Research Article

Localization of Click Trains and Speech by Cats: the Negative Level Effect

YAN GAI,¹ JANET L. RUHLAND,¹ AND TOM C. T. YIN¹

¹*Department of Neuroscience, University of Wisconsin-Madison, Medical Sciences Building, 251 SMI, 1300 University Avenue, Madison, WI 53706, USA*

Received: 14 January 2014; Accepted: 28 May 2014; Online publication: 19 June 2014

ABSTRACT

Although localization of sound in elevation is believed to depend on spectral cues, it has been shown with human listeners that the temporal features of sound can also greatly affect localization performance. Of particular interest is a phenomenon known as the negative level effect, which describes the deterioration of localization ability in elevation with increasing sound level and is observed only with impulsive or short-duration sound. The present study uses the gaze positions of domestic cats as measures of perceived locations of sound targets varying in azimuth and elevation. The effects of sound level on localization in terms of accuracy, precision, and response latency were tested for sound with different temporal features, such as a click train, a single click, a continuous sound that had the same frequency spectrum of the click train, and speech segments. In agreement with previous human studies, negative level effects were only observed with click-like stimuli and only in elevation. In fact, localization of speech sounds in elevation benefited significantly when the sound level increased. Our findings indicate that the temporal continuity of a sound can affect the frequency analysis performed by the auditory system, and the variation in the frequency spectrum contained in speech sound does not interfere much with the spectral coding for its location in elevation.

Keywords: localization, spectral cues, clicks, speech, level

INTRODUCTION

The mammalian binaural system relies on interaural time (ITDs) and level (ILDs) differences across the two ears to determine the sound location in azimuth (see review, Tollin and Yin 2009). In contrast, sound localization in elevation is based on spectral cues related to the direction-dependent acoustic filtering properties of the head and the pinnae (Gardner and Gardner 1973; Hebrank and Wright 1974). Due to the small size of the mammalian pinnae, the important spectral cues are largely above 7 kHz, and there are striking similarities between the transfer functions measured with humans (Roffler and Butler 1968; Shaw 1974; Bronkhorst 1995; King and Oldfield 1997; Carlile and Delaney 1999) and with cats (Musicant et al. 1990; Rice et al. 1992).

In agreement with the spectral cue theory is the finding that sound stimuli that have relatively flat and broad frequency spectra, such as white noise, are easier to localize than sound stimuli with narrower spectra, such as tones and narrow-band noise (Roffler and Butler 1968; Martin and Webster 1987; Middlebrooks 1992; Tollin et al. 2013) because the former allows a better representation of the spectral features associated with elevational locations. The fact that cats indeed use spectral notches is supported by behavioral studies using paired noise bursts (May and Huang 1996) and auditory illusions (Tollin and Yin 2003).

However, the fact that spectral cues dominate localization in elevation does not mean that temporal

Correspondence to: Yan Gai · Department of Neuroscience · University of Wisconsin-Madison · Medical Sciences Building, 251 SMI, 1300 University Avenue, Madison, WI 53706, USA. Telephone: +1-608-2626992; email: ygai@wisc.edu

features of the sound have no influence. Localization of an impulsive or short-duration sound in the medial sagittal plane has been shown to deteriorate with increasing sound level for humans (Hartmann and Rakerd 1993; Inoue 2001; Macpherson and Middlebrooks 2000; Vliegen and Van Opstal 2004) and cats (Gai et al. 2013). This observation is termed as a *negative level effect*. Interestingly, although both clicks and broadband noise have relatively flat and broad spectra, this effect is observed with clicks but not with noise. Brungart and Simpson (2008) further tested with human subjects the localization of what they called a random-phase click train, which was constructed by fixing the amplitude of the frequency spectrum of a click train and randomizing its phase. The resulting random-phase stimulus has a continuous time waveform similar to that of a noise. No negative level effect was observed with this random-phase stimulus when vertical polar errors were measured.

The present study evaluated the localization performance of domestic cats in terms of accuracy, precision, and response latencies over a large range of sound levels. Besides noise and click-like stimuli, we also tested human speech segments with our cats. It has been shown, for both free-field and virtual-space speech sound, that human listeners can localize well in azimuth, except when front-and-back judgments are required, but are worse in elevation (Begault and Wenzel 1993; Gilkey and Anderson 1995; Best et al. 2005). However, the level effect has not been systematically examined for speech sound, which is one of the main focuses of this study. Furthermore, in agreement with our previous study (Gai et al. 2013), we show that the flatness of sound-source spectrum is not a major factor that determines localization in elevation.

METHODS

Subjects and experimental setup

Four adult female cats were implanted with stainless steel coils (S170012A7-FEP, 137 Alan Baird Industries, Ho-Ho-Kus, NJ) for the eyes under aseptic surgical conditions. The coils were threaded subcutaneously and connected to connectors that were embedded in the dental acrylic cap on the skull. More details on the surgeries can be found in Populin and Yin (1998). All surgical and experimental procedures were approved by the University of Wisconsin Animal Care and Use Committee and the National Institutes of Health.

