

Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans

Vera U. Ludwig^{a,b,c,1,2}, Ikuma Adachi^{d,1,2}, and Tetsuro Matsuzawa^d

^aDepartment of Psychiatry and Psychotherapy, Charité - Universitätsmedizin Berlin, D-10117 Berlin, Germany; ^bBerlin School of Mind and Brain, Humboldt-Universität zu Berlin, D-10117 Berlin, Germany; ^cDepartment of Psychology, Humboldt-Universität zu Berlin, D-12489 Berlin, Germany; and ^dPrimate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan

Edited* by Nikos K. Logothetis, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, and approved November 3, 2011 (received for review August 28, 2011)

Humans share implicit preferences for certain cross-sensory combinations; for example, they consistently associate higher-pitched sounds with lighter colors, smaller size, and spikier shapes. In the condition of synesthesia, people may experience such cross-modal correspondences to a perceptual degree (e.g., literally *seeing* sounds). So far, no study has addressed the question whether nonhuman animals share cross-modal correspondences as well. To establish the evolutionary origins of cross-modal mappings, we tested whether chimpanzees (*Pan troglodytes*) also associate higher pitch with higher luminance. Thirty-three humans and six chimpanzees were required to classify black and white squares according to their color while hearing irrelevant background sounds that were either high-pitched or low-pitched. Both species performed better when the background sound was congruent (high-pitched for white, low-pitched for black) than when it was incongruent (low-pitched for white, high-pitched for black). An inherent tendency to pair high pitch with high luminance hence evolved before the human lineage split from that of chimpanzees. Rather than being a culturally learned or a linguistic phenomenon, this mapping constitutes a basic feature of the primate sensory system.

multisensory integration | audio-visual correspondences | comparative cognition | sound symbolism | language evolution

Humans share systematic, implicit preferences for pairing sounds with visual sensations (1, 2); for example, they consistently choose higher-pitched sounds to better fit lighter colors (3–5), smaller size (4, 6), spikier shape (7), and locations higher in space (7). Such phenomena are already observed in young children (4, 6–10), and they are also reflected in language. For example, the term *dunkler Ton* (“dark sound”) in German refers to a low-pitched sound, and both *kiroi koe* (“yellow voice”) in Japanese and *voces blancas* (“white voices”) in Spanish describe high-pitched voices. Some individuals may even experience cross-modal associations to a perceptual degree; for example, they may literally *see* sounds (5). This neurodevelopmental condition is known as *synesthesia* (11–13).

Cross-modal correspondences have been extensively described and analyzed, given their possible relevance for understanding the mechanisms of perception. However, so far nothing is known about when and why such correspondences (e.g., high pitch = bright) evolved. Associations might be mediated by language, cultural learning, and higher cognitive processes (1, 14), for example, by exposure to metaphors during development. If this is true for all associations, then humans might be the only species that show cross-modal correspondences and synesthesia. However, if cross-modal correspondences are a more basic, ancient feature of the perceptual system, then other animals should show such associations as well.

Here we tested whether chimpanzees (*Pan troglodytes*) share cross-modal correspondences with humans. We assessed pitch–

luminance mapping (i.e., the mapping of high pitch with high luminance), which is very well established in humans (3–6, 15). This association exists in both human nonsynesthetes and synesthetes (5); that is, although individual synesthetes experience different colors for sounds (e.g., a given sound might be green for one person and red for another), the lightness of the elicited color is strongly correlated with the pitch of the sound inducing the color (5). Chimpanzees—our closest living relatives—were the best candidates among nonhuman animals to assess. They share many perceptual mechanisms with us, including those related to sound and color perception (16–19). They also can detect real-world cross-modal relations; for example, some chimpanzees can match vocalizations of their conspecifics to still movies or to facial expression photos of the corresponding vocalizing chimpanzees (i.e., *amodal* matching) (17, 18). One chimpanzee learned to match sounds (e.g., the sound of a bell or the voice of a conspecific) to pictures of the object or chimpanzee that produces them (i.e., *arbitrary* cross-modal matching); however, this is a difficult task for chimpanzees (19–21).

