Perception, 2012, volume 41, pages 854 – 861

doi:10.1068/p7225

Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population

Carolyn Parkinson[1], Peter Jes Kohler[1], Beau Sievers[2], Thalia Wheatley[1]§
[1]Department of Psychological and Brain Sciences, Dartmouth College, 6207 Moore Hall, Hanover, NH 03755, USA; e-mail: thalia.p.wheatley@dartmouth.edu; [2]McIntire Department of Music, University of Virginia, 112 Old Cabell Hall, PO Box 400176, Charlottesville, VA 22904, USA
Received 31 January 2012, in revised form 6 July 2012

Abstract. Associations between auditory pitch and visual elevation are widespread in many languages, and behavioral associations have been extensively documented between height and pitch among speakers of those languages. However, it remains unclear whether perceptual correspondences between auditory pitch and visual elevation inform these linguistic associations, or merely reflect them. We probed this cross-modal mapping in members of a remote Kreung hill tribe in northeastern Cambodia who do not use spatial language to describe pitch. Participants viewed shapes rising or falling in space while hearing sounds either rising or falling in pitch, and reported on the auditory change. Associations between pitch and vertical position in the Kreung were similar to those demonstrated in populations where pitch is described in terms of spatial height. These results suggest that associations between visual elevation and auditory pitch can arise independently of language. Thus, widespread linguistic associations between pitch and elevation may reflect universally predisposed perceptual correspondences.

Keywords: universality, cross-modal correspondences, audiovisual interactions, auditory pitch, conceptual metaphor, spatial representation, SMARC effect

1 Introduction

The tendency to map auditory pitch onto visual elevation comprises one of the most robust cross-modal correspondences reported to date (Spence 2011). However, while mappings between pitch and elevation are widely reflected in behavior (eg Ben-Artzi and Marks 1995; Casasanto 2010; Evans and Treisman 2010; Melara and O'Brien 1987; Patching and Quinlan 2002; Pratt 1930; Rusconi et al 2006), they are also reflected in most languages (Stumpf 1883), rendering it difficult to determine whether this association stems from cultural convention or from deeper similarities in processing. To address this, we probed cross-modal mappings between pitch and height among members of a remote Kreung tribe in L'ak, Cambodia (figure 1) who describe "high" and "low" frequency sounds as "tight" and "loose", respectively.

The Kreung language is mutually unintelligible with the Cambodian national language, Khmer, and with other regional tribal languages. The cultural isolation of L'ak is further preserved geographically: L'ak is only accessible to outsiders by road during a brief, annual dry season. Participants completed a speeded classification task (Marks 2004) in which they reported whether a tone played through headphones was ascending or descending in pitch while viewing an animated red ball rise or fall on a computer screen.

Consistent with a cross-modal association, participants made fewer errors classifying pitch while viewing congruent motion (eg ascending pitch with rising ball; $M_{\text{error rate}} = 0.08$; SD = 0.14) compared to incongruent motion ($M_{\text{error rate}} = 0.15$; SD = 0.21), ($t_{31} = 2.09$, $p < 0.05$, $d = 0.38$). An analogous paradigm probed size–volume associations, which were

§Corresponding author.

