

Correspondence

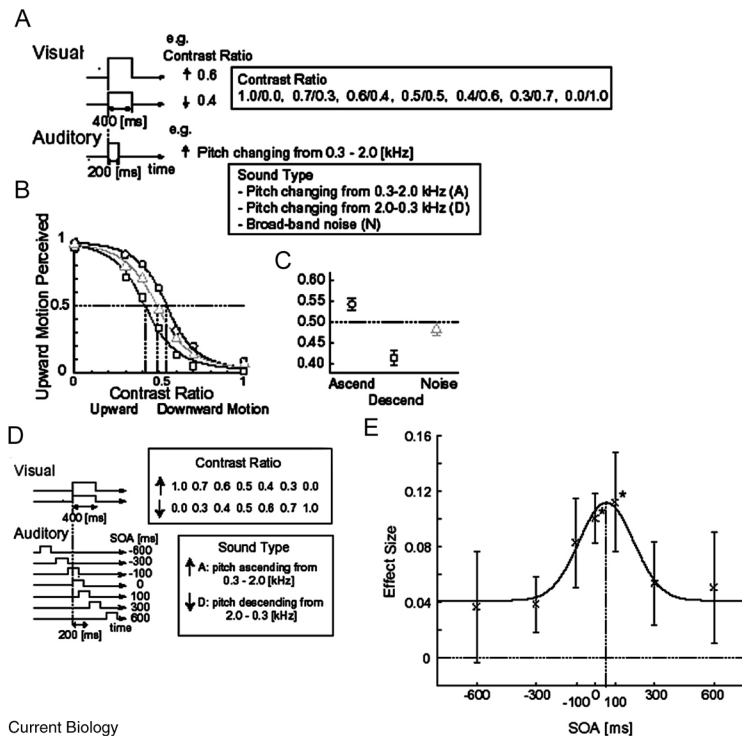
Changing pitch induced visual motion illusion

Fumiko Maeda^{1,2,*}, Ryota Kanai³, Shinsuke Shimojo^{1,4}

We often associate moving objects and changing pitch, e.g., falling stones with descending, and launching rockets with ascending pitch, even when these sounds do not happen in the real-world. The reason for this is unknown. Here we report an illusion in which auditory stimuli with no apparent spatial and motion information [1–3] alter human visual motion perception. Subjects made a two alternative forced choice (upward (V_{up}) or downward (V_{down}) visual motion perception) while presented with two superimposed, oppositely moving gratings (experiment 1), accompanied by either an ascending or a descending pitch of pure tone, or broad-band noise (Figure 1A). Gratings with ambiguous motion accompanied by ascending pitch were more likely to be perceived as an upward motion, those accompanied by descending pitch as a downward motion, whereas noise caused no directional bias.

In the first control experiment (experiment 2), we tracked eye positions to exclude possible confound of motion perception due to eye movement triggered by the sound as a cue. As expected, there was no significant eye movement during trials that could account for the perception of visual motion.

In the next experiment (experiment 3), we varied the stimulus-onset asynchrony (SOA) between gratings and sounds (Figure 1D) to exclude the possibility that the effect is due to semantic priming, cuing or other top-down influences. The maximum effect was observed when the onset of sound slightly lagged behind, but still completely overlapped with the presentation



Current Biology

Figure 1. Experimental design and results of the main experiment (experiment 1) and experiment 3.

(A) Main experimental design. (B,C) Results. (D) Experimental design of experiment 3. (E) Results varying stimulus-onset asynchrony (SOA). Psychometric curves and perceptual subjective equalities (PSEs) are shown for each sound type (ascendent pitch: circle, descendent pitch: square, noise: triangle; B). Error bars indicate a 95% confidence interval. There was a significant difference between all sound types (B,C). There was no significant difference in effect size (differences between ascending and descending PSE) between naïve and non-naïve subjects ($p > 0.05$). The mean and standard-error of the mean (SEM) of the PSE difference are plotted in (E). The peak-shift of the Gaussian curve by 56.2ms (E vertical dashed line), and the bootstrap method [19] (μ with a lower and upper limit of 26.54 and 71.77 ms, respectively), indicate that there is a significant shift to positive SOA that shows the maximum effect. SOAs of 0 and 100 ms (* $p = 0.01$) showed significant effects.

of the gratings (Figure 1E). This suggests that the illusion occurs at the perceptual level. The maximum effect was shifted slightly in the positive direction, which may be accounted for by the known auditory and visual neural delay [7,8].