We tested 6 chimpanzees and 33 humans in the exact same experimental settings in a speeded classification paradigm adapted from Melara (15) (Figs. 1 and 2 and [Movie S1](#)). This paradigm allowed us to assess pitch–luminance mapping without the use of self-report. Here participants are required to classify black and white stimuli as either black or white while simultaneously ignoring background sounds that are either high-pitched or low-pitched. In Melara’s study, pitch–luminance mapping in humans was illustrated by worse performance (i.e., slower latencies) in categorizing colors when the background sound was incongruent (high-pitched for black, low-pitched for white) than when it was congruent (high-pitched for white, low-pitched for black) (15). Thus, we predicted that if chimpanzees also experience pitch–luminance mapping, then they should perform worse (i.e., be slower and/or less accurate) on incongruent trials compared with congruent trials.

Results

Chimpanzees and humans had comparable reaction times (RTs) in this task (group means of the median RTs in correct trials: humans, 665 ± 102 ms; chimpanzees, 654 ± 42 ms). However, chimpanzees were less accurate (mean error rate: humans, 0.06 ± 0.18%; chimpanzees, 8.19 ± 7.33%). To control for speed–

Author contributions: V.U.L., I.A., and T.M. designed research; V.U.L. and I.A. performed research; V.U.L. and I.A. analyzed data; and V.U.L., I.A., and T.M. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

¹V.U.L. and I.A. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: vera.ludwig@charite.de or adachi@pri.kyoto-u.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1112605108/-DCSupplemental.

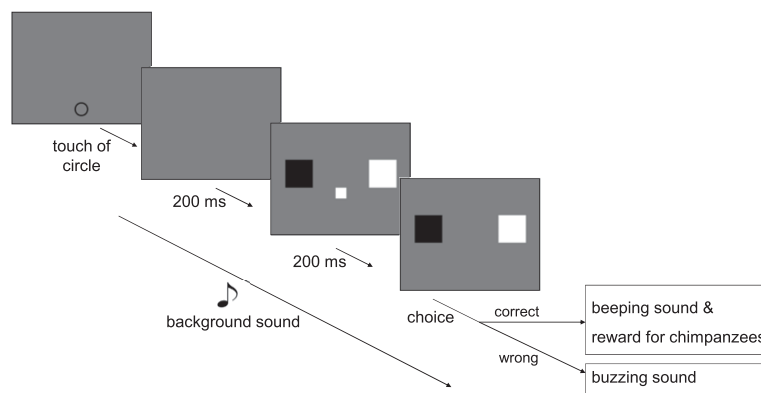


Fig. 1. Schematic representation of a trial. Touching a self-start circle initiated a high-pitched (1,047 Hz) or low-pitched (175 Hz) sound, followed 200 ms later by the sample (black or white; in the center) and two choice buttons (on the left and right). For half of the subjects, the left choice button was always black and the right choice button was white, and vice versa for the other half of the subjects. The sample disappeared after 200 ms. The subject had to touch the choice button with the same color as the sample. The sound persisted until a response was made (maximum, 2,000 ms).

accuracy trade-offs in the data, we calculated a standard combined performance measure of speed and accuracy for each condition and participant, the inverse efficiency (IE) score (22–24), as median RT divided by the proportion correct. A lower IE score indicates better performance.

Chimpanzees performed significantly better on congruent trials than on incongruent trials [mean IE, 706 ± 87 vs. 731 ± 98 ; $t(5) = -3.666$; $P = 0.01$, two-tailed paired-sample t test] (Fig. 3, *Left* and *Table S1*). The effect size was large: $r = 0.85$ (interpretation: $r = 0.10$, small; $r = 0.30$, medium; $r = 0.50$, large) (25). Humans also performed better on congruent trials than on incongruent trials [mean IE, 658 ± 105 vs. 673 ± 99 ; $t(32) = -3.938$, $P < 0.0005$] (Fig. 3, *Right*). Again, the effect size was large: $r = 0.57$. In a post hoc analysis, we found that the side of response (left or right) did not interact with the congruency effect (*SI Text*); that is, for both species, the effect was equally strong for trials requiring a response on the left side as for those requiring a response on the right side.