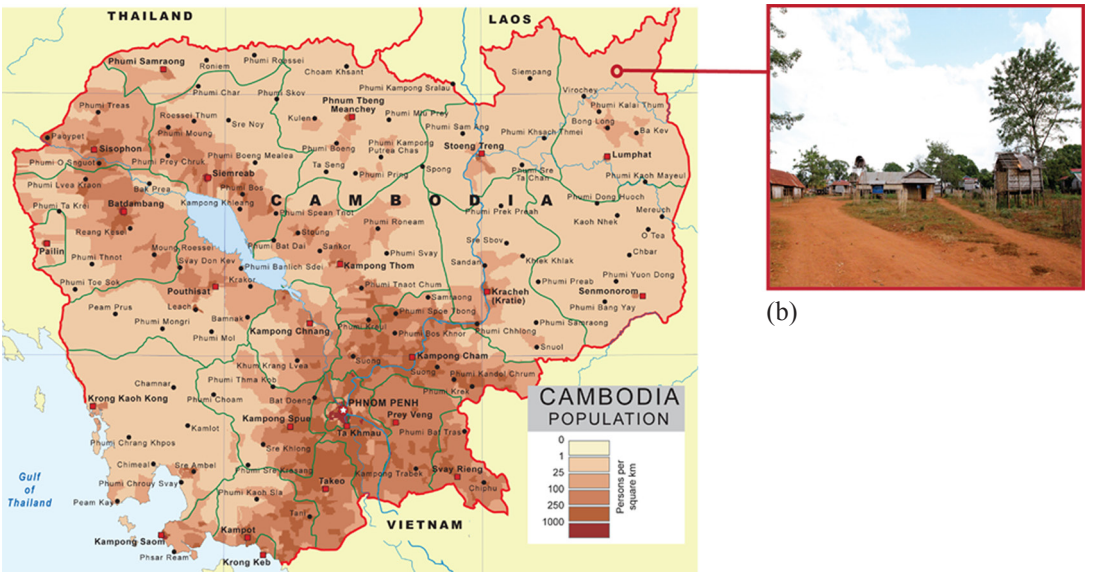

Figure 1. [In color online.] The experiment was conducted in the geographically, culturally, and linguistically isolated village of L'ak in Ratanakiri, Cambodia. (a) Population density map of Cambodia; L'ak is indicated by the [red] open circle. (b) Photograph of L'ak.

hypothesized to be universal because they may involve intensity matching (Stevens 1957) and are systematically coupled in the environment (Spence 2011). Again, the pattern of errors revealed a cross-modal association with lower mean error rates on congruent trials ($M_{\text{error rate}} = 0.05$; SD = 0.06) than on incongruent trials ($M_{\text{error rate}} = 0.10$; SD = 0.11), ($t_{31} = 2.57$, $p \leqslant 0.01$, $d = 0.52$). See figure 2.

The difference in participants' error rates on congruent and incongruent trials did not differ between blocks ($t_{31} = 0.61$, $p = 0.55$, $d = 0.14$), suggesting that the strength of volume–size and pitch-elevation associations were not markedly different in this sample.

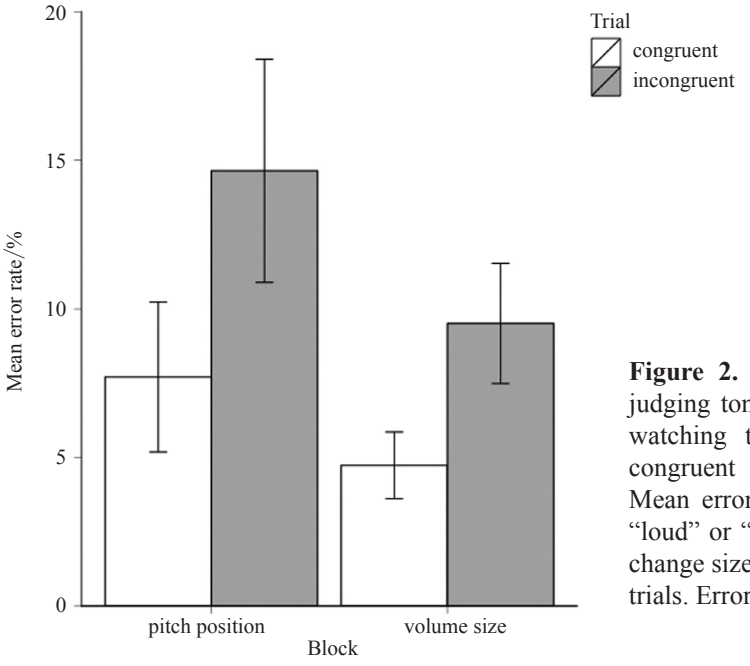

Figure 2. Left: Mean error rates while judging tones as "tight" or "loose" while watching the ball change position for congruent and incongruent trials. Right: Mean error rates while judging tones as "loud" or "quiet" while watching the ball change size for congruent and incongruent trials. Error bars indicate ± 1 SEM.

Additionally, reaction times (RTs) were not significantly different for correct responses to congruent and incongruent trials for either the pitch/position ($t_{31} = 0.45$, $p = 0.66$, $d = 0.05$) or volume/size blocks ($t_{31} = 0.71$, $p = 0.48$, $d = 0.06$).