Next, we examined the effect using actual words instead of sound to exclude the possibility of semantic influences (experiment 4). The experiment was similar to experiment 1 except that the sounds were: ‘ue’ (‘up’ in Japanese), ‘shita’ (‘down’ in Japanese), and N (experiment 4A). Even with non-Japanese subjects who had no knowledge of these words, there were no significant differences between the two groups, nor between either of the words and the control condition. These data suggest that the

auditory–visual interaction is unlikely to be mediated solely by semantic influences. Furthermore, by varying the SOA in the Japanese group (experiment 4B) we found that the effect was maximal and significant only when the voices were presented 400ms after the onset of the visual stimuli. This indicates that presentation of words affects the decision stage prior to when the motor responses are made rather than affecting visual motion perception directly.

Finally, in order to examine the specificity of the effect, we manipulated the orientation of the gratings, presenting them at either 0° (vertical motion), 45°/–45° (diagonal motion), or 90° (horizontal motion) (experiment 5). There was a systematic decrease in the effect, as motion became

more horizontal (Figures 2A–D). This further supports the perceptual origin of the illusion.

Taken together, these findings demonstrate that changing pitch can systematically affect the perception of visual motion. Our results suggest that the effect of these ‘metaphorically congruent’ sounds on vision is occurring predominantly at a perceptual level and could not be accounted for by eye-movement, cueing, semantic influences, or response bias. Previous studies [1,2,6,9–11] have shown congruency effects of auditory and visual stimuli. In addition, effects of vision on audition [12–14] and auditory stimuli capturing vision [15–18] has been reported. An illusion of ‘metaphorically congruent’ sounds capturing visual motion has never been reported. With regards to the neural mechanism underlying this illusion, future research using various neuroimaging techniques is warranted to elucidate this phenomenon.

Supplemental data

Supplemental data are available at <http://www.current-biology.com/cgi/content/full/14/23/R990/DC1/>

References

- Pratt, C.C. (1930). The spatial character of high and low tones. *J. experiment. Psychol.* 13, 278–285.
- Roffler, S.K., and Butler, R.A. (1968). Localization of tonal stimuli in the vertical plane. *J. Acoust. Soc. Am.* 43, 1260–1266.
- Timble, O.C. (1934). Localization of sound in the anterior-posterior and vertical dimensions of “auditory” space. *Br. J. Psychol.* 24, 320–334.
- Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *J. Exp. Psychol. Gen.* 106, 226–252.
- Ball, K., and Sekuler, R. (1981). Cues reduce direction uncertainty and enhance motion detection. *Percept. Psychophys.* 30, 119–128.
- Melara, R.D., and Marks, L.E. (1990). Dimensional interactions in language processing: investigating directions and levels of crosstalk. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 539–554.
- Luce, D.R. (1986). *Response time: their role in inferring elementary mental organization*, (New York: Oxford University Press).
- Stein, B.E., and Meredith, M.A. (1993). *The merging of the senses*, (Cambridge, MA: MIT Press).

