Multisensory Integration of Looming Signals by Rhesus Monkeys

Report

Joost X. Maier,[1] John G. Neuhoff,[2]
Nikos K. Logothetis,[1] and Asif A. Ghazanfar[1,*]
[1]Max Planck Institute for Biological Cybernetics
72076 Tuebingen
Germany
[2]Department of Psychology
The College of Wooster
Wooster, Ohio 44691

Summary

Looming objects produce ecologically important signals that can be perceived in both the visual and auditory domains. Using a preferential looking technique with looming and receding visual and auditory stimuli, we examined the multisensory integration of looming stimuli by rhesus monkeys. We found a strong attentional preference for coincident visual and auditory looming but no analogous preference for coincident stimulus recession. Consistent with previous findings, the effect occurred only with tonal stimuli and not with broadband noise. The results suggest an evolved capacity to integrate multisensory looming objects.

Introduction

The world is full of potentially dangerous approaching objects. These include predators, competitors, and abiotic sources. For our ancestors in the primate lineage, a raptor's stoop or pursuit by a dominant individual were clear and present dangers of everyday life. Such dangers remain for extant primates living in the wild (Gil-da-Costa et al., 2003; Kitchen et al., 2003). To survive in such a world, animals must escape or avoid such dynamic sources of danger by rapidly detecting and responding to the signals that specify such *looming* events. Because of their rapid transmission in the environment, visual and auditory signals can each independently be a reliable cue to fast approaching sources.

In the visual domain, many animals (including humans) produce adaptive avoidance responses to both real and simulated looming cues (Ball and Tronick, 1971; Schiff et al., 1962; Schiff, 1965; Tinbergen, 1951). For example, a rapidly expanding circular shadow elicits fear responses in adult and infant rhesus monkeys (Schiff et al., 1962) and human infants (Ball and Tronick, 1971), but rapidly contracting shadows do not. Studies in the auditory domain also suggest a perceptual bias for sound sources that are looming. When human listeners are asked to predict the arrival time of a moving sound source, they repeatedly err on the side of safety, expecting contact before the source actually arrives (Rosenblum et al., 1993; Schiff and Oldak, 1990). For such looming sound sources, intensity change is the dominant cue (Rosenblum et al., 1987). Indeed, adult and infant humans and rhesus monkeys have a strong perceptual bias toward rising- versus falling-intensity signals (Freiberg et al., 2001; Ghazanfar et al., 2002; Neuhoff, 1998; Stecker and Hafter, 2000). Taken together, responses to symmetrically expanding visual cues and rising-intensity auditory cues suggest an adaptive bias for perceiving looming sources.

Although many animals evolved parallel warning systems to escape undesirable encounters when one sensory modality failed (such as vision in dark environments), parallel sensory systems also confer enormous benefits when used in conjunction. The integration of multimodal signals enhances detection, discrimination, and learning of events across many different species (see Rowe, 1999, for review). Moreover, increases in bimodal localization performance with spatially coincident auditory and visual stimuli are multiplicative and exceed what would be predicted by performance in either modality alone, suggesting a true interaction of vision and audition (Stein et al., 1989).

It is unknown whether animals understand that signals from two modalities signify the same event and if so whether this ability is experience dependent. Our study, therefore, addressed the following question: can animals integrate the auditory and visual components of an artificial looming source into a unified percept? Using the preferential looking method with rhesus monkeys as subjects, we investigated whether videos of a rapidly expanding (looming) and a rapidly contracting (receding) disk could be matched with a rising (looming) or falling (receding) intensity complex tone. The artificiality of the stimuli and their conjunction preclude the subjects' prior experience as a factor, yet the stimuli retain the general fundamental features of looming/receding objects.

Results

We adopted the "preferential looking" technique, used by cognitive scientists to test intermodal perception in prelinguistic infants (e.g., Spelke et al., 1983), to investigate the natural capacity of rhesus monkeys to match visual looming cues with auditory looming cues. In this paradigm, subjects were seated in front of two LCD monitors where one monitor displayed a video of an expanding disk (the looming stimulus) while the other monitor displayed a contracting disk (the receding stimulus; Figure 1A). A rising- or falling-intensity complex tone (fundamental frequency: 400 Hz; 20 dB intensity change; Figure 1B) was played through a hidden loudspeaker placed between the two monitors. All videos and sounds were 1 s in duration, and the two videos and the sound were synchronized (see Experimental Procedures); these stimuli were played in repetition for 30 s. The subjects were not rewarded in any way for their performance during testing. The dependent measure was percentage of total looking time to the matching video.

