Responses of Auditory Cortical Neurons to Pairs of Sounds: Correlates of Fusion and Localization

BRIAN J. MICKEY AND JOHN C. MIDDLEBROOKS
Kresge Hearing Research Institute, University of Michigan, Ann Arbor, Michigan 48109-0506

Received 23 February 2001; accepted in final form 21 May 2001

Mickey, Brian J. and John C. Middlebrooks. Responses of auditory cortical neurons to pairs of sounds: correlates of fusion and localization. J Neurophysiol 86: 1333–1350, 2001. When two brief sounds arrive at a listener’s ears nearly simultaneously from different directions, localization of the sounds is described by “the precedence effect.” At inter-stimulus delays (ISDs) <5 ms, listeners typically report hearing not two sounds but a single fused sound. The reported location of the fused image depends on the ISD. At ISDs of 1–4 ms, listeners point near the leading source (localization dominance). As the ISD is decreased from 0.8 to 0 ms, the fused image shifts toward a location midway between the two sources (summing localization). When an inter-stimulus level difference (ISLD) is imposed, judgments shift toward the more intense source. Spatial hearing, including the precedence effect, is thought to depend on the auditory cortex. Therefore we tested the hypothesis that the activity of cortical neurons signals the perceived location of fused pairs of sounds. We recorded the unit responses of cortical neurons in areas A1 and A2 of anesthetized cats. Single broadband clicks were presented from various frontal locations. Paired clicks were presented with various ISDs and ISLDs from two loudspeakers located 50° to the left and right of midline. Units typically responded to single clicks or paired clicks with a single burst of spikes. Artificial neural networks were trained to recognize the spike patterns elicited by single clicks from various locations. The trained networks were then used to identify the locations signaled by unit responses to paired clicks. At ISDs of 1–4 ms, unit responses typically signaled locations near that of the leading source in agreement with localization dominance. Nonetheless the responses generally exhibited a substantial undershoot; this finding, too, accorded with psychophysical measurements. As the ISD was decreased from ~0.4 to 0 ms, network estimates typically shifted from the leading location toward the midline in agreement with summing localization. Furthermore a superposed ISLD shifted network estimates toward the more intense source, reaching an asymptote at an ISLD of 15–20 dB. To allow quantitative comparison of our physiological findings to psychophysical results, we performed human psychophysical experiments and made acoustical measurements from the ears of cats and humans. After accounting for the difference in head size between cats and humans, the responses of cortical units usually agreed with the responses of human listeners, although a sizable minority of units defied psychophysical expectations.

INTRODUCTION

The ability to localize sounds is found widely among animals, indicating its functional and evolutionary importance. Most vertebrates, including humans and cats, are thought to use the same acoustical cues for sound localization. Interaural time and intensity differences provide ample information about the azimuth of an isolated broadband sound source in the frontal hemisphere. Accordingly, humans and cats integrate these cues to localize single broadband sounds in a predictable way and with considerable accuracy (human: reviewed by Blauert 1997; Middlebrooks and Green 1991; cat: May and Huang 1996; Populin and Yin 1998). On the other hand, when multiple sounds arrive nearly simultaneously from different locations, the acoustical cues are confounded. Such sounds are generally not localized to their true locations. Investigation of how these sounds are represented in the brain might therefore provide valuable insight into the process of sound localization and stimulus coding.

When multiple sounds arrive in close succession, auditory mechanisms associated with the precedence effect are engaged (reviewed by Blauert 1997; Litovsky et al. 1999; Zurek 1987). If the inter-stimulus delay (ISD) separating two brief stimuli is below the echo threshold—about 3–8 ms for clicks, depending on the listener—only one sound is reported (Freyman et al. 1991; Thurlow and Parks 1961; Wallach et al. 1949). This perception is called fusion. Under most conditions, the fused spatial percept or image is fairly compact and localizable. The perceived location of a fused image depends systematically on the ISD. At ISDs in the range of ~1–5 ms, the reported location lies near the location of the leading sound (Chiang and Freyman 1998; Wallach et al. 1949); that phenomenon is called localization dominance. At those ISDs, the lagging sound is not heard as a separate sound and has relatively little influence on the location judgement. When the ISD is ~0.8 ms or less, in the domain of summing localization, both leading and lagging sounds strongly influence the perceived location (reviewed by Blauert 1997). The fused image falls between the two source locations and is biased toward the location of the leading sound. The presence of an inter-stimulus level difference (ISLD) also affects localization: at a given ISD, the location judgement is biased toward the more intense loudspeaker (Snow 1954). It follows that the shift due to an ISD can be compensated by applying an opposing ISLD—this balanc-

1 Although this particular observation is sometimes termed “the precedence effect,” we choose to follow the terminology of Litovsky et al. (1999) in which localization dominance is considered just one aspect of a broader set of phenomena called the precedence effect. Localization dominance is also commonly referred to as “the law of the first wavefront.”

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.
judgements had not been measured previously using these specific stimuli, we also performed human psychophysical experiments. We found that cortical units typically responded to paired sounds with a single burst of spikes. Spike patterns were analyzed with artificial neural networks to derive estimates of location that could be directly compared with psychophysical results. With notable exceptions, units signaled locations that, after accounting for the difference in head size between cats and humans, agreed with the responses of human listeners. Finally, we implemented a simple model that included peripheral filtering and interaural cross-correlation. The model results suggested that physiological correlates of localization dominance and time-intensity trading require central auditory processing beyond interaural cross-correlation.

METHODS

Animal preparation

Ten purpose-bred male young-adult cats (Harlan, Indianapolis, IN) with body weights ranging from 3.5 to 5.5 kg were used. All procedures complied with guidelines of the University of Michigan Committee on Use and Care of Animals. The animal preparation reviewed here was essentially identical to that detailed previously (Middlebrooks et al. 1998). Isoflurane anesthesia was used during surgery, and intravenous α-chloralose was used during unit recording. A skull opening ~1 cm in diameter exposed the middle ectosylvian gyrus of the right hemisphere. A plastic retainer was cemented to the ventral margin of the opening to create a recording chamber. The animal was positioned with its head in the center of the sound chamber, its body supported in a sling with a heating pad, and its head supported by a bar attached to a skull fixture. Thin wire supports held the pinnae symmetrically throughout the experiment. Experiments lasted 1–5 days and were ended when cortical responses became weak.

Physiological apparatus and stimulus generation

Physiological experiments were performed under free-field conditions with an apparatus that has been described previously (Middlebrooks et al. 1998). A sound-attenuating chamber (dimensions, 2.6 × 2.6 × 2.5 m) was lined with sound-absorbing foam to suppress reflections. A series of loudspeakers was positioned on a horizontal circular hoop. Loudspeakers were located 1.2 m from the cat’s head at various frontal azimuths (−80°, −60° to +60° in 10° steps, and +80°). The location directly ahead of the animal was assigned an azimuth of 0°, negative azimuths were to the left, and positive azimuths were to the right. Experiments were controlled by custom MATLAB software (The Mathworks, Natick, MA) running on a Pentium-based personal computer with instruments from Tucker-Davis Technologies (Gainesville, FL). Two-way coaxial loudspeakers (Pioneer TS-879 or JBL GT0302) were used. Computer-controlled two-channel D/A converters and multiplexers allowed sounds to be presented from single loudspeakers or from pairs of loudspeakers simultaneously, and two attenuators allowed the levels at the two loudspeakers to be varied independently.

Physiological experiments employed clicks, noise bursts, and pure-tone bursts. The maximal bandwidth of our system was 0.5–30 kHz. Because the loudspeakers generally differed in their detailed response properties, each loudspeaker was individually calibrated by obtaining an impulse response (Zhou et al. 1992). Click stimuli were created by convolution of a 100-μs rectangular pulse with the inverse impulse response of the intended loudspeaker. A 5-ms segment centered on the resulting transient was isolated, and 0.5-ms raised-cosine ramps were applied to the ends. About 80% of the energy of the click was concentrated within 100 μs. A few units were examined using 3-ms Gaussian noise bursts (with abrupt onsets and offsets), which also...
incorporated the loudspeaker correction. Each trial used an independently sampled token of noise; for paired noise bursts, the two tokens of noise were identical aside from the imposed ISD or ISLD. When stimulus pairs were presented with a nonzero ISD, an appropriate number of zeros was inserted in front of the waveform intended for the lagging loudspeaker. Stimulus waveforms were generated with 16-bit precision and a sampling rate of 100 kHz. During initial physiological characterization of units, we also delivered loudspeaker-corrected 80-ms Gaussian noise bursts (with abrupt onsets and offsets) and 80-ms pure-tone bursts (with 5-ms raised-cosine ramps applied to onsets and offsets).