We also analyzed RTs and error rates separately to see *how* exactly performance was disturbed on incongruent trials in chimpanzees and humans. Chimpanzees made more mistakes on incongruent trials than on congruent trials [mean error rate, $9.58 \pm 8.28\%$ vs. $6.81 \pm 6.40\%$; $t(5) = -3.400$; $P = 0.02$, two-tailed paired-samples t test]. The effect size was large: $r = 0.84$. *Table S2* lists individual error rates for each chimpanzee and condition. The mean median RTs for incongruent and congruent trials were not significantly different, however [655 ± 40 ms vs. 654 ± 43 ms; $t(5) = -0.257$; $P = 0.81$, two-tailed test].

For humans, there were only three mistakes in the entire dataset (two incongruent trials and one congruent trials), not allowing an analysis of error rates for humans. However, for humans the mean median RT was significantly higher on incongruent trials than on congruent trials [672 ± 99 ms vs. 658 ± 105 ms; $t(32) = -3.895$; $P < 0.0005$, two-tailed test]. The effect size was $r = 0.57$.

Thus, for chimpanzees the effect was evident in the error rates, whereas for humans it was evident in latencies. This difference can be explained by the differing speed–accuracy trade-offs in the two species. Because humans were trying to be maximally accurate at this relatively easy task, the incongruent sounds only interfered with humans’ speed; that is, error rates were too low to be a sensitive measure to the manipulation (a floor effect). In contrast, chimpanzees were doing the task more impulsively (i.e., quicker, at the risk of making more mistakes). This allowed the effect to be expressed through mistakes rather than latencies.

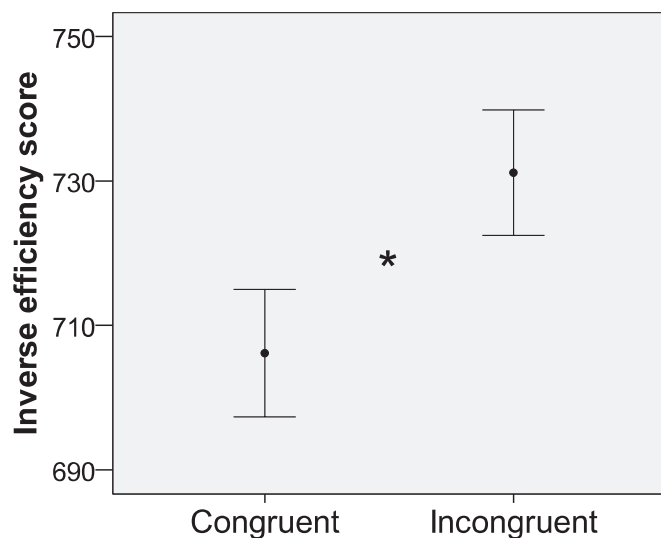
Discussion

Our results indicate that chimpanzees spontaneously associate high-pitched sounds with luminance (white) and low-pitched sounds with darkness (black), as do humans. In previous studies, teaching chimpanzees tasks involving auditory stimuli was found to be extremely difficult (19–21). For example, even after intensive training, only one out of six chimpanzees learned to press



Fig. 2. Experimental setup. Chimpanzee Ai (*Upper*) and a human participant (*Lower*) performing the task.