In condition 1, rhesus monkey subjects (n = 9) were
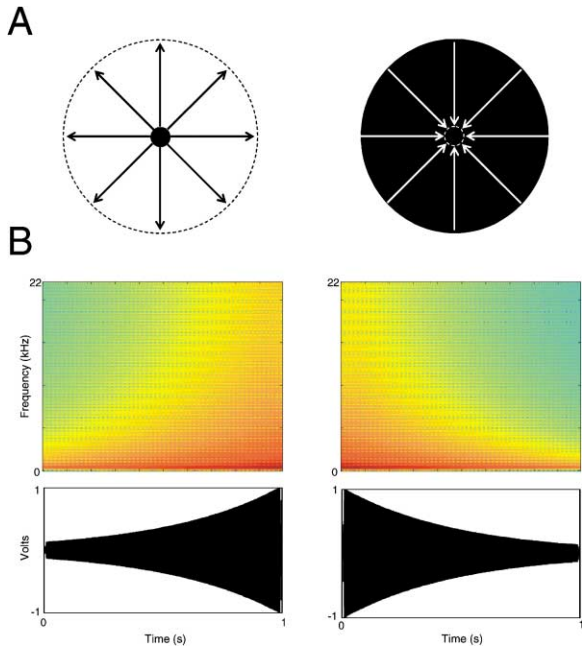
*Correspondence: asifg@tuebingen.mpg.de

Figure 1. Multimodal Looming and Receding Stimuli

(A) Visual stimuli. A black disk symmetrically expanding (looming stimulus) or contracting (receding stimulus) on a white background. Stimuli expanded/contracted over the stimulus duration of 1 s.
(B) Auditory stimuli. Spectrograms and time-amplitude waveforms of the rising-intensity (looming) and falling-intensity (receding) complex tones (based on a 400 Hz triangular waveform).

presented with the two dynamic visual stimuli and heard the looming version of the complex tone (rising in intensity from 55 to 75 dB). They looked significantly longer at the matching, looming video than at the receding video. The mean looking time toward the looming video was 8.73 ± 0.89 s (68% of the total looking time) versus 4.22 ± 0.66 s toward the receding video. This is significantly above chance level (t[8] = 5.76, p < 0.0001; Figure 2A). Furthermore, 9 out of 9 subjects looked longer at the looming visual stimulus (sign test, p = 0.002).

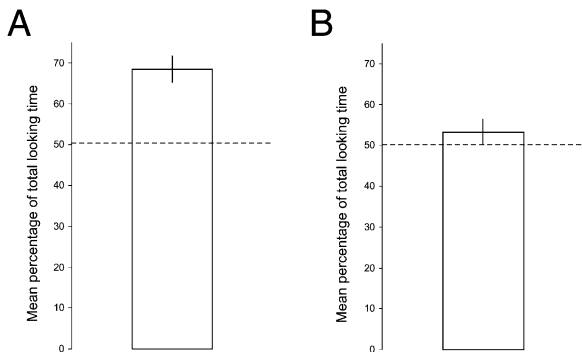It is possible that rhesus monkeys simply found the looming visual stimulus more salient than the receding

Figure 2. Percentage of Looking Time at the Matching Stimulus Presented with Rising- and Falling-Intensity Complex Tone Stimuli

Mean percentage of total looking time at the (A) looming and (B) receding stimulus in condition 1 and 2, respectively.
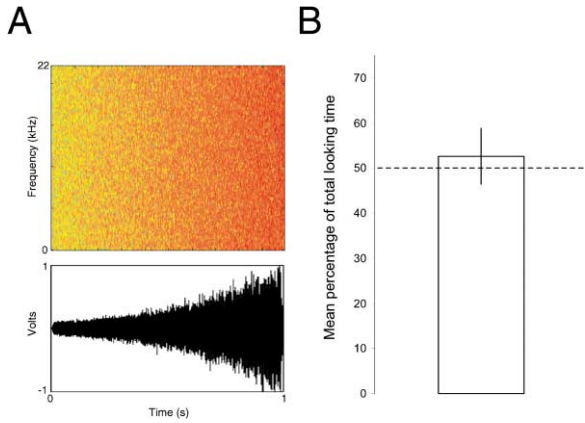

Figure 3. Percentage of Looking Time at the Matching Stimulus Presented with a Rising-Intensity White Noise Stimulus

