Spectral Cues Explain Illusory Elevation Effects With Stereo Sounds in Cats

Daniel J. Tollin and Tom C.T. Yin
Department of Physiology, University of Wisconsin-Madison, Madison, Wisconsin 53706

Submitted 5 February 2003; accepted in final form 1 March 2003

Tollin, Daniel J. and Tom C.T. Yin. Spectral cues explain illusory elevation effects with stereo sounds in cats. J Neurophysiol 90: 525–530, 2003; 10.1152/jn.00107.2003. Mammals localize sound sources in azimuth based on two binaural cues, interaural differences in the time of arrival and level of the sounds at the ears. In contrast, the cue for elevation is based on patterns of the broadband power spectra at each ear that result from the direction-dependent acoustic filtering properties of the head and pinnae. Although the exact form of this “spectral shape” cue is unknown, most attention has been directed toward a prominent direction-dependent energy minimum, or “notch,” because its location in frequency, for both humans and cats, moves predictably from low to high as a source is moved from low to high elevations. However, there is little direct evidence that these spectral notches are important elevational cues in animals other than humans. Here we demonstrate a striking illusion in the localization of sounds in elevation by cats using stimulus configurations that elicit summing localization and the precedence effect that can be explained by spectral shape cues.

INTRODUCTION

The perception of horizontal sound location is based on two binaural acoustical cues, interaural time and level differences (Stevens and Newman 1936), while the direction-dependent shape of the broadband spectrum provides the acoustical cue for sound elevation (Bloom 1977; Butler and Belendiiuk 1977; Hebrank and Wright 1974; Musicant et al. 1990; Rice et al. 1992; Shaw 1982; Watkins 1978). However, in natural environments, sound from a single source does not arrive at the ears in isolation because only a small amount of the total sound energy comes directly from the actual source. The remaining sounds are delayed reflections from surrounding objects. Acoustic reflections should interfere with the localization of the source because each could be interpreted as a new source of sound. Despite scores of reflections, observers typically experience only a single “fused” sound image near the location of the source, with little influence from the spatial attributes of the reflections. This perceptual phenomenon has been referred to as the precedence effect (Haas 1951; Wallach et al. 1949), and the mechanisms that produce it are thought to be responsible for the ability to accurately localize sounds in reverberant environments (Litovsky et al. 1999).

Psychophysical investigations of the precedence effect have typically mimicked reflective environments by presenting sounds from loudspeakers at two different locations but with a delay between their onsets (e.g., Fig. 1A, right); the delayed sound simulating a single reflection. We have been using such stimulus configurations to study the neural bases of sound localization (Litovksy and Yin 1998; Yin 1994). For over 70 yr, it has been appreciated that the resultant relative amplitudes and times of arrival at an observer’s two ears of sounds presented from two loudspeakers determines the apparent azimuth and this forms the basis for the perceptual illusion of stereo sound (Bauer 1961; Blauert 1997; Leakey 1959; Snow 1954). However, what about the apparent elevation of these sounds?

While there is behavioral evidence that humans use spectral cues for elevation localization (Bloom 1977; Butler and Belendiiuk 1977; Hebrank and Wright 1974; Middlebrooks 1992; Watkins 1978), the possibility that cats or any other nonhuman species could use spectra has only been inferred based on theoretical (Neti et al. 1992; Rice et al. 1992), neurophysiological (Imig et al. 2000; Xu et al. 1999; Young et al. 1992), and anatomical (May 2000) evidence. Here we show that cats not only experience stereo sound in two-dimensional space for both horizontally- and vertically-placed sources but also that the apparent elevation of such sounds is consistent with their use of the spectral shape cue.

METHODS

Under aseptic surgical conditions, three adult female cats were each outfitted with a head holder, eye coils, and coil connectors. During experimental sessions, cats were placed in the center of a dimly-lit to dark sound-attenuating chamber with their heads held fixed facing a bank of 15 loudspeakers (RadioShack model 40–1310B) arranged horizontally, vertically, and diagonally along an arc 62 cm from the head and hidden from view by a translucent cloth. A red LED was suspended over the center of each speaker, and when illuminated, could be easily seen through the cloth. Cats were trained using operant conditioning (Populin and Yin 1998a) to indicate via saccadic eye movements the apparent two-dimensional location of various auditory and visual targets placed within their ocular-motor range (approximately ±25°). Eye position was recorded using the scleral search coil technique. All procedures used were approved by the University of Wisconsin Animal Care and Use Committee and also complied with the National Institutes of Health guidelines for animal use.