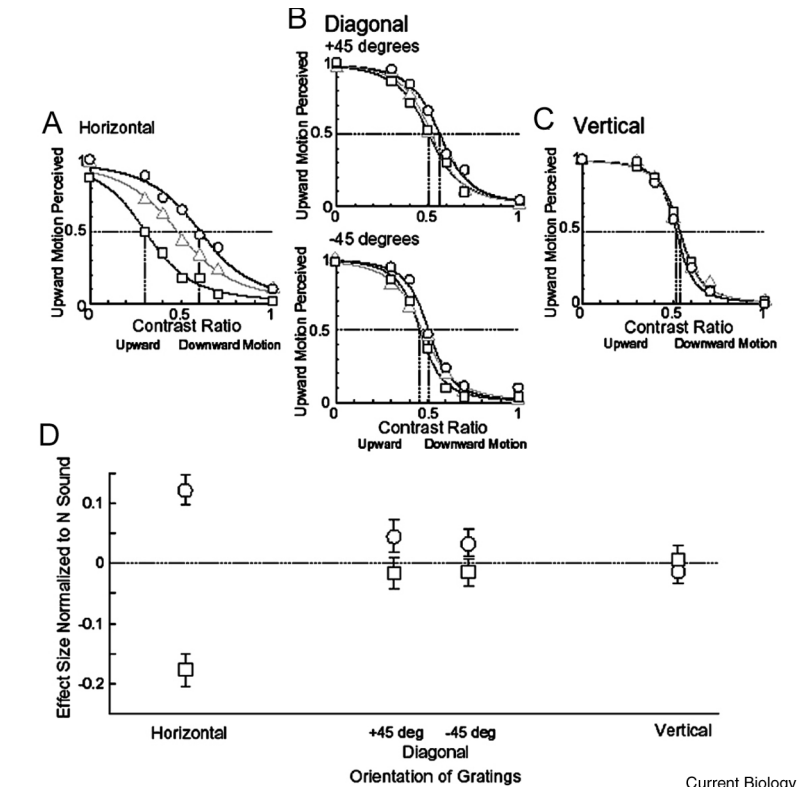


Figure 2. The effect of sound on gratings with different orientations and directions of motion.

(A–C) Psychometric curves for each condition (0°, +45°/–45°, and 90°). Further, after normalizing ascending and descending PSEs by subtracting noise PSE, normalized PSEs are plotted for each condition (0°, +45°/–45°, and 90°). Error bars represent the 95% confidence interval. There is a systematic decrease in the effect as the orientation becomes vertical and as visual motion becomes more horizontal.

- Bernstein, I.H., and Edelman, B.A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *J. Exp. Psychol.* 87, 241–247.
- Ben-Artzi, E., and Marks, L.E. (1995). Visual-auditory interaction in speeded classification: role of stimulus difference. *Percept. Psychophys.* 57, 1151–1162.
- Wagner, S., Winner, E., Cicchetti, D., and H.G. (1981). ‘Metaphorical’ mapping in human infants. *Child Dev.* 52, 728–731.
- Kitagawa, N., and Ichihara, S. (2002). Hearing visual motion in depth. *Nature* 416, 172–174.
- Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., and Kingstone, A. (2002). The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. *Brain Res. Cogn. Brain Res.* 14, 139–146.
- Sugita, Y., and Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound-arrival time. *Nature* 421, 911.
- Berger, T.D., Martelli, M., and Pelli, D.G. (2003). Flicker flutter: is an illusory event as good as the real thing? *J. Vis.* 3, 406–412.
- Meyer, G.F., and Wuerger, S.M. (2001). Cross-modal integration of auditory and visual motion signals. *Neuroreport* 12, 2557–2560.
- Sekuler, R., Sekuler, A.B., and Lau, R. (1997). Sound alters visual motion perception. *Nature* 385, 308.
- Shams, L., Kamitani, Y., and Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature* 408, 788.
- Efron, B., and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, (New York: Chapman and Hall).

¹Division of Biology / Computation and Neural Systems, California Institute of Technology, 139-74 Pasadena, California 91125, USA. ²Present address: Department of Psychiatry, Stanford University School of Medicine, Stanford, California 94305-5719, USA. ³Psychonomics Division, Helmholtz Research Institute, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands. ⁴NTT Communication Science Laboratories, 3-1, Morinosato Wakamiya Atsugi, Kanagawa, 243-0198, Japan.
*Email: fumiko@stanford.edu

Correspondence

Supplemental Data: Changing pitch induced visual motion illusion

Fumiko Maeda^{1,2,*}, Ryota Kanai³, Shinsuke Shimojo^{1,4}

Experimental Procedures

Subjects

Subjects were 12 healthy human volunteers, 4 female, 8 male, including the authors, with normal or normal-to-corrected vision and normal hearing. There were 12 subjects in experiment 1, and among the subjects who had an effect in experiment 1 (i.e., perceptual subjective equality (PSE) being significantly different for the ascending, D and N conditions using the bootstrap method), 4 subjects participated for each subsequent experiment. In experiment 4, there were two groups, i.e., Japanese ($n = 4$) and non-Japanese ($n = 4$). Written informed consent was obtained from all subjects.