In the past, it has been unclear whether behavioral associations between "higher" visual elevations and "higher" auditory pitches inform language, or merely reflect it, because natural couplings between these qualities are inconsistent. Often, "higher" visual positions and "lower" auditory pitches are associated in the environment. For example, in accordance with the Doppler effect, sounds emanating from an object *falling* from above *ascend* in pitch as they near the perceiver, whereas sounds emitted from an object travelling upwards, away from the perceiver, *descend* in pitch. Additionally, adult humans are generally taller ("higher") than children, yet have lower-pitched voices. Children are generally shorter ("lower") than adults, yet have higher-pitched voices. Other times, "high" pitch and "high" elevation are associated. For instance, the human larynx descends when producing lower pitched vocalizations, and rises while vocalizing with a higher pitch. Similarly, a body's resonant frequency depends on its mass, and heavier creatures are less likely to fly (Spence 2011). However, associations between "high" pitches and elevations are observed in infants (Walker et al 2010) likely to have more experience with other humans than with large land-bound mammals or small birds.

Although infants' pitch–position associations cannot be readily explained by couplings in nature, they could be explained by couplings in culturally determined aspects of their perceptual environment. Adult caregivers use infant-directed speech that evinces their own, culturally influenced, associations (Nygaard et al 2009; Stern et al 1982), and spontaneously couple higher pitch with rising vertical motion and lower pitch with falling vertical motion through speech and gestures (Shintel et al 2006). Infants are highly sensitive to conditional environmental probabilities (Aslin et al 1998), able to learn even arbitrary cross-modal mappings within short experimental sessions (Bahrick 2002). Thus, even very young infants may acquire the implicit, culturally determined associations of their caregivers. Therefore, although findings from preverbal infants strongly suggest that pitch–height mappings may not depend on language or other cultural conventions, the most stringent test of the pre-linguistic basis of these associations entails probing them in a culture without conventional pitch–height mappings. Data from the Kreung indicate that their mappings between pitch and position, and between size and volume, are the same as those observed in Western samples (Marks 2004).

Why particular mappings arise consistently across cultures remains an open question. Why does pitch seem to intuitively "go with" vertical space, rather than horizontal space, or even saltiness or softness? It cannot be an artifact of language in the case of the Kreung, because, linguistically, they do not map rising pitch onto rising elevation. Infant-synaesthesia theory (Spector and Maurer 2009) offers one explanation. Cortical connections present at birth may universally foster certain intersensory associations that influence language, perception, and associative learning. Intersensory associations not reflective of regular environmental couplings and that decrease during infant development have recently been demonstrated in infants (Wagner and Dobkins 2011), and mappings between auditory pitch and visual brightness have recently been documented in chimpanzees (Ludwig et al 2011). Thus, associations between auditory pitch and visual elevation may reflect innately predisposed mappings that arise independently of cultural learning, and that drive systematic associations between these qualities in many cultures (eg in language, musical notation, and prosody).

A second explanation is the "recycling" of cortical maps (Dehaene and Cohen 2007) or "exaptation" (Gould and Vrba 1982) hypothesis. Evolutionarily recent functions may reuse existing biological mechanisms (here, neural circuitry), a process suggested to underlie literacy, mathematics, tool use, and abstract reasoning (Dehaene and Cohen 2007; Srinivasan and Carey 2010). Like mapping number onto space (Dehaene et al 2008), the general tendency

to map pitch onto space appears to be a cross-cultural intuition that arises independently of language. This intuition may be the result of innate quirks of intersensory connectivity (infant-synaesthesia theory), or may confer evolutionary benefit (cortical recycling theory). Recycling cortical circuitry for spatial computations to operate on new domains of knowledge may support precise, flexible encoding and manipulations of this information (Bueti and Walsh 2009; Cantlon et al 2009). For pitch, the most obvious evolutionary benefit of this process is our capacity to imbue and extract meaning from an indeterminately large possibility space of tonal sequences (Krumhansl 1990). While other species communicate using pitch cues, these tend to be stereotyped and reliant on *absolute* pitch level. In contrast, nuanced and *relational* pitch patterns characterize melody and prosody across human history and cultures (Trehub 2003).