The levels of single stimuli are expressed relative to unit threshold for a single stimulus presented from 0° azimuth. In all but a few cases (10 units), we matched the levels of paired stimuli ($L_s$ and $L_l$, for right and left loudspeakers, in dB) to the level of a single stimulus ($L_0$) by equalizing the sum of the amplitudes of the paired stimuli ($A_r$ and $A_l$) and the amplitude of the single stimulus ($A_s$).

$$A_s = A_r + A_l$$

For a desired ISLD, $L_r - L_l$, the inter-stimulus amplitude ratio is

$$r = A_r/A_l = 10^{(L_r-L_l)/20}$$

Solving these two equations for $A_r$ and $A_l$, and converting to levels in dB, we get the following relations

$$L_s = L_0 + 20 \log_{10} (r/(r+1))$$
$$L_l = L_0 + 20 \log_{10} (1/(r+1))$$

This procedure achieved the desired ISLD and also roughly matched the subjective loudness of the paired stimulus to that of the single stimulus.

**Unit recordings and spike sorting**

Unit activity was recorded extracellularly with silicon-substrate 16-channel probes (Anderson et al. 1989) that were provided by the University of Michigan Center for Neural Communication Technology. Each probe had a single shank with 16 recording sites arranged linearly at intervals of 100 or 150 μm. After probe insertion, the recording chamber was filled with silicon oil or with warm agarose (2% in Ringer solution) that subsequently solidified. To improve unit stability, we waited ≥30 min after probe placement before the start of recording. The activity at each site was amplified, digitized with 16-bit precision and a sampling rate of 25 kHz, sharply low-pass filtered <6 kHz, resampled at 12.5 kHz, and stored on the computer hard disk.

Spike sorting was performed off-line using custom software (Furukawa et al. 2000) based on principal components analysis of spike shape. The quality of unit isolation was characterized based on scatterplots of weights of the first two principal components and on histograms of inter-spike intervals. In a minority of cases, distinct spike waveforms were inferred to be from single neurons, but more often we recorded unresolved spikes that were inferred to be from two or more neurons, i.e., multi-unit clusters (see Furukawa et al. 2000 for illustrations of unit isolation). In the present study, all such single- and multi-unit recordings are collectively referred to as “units.” During initial screening, units were eliminated from further analysis if the mean spike rate (number of spikes per trial) across all conditions varied by more than a factor of two during the recording or the mean spike rate across all conditions was <0.5 sps per trial. When more than one distinct unit was isolated at a site, only the best-isolated single unit was retained for further analysis. Spikes of one neuron sometimes appeared at two adjacent recording sites, as indicated by sharp peaks near zero in histograms of between-unit spike times. We eliminated one member of each such pair. This paper describes 151 units that survived this screening. Sixteen units (11%) were reliably identified as single units; the remaining 135 units (89%) were multi-unit clusters. Figures 2, 3, and 7 of the present study show data from single units; Figs. 4, 5, and 6 represent multi-unit recordings.

**Determination of cortical area**

We recorded from three distinct cortical areas: A1, the dorsal zone of A1, and A2. Categorization of a unit was based on three factors: the unit’s frequency bandwidth based on responses to 80-ms pure tones (tested frequency range, 0.5–30 kHz, 1/3-octave steps; duration, 80 ms; azimuth, 0° or −80°); consistency with the expected tonotopic organization of area A1 (Merzenich et al. 1975; Reale and Imig 1980); and the unit’s location relative to recognized sulci and to other characterized units. A unit was judged to be narrowly tuned if its bandwidth at half-maximal spike rate was ≤1.33 octave at a level ≥40 dB relative to threshold at the best frequency. All narrowly tuned units were assigned to area A1, although we were not able to rule out the inclusion of some high-frequency units from field AAF. Units were designated as broadly tuned if they had multi-peakd-frequency response areas or if they had bandwidths ≥1.67 octave at a level ≤40 dB relative to threshold. Broadly tuned units located dorsal or dorsocaudal to area A1 were judged to be in the dorsal zone of A1 (Middlebrooks and Zook 1983; those located ventral to area A1 were judged to be in area A2. For some units, a cortical area could not be assigned because the tuning bandwidth could not be determined due to a best frequency near the limits of the frequencies tested, the bandwidth of the rate-frequency function was ambiguous, or the unit responded only weakly to pure tones. Of 151 units, 41 (27%; 3 single units) were from A1, 74 (49%; 12 single units) were from A2, 20 (13%; all multi-unit) were from the dorsal zone of A1, and 16 (11%; 1 single unit) could not be assigned a cortical area. Most of our units responded best to pure tones of high frequencies. Among units recorded from area A1, the median best frequency was 9.5 kHz (range 1.3–24 kHz), and 85% of units had best frequencies >6 kHz. For broadly tuned units recorded from area A2 and the dorsal zone of A1, we defined the half-maximal frequency band as the band of pure-tone frequencies that elicited a spike rate >50% of maximum for a level 20 dB above the lowest recorded level that gave a reliable response to any frequency. By this definition, the half-maximal frequency bands of 90% of units extended >6 kHz, the bands of 67% of units included frequencies between 2 and 6 kHz, and the bands of 32% of units extended <2 kHz.

**Physiological procedure**

We recorded from neurons in the middle ectosylvian gyrus of the right hemisphere. The probe was inserted approximately tangential to the surface of the cortex with the goal of placing all recording sites in active cortical layers. The penetration was usually oriented dorsoventrally but sometimes rostrocaudally. The number of sites with usable unit activity ranged across probe placements from 1 to 13 (median 4). The number of probe placements per animal ranged from 1 to 7 (median 3), totaling 34 across the 10 animals.

Search stimuli were 80-ms broadband noise bursts, typically presented from an azimuth of 0° at 30 dB SPL. After initial characterization of frequency tuning, single stimuli were presented from 0° at various levels in 5-dB steps, and unit thresholds were estimated on-line. These estimates guided the choice of stimulus levels. Units’ actual thresholds for 0° azimuth were later determined to the nearest 5 dB by inspection of raster plots off-line. After estimating thresholds on-line, we presented single stimuli from various azimuths (−80°, −60° to +60° in 10° steps, and +80°); and paired stimuli, one stimulus from each of a pair of loudspeakers at −50° and +50°. For paired stimuli, we varied the ISD, the ISLD, or both. The ISD ranged from −4 ms (left loudspeaker leading) to +4 ms (right loudspeaker leading), and the ISLD ranged from −30 dB (left loudspeaker more intense) to +30 dB (right loudspeaker more intense). Stimuli were
presented at two or three levels in steps of 10 dB, \( \sim 20 \)–40 dB above unit threshold. Sounds were delivered every 1.1–1.5 s in pseudorandom order such that all stimulus conditions were tested once before repeating all stimuli again in a different pseudorandom order. Each stimulus condition was repeated a total of 10, 20, or 40 times. We typically presented a block of 20 repetitions of each single-stimulus condition and a block of 10 repetitions of each paired-stimulus condition and then repeated each block once more. The blocks were interleaved to reduce the effects of any potential variation of neuronal responsiveness during the 2–4 h stimulus set.

Physiological data analysis

Spike times were expressed relative to the onset of D/A conversion. Therefore latencies include 3.5 ms of acoustic travel time. For paired stimuli, spike times were expressed relative to the onset at the leading loudspeaker. Because we found stimulus-evoked responses only at poststimulus times between 10 and 50 ms, only spikes occurring within this range were included in the analysis.

We employed artificial neural networks to recognize spike patterns and associate them with particular azimuths using methods similar to those described previously (Middlebrooks et al. 1998). This approach has the advantages that it produces an output (estimated azimuth) that is directly comparable with psychophysical results and it does not require assumptions about the information-bearing features of spike patterns (e.g., spike rate, first-spike latency, or other features). The first step in the procedure was, for each unit, to train a naive network with responses evoked by single stimuli from azimuths in the range \(-80\) to \(+80\); odd-numbered trials were used for training. To validate the procedure and evaluate that unit’s ability to code azimuth, the trained network was then tested with responses to single stimuli collected on even-numbered trials. Finally the trained network was presented with responses to paired stimuli with various ISDs and ISLDs, and the outputs of the network were taken as the azimuths that were signaled by the unit’s responses.