(A) Spectrogram and time-amplitude waveform of the rising-intensity white noise stimulus.
(B) Mean percentage of total looking time at the matching looming stimulus in condition 3.

stimulus, independent of the sound that they heard. In condition 2, therefore, a new set of subjects (n = 16) was tested with the same visual stimuli but with the *receding* sound stimulus. If sound has no influence on the preferential looking, then we would expect that the monkeys would continue to prefer to look at the looming video. In contrast, if the sound does have an influence on looming perception, then in this condition they could either show a preference toward the *receding* video or show no preference at all. Our data support the latter hypothesis. Rhesus monkeys showed no preferential looking toward the receding video (4.62 ± 0.50 s) versus the looming video (4.09 ± 0.49 s) when they heard the receding sound stimulus. The mean percentage of total looking time to the matching, receding video was 53% (t[15] = 1.09, p = 0.292; Figure 2B), and the proportion of subjects looking longer overall toward the receding video was also not significant (sign test, p = 0.105).

Thus, despite the fact that duration, spectral content, and overall intensity change were identical between the looming and receding sound stimuli, only the looming sound biased the viewing behavior of rhesus monkeys toward the looming visual stimulus.

In humans and monkeys, the strong perceptual bias for looming sounds is specific for harmonically structured tonal sounds and is weakened or eliminated when broadband noise (unstructured) sounds are heard (Ghazanfar et al., 2002; Neuhoff, 1998). In condition 3, we investigated whether the influence of looming sounds on the looking preferences of rhesus monkeys toward the looming video was stimulus specific. We replaced the complex tone used in condition 1 with white noise, but with an identical rising intensity change (from 55 dB to 75 dB; Figure 3A). Under these conditions, subjects did not show any preference for the looming visual stimulus. Fifty-three percent of the total looking time was toward the matching, looming video (n = 8; t[7] = 0.421, p = 0.686; Figure 3B), and the proportion of animals looking longer toward the match was also not significant (sign test, p = 0.363). Mean looking time at the looming

stimulus was 6.30 ± 1.75 s versus 4.47 ± 0.67 s for the receding visual stimulus. These data demonstrate that the ability to integrate auditory and visual looming signals is dependent on the spectral characteristics of the looming sound.

Discussion

The present study demonstrates the capacity for multisensory integration of looming but not receding signals in a nonhuman animal. This matching ability is due solely to the salience of symmetrically expanding visual signals combined with the rising intensity of structured sounds; temporal cues (such as duration) and spatial cues could not have been used to make the match (as the visual signals were displaced from the auditory signal). These findings suggest that rhesus monkeys (and likely other animals, including humans) have evolved perceptual mechanisms to detect bimodally specified rapidly approaching objects.

For the looming condition, the magnitude of the behavioral bias is very large: 9 out of 9 subjects demonstrated the effect and their percentage of total looking time to the match screen was also highly significant (68% versus chance at 50%). The magnitude of our statistical effects is similar to those reported in preferential looking studies of human infants. For instance, for infant studies of crossmodal perception, Walker-Andrews and Lennon (1985) reported 60% looking time to the match screen in a crossmodal distance perception study and Kuhl and Meltzoff (1982) reported 73% looking time to the matching screen in a crossmodal speech perception study. The present results are also consistent with a previous study of face-voice integration by monkeys in our lab using the same technique (Ghazanfar and Logothetis, 2003). Indeed, the results of the monkey studies are actually more robust than the infant studies because we obtain a highly significant statistical outcome with far fewer subjects. The results for the receding and white noise conditions were similarly very robust in favoring the null statistical hypothesis of no preferential looking to either screen.

Multisensory looming integration was dependent on the *spectral structure* of the rising intensity sound. White noise of identical duration and intensity change failed to elicit integration. This is consistent with previous auditory looming studies (Ashmead et al., 1995; Neuhoff, 1998, 2001). Tonal sounds are generally more meaningful than noise because their ordered harmonics undergo correlated changes in intensity and can provide information about the nature and location of the source, whereas the uncorrelated changes present in the components of noise do not provide such information (Bregman, 1990). Under natural conditions, tonal sounds are much more reliable markers for ecologically important individual sound sources than broadband noise (Neuhoff, 2001).