Such an explanation is consistent with suggestions that similar distance effects when comparing numbers and auditory pitches (ie longer RTs for more similar stimuli) reflect the recruitment of a domain-general sensorimotor transformation for both tasks (Cohen Kadosh et al 2008), and with the finding that amusia, characterized by deficits in fine-grained pitch discrimination and processing pitch sequences, is associated with impaired visuospatial abilities and a lack of interference effects between pitch and vertical space (Douglas and Bilkey 2007). Similarly, musicians show above-average visuospatial reasoning abilities (Sluming et al 2007), most pronounced in the vertical dimension (Brochard et al 2004). Additionally, while pitch primarily signifies identity (Van Dommelen 1990), posterior parietal areas devoted to sound localization (Zatorre et al 2002) process pitch, especially for tasks involving manipulations of tonal patterns (Foster and Zatorre 2010), further suggesting that resources initially for spatial computations were co-opted to process the complex, relational and evolutionarily recent (Trehub 2003) tonal patterns that characterize prosody and melody.

Interestingly, whereas the specific orientation of number–space mappings depends on culturally variable aspects of spatial experience (eg reading direction—Zebian 2005), the direction of pitch–space mappings among the Kreung was consistent with those that have been previously documented among Westerners (ie "higher" location with "higher" pitch). The direction of these associations may be learned from universalities in physical experience while speaking: the human larynx rises when producing vocalizations of a higher fundamental frequency and descends while vocalizing at a lower fundamental frequency. Similarly, raising the fundamental frequency of one's voice often involves tensing the intrinsic laryngeal muscles, whereas lowering one's voice often involves relaxing these muscles (Atkinson 1978). Thus, tension and elevation may become systematically associated with each other and with auditory pitch through implicitly learned associations while vocalizing.

Previously, it has been difficult to determine the basis of spatial metaphors for pitch as past work was conducted in samples where culturally learned metaphors color the perceptual environment through sound–meaning associations in language, gestures, and prosodic cues (Nygaard et al 2009; Shintel et al 2006). The current results suggest that, like number (Andres et al 2004), associations between pitch and other dimensions are not limited to those mappings that are reflected in linguistic metaphors. More specifically, pitch appears to be mapped onto vertical space in the absence of shared verbal labels for visual elevation and auditory pitch. This mapping may stem from aspects of physiology that are universal to all humans. That is, consistent pitch–space mappings across cultures may arise from implicitly learned associations between the position and tension of one's laryngeal muscles and the fundamental frequency of subsequent vocalizations. Such a mapping could also stem from how the brain represents this information, in line with both infant-synaesthesia theory and cortical recycling theory. While future work may further clarify the physical basis of this mapping, the current results suggest that it arises independently of language or other cultural conventions.

2 Methods

2.1 *Participants*

Participants were thirty-two inhabitants (twenty-four female) of L'ak, a village of approximately three hundred Kreung tribe members in the sparsely populated province of Ratanakiri, northeastern Cambodia (see figure 1). L'ak is inaccessible to outsiders for most of the year. During the brief dry season, the monolingual tribe remains isolated by virtue of their tribal language (Kreung) that is mutually unintelligible with other regional languages. Visits were facilitated by English–Khmer translators and with the Ratanakiri Department of Culture who provided Khmer–Kreung translators.

The Kreung do not formally document age; participants ranged in age from late adolescence to older adults. Parental consent was obtained for participants believed to be younger than 18 years. Participants were asked not to discuss the experiments until after a village-wide debriefing session, and were compensated with local currency equivalent of a full day of farm labor. To compensate the village, a donation was made through the NGO Ockenden Cambodia to fund the construction of a water well.

2.2 *Task*

Participants completed a speeded classification task in which they were instructed to identify sounds heard through headphones while viewing animations of a rising or falling ball presented on a 13 inch laptop running SuperLab 4.0. The experiment consisted of two blocks: volume/size and pitch/position. Block order was counterbalanced across participants. Each block contained 32 trials lasting 2.5 s each. There were 8 repetitions of each audiovisual pairing, yielding 16 congruent and 16 incongruent trials per block.