The acoustic stimuli consisted of a train of five identical broadband (approximately 1.5–40 kHz) noisebursts, each 10 ms in duration, and gated by a rectangular window, presented at a rate of 5 Hz for a total of 1 s. This stimulus was presented either from single speakers (single source condition) or from two different speakers connected in phase (paired source condition) but with a delay, the inter-stimulus delay (ISD), between the onsets. In the paired source condition, the two stimuli were delivered at the same level. The overall level of each

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Address for reprint requests: D. J. Tollin, Dept. of Physiology, Rm. 290 Medical Sciences Bldg., Univ. of Wisconsin-Madison, 1300 University Ave., Madison, WI 53706 (E-mail: tollin@physiology.wisc.edu).

www.jn.org 0022-3077/03 $5.00 Copyright © 2003 The American Physiological Society
The dependent variable was the final two-dimensional eye position after the eye movement to the acoustic target. Data were taken from only those trials for which the initial LED was at (0°,0°) to ensure that the eyes were centered in the orbit and that the pinnae were in a standard “ready” position (Populin and Yin 1998b). To initiate a trial, the cats visually fixated the initial LED within ±4° for a variable time (approximately 500–1,000 ms) whereupon the LED would be extinguished, and the acoustic target, in either the single or paired source configuration, would be presented. The cats were required to shift their gaze to the apparent location of the target. For single sources, if the eye position was maintained for 600–900 ms within a square electronic acceptance window of approximately ±8–16° around the target, the cat was presented with a food reward consisting of a puree of canned and soft cat food. On a small percentage (<5–10%) of the total number of trials on any given day, we presented the stimuli from paired sources. The paired source condition required us to alter our reward contingencies because we had no a priori expectations as to where in space the cats would orient. To encourage normal responding, we rewarded the cat approximately 1 s after the onset of the stimuli in the paired source conditions. There was a variable inter-trial interval of 10–15 s between each trial.

We used a velocity criterion (Populin and Yin 1998a) to determine when the eye movements began and ended by determining the time at which the magnitude of the movement velocity departed by 2 SD of the mean velocity computed during the fixation of the initial LED over a window spanning from 100 ms before to 30 ms after the onset of the acoustic target. Corrective saccades were considered provided they occurred within approximately 200 ms of the endpoint of the initial eye movement (Populin and Yin 1998a). All trials were used in the analysis of the data even if the cat was not “correct” as determined by our criteria, ensuring that the accuracy of the responses was not confounded by the size of the acceptance windows. For each cat, each data point for the single source conditions was based on an average of 71 ± 25 trials, while data points in the paired source conditions were based on an average of 19 ± 12 trials. The head-related transfer functions (HRTFs) used for the modeling were from animal A8727 (Musicant et al. 1990).

### RESULTS

Figure 1A (top) shows that cats perceive the horizontal aspect of stereo sounds similarly to humans (Bauer 1961; Blauert 1997; Leakey 1959; Mickey and Middlebrooks 2001; Snow 1954; Warncke 1941). The mean apparent location in azimuth, or response azimuth, for paired sources was dependent on the ISD provided the ISD was between about ±400 μs. For the smallest ISD, ±50 μs, response azimuths were near the midline horizontally but biased toward the leading loudspeaker. As ISD increased, response azimuths increased systematically toward the leading speaker until approximately 400 μs, where no further increase in response azimuth was apparent. The perception of a phantom image whose apparent location can be changed by adjusting the amplitudes and/or arrival times of similar sounds from two sources is often called summing localization (Warncke 1941). Since the other two cats tested performed similarly for each of the conditions, only data from one cat are shown for simplicity. The ISD range for summing localization in our cats was approximately ±400 μs and is indicated by the shaded regions in Fig. 1.