Experimental set-up and stimuli

A monitor controlled by a Power Macintosh 9500/180MP (Apple Inc., CA, USA) computer running MATLAB (Mathworks Inc, MA, USA.) was used to present stimuli. Subjects sat comfortably 57 cm apart from the 21" monitor (Electron21, LaCie, OR, USA) with their heads partially immobilized in a chinrest (Handaya Co., Japan).

The visual stimuli were two superimposed patches of spatially enveloped (sigma of spatial Gaussian 1.07°) horizontally oriented sinusoidal luminance gratings (contrast 0.05/patch, spatial frequency 2.5 cycles/deg, temporal frequency 6.25 Hz). Two gratings with various contrast ratios (upward/downward component: 1.0/0.0, 0.7/0.3, 0.6/0.4, 0.4/0.6, 0.3/0.7, 0.0/1.0) drifted in opposing vertical

directions for a duration of 400 ms. Contrast ratios of 0.5/0.5 produced ambiguous motion [S1].

In the *Orientation* experiment (experiment 5), however, gratings were either in the horizontal orientation as in experiment 1 (0°), diagonal (clockwise $+45^\circ$, counter-clockwise -45°), or vertical (90°) orientation and drifted in their respective perpendicular directions.

The auditory stimuli were either a tone changing from 0.3 to 2.0 kHz (A), 2.0 to 0.3 kHz (D), or a broad-band noise (N) of approximately 65–70 dB (ascending) SPL (Bruel and Kjaer, GA, USA) delivered through headphones (MDR-CD270, Sony Inc., Japan) binaurally for 200 ms with the onset synchronized with visual stimuli. In the *Stimulus-onset-asynchrony (SOA) Experiment* (experiment 3), the SOA was varied and there were no noise conditions. In the semantic experiment (experiment 4A), recorded voices of the words 'ue' ('up' in Japanese) and 'shita' ('down' in Japanese) were replaced with A and D. SOAs used in the semantic experiment (experiment 4B) were -400, 0, 400 ms (negative values indicate the onsets of auditory stimuli preceding visual stimuli). The values ± 400 ms were chosen to avoid any overlap with the auditory and visual stimuli.

Experimental paradigm and data analyses

The trial started with a fixation point followed by the gratings after a random time (1–2 s) both presented at the center of the screen. Subjects judged whether the gratings were drifting in an upward (V_{up} ; spatially compatible '4' on the keyboard) or downward (V_{down} ; '1' on keyboard) direction and were told to respond as accurately as possible. The keyboard was rotated to maintain spatial compatibility when the gratings were in the diagonal and horizontal orientations in experiment 5.

There were 20 trials per condition (experiment 1: 420 trials total, experiment 3, 5: 1960 trials, experiment 4A: 420, experiment 4B: 840 trials) except for the *Eye*

Movement Experiment

(experiment 2). In experiment 2, there were 100 trials (480 trials total) of gratings with a contrast ratio of 0.5/0.5, and 50 trials each for 1.0/0 and 0/1.0, as well as 10 trials each with other contrast ratios. To track eye movement, EyeLink II (SMI USA, MA, USA) was used at a sampling frequency of 500 Hz. All analysis was performed off-line using MATLAB.

In each experiment, percent of perceived upward motion (V_{up}) was calculated for each auditory condition (A, D, 'ue', 'shita', and N). A psychometric curve was fitted for each sound type (A, D, 'ue', 'shita', and N, at each SOA, depending on the experiment) and perceptual subjective equality (PSE) was calculated. Error bars of the psychometric curves indicate 95% confidence interval after 2,000 simulations obtained by the bootstrap method [S2].