Sensory information from different modalities can be transmitted in one of two ways: it can specify the identical stimulus property and thus be *amodal*, or each modality can have unique features that together then form specific arbitrary relations. Amodal equivalence is usually temporally specified along such stimulus characteristics as duration, rhythm, and/or intensity (see Walker-

Andrews, 1994, for review). For example, the sight and sound of a bouncing ball can indicate that the visual and auditory events come from the same source, and the amodal information separates them from co-occurring events that do not share the same temporal structure. The multisensory integration of looming signals shown by rhesus monkeys *cannot* be based on amodal equivalency because temporal cues such as duration and synchrony were present for both visual signals and the looming sound.

In the case of unimodal looming signals, there is strong evidence that the perceptual bias is an evolved trait (Neuhoff, 2001; Ghazanfar et al., 2002; Schiff et al., 1962; Schiff, 1965). In the present case, our monkey subjects had no prior experience with computer-generated expanding disks, with rising intensity complex tones, or their association. The synchrony and duration provided amodal cues, but the receding disk was a conflicting signal and the integration was spectrally dependent. The monkey subjects, therefore, made an arbitrary association of visual expansion with a rising-intensity tone that was immediately salient. This suggests that monkeys (and likely humans and other animals) have an evolved capacity to integrate bimodal looming signals, but whether this capacity manifests itself innately or requires experience is not known. Regardless, they are able to extract the relevant looming cues independent of the unique features of an arbitrary approaching object.

Our understanding of the neuroanatomical and neurophysiological bases for sensory convergence has advanced much more rapidly than our understanding of the role of polysensory areas in behavior (see Ettlinger and Wilson, 1990, for review). For natural, behaviorally relevant multimodal signals, virtually nothing is known about the neural level integration. For visual looming, many taxa show similar avoidance responses (Schiff, 1965), but the neural architecture that underlies this behavior may be vastly different. In monkeys, a large proportion of neurons in the upper bank of the superior temporal sulcus (STS) have a strong bias toward looming visual motion signals (Anderson and Siegel, 1999; Hietanen and Perrett, 1996; Mistlin and Perrett, 1990). For auditory signals, the human STS is more responsive to looming signals than to receding or static sounds (Seifritz et al., 2002). Taken together with the known polysensory properties of STS (Desimone and Gross, 1979; Bruce et al., 1981; Hikosaka et al., 1988), these studies suggest that the STS is likely a site for the neural integration of multimodal looming signals. The ventral intraparietal area (VIP) of the posterior parietal cortex may be another possible site of integration. VIP also has a representational bias for expanding optic flow stimuli (Bremmer et al., 2002), and VIP neurons are often polysensory, responding to visual, tactile, and auditory stimuli (Colby et al., 1993; Bremmer et al., 2001). Furthermore, a recent study revealed that microstimulation of VIP elicits defensive movements (Cooke et al., 2003), a feature compatible with a putative role in looming processing. Together, the polysensory properties of STS and VIP and their sensitivity to looming signals strongly suggest them as candidate sites for the neuronal integration of bimodal looming signals.

Experimental Procedures

Subjects

We tested male rhesus macaques (age range: 4–12 years) who are part of a large colony housed at the Max Planck Institute for Biological Cybernetics. All animals are socially housed and provided with enrichment (toys, hammocks, ropes, etc.). All experimental procedures were in accordance with the local authorities (Regierungspraesidium) and the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals.

Stimuli

Visual stimuli were generated in Matlab (www.mathworks.com), using the Psychophysics Toolbox extensions (www.psychtoolbox.org; Brainard, 1997). The visual looming stimulus consisted of a black disk symmetrically expanding (ranging from 0.68 degrees [fully contracted] to 9.94 degrees [fully expanded] eccentricity) on a white background. The receding stimulus was a black disk symmetrically contracting (same dimensions as the looming stimulus) on a white background. Visual stimuli expanded or contracted over a 1 s duration. Auditory stimuli were 400 Hz complex tones composed of a triangular waveform and generated in Cool Edit Pro software (Syntrillium Software Corp., www.syntrillium.com). The tones were 1 s in duration and either rose in intensity from 55 to 75 dB (looming stimulus) or fell from 75 to 55 dB (receding stimulus). They were sampled at 44.1 kHz, had 10 ms onset and offset ramps, and changed 20 dB in intensity from start to end. A white noise stimulus with identical intensity and temporal characteristics as the looming sound was also generated.