The behavioral experiment was conducted in a dimly lit sound-attenuated chamber (2.2×2.5×2.5 m). The eye coil was connected to a magnetic search coil system that enabled the vertical and horizontal components of the eye movements to be monitored

with a resolution of $<1^\circ$. Speakers (RadioShack 4-in. Super Tweeters, model 40-1310B; Fort Worth, TX) and LEDs (2.0-mm-diameter red light emitting diodes, $\lambda_{\text{max}}=635$ nm) were placed in the frontal hemifield between $\pm 50^\circ$ in azimuth and $\pm 30^\circ$ in elevation. The 17 speakers included a central speaker ($0^\circ, 0^\circ$), six speakers in the horizontal plane, six speakers in the sagittal plane, and four speakers along the diagonals ($\pm 20^\circ, \pm 20^\circ$). Speakers were hidden from view by a black translucent cloth. At the center of each speaker, an LED was suspended which, when illuminated, could be easily seen through the cloth. A Tucker-Davis Technologies III system was used for stimulus generation and data collection.

Behavioral paradigm and acoustic stimuli

Cats were trained by operant conditioning with food reward. Eye coils were calibrated with a behavioral procedure (Populin and Yin 1998) that relied on the natural instinct of the cat to look at a light source that suddenly appears in the visual field. The output of the coil system was recorded when the eye assumed a stationary position at the end of movements to known locations. The vertical and horizontal components of eye positions were fit separately with linear equations. The coefficients obtained from the fitting procedure were used to convert the voltage output of the coil system to degrees of visual angle.

All cats had participated in previous auditory and visual tasks for more than a year. No specific training was found necessary for the present study. Cats performed the experiment with their heads unrestrained. A feeding tube was attached to the head post. In each trial, the cat was required to fixate an LED presented from straight ahead ($0^\circ, 0^\circ$) and maintain gaze fixation within the acceptance window for a variable period of time (600–1,000 ms). Immediately after the LED was extinguished, a target sound was presented by 1 of the 17 speakers. The cat was required to make a gaze saccade to the perceived location of the target. If within 2,000 ms following the offset of the LED, the gaze entered a specified acceptance window ($\pm 13^\circ, \pm 13^\circ$) around the target location and remained there for ≥ 650 ms, the cat was given a food reward. In the rare case when a reward was given during stimulus presentation, the stimulus would be truncated at the time of the reward. Data were analyzed regardless of whether a reward was received. Trials presented in this study were randomly mixed with other types of trials (e.g., random-noise or visual-target trials) so that the cat could not predict the upcoming trial.

Several types of stimuli were used as target sounds, which are shown in the figures above each corresponding condition. The click train had an overall

duration of 550 ms, consisting of five 40- μ s clicks with 110-ms inter-click intervals (Fig. 1, top inset). We chose this value following Hartmann and Rakerd (1993), but reduced the number of clicks from eight to five so that no click train was truncated at the time of a food reward (≥ 650 ms). To create the random-phase stimulus, a Fourier Transform was first applied to the click train. The amplitude of the frequency spectrum was fixed when the Inverse Fourier Transform was performed, while the phase of the transform was randomized (Brungart et al. 2008). The phase randomization produced a signal with the same frequency content (theoretically broad and flat) as the click train but a continuous temporal structure (Fig. 2B, inset). A 5-ms $\cos^2$ ramp was applied to the beginning and the end of the random-phase stimulus. No ramp was applied to the click train following Hartmann and Rakerd (1993), although ramps were applied to the click trains used by Brungart and Simpson (2008). We chose not to have the ramp so that more direct comparisons can be made with the single-click stimulus (Fig. 5, top inset), which was just a

single 40- μ s click presented when the fixation LED was turned off.

Two female-speaker speech tokens truncated from standard Hearing in Noise Test (HINT) sentences were used in the study. The first sentence segment (“Her shoes”) was 680-ms long (Fig. 6, top left). The second token (“A boy fell from”) was 1-s long (Fig. 6, top right). Note that on some trials, the sentence actually played out could be shorter if the animal responded quickly and received a reward before the end of stimulus presentation but not sooner than 650 ms. The second speech was further tested with only the first 500 ms (“A boy”) to examine the duration effect on localization of speech.

The sound level varied over a large range for each stimulus condition. Data will be presented as a parameter of sound attenuation with one measurement of the actual sound level at a certain attenuation value (Hartmann and Rakerd 1993). When the attenuation was set to 10 dB, the peak value of the sound-pressure meter (Extech Instruments; C-weighting) read 54.7, 64.5, 59.1, and 57.4 dB SPL, for the click train

FIG. 1. Examples of localization performance for the click train obtained with cat 38. **A** Scatter plots of the final gaze positions for targets at soft (left), medium (middle), and high (right) levels. The large open symbols represent target locations, and the small filled symbols represent gaze positions. Different colors represent different

locations. **B** Gaze shifts vs. motor error. The gain and δ derived from the regression analysis (straight lines) are indicated by the two values in brackets computed separately for horizontal (blue circles and first value) and vertical (red squares and second value) gaze movements.

