991)*, 14(4), 452–465.
- Nichols, E. S., & Grahn, J. A. (2016). Neural correlates of audiovisual integration in music reading. *Neuropsychologia*. <https://doi.org/10.1016/j.neuropsychologia.2016.08.011>
- Orchard-Mills, E., Van der Burg, E., & Alais, D. (2015). Crossmodal correspondence between auditory pitch and visual elevation affects temporal ventriloquism. *Perception*, 45(4), 409–424. <https://doi.org/10.1177/0301006615622320>
- Oxenham, A. J. (2012). Pitch perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(39), 13335–13338. <https://doi.org/10.1523/JNEUROSCI.3815-12.2012>
- Paraskevopoulos, E., Kraneburg, A., Herholz, S. C., Bamidis, P. D., & Pantev, C. (2015). Musical expertise is related to altered functional connectivity during audiovisual integration. *Proceedings of the National Academy of Sciences of the United States of America*, 112(40), 12522–12527. <https://doi.org/10.1073/pnas.1510662112>
- Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Musical expertise induces audiovisual integration of abstract congruency rules. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(50), 18196–18203. <https://doi.org/10.1523/JNEUROSCI.1947-12.2012>
- Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research*, 29(1/3), 7–28.
- Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences of the United States of America*, 111(16), 6104–6108. <https://doi.org/10.1073/pnas.1322705111>
- Parise, C. V., Spence, C., & Deroy, O. (2016). Understanding the correspondences: Introduction to the special issue on crossmodal correspondences. *Multisensory Research*, 29(1/3), 1–6. <https://doi.org/10.1163/22134808-00002517>
- Patching, G. R., & Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception and Performance*, 28(4), 755–775. <https://doi.org/10.1037/0096-1523.28.4.755>
- Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, 13(3), 278.
- Proulx, M. J., Brown, D. J., Pasqualotto, A., & Meijer, P. (2014). Multisensory perceptual learning and sensory substitution. *Neuroscience and Biobehavioral Reviews*, 41, 16–25. <https://doi.org/10.1016/j.neubiorev.2012.11.017>
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex (New York, NY: 1991)*, 17(5), 1147–1153. <https://doi.org/10.1093/cercor/bhl024>
- Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour*, 58(5), 921–931. <https://doi.org/10.1006/anbe.1999.1242>
- Rowland, B. A. (2012). Commentary: Computational models of multisensory integration: Bayesian frameworks, development, and timing. In B. E. Stein (Ed.), *The new handbook of multisensory processing* (pp. 559–511). Cambridge, MA: MIT Press.
- Senkowski, D., Saint-Amour, D., Hofle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage*, 56(4), 2200–2208. <https://doi.org/10.1016/j.neuroimage.2011.03.075>
- Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, 28(2), 61–70.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception & Psychophysics*, 73(4), 971–995. <https://doi.org/10.3758/s13414-010-0073-7>
- Stein, B. E., Laurienti, P. J., Wallace, M. T., & Stanford, T. R. (2002). Multisensory integration. In V. S. Ramachandran (Ed.), *Encyclopedia of the human brain* (pp. 227–241). New York, NY: Academic Press.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.
- Stumpf, C. (1883). *Tonpsychologie*. I. Leipzig, Germany: Hirzel.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26(2), 212. <https://doi.org/10.1121/1.1907309>
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schroger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, 161(1), 1–10. <https://doi.org/10.1007/s00221-004-2044-5>
- Van Engen, K. J., Phelps, J. E., Smiljanic, R., & Chandrasekaran, B. (2014). Enhancing speech intelligibility: Interactions among context, modality, speech style, and masker. *Journal of Speech, Language, and Hearing Research : JSLHR*, 57(5), 1908–1918. <https://doi.org/10.1044/JSLHR-H-13-0076>
- Vroomen, J., & Gelder, B. D. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5), 1583–1590. <https://doi.org/10.1037/0096-1523.26.5.1583>
- Vuust, P., Brattico, E., Seppänen, M., Naatanen, R., & Tervaniemi, M. (2012). The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia*, 50(7), 1432–1443. <https://doi.org/10.1016/j.neuropsychologia.2012.02.028>
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21(1), 21–25. <https://doi.org/10.1177/0956797609354734>
- Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The musical ear test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, 20(3), 188–196. <https://doi.org/10.1016/j.lindif.2010.02.004>
- Williamson, V. J., Liu, F., Peryer, G., Grierson, M., & Stewart, L. (2012). Perception and action de-coupling in congenital amusia: Sensitivity to task demands. *Neuropsychologia*, 50(1), 172–180.