As participants were unfamiliar with computers, vocal responses were used to assess RT and accuracy. Before each block, translators explained the task, showed the two possible variations of the target stimulus and elicited the appropriate verbal responses (the Kreung word for the relevant stimulus dimension for that trial—"taut"/"loose" for pitch changes; "loud"/"quiet" for volume changes). The Kreung describe loud and soft sounds as "*khlang*" and "*khsaoy*", and big and small shapes as "*tih*" and "*keh*", respectively. High and low auditory pitches are referred to as "*nyang*" and "*you*", while relatively high and low elevations are called "*jrung*" and "*dap*", respectively. RTs were measured with the computer's internal microphone. To assess accuracy, an experimenter blind to trial content transcribed responses, which were later compared to SuperLab output, noting trials when animals or other ambient noises triggered the microphone, or when vocal responses were insufficiently loud to trigger the microphone. Across participants, 15 out of 1024 pitch/position trials were discarded and 40 of 1024 size/volume trials were discarded.

2.3 *Stimuli*

In the pitch/position block, participants viewed a red ball 65 pixels in diameter against a gray background. The ball remained stationary in the center of the screen for 1 s, then moved 300 pixels upwards or downwards at a constant rate for 1.5 s. Simultaneously, a synthesized tone with high-frequency formants was played at a fundamental frequency of 261.63 Hz (C4) for 1 s, then ascended or descended to 523.25 Hz (C5) or 130.81 Hz (C3). See figure 3.

In the volume/size block, a red ball 65 pixels in diameter, remained stationary on a gray background for 1 s, then contracted or expanded by 32 pixels at a constant rate until the end of the trial. Simultaneously, a 261.63 Hz tone played at a constant volume for 1 s, then increased or decreased in volume by 9.1 dB at a constant rate (figure 4).

Lacking individual equal-loudness curves for participants and a controlled listening environment, an average equal-loudness curve was approximated using a parametric equalizer to decrease the volume of frequencies around 3000 Hz by 4.5 dB.

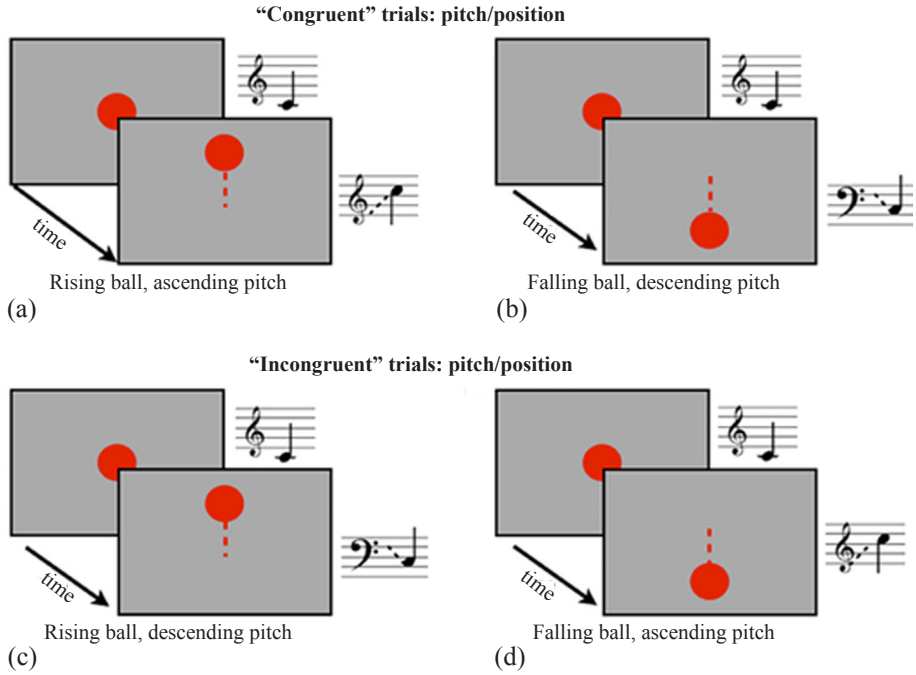

Figure 3. [In color online.] Schematic diagrams of congruent (a, b) and incongruent (c, d) pitch/position trials. The first second of every trial was identical: the ball did not move and a constant 261.63 Hz (middle C) tone played. For the remaining 1.5 s of the trial, the ball either ascended or descended at a constant rate while the frequency of the tone either increased or decreased at a constant rate until one octave higher (523.25 Hz) or lower (130.81 Hz) than the initial frequency. Dashed lines indicate the path of auditory and visual stimulus changes.