Stimulus Presentation and Testing Procedure

The looming and the receding visual stimuli were played simultaneously on side-by-side 15 inch LCD monitors (Acer FP559, www.global.acer.com). Audio tracks were synchronized with both videos and played through a hidden speaker (a self-powered Advent AV750 speaker) placed directly between and slightly behind the monitors. The RadLight 3.03 Special Edition software video player (www.radlight.net) was used to play the videos in synchrony. Sounds were presented at an intensity of 55–75 dB (A-weighted) SPL as measured using a Brüel & Kjær 2238 Mediator sound level meter (www.bksv.com) at a distance of 75 cm. For testing, a subject was seated in a primate chair, brought to the testing room, and placed in front of the two monitors at a distance of 72 cm. The monitors were 65 cm apart (center-to-center distance) and at eye-level with the subject. All trials were videotaped using a digital video camera placed above and between the monitors. All equipment was concealed by a thick black curtain except for the monitor screens and the lens of the camera. The experimenter monitored subject activity from outside of the room. During this time, subject's attention was directed to the center by flashing a 1.2 W light placed centrally between the two monitors. A test session began when the subject looked centrally. A trial consisted of the two videos played in a continuous loop for 30 s with one of the two sounds also played in a loop through the speaker. The left-right position of the two dynamic visual stimuli was counterbalanced. Each subject was only tested once, and all trials were recorded on digital video. We used a between-groups design because, as in all studies that examine the spontaneous behavior of animals and prelinguistic human infants, the subjects quickly habituate to the testing environment. No reward or training is provided. This precludes us from testing them on more than one condition.

Video Scoring

We collected high-quality, close-up digital videos of the subjects' behavior with a JVC GR-DVL805 digital camera (www.jvc.com). Videos were acquired at 30 frames/s (frame size: 720 × 480 pixels) onto a PC using an IEEE 1394a input and Adobe Premiere 6.0 software (www.adobe.com). The audio tracks were acquired at a 32 kHz sampling rate and 16 bit resolution. Clips for analysis were edited down to 30 s by identifying the onset of the auditory track and adding 900 frames after this point. Two observers scored each trial; one observer was blind to the experimental condition. For each trial, video scorers judged, frame by frame, whether the subject was looking toward the right screen, the left screen, or away from both. The total duration of a subject's looking toward each video (left or right) or away from both was recorded and expressed as the proportion of total looking time.

Scoring which of the screens the monkey subjects were looking toward was absolutely unambiguous. The screens are far apart in the horizontal dimension, fairly close to the monkey's face, and at eye level. Thus, the monkey has to make large eye movements/head movements to look to one screen or the other, and it is similarly clear when he is not looking at either screen. To validate this, we measured interobserver reliability. The mean difference in the percent fixation scores was 1.0%. Observer agreement was 0.985 (p < 0.0001) as measured by a Pearson r test.

References

Anderson, K.C., and Siegel, R.M. (1999). Optic flow selectivity in the anterior superior temporal polysensory area, STPa, of the behaving monkey. J. Neurosci. *19*, 2681–2692.

Ashmead, D.H., Davis, D.L., and Northington, A. (1995). Contribution of listeners' approaching motion to auditory distance perception. J. Exp. Psychol. Hum. Percept. Perform. *21*, 239–256.

Ball, W., and Tronick, E. (1971). Infant responses to impending collision—optical and real. Science *171*, 818–820.

Brainard, D.H. (1997). The psychophysics toolbox. Spat. Vis. *10*, 433–436.

Bregman, A.S. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound (Cambridge, MA: MIT Press).

Bremmer, F., Schlack, A., Shah, N.J., Zafiris, O., Kubischik, M., Hoffman, K.-P., Zilles, K., and Fink, G.R. (2001). Polymodal motion processing in the posterior parietal and premotor cortex: a human fMRI study strongly implies equivalences between humans and monkeys. Neuron *29*, 287–296.

Bremmer, F., Klam, F., Duhamel, J.-R., Ben Hamed, S., and Graf, W. (2002). Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). Eur. J. Neurosci. *16*, 1569–1586.

Bruce, C., Desimone, R., and Gross, C.G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J. Neurophysiol. *46*, 369–384.

Colby, C.L., Duhamel, J.-R., and Goldberg, M.E. (1993). The ventral intraparietal area (VIP) of the macaque: anatomical location and visual response properties. J. Neurophysiol. *69*, 902–914.

Cooke, D.F., Taylor, C.S.R., and Graziano, M.S.A. (2003). Complex movements evoked by microstimulation of the ventral intraparietal area. Proc. Natl. Acad. Sci. USA *100*, 6163–6168.