Networks were implemented with the MATLAB Neural Network Toolbox. Input to the networks consisted of bootstrap-averaged spike density functions (Middlebrooks et al. 1998), with four samples (trials) per bootstrap average. For analysis of individual units, 100 average spike density functions were created for each unit. For analysis of ensembles of units, a subset of units was drawn from the population (described in RESULTS), the spike density functions of these units were concatenated, and 40 average spike density functions were created. Network architecture and training were the same as described previously for individual units (Middlebrooks et al. 1998) and ensembles of units (Furukawa et al. 2000). Briefly, a single hidden layer contained four or eight units with hyperbolic tangent sigmoid transfer functions. The output layer had two units, representing the sine and cosine of azimuth, with linear transfer functions. The network was trained by backpropagation algorithm to adapt network weights and biases. The training was repeated three times, and the network with the smallest mean-squared error performance function and the resilient backpropagation algorithm to adapt network weights and biases. The training was repeated three times, and the network with the smallest centroid error (defined in the following text) was retained. This trained network was then presented with average spike density functions created from responses to paired stimuli at various levels, ISDs, and ISLDs, resulting in multiple estimates of azimuth for each paired-stimulus condition.

Since we were interested in aspects of azimuth coding that are largely independent of stimulus intensity, we analyzed unit responses to sounds that varied in level. Unit responses at two or three levels, 10 dB apart, were used to train networks. Networks were tested with responses at levels that were similar to the training levels. Levels of paired stimuli were matched to those of single stimuli as described under Physiological apparatus and stimulus generation.

Estimates of azimuth were characterized in the same way for physiological signaling of location and for psychophysical responses (described in Psychophysical methods). The central tendency of multiple azimuth estimates was represented by the centroid (i.e., the circular mean), which was computed by treating each estimate as a unit vector, forming the vector sum, and finding the direction of the resultant. To characterize the spread or variability of the data, we calculated the quartile deviation by expressing azimuth estimates as values within \( \pm 180\)° of the centroid, and finding the 25th and 75th percentile values of the distribution. Azimuth estimates falling within the quartile deviation constituted the central 50% of the data. When evaluating the accuracy of responses to single stimuli, we calculated the centroid error, which is the unsigned difference between the centroid and the true source azimuth, averaged across source azimuth. The centroid error serves as a single measure of overall accuracy but does not indicate bias in responses or variation of errors across azimuth. For psychophysical data, the centroid error was calculated over the source azimuth range \(-70\) to \(+70\)°. For psychophysical data, centroid error was calculated over a narrower source azimuth range of \(-60\) to \(+60\)° because artificial neural networks were almost always less accurate at the extremes of the training range (i.e., \(-80\) and \(+80\)°). Network estimates tended to fall near \(00\)° in the face of uncertainty (instead of, e.g., falling uniformly between the extremes of the training range), so the chance-level centroid error was \(32.3\)° (the mean of the absolute values of the azimuths tested). When evaluating responses to paired stimuli, we calculated the centroid difference, which is the unsigned difference between the centroid estimate and a psychophysical template (described in Acoustical measurements, computational model, and psychophysical templates), averaged across a specified range of ISD or ISLD. Like the centroid error, the centroid difference characterizes a unit’s responses with a single measure but does not indicate bias in responses or variation across stimulus conditions.

Psychophysical methods

For human psychophysical experiments, five paid listeners (age 18–30, 3 female, 2 male) were recruited from students and staff of the University of Michigan. All had normal hearing as determined by standard audiometric screening. Two of the listeners (S75 and S79) had brief previous experience with psychoacoustic tasks.

Psychophysical experiments were performed under free-field conditions using an apparatus similar to that described previously (Middlebrooks 1999). Each listener stood on a platform in a sound-attenuating anechoic chamber (dimensions 2.6 \( \times \) 3.7 \( \times \) 3.2 m). The chamber walls, floor, and ceiling were lined with fiberglass wedges and sound-absorbing foam. A headrest was positioned directly below the listener’s chin. Sounds were delivered from five two-way coaxial loudspeakers. A computer-controlled movable hoop with a radius of 1.2 m was equipped with two loudspeakers that could be positioned nearly anywhere on a spherical surface around the listener. In addition to the two movable loudspeakers, three stationary loudspeakers were located on the horizontal plane (i.e., 0° elevation); one at a distance of 1.6 m and an azimuth of 0° and two at a distance of 1.8 m and azimuths of \(-46\) and \(+46\)°. The latter two loudspeakers were hidden behind acoustically transparent black cloth so that listeners would not be aware of them. Experiments were controlled by custom MATLAB software running on a Pentium-based personal computer with instruments from Tucker-Davis Technologies. Computer-controlled two-channel D/A converters and multiplexers allowed sounds to be presented from single loudspeakers or from pairs of loudspeakers simultaneously, and two attenuators allowed the levels at the two loudspeakers to be varied independently. Click stimuli were generated as described above for physiological experiments except that the passband was 0.3–18 kHz for human experiments.

Listeners reported the apparent location of sounds by orienting their heads. An electromagnetic tracking system (Polhemus Fastrak, Colchester, VT) measured head orientation. Prior to participating in localization experiments, each listener was trained in the localization task. This procedure is detailed elsewhere (Macpherson and Middle-
and cats to characterize the proximal stimulus created at the loudspeakers at complete darkness. They were not told of the hidden stationary trading. Listeners stood in the center of the anechoic chamber in sessions designed to measure summing localization and time-intensity trading. The sound level varied randomly from trial to trial. To eliminate adventitious cues about the stimulus location, the stimulus, either a single click from one of the movable hoop loudspeakers or a click pair from the two stationary loudspeakers. The conditions under which this percept occurred were unclear since we didn’t systematically collect this information from listeners. Each trial was initiated when a continuous noise was presented from the centering loudspeaker. The noise cue the listener to face that loudspeaker, position his or her head on the chin rest, and press a button on a hand-held response box. The button press terminated the noise. One second following the button press, the listener was presented with the stimulus, either a single click from one of the movable hoop loudspeakers or a click pair from the two stationary loudspeakers. The listener then oriented his or her head to face the perceived location of the sound and pressed the response button, which triggered measurement of head orientation. The hoop was then positioned for the next trial. To eliminate adventitious cues about the stimulus location, the hoop was moved after each trial, even when the hoop position was the same for two consecutive trials. Following hoop movement, noise was presented from the centering loudspeaker and the cycle began again.

Within each session, the stimulus set consisted of single clicks and click pairs interleaved in pseudo-random order with 63–67% of stimuli being click pairs. The sound level varied randomly from trial to trial within a range of 40–50 dB above threshold. Single clicks were presented from azimuths of −70 to +70° in 10° steps. To avoid obstruction of the centering loudspeaker by the hoop, it was necessary to present single clicks from an elevation of 0° above the horizon. Click pairs were presented from loudspeakers at azimuths of −46° and +46° (1 click from each loudspeaker) with a variable ISD, a variable ISLD, or both. Each session employed one of four stimulus sets: variable ISD (range −0.8 to +0.8 ms) with ISLD = 0; variable ISD (range −1.4 to +1.4 ms) with ISLD = 0; variable ISLD (range −27 to +27 ms) with ISD = 0; and variable ISD (range −0.8 to +0.8 ms) and variable ISLD (−5, 0, and +5 ms). Each session lasted ~10 min, and listeners completed three to six sessions per day. Each listener completed 3–12 repetitions of each stimulus set for a total of 12–27 sessions.

**Acoustical measurements, computational model, and psychophysical templates**

We made physical acoustical measurements from the ears of humans and cats to characterize the proximal stimulus created at the eardrums by paired clicks at ISDs <1 ms. At these short delays, incident sound waves from the two sources superposed as they interacted physically with the head and pinnae, resulting in complex interaural time and level differences. We measured the directional impulse response for each ear and each source location by presenting broadband sounds and recording from the ear canals with miniature microphones, essentially as previously described (Middlebrooks and Green 1990; Xu and Middlebrooks 2000). Four hundred uniformly spaced source locations at various elevations and azimuths were used for the human measurements; 24 source azimuths confined to the horizontal plane (20° spacing for rear locations, 10° spacing for frontal locations) were used for the cat measurements. The proximal stimulus at each ear (of the stimulus at each eardrum was simulated by convolving the directional patterns with theGammatone filterbank (Slaney 1993), using center frequencies of 625 Hz to 2 kHz in 1/3-octave steps. High-frequency channels (>1 kHz) were full-wave rectified; all channels were then low-pass filtered (<1 kHz using a fourth-order Butterworth filter. Finally, for each critical band, the signals from the two ears were cross-correlated over a lag range of −20 to +20 ms, and the lag of the maximum of the cross-correlation function was taken as the output of the model. At the lowest center frequencies, cross-correlation functions often had multiple prominent peaks, which led to discontinuities in model output such as those shown in Fig. 10B. This computational model is based on previous models that employ interaural cross-correlation (reviewed by Stern and Trahoitis 1997) and resembles models developed to describe free-field localization of paired sounds (Blauert and Cobben 1978; MacPherson 1991; Pulkki et al. 1999). It differs somewhat from binaural models developed to describe lateralization of stimuli presented over headphones because those models do not include filtering by the head and pinnae (Gaskell 1983; Lindemann 1986; Tollin and Henning 1999; Zurek 1980).