In experiment 2, instantaneous velocity of eye movement for all illusion and non-illusion trials were analyzed at each sample point collapsed across V_{up} and V_{down} . We contrasted the instantaneous velocity of illusion trials at each sample point and zero velocity, and the P-values of t tests are plotted against time from the onset of stimulus.

In experiment 3, the mean and standard error of the mean (SEM) of the differences between A and D PSE (effect size) for each SOA were calculated, and a Gaussian curve was fitted. 95% confidence interval after 2,000 simulations using the bootstrap method [2] was also calculated. 2-tailed pairwise t tests were then calculated for each SOA between A and D.

Illusion trials were defined as trials in which the subject perceived ambiguous gratings (contrast ratio of 0.5/0.5) as drifting in 'accord' with changing pitch, i.e., gratings perceived as moving upward accompanied by A, and gratings perceived as moving downward accompanied by D. Effect size was defined as the PSE difference between A and D. The significance level was set as $p=0.05$ unless specifically stated.

Instructions on the demo

1. On a PC or a Mac, open program 'Internet Explorer' and go to <http://neuro.caltech.edu/~fumiko/demo.swf>.
2. Click on arrow and fixate on the visual stimuli (or you can fixate on the arrow instead).
3. The direction of motion of the visual stimuli, i.e., ambiguous counter-phase grating, should be perceived in accord with the pitch change; descending pitch should be associated with downward visual motion perception, and ascending pitch with upward visual motion perception.
4. Please make sure you space each trial out by around 2 s, to ensure no carry-over effect.

Other Notes

Experiment 3 (Varying SOA)

We varied the stimulus-onset-asynchrony (SOA) between the gratings and sounds to exclude the possibility that our effect was not due to semantic priming, cuing or other top-down influences, which then should be observed when sound precedes gratings [S3,S4], or effect on the decision process, which then should be observed when the sound follows the gratings yet before the motor response [S5].

Past Studies On Auditory And Visual Stimuli Sharing Same Verbal Labels ('high/low')

Several studies showed congruency and interference effects in adults [S5–S7], as well as interaction in infants and children before the acquisition of language and/or concept of these labels [S8–S10]. Hence, there is some evidence that raises the possibility that sound and visual stimuli sharing common verbal labels interact at a prelinguistic and perceptual level.

Past Literature On Interactions Between Sound Of Different Frequencies And Visual Spatial / Motion Perception

With regards to the neural mechanism underlying this illusion, it is premature to draw any conclusions about the brain region or at which processing stage this illusion occurs based

on studies of auditory and visual motion perception [S11–S16], cross-modal change perception [S17], direction perception of ambiguous motion [S18], and cross-modal illusion [S19–S21]. To our knowledge, there is no clear neural evidence of direct interactions between pure tones of changing pitch and visual motion perception. It may be worth noting that narrow-band sounds of higher frequencies induce spike patterns of auditory cortical neurons based on their spectral frequencies that signal incorrect spatial locations, which predicts and explains spatial illusions (misplacement of sound source information) experienced in human listeners [S22–S25]. Their results, however, were not generalized to all sounds including sounds used here. Our pilot data indicated that there were no auditory and visual parameters, such as rate of change size of stimuli that systematically changed the effect size. In addition, these studies explain spatial illusions but do not offer explanation about motion illusions nor how mislocalization extends to the visual domain as shown here.

Acknowledgments

F.M. conceived the experiment, and together with R.K. carried out the experiment and data analysis; S.S. gave comments throughout the course of the research; F.M., R.K. and S.S. co-wrote the paper. We thank Lawrence Parsons, Noa Ofen, and Olie & Ivan Bodis-Wollner for commenting on an earlier draft of this paper. We also thank Ryusuke Hayashi and Daw-An Wu for comments and assistance during the course of this research.