Desimone, R., and Gross, C.G. (1979). Visual areas in the temporal cortex of the macaque. Brain Res. *178*, 363–380.

Ettlinger, G., and Wilson, W.A. (1990). Cross-modal performance: behavioural processes, phylogenetic considerations and neural mechanisms. Behav. Brain Res. *40*, 169–192.

Freiberg, K., Tually, K., and Crassini, B. (2001). Use of an auditory looming task to test infants' sensitivity to sound pressure level as an auditory distance cue. British Journal of Developmental Psychology *19*, 1–10.

Ghazanfar, A.A., and Logothetis, N.K. (2003). Facial expressions linked to monkey calls. Nature *423*, 937–938.

Ghazanfar, A.A., Neuhoff, J.G., and Logothetis, N.K. (2002). Auditory

looming perception in rhesus monkeys. Proc. Natl. Acad. Sci. USA *99*, 15755–15757.

Gil-da-Costa, R., Palleroni, A., Hauser, M.D., Touchton, J., and Kelley, J.P. (2003). Rapid acquisition of an alarm response by a neotropical primate to a newly introduced avian predator. Proc. R. Soc. Lond. B Biol. Sci. *270*, 605–610.

Hietanen, J.K., and Perrett, D.I. (1996). A comparison of visual responses to object- and ego-motion in the macaque superior temporal polysensory area. Exp. Brain Res. *108*, 341–345.

Hikosaka, K., Iwai, E., Saito, H.A., and Tanaka, K. (1988). Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. J. Neurophysiol. *60*, 1615–1637.

Kitchen, D.M., Seyfarth, R.M., Fischer, J., and Cheney, D.L. (2003). Loud calls as indicators of dominance in male baboons (*Papio cynocephalus ursinus*). Behavioural Ecology and Sociobiology *53*, 374–384.

Kuhl, P.K., and Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. Science *218*, 1138–1141.

Mistlin, A.J., and Perrett, D.I. (1990). Visual and somatosensory processing in the macaque temporal cortex: the role of 'expectation'. Exp. Brain Res. *82*, 437–450.

Neuhoff, J.G. (1998). Perceptual bias for rising tones. Nature *395*, 123–124.

Neuhoff, J.G. (2001). An adaptive bias in the perception of looming auditory motion. Ecological Psychology *13*, 87–110.

Rosenblum, L.D., Carello, C., and Pastore, R.E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. Perception *16*, 175–186.

Rosenblum, L.D., Wuestefeld, A.P., and Saldana, H.M. (1993). Auditory looming perception—influences of anticipatory judgments. Perception *22*, 1467–1482.

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. Anim. Behav. *58*, 921–931.

Schiff, W. (1965). Perception of impending collision—a study of visually directed avoidant behavior. Psychol. Monogr. *79*, 1–26.

Schiff, W., and Oldak, R. (1990). Accuracy of judging time to arrival—effects of modality, trajectory, and gender. J. Exp. Psychol. Hum. Percept. Perform. *16*, 303–316.

Schiff, W., Caviness, J.A., and Gibson, J.J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of "looming". Science *136*, 982–983.

Seifritz, E., Neuhoff, J.G., Bilecen, D., Scheffler, K., Mustovic, H., Schachinger, H., Elefante, R., and Di Salle, F. (2002). Neural processing of auditory looming in the human brain. Curr. Biol. *12*, 2147–2151.

Spelke, E.S., Born, W.S., and Chu, F. (1983). Perception of moving, sounding objects by four-month-old infants. Perception *12*, 719–732.

Stecker, G.C., and Hafter, E.R. (2000). An effect of temporal asymmetry on loudness. J. Acoust. Soc. Am. *107*, 3358–3368.

Stein, B.E., Meredith, M.A., Huneycutt, W.S., and McDade, L. (1989). Behavioural indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. J. Cogn. Neurosci. *1*, 12–24.

Tinbergen, N. (1951). The Study of Instinct (London: Clarendon Press).

Walker-Andrews, A.S. (1994). Taxonomy for intermodal relations. In The Development of Intersensory Perception: Comparative Perspectives, D.J. Lewkowicz and R. Lickliter, eds. (Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers), pp. 39–56.

Walker-Andrews, A.S., and Lennon, E.M. (1985). Auditory-visual perception of changing distance by human infants. Child Dev. *56*, 544–548.