Chimpanzees



Humans

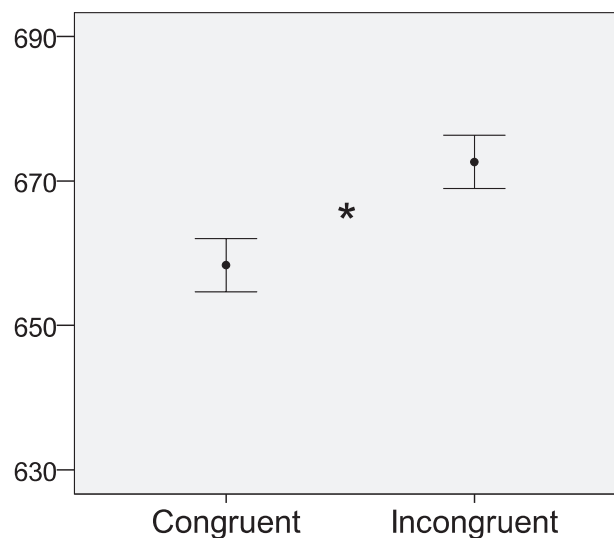


Fig. 3. Error bar graphs of the adjusted (61) IE scores (i.e., median RT for correct trials divided by proportion of correct responses) (22–24) for chimpanzees (*Left*) and humans (*Right*) for congruent (white and high pitch, black and low pitch) and incongruent trials (white and low pitch, black and high pitch). A smaller IE indicates better performance. Error bars represent the 95% confidence interval of the mean. * $P < 0.05$ (two-tailed).

a button on a particular side (left or right) in response to hearing a certain auditory stimulus (human or chimpanzee voice) (26). Thus, it is striking that in the present study, the chimpanzees' performance was systematically disturbed by sounds that were irrelevant to the visual matching task, with no previous training for the association of high pitch with luminance or of low pitch with darkness.

Our findings demonstrate that cross-modal correspondences are not a uniquely human or a purely linguistic phenomenon. Instead, at least the mapping of pitch-to-luminance seems to constitute a basic feature of the perceptual system that evolved before the human lineage split from that of chimpanzees. It is unlikely that chimpanzees *learned* the pitch-luminance association, given the lack of corresponding regularities in the real world (e.g., light objects do not make higher-pitched sounds than dark objects) (9). Cross-modal correspondences might be a natural byproduct of the way in which the primate brain processes multisensory information, such as when binding different sensory attributes of a percept (27). The fact that some human languages metaphorically use terms pertaining to bright color to refer to high-pitched sounds might stem directly from this evolutionarily old feature of the perceptual system. An overexpression of the mechanisms underlying pitch-luminance mapping and similar cross-modal associations might lead to synesthesia (5, 28). Our findings open up the interesting possibility that the condition of synesthesia may exist in nonhuman animals as well.

Cross-modal correspondences of the type shown here (i.e., pitch-luminance) have not been previously documented in nonhuman animals. Nevertheless, our study adds to a large body of work on nonhuman multisensory integration that suggests possible sharing of various other cross-modal processes and mechanisms across species (29–32). For example, dogs link their owner's voice to a visual representation of the owner (33), and horses (34) and rhesus monkeys (35, 36) know which calls belong to which individual conspecifics. Rhesus monkeys also are able to match monkey calls to the corresponding facial expressions of their conspecifics (37), as well as to show similar eye movement patterns as humans while viewing vocalizing conspecifics (38). Moreover, similar neural structures seem to underlie face-voice

integration in rhesus monkeys and audiovisual speech perception in humans (39–41).

Importantly, our findings may have implications for understanding the evolution of language. It has been proposed, in different variants, that when language first emerged, it was largely sound-symbolic (42–44). *Sound symbolism* refers to the idea that the relationship between words and their referents is not arbitrary (45, 46). For example, Sapir (47) and Newman (48) found that humans preferentially assign words containing high vowels (e.g., *mil*) to white and small objects and assign words containing low vowels (e.g., *mol*) to black and large objects. Moreover, Köhler (49, 50) demonstrated that humans prefer to use the label *baluma* or *maluma* for a rounded shape, and to use the label *takete* for an angular shape (see also refs. 42, 51, and 52). The presence of such systemic shared preferences in early humans would have facilitated the emergence of a first vocabulary. This possibility was supported by work of Ramachandran and Hubbard (42; see also ref. 44), and related earlier accounts can be found in the literature (43, 53, 54). Systematic cross-modal mappings might have constrained the way in which words were first mapped onto referents (42). Our findings in the present study suggest that natural tendencies to systematically map certain dimensions (here, pitch-luminance) were already present in our nonlinguistic ancestors. Thus, such cross-modal mappings might indeed have influenced the emergence of language.