For ISDs from 400 μs ≤ 2 ms, the cats again performed similarly. However, unlike that found during summing localization their orienting responses were relatively independent of changes in ISD. Figure 1A shows that response azimuths (top) and elevations (bottom) were similar to the response azimuths and elevations for the single-source condition where the stimuli were presented at the “leading” locations alone. Although two sounds were actually presented, each of which was localized easily by each cat when presented from single sources (Fig. 1A, solid horizontal lines), the orienting responses suggested that the cats perceived the paired-source stimulus at only a single

---

**FIG. 1.** Apparent azimuths and elevations of sounds presented from single and paired sources. A: symbols and error bars show response azimuth (top) and elevation (bottom), ± SD, respectively, for 1 cat as a function of inter-stimulus delay (ISD) for horizontally placed paired sources at ±18°. Solid and dashed horizontal lines indicate mean response azimuth (top) and elevation (bottom) ± SD for the 2 single source locations. B: mean response elevations for sounds presented from single and paired sources in elevation on the median-sagittal plane at the indicated elevations (right). Symbols and lines as in A. Cats experienced the precedence effect for paired sources in azimuth and elevation, but summing localization (indicated by shaded regions) only in azimuth.
location in the vicinity of the leading source consistent with the precedence effect.

While the response azimuths for the horizontally placed paired sources were expected (e.g., Cranford 1982; Kalmykova 1993; Populin and Yin 1998a), we found a striking and unexpected phenomenon that has not been reported in cats and only rarely in human experiments, which was present only for ISDs encompassing summing localization and that disappeared for ISDs that elicited the precedence effect. While the response azimuths to the paired sources were consistent with the general expectations of summing localization and stereophonic perception, the apparent elevations were not (Fig. 1A, bottom). That is, the paired-source stimuli were not localized at the elevations of the sources themselves (e.g., 0° elevation), but instead the cats reported the paired sources as significantly elevated compared with the single sources along the horizontal plane. All cats showed this same illusion: the mean response elevations computed across the three cats in the paired source condition at the two smallest ISDs tested bounding 0 ms was 7.8 ± 1.47°, which was significantly higher than the −0.45 ± 2.59° averaged across cats for the two single sources [t(10) = 6.81, P < 0.0001].

When the paired sources were located in elevation along the median-sagittal plane with the same ISDs that encompassed horizontal summing localization, the cat showed a similar unexpected tendency to localize the sound above the location midway between the sources (e.g., 0° elevation), but instead the cats reported the paired sources as significantly elevated compared with the single sources along the horizontal plane. All cats showed this same illusion: the mean response elevations computed across the three cats in the paired source condition at the two smallest ISDs tested bounding 0 ms was 14.7 ± 1.88°. Interestingly, these data reveal for the first time that cats do not experience summing localization for sources in elevation. Rather, the cats always oriented toward higher elevations. Figure 1B also shows that the cats experienced the precedence effect with sources along the median sagittal plane for ISDs > 1 ms in that they oriented toward the azimuth (0°, data not shown) and elevation of only the leading source. The results with horizontally and vertically placed paired sources suggest that, over the range of ISDs tested here (0.4–2.0 ms), the precedence effect in cats is a general phenomenon of spatial hearing that can be elicited by both binaural disparities and spectral cues.

For small ISDs in both horizontal and vertical paired source conditions, the cats did not look to elevations in between the speakers as expected, but rather looked to more elevated locations. Here we present an explanation of this illusion based on the broadband spectral shape cues at the two ears in the paired source conditions. Figure 2A shows examples of the patterns of broadband spectra, often called HRTFs, at the left and right ears for three locations along the horizontal plane as recorded by Musicant et al. (1990). Figure 2B shows the HRTFs at the right ear for four elevations along the median-sagittal plane; for simplicity, we assumed the ears were symmetrical and plotted only the right-ear HRTFs. Note the presence of the deep spectral “notch” and how its location in frequency increases from approximately 10 to approximately 13 kHz as elevation changes from −13.5° to 18°. In cats, the frequency of this first prominent notch [F(N)F, Rice et al. (1992)] has been shown to vary not only with elevation (Fig. 2B), but also with changes in source azimuth (Musicant et al. 1990; Rice et al. 1992). For
example, Fig. 2A shows that as source azimuth is changed from \((-18^\circ, 0^\circ)\) to \((+18^\circ, 0^\circ)\), the FNf at the right ear (solid lines) increases from approximately 10 to approximately 12 kHz and in an opposing fashion at the left ear (dotted lines). Rice et al. (1992) first brought attention to the orderly progression of FNf with location changes in both azimuth and elevation in their HRTF measurements and suggested that FNf could potentially be used for localization. However, direct behavioral evidence that cats, or any other nonhuman species, use spectral cues, like FNf, for localization has heretofore been elusive.