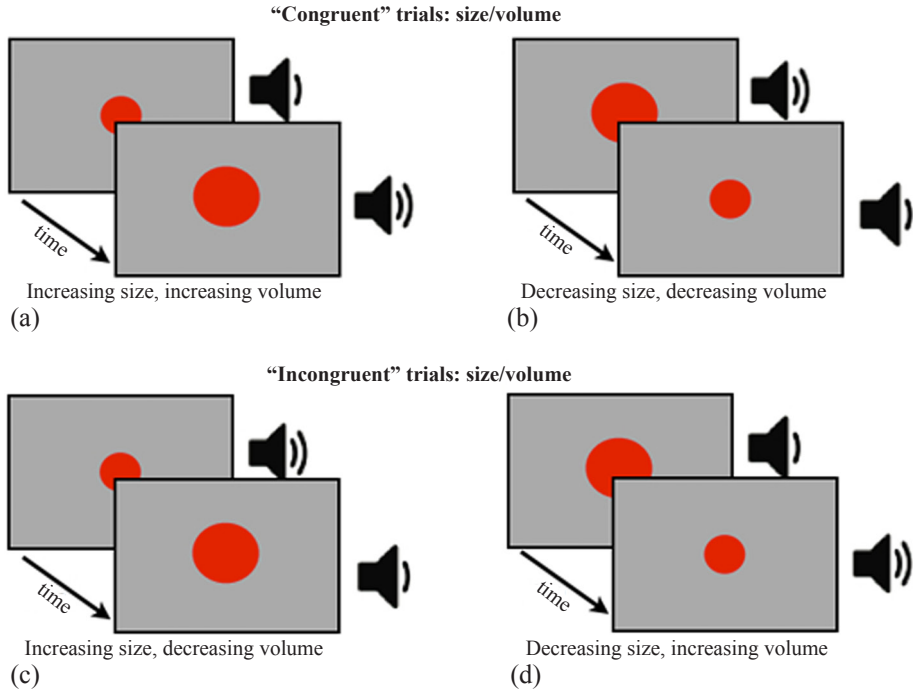

Figure 4. [In color online.] Schematic diagrams of congruent (a, b) and incongruent (c, d) volume/size trials. The first second of every trial was identical: the ball did not move and a constant 261.63 Hz (middle C) tone played. For the remaining 1.5 s of the trial, the ball either increased or decreased in size at a constant rate while the volume of the tone either increased or decreased at a constant rate until 9.1 dB louder or quieter than the initial volume.

Acknowledgments. We thank the Ratanakiri Department of Culture, Ockenden Cambodia and Cambodia Living Arts for facilitating visits to L'ak and for assistance with Khmer–Kreung translation, as well as Trent Walker for English–Khmer translation. This research was supported in part by a McNulty Grant from the Nelson A Rockefeller Center (TW), a Foreign Travel Award from the John Sloan Dickey Center for International Understanding (TW), and a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship (CP).

References

Andres M, Davare M, Pesenti M, Olivier E, Seron X, 2004 "Number magnitude and grip aperture interaction" *NeuroReport* **15** 2773–2777

Aslin R N, Saffran J R, Newport E L, 1998 "Computation of conditional probability statistics by 8-month-old infants" *Psychological Science* **9** 321–324

Atkinson J E, 1978 "Correlation analysis of the physiological factors controlling fundamental voice frequency" *Journal of the Acoustical Society of America* **63** 211–222

Bahrick L E, 2002 "Generalization of learning in three-and-a-half-month-old infants on the basis of amodal relations" *Child Development* **73** 667–681

Ben-Artzi E, Marks L E, 1995 "Visual–auditory interaction in speeded classification: role of stimulus difference" *Perception & Psychophysics* **57** 1151–1162

Brochard R, Dufour A, Depres O, 2004 "Effect of musical expertise on visuospatial abilities: evidence from reaction times and mental imagery" *Brain and Cognition* **54** 103–109

Bueti D, Walsh V, 2009 "The parietal cortex and the representation of time, space, number and other magnitudes" *Philosophical Transactions of the Royal Society of London, B* **364** 1831–1840