We sought to compare quantitatively our physiological results to the responses of listeners. Ideally, one would compare cat physiological results with cat behavioral responses obtained under similar stimulus conditions. Such data are available (Populin and Yin 1998), but they are not sufficiently detailed for our purposes. We therefore chose to compare our physiological results to human psychophysical data, which are more readily available. To make this comparison, we used human psychophysical data to construct psychophysical standard curves, or templates, of azimuth versus ISD and azimuth versus ISLD. First, mean responses at each value of ISD or ISLD were averaged across the five listeners. The data were then symmetrized by averaging responses to stimuli that were symmetric with respect to the midline (e.g., the absolute value of the response at ISD = +0.4 ms was averaged with the absolute value of the response at ISD = −0.4 ms). The averaged symmetrized data were then fit with a logistic function

\[
\sin(\theta) = \sin(x)(10^a - 1)/(10^a + 1)
\]

where \(\theta\) is estimated azimuth, \(x\) is either ISD or ISLD, and \(a \) and \(b\) are fit parameters. This fitting procedure was chosen for three reasons: phasor analysis suggests a physical basis for this functional form for simultaneous low-frequency sounds with a variable ISLD (Bauer 1961); our psychophysical data showed a dependence on ISD and ISLD with the same general shape; and the procedure resulted in smooth curves that could be scaled to account for acoustical differences between cat and human (see following text). Fitting was performed in MATLAB using nonlinear optimization to minimize the squared error of \( \theta \). The resulting fit parameters based on human...
psychophysical responses were \( a = 31.2^\circ \) and \( b = 0.477 \) ms for azimuth versus ISD, and \( a = 39.8^\circ \) and \( b = 13.6 \) dB for azimuth versus ISLD. Finally, to create psychophysical templates for comparison with cat physiological data, two adjustments were made to these curves: the estimated azimuth was multiplied by a factor of \( \sin (50^\circ)/\sin (46^\circ) \) to account for the slightly different loudspeaker locations used in the cat experiments and for the ISD curve, the fit parameter \( b \) was divided by a factor of 1.64 to account for the smaller effective head size of a cat.

We calculated the factor of 1.64 as follows. Lag values were computed for azimuths of \(-60^\circ\) to \(+60^\circ\) in 10° steps using the cross-correlation model described in the preceding text. For each frequency band, the best-fitting line was determined by least-squares fitting; the slope (\( \mu s/° \)) and its variance were retained. Slopes were determined from both human and cat acoustics, and a ratio of slopes was calculated within each frequency band. This ratio varied irregularly with frequency. Finally, a weighted-mean ratio was computed across frequency bands; weighting was based on the variance of the slope obtained during least-squares fitting. Using acoustical data from one 4.5-kg cat, we found weighted-mean ratios of 1.62 and 1.67 for listeners S75 and S79; the average was 1.64. Given the inter-subject variability of interaural delays found among humans (Middlebrooks 1999) and cats (Roth et al. 1980), the factor of 1.64 should be viewed as a rough estimate. Nonetheless, considering that relatively large male cats were used in our study, this value is consistent with previous acoustical measurements of interaural delays in cats (Roth et al. 1980) and humans (Middlebrooks 1999; Middlebrooks and Green 1990).

**Results**

We performed human psychophysical experiments and cat physiological experiments under similar stimulus conditions. We first present the human psychophysical results and thereby demonstrate summing localization, time-intensity trading, and localization dominance using broadband click stimuli. Then we describe the responses of cortical neurons to the same stimuli and analyze these responses in a way that permits comparison to the psychophysical responses. Finally, we use a simple computational model to investigate the extent to which our physiological results might be explained by peripheral filtering and interaural cross-correlation.

**Psychophysics**

Each listener participated in a localization task in which single clicks and click pairs were presented and the listener oriented his or her head to face the perceived location of the sound. Figure 1 shows the responses of two individual listeners (Fig. 1, top and middle) and mean responses across five listeners (Fig. 1, bottom). As expected, when single clicks were presented from various frontal azimuths, listeners localized the sounds with considerable accuracy (Fig. 1, 1st column). To quantify localization accuracy, we calculated the centroid error, which is the unsigned error of the mean response at each target location, averaged across locations. The centroid error ranged from 3.8 to 5.8° among the five listeners tested. The quartile deviation, which is the range of azimuth that included half of a listener’s responses (see Methods), was used to characterize response variability (gray areas in Fig. 1). Values for single clicks, averaged across source azimuth, ranged from 9 to 16° among the five listeners.

Trials using paired clicks were randomly interspersed with trials using single clicks. Paired clicks were always presented from two loudspeakers located 46° to the left and right of midline; one click was presented from each loudspeaker. The ISD, the ISLD, or both ISD and ISLD were varied. When click pairs were presented at equal intensity with a variable ISD, listeners localized click pairs to intermediate azimuths (Fig. 1, 2nd column). When the clicks were simultaneous (ISD = 0), listeners pointed near 0°. As the magnitude of the ISD was increased, listeners’ judgements shifted laterally, reaching a maximum at an ISD of \( \pm 0.8 \) ms. At ISDs of 1.0–1.4 ms, listeners pointed near the leading loudspeaker, but all exhibited an appreciable undershoot. After compensating for small biases in responses to single clicks at azimuths of \( \pm 40^\circ \) and \( \pm 50^\circ \), the mean undershoots were 7, 22, 18, and 18° for the four listeners tested at these delays. The variability in listeners’ responses with this stimulus set, as measured by the quartile deviation averaged across stimulus conditions and listeners, was 13.8° for click pairs compared with 11.6° for single clicks tested during the same sessions.

When click pairs were presented simultaneously and only the ISLD was varied (Fig. 1, 3rd column), listeners’ responses depended systematically on the ISLD. When the ISLD was zero, listeners pointed near 0°. As the absolute ISLD was increased, listeners pointed to increasingly lateral azimuths, in the direction of the more intense loudspeaker. Beyond a level difference of \( \pm 15 \) dB, listeners’ estimates fell near the more intense source. At ISLDs beyond 15 dB, listeners commonly undershot the more intense source, but the undershoot was somewhat smaller than it was when the ISD was varied. Curiously, the averaged quartile deviation with this stimulus set was smaller for click pairs (9.3°) than for single clicks tested during the same sessions (13.4°).

When both a delay and a level difference were imposed, both variables influenced listeners’ responses. When an ISLD of \( +5 \) dB was superposed on an ISD, so that the loudspeaker to the right (\( +46^\circ \)) was more intense (Fig. 1, 4th column, top curve), judgements shifted toward the right (upward in the figure). Similarly, when the ISLD was \(-5 \) dB (bottom curve), so that the left loudspeaker (\(-46^\circ \)) was more intense, judgements shifted toward the left (downward). Thus at a given ISD, a nonzero ISLD biased listeners’ responses toward the more intense loudspeaker. Alternatively, at each ISLD tested (each curve in the figure), a nonzero ISD biased responses toward the leading loudspeaker. Although responses varied among listeners, when the ISD was 0.6–0.8 ms, an opposing ISLD of 5 dB generally shifted judgements back to the midline (Fig. 1L). The averaged quartile deviation with this stimulus set was 18.8° for click pairs compared with 13.2° for single clicks tested during the same sessions.

**Physiology**

We recorded spike activity of single units and multi-units from areas A1 and A2 on the right side of anesthetized cats while presenting click stimuli. Single clicks were presented from various frontal azimuths; paired clicks were presented from \(-50^\circ \) and \(+50^\circ \) while varying the ISD, the ISLD, or both.