Further research is needed to investigate how widely cross-modal correspondences are shared in the animal kingdom, how these correspondences evolved, and what mechanisms underlie them. Besides pitch-luminance mapping, species should be tested for other correspondences, such as sound-shape (7, 42, 52), pitch-size (4, 6), and pitch-location mapping (7). Some of these associations might turn out to exist uniquely in humans, being related to learning, culture, and language, but others may be natural features of the primate perceptual system, like pitch-luminance mapping. Further research on cross-modal correspondences might provide insight into the mechanisms and evolution of multisensory perception, synesthesia, and language.

Materials and Methods

Participants. We tested six female chimpanzees: Ai (age 32 y), Pendesa (31 y), Pal (8 y), Pan (25 y), Chloe (28 y), and Cleo (8 y). The chimpanzees were never water-deprived or food-deprived, and all participated in the experiment voluntarily (55). Our human sample comprised 33 right-handed volunteers (20 females), ranging in age from 18 to 35 y (mean, 24.76 y). One more participant was tested but was excluded because of technical problems during the procedure. The chimpanzees were housed in social groups in an enriched indoor and outdoor enclosure at the Kyoto University's Primate Research Institute. All chimpanzees had extensive previous experience with cognitive experiments using a touch-screen (56). The human participants were naive to the purpose of the experiment, and all signed an informed consent form before the experiment, in which they also agreed to be videotaped during the session. All but two participants (one from France and one from Myanmar) were Japanese.

Apparatus. Both chimpanzees and humans were tested on the same 17-inch touch-sensitive monitor in the same experimental booth. The stimulus presentation and the experimental devices were controlled by a program written in Microsoft Visual Basic. After correct responses, a feeder (Bio Medica Co. Ltd.) automatically delivered a piece of apple or a raisin to the chimpanzee, accompanied by a 0.5-s beeping sound. For human participants the sound and the activation of the feeder occurred after correct responses, but without delivery of fruit. After incorrect responses, a buzzing signal occurred. The intertrial interval was 2,600 ms. Bose loudspeakers, located immediately below the touch screen and centered in front of the subject, produced all sounds.

Visual Stimuli. The sample stimulus was a 99×99 pixel square (2.5×2.5 cm) presented at the center of the screen on a half-gray background [hue, saturation, lightness (HSL): 170, 0, 128]. Depending on the trial, this square was either black (HSL: 170, 0, 0) or white (HSL: 170, 0, 255). The two choice buttons were each 250×250 pixels (6.5×6.5 cm). For 17 humans and chimpanzees Ai, Chloe, and Cleo, the left choice button was always black; for 16 humans and chimpanzees Pal, Pan, and Pendesa, the left choice button was white. The self-start button at the beginning of each trial was a 3-cm-diameter blue circle.

Auditory Stimuli. The sounds were continuous computer-generated tones of either 1,047-Hz (high-pitched) or 175-Hz (low-pitched) frequency. These frequencies have been shown to produce a congruence effect in humans (15). Sound pressure level was set at ~ 76 dB, measured at the subjects' ears with a RION NL-22 sound level meter.

Procedure. Chimpanzees were first trained in the speeded classification paradigm without any background sounds. To learn the task, chimpanzees started by practicing a version of the task familiar to them, that is, a simple matching-to-sample task with white and black sample squares as big as the choice buttons. By decreasing size, duration of sample presentation, and other parameters, we gradually approached the final experimental paradigm in which the sample stimulus is small (99×99 pixels) and shown for only 200 ms. After chimpanzees had understood the final training task, they proceeded to the testing stage.