Cantlon J F, Platt M L, Brannon E M, 2009 "Beyond the number domain" *Trends in Cognitive Sciences* **13** 83–91

Casasanto D, 2010 "Space for thinking", in *Language, Cognition and Space: the State of the Art and New Directions* Eds V Evans, P Chilton (London: Equinox) pp 453–478

Cohen Kadosh R, Brodsky W, Levin M, Henik A, 2008 "Mental representation: what can pitch tell us about the distance effect" *Cortex* **44** 470–477

Dehaene S, Cohen L, 2007 "Cultural recycling of cortical maps" *Neuron* **56** 384–398

Dehaene S, Izard V, Spelke E, Pica P, 2008 "Log or linear? Distinct intuitions of the number scale in western and Amazonian indigene cultures" *Science* **320** 1217–1220

Douglas K M, Bilkey D K, 2007 "Amusia is associated with deficits in spatial processing" *Nature Neuroscience* **10** 915–921

Evans K K, Treisman A, 2010 "Natural cross-modal mappings between visual and auditory features" *Journal of Vision* **10** 1–12

Foster N E V, Zatorre R J, 2010 "A role for the intraparietal sulcus in transforming musical pitch information" *Cerebral Cortex* **20** 1350–1359

Gould S, Vrba E, 1982 "Exaptation—A missing term in the science of form" *Paleobiology* **8** 4–15

Krumhansl C, 1990 *Cognitive Foundations of Musical Pitch* (New York: Oxford University Press)

Ludwig V U, Adachi I, Matsuzawa T, 2011 "Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans" *Proceedings of the National Academy of Sciences of the USA* **108** 20661–20665

Marks L E, 2004 "Cross-modal interactions in speeded classification", in *Handbook of Multisensory Processes* Eds G Calvert, C Spence, B E Stein (Cambridge, MA: MIT Press) pp 85–106

Melara R D, O'Brien T P, 1987 "Interaction between synesthetically corresponding dimensions" *Journal of Experimental Psychology: General* **116** 323–336

Nygaard L C, Herold D S, Namy L L, 2009 "The semantics of prosody: acoustic and perceptual evidence of prosodic correlates to word meaning" *Cognitive Science* **33** 127–146

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** 755–775

Pratt C C, 1930 "The spatial character of high and low tones" *Journal of Experimental Psychology* **13** 278–285

Rusconi E, Kwan B, Giordano B L, Umiltà C, Butterworth B, 2006 "Spatial representation of pitch height: the SMARC effect" *Cognition* **99** 113–129

Shintel H, Nusbaum H C, Okrent A, 2006 "Analog acoustic expression in speech communication" *Journal of Memory and Language* **55** 167–177

Sluming V, Brooks J, Howard M, Downes J J, Roberts N N, 2007 "Broca's area supports enhanced visuospatial cognition in orchestral musicians" *Journal of Neuroscience* **27** 3799–3806

Spector F, Maurer D, 2009 "Synesthesia: a new approach to understanding the development of perception" *Developmental Psychology* **45** 175–189

Spence C, 2011 "Crossmodal correspondences: a tutorial review" *Attention, Perception, & Psychophysics* **73** 971–995

Srinivasan M, Carey S, 2010 "The long and the short of it: On the nature and origin of functional overlap between representations of space and time" *Cognition* **116** 217–241

Stern D N, Spieker S, MacKain K, 1982 "Intonation contours as signals in maternal speech to prelinguistic infants" *Developmental Psychology* **18** 727–735

Stevens S S, 1957 "On the psychophysical law" *Psychological Review* **64** 153–181

Stumpf C, 1883 *Tonpsychologie* (Leipzig: S Hirzel)

Trehub S E, 2003 "The developmental origins of musicality" *Nature Neuroscience* **6** 669–673

Van Dommelen W, 1990 "Acoustic parameters in human speaker recognition" *Language and Speech* **33** 259–272

Wagner K, Dobkins K R, 2011 "Synaesthetic associations decrease during infancy" *Psychological Science* **22** 1067–1072

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** 21–25

Zatorre R J, Bouffard M, Ahad P, 2002 "Where is 'where' in the human auditory cortex?" *Nature Neuroscience* **5** 905–909

Zebian S, 2005 "Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the REVERSE SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers" *Journal of Cognition and Culture* **5** 165–190