We first characterized each unit’s responses to single clicks. Most units exhibited some sensitivity to sound-source azimuth. Figure 2A shows a raster plot of responses of one such unit. This unit responded more strongly to clicks from locations contralateral to the recording site, and less strongly to ipsilateral clicks (Fig. 3A). To directly compare neuronal responses to...
the responses of psychophysical listeners in a localization task, we analyzed spike patterns using an artificial neural network as a general-purpose pattern-recognition algorithm. The network recognized location-specific spike patterns and produced estimates of sound-source location. For each unit, a naive network was trained to associate spike patterns recorded on odd-numbered trials with various frontal azimuths. Then the trained network, when presented with the spike patterns recorded on even-numbered trials, produced estimates of source azimuth. Such an analysis of the unit represented in Figs. 2A and 3A resulted in the estimates of azimuth shown in Fig. 3B. Network estimates fell near the perfect performance line for most locations, indicating that this unit’s responses reliably signaled the locations of single clicks. The centroid error was 8.2°—worse than psychophysical values but significantly better than the chance level of 32.3° (see METHODS).

This analysis was applied to each of the 149 units of our unit population that were tested with click stimuli. For the 61 units examined at two stimulus levels, the centroid error ranged from 7.4° (near psychophysical values) to 32.5° (near chance levels) with a median value of 18.3°. For the 88 units examined at three stimulus levels, the centroid error ranged from 7.3 to 32.3° with a median value of 16.4°. These distributions did not differ significantly (P = 0.56, χ² test), so we pooled the two groups of units together for subsequent analyses. Centroid errors for single units (median, 14.9°; range, 7.3 to 32.5°) were similar to those for multi-unit recordings (median, 17.2°; range, 7.9 to 32.5°).

Unit responses to click pairs at various ISDs resembled the responses to single clicks at various azimuths. The unit described in Figs. 2 and 3, for example, typically responded with a single spike or burst of spikes when paired clicks were presented from -250 and +250° (Fig. 2B). At negative ISDs (-50° loudspeaker leading), this unit responded more strongly, resembling the response to a single click from a contralateral location (Fig. 3C). At ISDs near zero, the unit responded with fewer spikes. At large positive ISDs (+50° loudspeaker leading), the spike rate was reduced, resembling the response to a

FIG. 1. Psychophysical demonstration of summing localization, time-intensity trading, and localization dominance. Human subjects listened to single clicks from various frontal azimuths and click pairs from loudspeakers at -46 and +46° with a variable inter-stimulus delay and level difference (ISD and ISLD). The ordinate of each panel represents listeners’ judgements of location.

○, the centroid (circular mean) response; □, quartile deviations, which contain 50% of the responses. Horizontal lines indicate the source locations for paired clicks. First column: responses to single clicks as a function of azimuth. Second column: responses to paired clicks with a variable ISD and 0 ISLD. Third column: responses to simultaneous paired clicks with a variable ISLD. Fourth column: both the ISD and the ISLD of paired clicks were varied: ISLDs of -5 dB (bottom curve), 0 dB (middle curve), and +5 dB (top curve) were superposed at each ISD. (Quartile deviations have been omitted for clarity.) Top and middle show responses for 2 of 5 listeners. Bottom shows the mean of centroid responses across 5 listeners; error bars represent SDs. For clarity, error bars have been omitted for non-0 ISLDs in L. The asymmetry evident in K was not found consistently across subjects.
single click from an ipsilateral location. That is, a leading click from +50° suppressed the response to a lagging click from −50°. We analyzed the responses to click pairs using the artificial neural network that had been previously trained with 3-ms noise bursts and noise-burst pairs. Six of the eight units examined with noise bursts were among those examined with clicks. In response to single or paired noise bursts, units typically fired a brief burst of spikes. Responses to single noise bursts signaled source location with accuracy similar to that for clicks: among the eight units, centroid errors ranged from 9.4 to 26.0° (median, 16.0°). Units generally responded to paired noise bursts in a manner consistent with summing localization and localization dominance; analysis of one such unit is shown in Fig. 5. This finding indicated that units’ signaling of location generalized to another type of broadband stimulus.

In addition to sensitivity to ISD, cortical units showed sensitivity to the ISLD of click pairs. Figure 6 shows artificial-neural-network analysis of one unit. When click pairs were presented simultaneously (ISD = 0) with a variable ISLD, the unit signaled locations near the midline at small ISLDs, and locations closer to the more intense loudspeaker at greater absolute ISLDs (Fig. 6B). Conversely, when click pairs were presented at equal intensity (ISLD = 0) with a variable ISD, the unit signaled locations on the side of the leading loudspeaker (Fig. 6C, central curve). When a nonzero ISLD was superposed at a given ISD, both the ISD and the ISLD influenced network estimates (Fig. 6C). An ISLD biased network estimates toward the more intense loudspeaker. That is, the curve shifted downward (toward the left loudspeaker) when the ISLD was negative (left loudspeaker more intense) and upward when the ISLD was positive. Alternatively, at each ISLD (for each curve), introducing a nonzero ISD generally shifted network estimates toward the leading loudspeaker. The shift was particularly evident at ISDs between −1 and +1 ms. The flattening of the curves in Fig. 6C may be attributed to the severe undershoot seen in response to single clicks at ±50° (Fig. 6A); this effectively reduced the range of azimuth that was accessible to the network. Nonetheless the activity of this unit was at least qualitatively consistent with time-intensity trading.

Among units that accurately localized single stimuli, a sizable minority responded to paired stimuli in ways inconsistent with localization dominance and summing localization. As described in the preceding text, the responses of the unit represented in Fig. 4 agreed with psychophysical results at most delays but not when the ISD was between +1.4 and +4 ms (Fig. 4D). An even clearer contrary example is shown in Fig. 7. This unit responded to single clicks from all azimuths tested with a preference for contralateral locations, and accurately signaled source location, with a centroid error of 7.9° (Fig. 7, A and B). Nonetheless, the unit showed little sensitivity to the ISD of click pairs (Fig. 7, C and D). Furthermore, when click pairs were presented simultaneously with a variable ISLD, this unit responded strongly at most ISLDs tested, showing a decrease in response only at ISLDs greater than +15
The results are shown in the scatterplots of Fig. 8. In each steps. The centroid difference for each unit was then plotted against the unit’s centroid error for localizing single stimuli. The centroid error for this unit was 8.2°.

We used the following procedure to quantify the extent to which each unit’s responses agreed with psychophysical measurements of summing localization and localization dominance. First, we constructed psychophysical templates of azimuth versus ISD and azimuth versus ISLD (solid curves, Fig. 8, A–C, insets). The templates were based on our human psychophysical results and scaled according to physical acoustical measurements from humans and cats (see METHODS). Then for each unit, we found the centroid difference—the unsigned difference of the mean physiological response from the psychophysical template at specified values of ISD or ISLD (○, Fig. 8, A–C, insets), averaged across ISD or ISLD. The centroid difference is a measure of the overall disagreement of average physiological responses from the psychophysical template. ISD-based summing localization was evaluated at ISDs of −0.4, −0.2, 0, +0.2, and +0.4 ms; localization dominance at ISDs of −3, −2, −1, +1, +2, and +3 ms; and ISLD-based summing localization at ISLDs from −18 to +18 dB in 3-dB steps. The centroid difference for each unit was then plotted against the unit’s centroid error for localizing single stimuli. The results are shown in the scatterplots of Fig. 8. In each panel, the various types of symbol represent units recorded from specific cortical fields; ▲, ○, ■, ● signify units used as examples in other figures. Each of the centroid-difference measures showed a significant correlation with the centroid error (ISD-based summing localization: $r^2 = 0.085$, $P < 0.01$; localization dominance: $r^2 = 0.61$, $P < 0.001$; ISLD-based summing localization: $r^2 = 0.16$, $P < 0.02$). The positive correlation of the two measures indicates that units that localized single stimuli most accurately also tended to show the strongest summing localization and localization dominance. Nonetheless, among units that accurately localized single sounds, a sizable minority showed weak ISD-based summing localization, as indicated by symbols in the upper left quadrant of Fig. 8A. This minority of units contributed to the weak correlation between centroid difference and centroid error noted in the preceding text. A smaller proportion of units that accurately localized single sounds failed to show localization dominance (Fig. 8B, top left quadrant) or ISLD-based summing localization (Fig. 8C, top left quadrant). By this analysis, no consistent differences were found between cortical areas or between responses to clicks and responses to 3-ms noise bursts (also plotted in Fig. 8, A and B).