The procedure on the test day was almost exactly the same for humans and chimpanzees. Humans first received written instructions in English and Japanese, which explained the task and asked participants to be as quick and as accurate as possible and to use only the dominant hand. Both chimpanzees and humans first underwent 16 practice trials in which no experimental

sounds were played. If they made a mistake, they were required to repeat the practice session to achieve 16 correct trials. This occurred only once in humans. A chimpanzee or human participant who could not achieve perfect performance on the second practice session could not proceed to the testing sessions. This occurred only in chimpanzees. In this case, the chimpanzee received further training (without sound) for ~ 180 trials, and had another opportunity to achieve perfect performance again on the next experimental day. After the practice trials, humans were informed in writing that they should focus on their visual matching task even though sounds would be played back in subsequent tests. Both high-pitched and low-pitched sounds were briefly played to the participants of both species before the start of each session.

The first four experimental trials with sounds were habituation trials containing all possible sound-color combinations in random order. These trials and the practice trials were not analyzed. On an experimental day, participants completed two experimental sessions separated by a short break. A session consisted of 80 trials. Humans were tested only on one experimental day (two sessions, 160 trials total), whereas chimpanzees were tested on three experimental days (six sessions, 480 trials total). This schedule was used to compensate for the low statistical power in chimpanzees because of the small sample size. Due to a technical issue, chimpanzee Pan completed the six sessions across four instead of three experimental days.

A session consisted of an equal number of each possible color-sound combination, that is, 40 congruent trials (20 white sample and high pitch, 20 black sample and low pitch) and 40 incongruent trials (20 white sample and low pitch, 20 black sample and high pitch). The order of trials was balanced in the following way. Each four consecutive trials (the first four, the second four, etc.) contained one of each possible color-sound combination. Within these blocks of four trials, order was randomized. This was done in such a way that in the overall sequence, there were never consecutive sequences of more than three times the same pitch, the same color, or the same congruency type (congruent or incongruent). Simple repetitive patterns with more than three recurrences (e.g., black, white, black, white, black, white, black, white) were discarded. Both humans and chimpanzees were videotaped during the experiments.

Statistical Analysis. Our outcome measure, the IE score, is commonly used to control for possible speed-accuracy trade-offs in data (23, 24, 57-59). IE score is calculated by dividing the median RT for correct trials for each condition by the proportion correct for that condition for each participant (24, 57). We used the median RT instead of the mean as a measure of central tendency on the individual level because it is robust to outliers. All reported effect sizes were calculated following Rosenthal (60, p. 19) using the following equation:

$$r = \sqrt{t^2 / (t^2 + df)}.$$

The r values were interpreted as follows: $r = 0.10$, small; $r = 0.30$, medium; $r = 0.50$, large (25).

ACKNOWLEDGMENTS. We thank J. Sackur for discussions on the data and staff members of the Kyoto University Primate Research Institute's Center for Human Evolution Modeling for veterinary and daily care of the chimpanzees. This study was supported by Japanese Ministry of Education, Culture, Sports, Science, and Technology Grants 16002001 and 20002001; Global Centers of Excellence Grants A06 and D0; and a grant from the Kyoto University Primate Research Institute HOPE program (all to T.M.). Moreover, this study was supported by a Grant-in-Aid for JSPS Fellows and Grant-in-Aid for Young Scientists (B) 22700270 (both to I.A.).