To test the spectral cue hypothesis based on FNf, we first show in Fig. 3a the mean response elevations computed across three cats for single sources at four different elevations along the median-sagittal plane and the corresponding FNfs computed from the right-ear HRTFs at those same elevations. As the elevation of the single sources increases from \((0^\circ, -23^\circ)\) to \((0^\circ, 18^\circ)\), FNf increases from 8.6 to 13.13 kHz, and the response elevations of the cats increased accordingly. The clear correlation between FNf and response elevation for single sources in Fig. 3A follows logically from the spectral cue hypothesis and the general tendency of the cats to localize accurately but, on its own, does not necessarily implicate the FNf as the cue for elevation. However, the elevated percepts during summing localization provide a unique opportunity to test the hypothesis that cats use spectra for sound elevation because the spectra at the ears in both paired source conditions were different from those from either of the two different sources alone that were actually emitting the sounds.

Our explanation for the unexpected elevation effects can be most easily demonstrated by considering the spectra at the ears that result with an ISD of 0 ms in the horizontal paired source configuration (Fig. 3B). As illustrated in Fig. 3B, the resultant sounds arriving at any one ear in the paired source conditions come not from one but from two different source locations. The broadband spectral cues at that ear that would result were computed by simply adding, linearly in the time domain, the impulse responses of the HRTFs from the source ipsilateral to that ear (e.g., left ear, from speaker B) and that arriving a short time later (due to the interaural delay) from the source contralateral to that ear (e.g., left ear, from speaker A). The impulse responses interacted in phase and amplitude resulting in a spectrum unlike that of the two inputs themselves and Fig. 3B shows the Fourier transform of the sum plotted as the resultant gain for each ear. As expected, the binaural cues are

\[ \text{FIG. 3. Illusory apparent elevations for paired-sources are predicted by resultant spectral cues.} \]

A: mean response elevation computed across 3 cats to single-source sounds at the 4 different indicated locations on the median-sagittal plane as a function of the FNfs at those locations. B and C: resultant spectral cues in the horizontal and vertical paired-source conditions, respectively, with an ISD equal to 0 ms. Note that due to our assumption of symmetry, the left and right ear spectra overlap. The spectral shape cues that result, and the FNfs, are indicative of elevated sound sources. Vertical reference lines as in Fig. 2.
minimal indicating a source at the midline in azimuth. Consistent with this, Fig. 1A (top) shows that as ISD approached 0 ms, response azimuths also approached the midline. The spectral cues, most notably the FNf, however, do not specify a source at 0° elevation as expected given the physical locations of the sources themselves, but rather a more elevated source slightly below the source at +9° elevation, as shown in Fig. 2B. A similar analysis was performed for the paired sources in elevation (Fig. 3C) and the resultant spectra were also indicative of an elevated source above the horizontal plane near, but slightly above, the source located at +9° elevation.