To determine the locations that were signaled by our unit population as a whole, we performed artificial-neural-network analyses of small ensembles of units. Similar analyses have shown that ensemble networks more accurately classify unit responses to single broadband noise bursts than do individual-
FIG. 4. Responses of a unit to single clicks and paired clicks. The conventions of Fig. 3 are used. A: the spike rate is plotted vs. azimuth for 40 repetitions at a level 20 dB above threshold. B: spike patterns obtained at 10, 20, and 30 dB above threshold were analyzed using an artificial neural network (centroid error: 12.9°). C: the spike rate is plotted vs. ISD for 40 repetitions at a level 20 dB above threshold. D: spike patterns obtained at 10, 20, and 30 dB above threshold were analyzed using an artificial neural network. This unit responded only weakly to pure tones and was located ventral to area A1. Thus the unit was likely in area A2, but by our criteria, the cortical area could not be determined.

FIG. 5. Responses of a unit to single 3-ms noise bursts and paired 3-ms noise bursts. The conventions of Fig. 3 are used. A: the spike rate is plotted versus azimuth for 40 repetitions at a level 25 dB above threshold. B: spike patterns obtained at 25, 35, and 45 dB above threshold were analyzed using an artificial neural network (centroid error: 9.4°). C: the spike rate is plotted vs. ISD for 20 repetitions at a level 25 dB above threshold. D: spike patterns obtained at 25, 35, and 45 dB above threshold were analyzed using an artificial neural network. This unit was located in cortical area A2; it responded most strongly to pure-tone frequencies of 7.5 to 19 kHz.
unit networks (Furukawa et al. 2000). Furthermore, we have found that ensemble networks are particularly accurate for azimuthal targets near the extremes of the training range (data not shown). Note that ensemble analysis takes into account all units in the ensemble, including those that contradict psycho-physical predictions such as the unit in Fig. 7. Among our unit population, 123 units were tested with a variable ISD only, 16 were tested with a variable ISD and a variable ISLD, and 46 were tested with a variable ISLD only. For each of these three subpopulations, we selected the 25% of the subpopulation with the lowest centroid errors (derived from responses of individual units to single clicks) for use in ensemble analysis. This selection process resulted in three ensembles consisting of 30, 4, and 11 units, respectively. Networks were trained with a set of ensemble responses to single clicks, and validated by testing with an independent set of ensemble responses to single clicks (Fig. 9, A–C, insets). We found that unit ensembles signaled the locations of single clicks with considerable accuracy: centroid errors for the respective ensembles were 8.1, 8.2, and 8.3°. When the first ensemble was tested with paired clicks at...

**Fig. 6.** Artificial-neural-network analysis of the responses of a unit to single clicks and paired clicks. The conventions of Fig. 3, B and D, are used. A: unit responses to single clicks at 15, 25, and 35 dB above threshold were analyzed (centroid error: 12.9°). B: analysis is shown for unit responses to simultaneous paired clicks at 15, 25, and 35 dB above threshold. Network output is plotted vs. ISLD. C: unit responses to paired clicks at 15, 25, and 35 dB above threshold with various ISDs and ISLDs were analyzed. The centroid of the network output is shown as a function of ISD; quartile deviations have been omitted for clarity. From top to bottom, the 5 curves correspond to ISLDs of +18, +9, 0, −9, and −18 dB. This unit is the same one depicted in Fig. 5.

**Fig. 7.** Responses of a unit to single clicks and paired clicks. The conventions of Fig. 3 are used. Left: responses to single clicks presented from various frontal azimuths. Middle: responses to paired clicks with a variable ISD (and 0 ISLD). Right: responses to paired clicks with a variable ISLD (and 0 ISD). A: the spike rate is plotted vs. azimuth for 20 repetitions at a level 25 dB above threshold. B: spike patterns obtained at 15, 25, and 35 dB above threshold were analyzed using an artificial neural network (centroid error: 7.3°). C: the spike rate is plotted vs. ISD for 20 repetitions at a level 25 dB above threshold. D: spike patterns obtained at 15, 25, and 35 dB above threshold were analyzed using an artificial neural network. E: the spike rate is plotted vs. ISLD for 20 repetitions at a level 25 dB above threshold. F: spike patterns obtained at 15, 25, and 35 dB above threshold were analyzed using an artificial neural network. This unit was located in cortical area A1; the best frequency was 24 kHz.
various ISDs, the signaled locations shifted from the midline toward the leading loudspeaker as the magnitude of the ISD was increased from 0 to 0.4 to 0.6 ms (Fig. 9A). At greater delays, network estimates fell near, but somewhat short of, the location of the leading loudspeaker. This undershoot was prominent at most delays, it was greater when the ipsilateral source led, and it could not be accounted for by an undershoot in response to single clicks (Fig. 9A, inset). The thin curve in Fig. 9A represents the psychophysical template (described in the preceding text); the network estimates fell close to the predicted at negative ISDs (contralateral leading) but less so at positive ISDs. The second ensemble was tested with click pairs over a range of ISDs at five ISLDs; each ISLD is represented by a distinct curve in Fig. 9B. At a given ISD, superposition of an ISLD biased network estimates toward the more intense source. The bias was notably asymmetric, being stronger at negative ISDs than at positive ISDs. The third ensemble, in response to paired clicks at various ISLDs, signaled locations spanning −50 to +50° (Fig. 9C). Network estimates reached an asymptote when the absolute ISLD reached 15–20 dB, and they showed little undershoot at extreme ISLDs. This ensemble signaled locations that agreed well with the psychophysical template (Fig. 9C, thin curve).

**Acoustical modeling**

When a listener is presented with paired sounds such as those used in the current study, rather complex acoustical cues can result. Several factors transform the signal before it even reaches the nervous system: summation of sound waves in the air, interaction of the resultant sound waves with the head and pinnae, and critical-band filtering in the cochleae. Under these conditions, it is important to distinguish the proximal interaural time differences (ITDs) and interaural level differences (ILDs) from the applied ISDs and ISLDs. For instance, because of phasor addition, paired sounds presented simultaneously (ISD = 0) with a nonzero ISLD will produce a nonzero ITD at low frequencies (Bauer 1961; Blauert 1997). Given such complications, we wondered to what extent the phenomena we studied might be explained by relatively well-described auditory processes such as acoustical filtering by the head and pinnae, cochlear filtering, and sensitivity of brain stem circuits to ITDs. We expected that summing localization would be explicable by such mechanisms, but we wondered, in particular, whether time-intensity trading and localization dominance might also be explained in this way.

We investigated the degree to which physiological correlates of summing localization, time-intensity trading, and localization dominance could be explained by a simple computational model. The model was based on simple superposition of sounds in the sound field, peripheral filtering, and delay-line cross-correlation of inputs to the two ears (see METHODS). We reasoned that after accounting for these factors, any physiological results that remained unexplained by the model were likely to arise from central auditory processing beyond interaural cross-correlation. For given binaural input waveforms,
the model produced a frequency-dependent interaural lag. Figure 10 shows model output for three representative frequency bands. As expected, single clicks produced an interaural lag that shifted systematically with the azimuth of the source, regardless of the frequency band (Fig. 10, left).

Paired clicks, on the other hand, produced outputs that were more varied and frequency dependent. Three frequency domains, with boundaries at ~2 and ~6 kHz, were distinguished on the basis of the model results. At the lowest frequencies (0.6–2 kHz), the interaural lag showed a systematic dependence on ISLD (Fig. 10C) as expected from the phase differences that are generated by ISLDs at low frequencies (Bauer 1961; Blauert 1997). In contrast, when the ISD was varied, the interaural lag tended to jump unstably at low frequencies (Fig. 10B) due to the presence of multiple nearly equal peaks in cross-correlation functions. At middle frequencies (2–6 kHz), the ISD influenced the interaural lag in a rather unstable manner, the dependence on ISLD was somewhat less than the dependence at low frequencies, and a weak effect that resembled time-intensity trading was appreciated (Fig. 10, E and F). At high frequencies (6–20 kHz), the ISD dominated and the ISLD had relatively little influence on interaural lag (Fig. 10, H and I).

Because nearly all the units we studied were most sensitive to middle or high frequencies, the model appeared to account for physiological correlates of ISD-based summing localization (Fig. 10, E and H). In contrast, the model did not appear to explain time-intensity trading. Many of the units we studied—and in particular, the units in Figs. 6 and 9B that showed correlates of time-intensity trading—did not respond to pure tones <6 kHz. Assuming that these units were not influenced by frequencies <6 kHz, if unit responses were based only on peripheral filtering and interaural cross-correlation, the responses would be largely insensitive to ISLD, as shown in Fig. 10, H and I. Because these units did show ISLD sensitivity, to explain their responses one must invoke brain mechanisms beyond interaural cross-correlation (e.g., processing of interaural level differences). The model also failed to account for localization dominance. At ISDs beyond ~0.4 ms, the interaural lag did not approach an asymptote in any frequency band, but continued to increase beyond lag values encountered with single clicks (data not shown).