- Martino G, Marks LE (2001) Synesthesia: Strong and weak. *Curr Dir Psychol Sci* 10: 61-65.
- Evans KK, Treisman A (2010) Natural cross-modal mappings between visual and auditory features. *J Vis* 10:6-12.
- Hubbard TL (1996) Synesthesia-like mappings of lightness, pitch, and melodic interval. *Am J Psychol* 109:219-238.
- Marks LE, Hammel RJ, Bornstein MH (1987) Perceiving similarity and comprehending metaphor. *Monogr Soc Res Child Dev* 52:1-102.
- Ward J, Huckstep B, Tsakanikos E (2006) Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex* 42:264-280.
- Mondloch CJ, Maurer D (2004) Do small white balls squeak? Pitch-object correspondences in young children. *Cogn Affect Behav Neurosci* 4:133-136.
- Walker P, et al. (2010) Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychol Sci* 21:21-25.
- Spector F, Maurer D (2008) The colour of Os: Naturally biased associations between shape and colour. *Perception* 37:841-847.
- Spector F, Maurer D (2009) Synesthesia: A new approach to understanding the development of perception. *Dev Psychol* 45:175-189.
- Wagner K, Dobkins KR (2011) Synaesthetic associations decrease during infancy. *Psychol Sci* 22:1067-1072.
- Palmeri TJ, Blake R, Marois R, Flanery MA, Whetsell W, Jr. (2002) The perceptual reality of synesthetic colors. *Proc Natl Acad Sci USA* 99:4127-4131.
- Simner J, et al. (2006) Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception* 35:1024-1033.
- Simner J, Ward J (2006) Synaesthesia: The taste of words on the tip of the tongue. *Nature* 444:438 (abstr).
- Morgan GA, Goodson FE, Jones T (1975) Age differences in the associations between felt temperatures and color choices. *Am J Psychol* 88:125-130.
- Melara RD (1989) Dimensional interaction between color and pitch. *J Exp Psychol Hum Percept Perform* 15:69-79.
- Hashiya K, Kojima S (2001) *Primate Origins of Human Cognition and Behavior*, ed Matsuzawa T (Springer, Tokyo), pp 155-189.