To predict quantitatively the apparent elevations expected based on the resultant spectra in the two paired source conditions, we fit a regression line to the data in Fig. 3A relating the FNf of single sources in elevation to the psychophysically measured response elevations obtained at those same locations. Using the coefficients of the fit along with the FNf measured from the horizontally and vertically placed paired sources, 12.03 and 12.5 kHz, respectively, we computed response elevations estimates of 5.9° and 9.1°, respectively. While the actual mean response elevations across cats for the two paired source conditions were somewhat larger at 7.8° and 14.7°, the overall trends are consistent with the spectral cue model predictions. One likely reason for the discrepancy in the absolute values of the predictions is that broadband spectra and associated FNfs at the source locations used in our experiments (Fig. 2) actually vary from cat to cat (Rice et al. 1992; Xu and Middlebrooks 2000), yet we used HRTFs from only one cat in the model. But this fact does not compromise our explanation because the global patterns of FNf, in particular the orderly dependence of FNf on source azimuth and elevation, are very similar across cats (Musicant et al. 1990; Rice et al. 1992). Spectral cues in addition to FNf may also contribute, such as peaks and local slopes (Middlebrooks 1992). However, the important findings here are that 1) the cats oriented reliably to single sources at different elevations along the median-sagittal plane, 2) the FNf changes predictably for these single source locations, 3) the cats oriented to unexpected elevations in the paired source conditions, and finally, 4) the general trends in their behavior could be predicted based on spectral cues.

**DISCUSSION**

These findings are the first to support directly the hypothesis that cats use the particular direction-dependent shapes of the broadband spectrum at the ears to determine sound source elevation. More specifically, our findings are consistent with the hypothesis that the location in frequency of the first spectral “notch” is a critical cue for elevation and are in accord with previous observations that cats experience deficits in sound localization in elevation with spectrally restricted stimuli containing no energy in the frequency bands encompassing the FNfs (Huang and May 1996). While similar “elevation effects” have been reported in human psychophysical studies using similar stimulus configurations, explanations have generally not considered spectral cues (Bauer 1961; Blauert 1997; Damase 1969; de Boer 1947; Leakey 1959). Interestingly, unlike in the cat, where FNfs move prominently for changes in both source azimuth and elevation, the FNfs in humans increase rapidly primarily with increasing source elevation but change hardly at all from the FNf at 0° azimuth and elevation with increases in source azimuth ipsilaterally along the horizontal plane (Mehrgardt and Mellert 1977; Shaw 1982). Therefore while such illusory elevation effects are predicted for cats, they might be expected to be less apparent for human observers given these acoustical differences.

Spectral cues can account for the response elevations and azimuths for the simple conditions of single sources along the median-sagittal plane (Fig. 3A) and paired sources with an ISD of 0 ms. Both conditions result in minimal binaural difference cues that simplifies the problem by restricting the predictions to positions in elevation along the midline. But for larger ISDs, not only do the resultant binaural difference cues change, which is the basis for summing localization, but so too do the spectral cues due to the delay-and-add comb filtering effect resulting in not only different spectral shape cues at the two ears but also much more complex spectra (Tollin and Henning 1999). Since the exact “rules” by which observers, cat or human, combine the binaural and the spectral difference cues for two-dimensional location with broadband stimuli are not yet known (Middlebrooks 1992), particularly for the paired-source conditions studied here, complete explanations of our findings at all ISDs are not yet possible. Despite this limitation, our findings that the “elevation effects” are eliminated for ISDs beyond those encompassing summing localization are wholly consistent with prior phenomenological explanations of the precedence effect (Blauert 1997; Haas 1951; Litovsky et al. 1999; Wallach et al. 1949) in that the horizontal and, as we’ve shown here, the vertical spatial-location information of a sound’s source, based on binaural and spectral cues, respectively, is preserved in the face of later arriving and potentially conflicting reflections. And for the smaller ISDs encompassing summing localization, while the azimuth of the “phantom” image is determined by the interaural cues that result at the ears from the two sources, as we have also shown here the resultant spectral cues determine the elevation. The mechanisms that produce the precedence effect may have evolved to spare the spatial auditory system from having to interpret the complex spectral cues that result in the presence of acoustic reflections.

We are grateful to Drs. Fred Wightman and Micheal L. Dent for carefully reading and commenting on earlier drafts of this manuscript. The comments of two anonymous reviewers were also very helpful. We also acknowledge R. Kochhar and J. Sekulski for designing and implementing the analysis software and the data collection software, respectively.

This work was supported by National Institutes of Deafness and other Communicative Disorders Grants DC-00116 and DC-02840 to T.C.T. Yin and an Individual National Research Service Award DC-00376 to D. J. Tollin.

**REFERENCES**