Discussion

We have studied spatial signaling by cortical neurons in response to pairs of sounds that are known to fuse perceptually. With notable exceptions, most units signaled locations that would be reported by a listener. We begin this section by discussing the psychophysical findings in our study and in studies by other groups. We then compare our physiological results to the psychophysical predictions and to the results of previous physiological studies.

Psychophysics

Because many aspects of the precedence effect appear to be sensitive to the type of stimulus used (Blauert 1997; Litovsky et al. 1999), we felt it was important to compare our physiology to psychophysical results obtained in a localization task using the same stimuli. Therefore we performed human localization experiments using the same free-field click stimuli as were used in the physiological experiments. Furthermore the listeners performed a task in which they made absolute judgements of location (rather than a task based on discrimination). Thus we were able to directly compare artificial-neural-network
estimates to psychophysical listeners’ estimates because both are continuous measures in degrees of azimuth.

Our results largely support previous psychophysical studies of summing localization (reviewed by Blauert 1997). Of past studies of summing localization, the most comparable is one by Wendt (1963). He presented broadband impulses from loudspeakers at $-30^\circ$ and $+30^\circ$ azimuth and varied the ISD or ISLD. Wendt found that as the ISD was varied from 0 to $0.5\text{ ms}$, the reported azimuth progressed from 0 to $20^\circ$. At delays of 0.5–1 ms, the image location moved slightly more laterally toward $-25^\circ$. Consistent with these results, we found that an ISD in the range of 0–0.8 ms systematically shifted location judgements toward the leading source. When Wendt presented impulses simultaneously with a variable ISLD, the perceived azimuth shifted systematically from the midline toward the more intense loudspeaker, reaching nearly 90% of the distance to the more intense source at an ISLD of 20 dB. Our ISLD curves are in close agreement with these results.

Previous studies of time-intensity trading have generally used a procedure whereby the ISLD of a stimulus pair is adjusted to offset a constant ISD, thereby moving the sound image to the midline (e.g., Chiang and Freyman 1998; Snow 1954). In contrast, we measured listeners’ judgements of location when an ISLD was superposed at a given ISD. Nonetheless, some rough comparisons are possible. A report by Snow (1954) used paired click trains with loudspeakers arranged at $-27^\circ$ and $+27^\circ$. He found that ISLDs of $-5$ to $-8$ dB were required to offset ISDs of $+0.45$ to $+1.8$ ms (time-intensity trading ratio, $0.06–0.40$ ms/dB). In a more recent study, Chiang and Freyman (1998) found that an ISD of $+2$ ms between pairs of 4-ms broadband noise bursts was offset by an ISLD ranging from $-9$ to $-19$ dB among five listeners (trading ratio, $0.11–0.22$ ms/dB). When our listeners were presented with click pairs at an ISD of 0.6–0.8 ms, an opposing level difference of 5 dB typically brought listeners’ responses back near the midline (trading ratio, $0.12–0.16$ ms/dB). Thus our results fell within the range of previous free-field measurements of time-intensity trading.

Our results support previous studies of localization dominance at ISDs of 1–4 ms. Since at least the 1930s, studies have reported the basic qualitative observation that when a sound is delivered from two loudspeakers with a delay of a few milliseconds between the loudspeakers, the sound is localized near the leading source (reviewed by Blauert 1997; Gardner 1968; Litovsky et al. 1999; Zurek 1987). Our results essentially replicated this observation: listeners’ responses were strongly biased toward the leading loudspeaker at delays of 1.0–1.4 ms. Although the leading sound dominated localization, listeners’ responses uniformly undershot the leading location, demonstrating an appreciable influence of the lagging sound at delays of 1.0–1.4 ms. Similar findings have been reported previously. Headphone studies that have measured lateralization of lead-lag pairs of dichotic clicks or noise bursts (e.g., Shinn-Cunningham et al. 1993; Wallach et al. 1949; Yost and Soderquist 1984; Zurek 1980) have shown that a click is lateralized less strongly when followed 1–4 ms later by a...
lagging click than when presented alone. Free-field studies have also indicated that a lagging sound influences localization. Wendt (1963) found an undershoot of \(~20\%\) at an ISD of \(1\) ms in a localization task. Litovsky and Macmillan (1994) measured minimum-audible-angle thresholds for pairs of 6-ms noise bursts; their analysis showed that the lag stimulus had significant weight in listeners’ azimuth discriminations when the ISD was \(4\) ms. Chiang and Freyman (1998) used a localization task in which listeners compared the apparent location of a test stimulus (a train of paired 4-ms noise bursts with a fixed ISD of \(2\) ms) with that of a movable reference stimulus (a train of single noise bursts). At sensation levels near the levels we used (\(40–50\) dB), they found undershoots that ranged up to \(~20\%\) among four listeners with a fifth listener exhibiting an overshoot of a few degrees (Chiang and Freyman 1998, Table 1). Compared with these studies, the undershoots we found are somewhat greater for most listeners. The difference might be explained by differences in the specific task performed or in the stimuli used.

The undershoot that we and others have found supports the notion that the lagging stimulus influences localization, even at delays at which the precedence effect is strongest. The lagging stimulus could contribute to localization judgments in several ways. It might simply shift the image of the leading stimulus toward the lagging location without affecting the quality or extent of the image. Alternatively, the lagging stimulus might broaden the image so that the image extends toward the lagging location; broadening would move the center of the image closer to the lagging location. Our data appear to favor image broadening over simple shifting because listeners’ responses showed greater dispersion for paired clicks than for single clicks. A third explanation, however, cannot be ruled out: listeners might have perceived a complex (and possibly confusing) amalgam of the two sources and then pointed to a single location according to some as-yet-unknown cognitive strategy.

The human psychophysical results of the current study qualitatively agree with behavioral studies in cats. Cranford (1982) trained cats to lateralize single clicks and then tested the cats with paired clicks at various ISDs. At ISDs of \(~0.1–5\) ms, responses were significantly biased toward the side of the leading sound with the strongest effect at ISDs of \(~0.5–2.0\) ms. This result is consistent with localization dominance found with human listeners. Populin and Yin (1998) trained cats to respond to single clicks with an oculomotor saccade. When tested with click pairs, the cats moved their eyes toward the leading source at ISDs of \(0.1–1\) ms. At ISDs less than \(~0.3\) ms, saccades fell somewhat closer to the midline; at ISDs greater than \(~0.3\) ms, they fell more laterally but considerably short of the location of the leading source. Furthermore when an ISLD of \(5\) or \(10\) dB was imposed, eye movements shifted toward the more intense speaker. Thus the cats responded with saccades that were consistent with localization dominance, summing localization, and time-intensity trading seen in human listeners.

**Limitations of the physiological findings**

When comparing our cat physiological results to human psychophysical results, two aspects of our methodology should be considered: species differences and anesthesia. Cats are thought to use the same acoustical cues as humans to localize sounds, and behavioral studies have shown that cats localize isolated broadband sounds with an accuracy similar to that of humans (May and Huang 1996; Populin and Yin 1998). Furthermore, cats localize paired clicks in a manner qualitatively similar to humans, suggesting that cats also experience the precedence effect over a similar range of ISDs (Cranford 1982; Populin and Yin 1998). Since a cat’s head is smaller than a human head, one would expect differences in the time scale of ISD-based summation localization because this aspect of the precedence effect depends on precise differences in arrival times at the two ears. The existing behavioral data are consistent with this expectation (Populin and Yin 1998). In comparing our physiological results to human psychophysical results, we compensated for this expected difference by scaling the human psychophysical results by an amount that we determined by physical acoustical measurements. This scaling procedure is ad hoc but reasonable, and the results do essentially agree with the cat behavioral data that are available (Populin and Yin 1998).