17. Izumi A, Kojima S (2004) Matching vocalizations to vocalizing faces in a chimpanzee (*Pan troglodytes*). *Anim Cogn* 7:179–184.
18. Parr LA (2004) Perceptual biases for multimodal cues in chimpanzee (*Pan troglodytes*) affect recognition. *Anim Cogn* 7:171–178.
19. Hashiya K, Kojima S (2001) Acquisition of auditory–visual intermodal matching to sample by a chimpanzee (*Pan troglodytes*): Comparison with visual–visual intramodal matching. *Anim Cogn* 4:231–239.
20. Martinez L, Matsuzawa T (2009) Effect of species-specificity in auditory–visual intermodal matching in a chimpanzee (*Pan troglodytes*) and humans. *Behav Processes* 82:160–163.
21. Matsuzawa T (2009) Symbolic representation of number in chimpanzees. *Curr Opin Neurobiol* 19:92–98.
22. Townsend JT, Ashby FG (1983) *Stochastic Modelling of Elementary Psychological Processes* (Cambridge Univ Press, New York).
23. Occelli V, Spence C, Zampini M (2009) Compatibility effects between sound frequency and tactile elevation. *Neuroreport* 20:793–797.
24. Putzar L, Goerendt I, Lange K, Rösler F, Röder B (2007) Early visual deprivation impairs multisensory interactions in humans. *Nat Neurosci* 10:1243–1245.
25. Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences* (Academic, New York), 2nd Ed.
26. Martinez L, Matsuzawa T (2009) Visual and auditory conditional position discrimination in chimpanzees (*Pan troglodytes*). *Behav Processes* 82:90–94.
27. Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. *Neuroreport* 16:419–423.
28. Cohen Kadosh R, Henik A (2007) Can synaesthesia research inform cognitive science? *Trends Cogn Sci* 11:177–184.
29. Stein BE, Meredith MA (1993) *The Merging of the Senses* (MIT Press, Cambridge, MA).
30. Calvert GA, Spence C, Stein BE (2004) *The Handbook of Multisensory Processes* (MIT Press, Cambridge, MA).
31. Murray MM, Wallace MT (2011) *The Neural Bases of Multisensory Processes* (CRC Press, Boca Raton, FL).
32. Seyfarth RM, Cheney DL (2009) Seeing who we hear and hearing who we see. *Proc Natl Acad Sci USA* 106:669–670.
33. Adachi I, Kuwahata H, Fujita K (2007) Dogs recall their owner's face upon hearing the owner's voice. *Anim Cogn* 10:17–21.
34. Proops L, McComb K, Reby D (2009) Cross-modal individual recognition in domestic horses (*Equus caballus*). *Proc Natl Acad Sci USA* 106:947–951.
35. Sliwa J, Duhamel J-R, Pascalis O, Wirth S (2011) Spontaneous voice–face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proc Natl Acad Sci USA* 108:1735–1740.
36. Adachi I, Hampton RR (2011) Rhesus monkeys see who they hear: Spontaneous cross-modal memory for familiar conspecifics. *PLoS ONE* 6:e23345.
37. Ghazanfar AA, Logothetis NK (2003) Neuroperception: Facial expressions linked to monkey calls. *Nature* 423:937–938.
38. Ghazanfar AA, Nielsen K, Logothetis NK (2006) Eye movements of monkey observers viewing vocalizing conspecifics. *Cognition* 101:515–529.
39. Ghazanfar AA, Chandrasekaran C, Logothetis NK (2008) Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *J Neurosci* 28:4457–4469.
40. Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of cross-modal binding in the human heteromodal cortex. *Curr Biol* 10:649–657.
41. Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cogn Neurosci* 17:377–391.
42. Ramachandran VS, Hubbard EM (2001) Synaesthesia: A window into perception, thought and language. *J Conscious Stud* 8:3–34.
43. Thorndike EL (1943) The origin of language. *Science* 98:1–6.
44. Kita S, Kantartzis K, Imai M (2010) in *Evolution of Language: The Proceedings of the 8th International Conference*, eds Smith ADM, De Boer B, Schouwstra M (World Scientific, Singapore), pp 206–213.
45. Hinton L, Nichols J, Ohala JJ (1994) *Sound Symbolism* (Cambridge Univ Press, Cambridge, UK).
46. Nuckolls JB (1999) The case for sound symbolism. *Annu Rev Anthropol* 28:225–252.
47. Sapir E (1929) A study in phonetic symbolism. *J Exp Psychol* 12:225–239.
48. Newman SS (1933) Further experiments in phonetic symbolism. *Am J Psychol* 45: 53–75.
49. Köhler W (1947) *Gestalt Psychology* (Liveright, New York), 2nd Ed.
50. Köhler W (1929) *Gestalt Psychology* (Liveright, New York).
51. Davis R (1961) The fitness of names to drawings: A cross-cultural study in Tanganyika. *Br J Psychol* 52:259–268.
52. Maurer D, Pathman T, Mondloch CJ (2006) The shape of boubas: Sound-shape correspondences in toddlers and adults. *Dev Sci* 9:316–322.
53. LeCron Foster M (1978) in *Human Evolution: Biosocial Perspectives*, eds Washburn SL, McCown ER (Benjamin Cummins, Menlo Park, CA), pp 77–122.
54. Paget R (1930) *Human Speech* (Routledge and Kegan Paul, London).
55. Matsuzawa T, Tomonaga M, Tanaka M (2006) *Cognitive Development in Chimpanzees* (Springer, Tokyo).
56. Matsuzawa T (2003) The Ai project: Historical and ecological contexts. *Anim Cogn* 6: 199–211.
57. Brozzoli C, et al. (2008) Touch perception reveals the dominance of spatial over digital representation of numbers. *Proc Natl Acad Sci USA* 105:5644–5648.
58. Shore DI, Barnes ME, Spence C (2006) Temporal aspects of the visuotactile congruency effect. *Neurosci Lett* 392:96–100.
59. Spence C, Kingstone A, Shore DI, Gazzaniga MS (2001) Representation of visuotactile space in the split brain. *Psychol Sci* 12:90–93.
60. Rosenthal R (1991) *Meta-Analytic Procedures for Social Research* (Sage, Newbury Park, CA), Rev. Ed.
61. Loftus GR, Masson MEJ (1994) Using confidence intervals in within-subject designs. *Psychon Bull Rev* 1:476–490.