References

- S1. Gorea, A., and Lorenceau, J. (1984). Perceptual bistability with counterphase gratings. *Vision Res.* 24, 1321–1331.
- S2. Efron, B., and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, (New York: Chapman and Hall).
- S3. Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *J. Exp. Psychol. Gen.* 106, 226–252.
- S4. Ball, K., and Sekuler, R. (1981). Cues reduce direction uncertainty and enhance motion detection. *Percept. Psychophys.* 30, 119–128.
- S5. Melara, R.D., and Marks, L.E. (1990). Dimensional interactions in language processing: investigating directions and levels of crosstalk. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 539–554.
- S6. Bernstein, I.H., and Edelman, B.A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *J. Exp. Psychol.* 87, 241–247.
- S7. Ben-Artzi, E., and Marks, L.E. (1995). Visual-auditory interaction in speeded classification: role of stimulus difference. *Percept. Psychophys.* 57, 1151–1162.
- S8. Pratt, C.C. (1930). The spatial character of high and low tones. *J. Exp. Psychol.* 13, 278–285.
- S9. Roffler, S.K., and Butler, R.A. (1968). Localization of tonal stimuli in the vertical plane. *J. Acoust. Soc. Am.* 43, 1260–1266.
- S10. Wagner, S., Winner, E., Cicchetti, D., and H.G. (1981). 'Metaphorical' mapping in human infants. *Child Dev* 52, 728–731.
- S11. Macko, K.A., Jarvis, C.D., Kennedy, C., Miyaoka, M., Shinohara, M., Sololoff, L., and Mishkin, M. (1982). Mapping the primate visual system with [2-14C]deoxyglucose. *Science* 218, 394–397.
- S12. Mazzoni, P., Bracewell, R.M., Barash, S., and Andersen, R.A. (1996). Spatially tuned auditory responses in area LIP of macaques performing delayed memory saccades to acoustic targets. *J. Neurophysiol.* 75, 1233–1241.
- S13. Poremba, A., Saunders, R.C., Crane, A.M., Cook, M., Sokoloff, L., and Mishkin, M. (2003). Functional mapping of the primate auditory system. *Science* 299, 568–572.
- S14. Bremner, F., Schlack, A., Shah, N.J., Zafiris, O., Kubischik, M., Hoffmann, K., Zilles, K., and Fink, G.R. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron* 29, 287–296.
- S15. Lewis, J.W., Beauchamp, M.S., and DeYoe, E.A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb. Cortex* 10, 873–888.
- S16. Lewis, J.W., and Van Essen, D.C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* 428, 112–137.
- S17. Downar, J., Crawley, A.P., Mikulis, D.J., and Davis, K.D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nat. Neurosci.* 3, 277–283.
- S18. Williams, Z.M., Elfar, J.C., Eskandar, E.N., Toth, L.J., and Assad, J.A. (2003). Parietal activity and the perceived direction of

- ambiguous apparent motion. *Nat. Neurosci.* 6, 616-623.
- S19. Bushara, K.O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., and Hallett, M. (2003). Neural correlates of cross-modal binding. *Nat. Neurosci.* 6, 190-195.
- S20. Shams, L., Kamitani, Y., Thompson, S., and Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport* 12, 3849-3852.
- S21. Stekelenburg, J.J., Vroomen, J., and de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neurosci. Lett.* 357, 163-166.
- S22. Middlebrooks, J.C. (1992). Narrow-band sound localization related to external ear acoustics. *J. Acoust. Soc. Am.* 92, 2607-2624.
- S23. Xu, L., Furukawa, S., and Middlebrooks, J.C. (1999). Auditory cortical responses in the cat to sounds that produce spatial illusions. *Nature* 399, 688-691.
- S24. Butler, R.A. (1999). The unfolding of an auditory illusion. *Perspect. Biol. Med.* 42, 157-173.
- S25. Tollin, D.J., and Yin, T.C. (2003). Spectral cues explain illusory elevation effects with stereo sounds in cats. *J. Neurophysiol.* 90, 525-530.

¹Division of Biology / Computation and Neural Systems, California Institute of Technology, 139-74 Pasadena, CA 91125, USA. ²Present address: Department of Psychiatry, Stanford University School of Medicine, Stanford, CA 94305-5719, USA. ³Psychonomics Division, Helmholtz Research Institute, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands. ⁴NTT Communication Science Laboratories, 3-1, Morinosato Wakamiya Atsugi, Kanagawa, 243-0198, Japan.
*Email: fumiko@stanford.edu