The animals that we recorded from were anesthetized, so the cortical responses obtained are likely to differ from those of a human or cat that is actively localizing sounds. The effects of anesthesia on cortical responses are well appreciated but generally not understood. It is unknown whether anesthesia alters spike patterns in important ways, but anesthesia is known to decrease overall response magnitude, and studies of responses to paired sounds using ISDs beyond echo threshold indicate that anesthesia increases the suppression of responses to lagging sounds (Fitzpatrick et al. 1999; Mickey and Middlebrooks 2000; Reale and Brugge 2000). Cortical responses may also be influenced by behavioral state—e.g., whether the listener is localizing sounds (Benson et al. 1981). Such effects cannot be addressed with an anesthetized preparation such as ours. Thus it is possible that we have failed to appreciate important aspects of cortical responses that would be apparent in the absence of anesthesia. Nevertheless close parallels found between the present physiological results and psychophysical measurements attest to the robustness of cortical neuronal sensitivity, which survives even the depressive effects of anesthesia.

**Cortical responses and perceptual fusion**

When two similar, brief stimuli are presented at ISDs below \(~5\) ms, listeners usually report hearing a single fused sound. At these ISDs, we found that a unit’s response to a paired click typically consisted of a single burst of spikes, much like the response to a single click. This finding is corroborated by other cortical studies (Fitzpatrick et al. 1999; Reale and Brugge 2000). This unified response to paired stimuli suggests a correspondence with perceptual fusion. Of course, the correspondence would be complete only if units were to show a discrete lagging response at ISDs just beyond echo threshold (\(~5\) ms for clicks). Such responses are rarely seen. The great majority of cortical units remains suppressed at ISDs of tens to hundreds of milliseconds; only a small proportion of units show discrete lagging responses at shorter ISDs, even in the absence of anesthesia (awake rabbit: Fitzpatrick et al. 1999; anesthetized cat: Mickey and Middlebrooks 2000; Reale and Brugge 2000). That is, cortical neuronal “echo thresholds”—usually defined as the ISD at which lagging responses recover to \(50\%\) of the maximal response—are nearly always greater than psycho-
when the azimuth of a single click was varied from 1 ms, the magnitude of unit responses changed much as it did predicted by cross-correlating the waveforms at the two ears.

and ISD were chosen as to position the fused image within the paired 100-ms noise bursts in the barn owl inferior colliculus. 45°. Keller and Takahashi (1996) recorded responses to responses consistent with summing localization. Yin (1994) the difference in head size, in humans (Wendt 1963; the current qualitatively, with summing localization seen psychophysi- two loudspeaker locations with a more-or-less systematic pro- 1 0.4 and 2 ms, most units signaled locations between the ears, whereas uncorrelated bursts produce two foci corresponding to the locations of the two sources (Takahashi and Keller 1994). A cortical version of this population code may exist, but one would expect it to be less obvious since cortical neurons appear to signal source locations over large regions of space and since no topographical map of space is known to exist in the cortex (Middlebrooks et al. 1998). To address this question, cortical studies using dissimilar sound sources, as well as models of populations of neurons, will be needed.

Localization of fused sounds by cortical neurons

In this study, we focused on the spatial aspects of cortical responses at delays below the echo threshold where the precedence effect is strongest. We found physiological correlates of two distinct effects: summing localization at the smallest ISDs (below ~0.5 ms in cats) and localization dominance at ISDs of ~1–4 ms.

When we presented paired clicks at ISDs between about −0.4 and +0.4 ms, most units signaled locations between the two loudspeaker locations with a more-or-less systematic progression from left to right. This observation agrees, at least qualitatively, with summing localization seen psychophysically in cats (Populin and Yin 1998) and, after accounting for the difference in head size, in humans (Wendt 1963; the current study). Studies of lower auditory areas have also described unit responses consistent with summing localization. Yin (1994) described several units recorded from the cat inferior colliculus. As the ISD of click pairs was varied from roughly −1 to +1 ms, the magnitude of unit responses changed much as it did when the azimuth of a single click was varied from −45 to +45°. Keller and Takahashi (1996) recorded responses to paired 100-ms noise bursts in the barn owl inferior colliculus. They found high spike rates when the loudspeaker azimuths and ISD were chosen as to position the fused image within the unit’s receptive field; response profiles were successfully predicted by cross-correlating the waveforms at the two ears.

We found that cortical responses to click pairs also depended on the ISLD. When the ISLD of simultaneous clicks was varied from about −20 to +20 dB, units signaled locations from −50 to +50°. This result closely mirrors the psychophysical results of the present study and previous studies (Bauer 1961; Wendt 1963). Furthermore we found that superposition of a nonzero ISLD at a given ISD shifted network estimates toward the more intense loudspeaker. When the ISLD opposed the ISD, the shift was smaller, consistent with time-intensity trading. This physiological time-intensity trading likely requires synthesis of ITDs with ILDs in the central auditory pathway. Using a model based on cross-correlation of inputs to the two ears, we showed that, at frequencies above ~6 kHz, the interaural lag did not change appreciably with moderate ISLDs. Because the units we studied (in particular, those shown in Figs. 6 and 9B) responded only to higher frequencies and because these units were nonetheless sensitive to moderate ISLDs, we conclude that their responses resulted from combination of ITDs and ILDs by the central auditory system. This synthesis may be complete as early as the inferior colliculus: Yin (1994) described a collicular unit with responses that were reminiscent of time-intensity trading. How might ITD and ILD cues be integrated? Yin and colleagues (1985) have proposed a mechanism whereby a change of ILD would alter a neuron’s spike latency. This mechanism is appealing because ILDs could be processed by the same coincidence-detection circuits that presumably process ITDs.

At ISDs of ~1–4 ms, we found that most units signaled locations near the lead location in agreement with psychophysical demonstrations of localization dominance. Notably, network estimates fell slightly short of the lead location, much as listeners’ estimates fell short. In neither case could the undershoot be accounted for by a general response bias toward undershooting because responses to single clicks showed no such tendency. Psychophysically such undershoots may be attributed to the influence of the lagging sound on the perceived location of the fused image. Physiologically the undershoot might arise from incomplete forward suppression of the lagging response by the leading stimulus (e.g., Figs. 2B and 3, C and D, positive ISDs). In other cases, a backward suppression of the leading response by the lagging stimulus might account for the undershoot (e.g., Fig. 4, C and D, positive ISDs). The cortex is thought to be required for localization dominance (Cranford and Oberholtzer 1976; Cranford et al. 1971; Whitfield et al. 1972), but much of the underlying processing is likely to occur subcortically. Yin (1994) described unit responses in the inferior colliculus that were consistent with localization dominance. At ISDs of ~1–5 ms, units typically responded strongly to click pairs if the leading stimulus evoked a strong response when presented in isolation and responded weakly to click pairs if the leading stimulus evoked a weak response when presented in isolation.

Our finding that cortical responses generally signaled the locations reported by listeners was not a foregone conclusion. Indeed, we found a small number of units whose responses clearly disagreed with psychophysical measurements of localization dominance and summing localization even though these same units accurately signaled the locations of single clicks (Figs. 7 and 8). What role might these neurons play in the perception of fused sounds? They may play no role at all in sound localization. On the other hand, these neurons might
Contribute to the behavioral undershoot in response to paired sounds. They might also contribute to other qualities of the fused percept such as spatial extent, timbre, or loudness—qualities in which single sounds and paired sounds generally differ (Blauert 1997).

Conclusion

In summary, when presented with pairs of sounds that fuse perceptually, most auditory cortical neurons responded in a manner consistent with the spatial judgements of human listeners. A number of questions remain unanswered. How might cortical neurons represent simultaneous non-fused sounds? Do cortical neurons exhibit a correlate of localization dominance when both sound sources lie in the median sagittal plane, where interaural cues are negligible? At delays near behavioral echo threshold, can cortical neurons localize lagging stimuli as accurately as psychophysical listeners do? Do dynamic aspects of the precedence effect (Clifton 1987; Freyman et al. 1991) have a parallel in cortical responses? We hope to address these issues in future studies.

We thank E. Macpherson for illuminating discussions throughout the course of this work, S. Furukawa and E. Macpherson for assistance with physiological experiments and for valuable comments on an earlier version of the manuscript, and Z. Onsan for assistance with psychophysical experiments and with preparation of the manuscript.

This work was supported by National Institutes of Health (NIH) Grants RO1 DC-00420, T32 GM-07863, T32 DC-00011, and RO1 RR-13619 and by the Scottish Rite Schizophrenia Research Program. Multichannel silicon probes were kindly provided by the University of Michigan Center for Neural Communication Technology, sponsored by NIH Grant P41 RR-09754.

REFERENCES


MIDDLEBROOKS JC, XU L, FURUKAWA S, AND MICKEY BJ. Location signaling by cortical neurons. In: Integrative Functions in the Mammalian Auditory