FIG. 2. Localization accuracy measured with the gain for the click train (A) and the random-phase noise (B) as a function of target sound level. The error bars are 95 % confidence intervals derived with the bootstrap method. Each horizontal bracket indicates a significant decrease in the gain relative

to the maximum ($p<0.01$). For cat 36 and cat 38, the data at each sound level had $N>76$ trials across all locations. For cat 33, no data was obtained at the lowest sound level, and N ranged from 19 to 133 trials at other levels.

(550 ms), the random-phase stimulus (550 ms), the speech token #1 (680 ms), and the speech token #2 (1,000 ms), respectively.

Data analysis

Horizontal and vertical gaze positions were determined separately by a velocity criterion (Populin and Yin 1998; Tollin et al. 2005). The beginning of the gaze movement was marked as the point in time when steady fixation ended, i.e., when the magnitude of the velocity exceeded 2 standard deviations (SDs) of the mean velocity (near zero) computed during the initial steady fixation. The latency was computed as the time when the fixation of the gaze ended, which was also when the gaze movement began, relative to beginning of the target. The final gaze position (examples shown in Fig. 1A) was determined at the time when the magnitude of the velocity returned to ≤ 2 SDs of the baseline mean velocity. For most of the subjects, a saccade was considered valid and included in the analysis only when the beginning of gaze movement occurred within 600 ms after stimulus onset, even though later responses could still trigger a reward. An exception had to be made for cat 39, which will be

discussed later with the localization of speech data. For trials with no saccade, the start and the end of the “saccade” were chosen to be 200 and 600 ms, respectively, to get a final gaze position for that trial (usually close to 0°); those trials were excluded in computing the response latency.

The localization performance was measured by comparing the actual gaze shift with the motor error (see Tollin et al. 2005 for details). The motor error was the difference between the target-in-space position and initial gaze positions at the time when the target was turned on (Fig. 1B, x-axis). For most trials, the gaze motor error was nearly equivalent to the position of the target in space since the animal was initially fixating at or near (0°, 0°). A linear function was fit to the data separately for the azimuthal and/or the elevational components of target locations (Fig. 1B, straight lines). Sound localization accuracy was measured by the slope of the function, which was referred to as the gain (Tollin et al. 2005). A gain of 1.0 indicates that, on average across all trials and target positions, the cat locates the targets to actual positions, while gains <1.0 indicate that target locations are underestimated. Gains were computed along both dimensions for the central and the four diagonal

speaker locations. Only the horizontal gain was computed for targets in the horizontal plane, and only vertical gain for targets in the sagittal plane. Standard statistical bootstrapping techniques (Efron and Tibshirani 1986) were used to obtain an estimate of the 95 % confidence intervals of the gain.

Using the empirical data set, the standard deviation of the residuals of the fitted function was computed to measure the distribution of behavioral responses about the mean gain. This value, referred to as δ , provides a numerical estimate of the precision (or consistency) of the localization responses (see also Moore et al. 2008). Bootstrapping was also used to obtain an estimate of the 95 % confidence intervals of δ .

RESULTS

Localization performance with click stimuli, including the click train, its random-phase version, and the single click obtained with three cats (cats 33, 36, and 38), will be presented first, followed by performance with speech segments obtained with two cats (cats 38 and 39).

Temporal features can greatly affect localization in elevation

Figure 1 shows examples of the localization performance obtained at three sound levels with the click train for cat 38. At the lowest sound level (90-dB attenuation), final gaze positions generally clustered to the left of the front center or at the lower right corner (Fig. 1A, *left*), indicating that the cat was unable to localize the sound for most of the trials. The failure of localization is further reflected when gaze shifts are plotted against motor errors (Fig. 1B, *left*). The slopes of the fitted lines are very shallow for both azimuth and elevation, corresponding to near-zero gains. There was also a large variation in the responses, as quantified by large values of δ . Since we did not do a detection task, it is unclear whether or not the cat could hear the sound at all.

At an intermediate sound level (45-dB attenuation), the final gaze positions were close to the actual target positions along both azimuthal and elevational axes (Fig. 1A, *middle*). Gains were close to 1 and δ values were generally small (Fig. 1B, *middle*). However, as the sound level further increased (10-dB attenuation), although the azimuthal performance remained relatively unaffected, the perceived elevational locations of the targets were compressed towards the horizontal plane (Fig. 1A, *right*). Correspondingly, the elevational gain for the click train with a 10-dB attenuation decreased to 0.14 (Fig. 1B, *right*), almost as bad as when the attenuation was 90 dB (Fig. 1B,

left). In summary, a profound negative level effect was observed for localization of the click train varying in elevation.

Figure 2A summarizes the gain values obtained with the same click train for three cats. For targets in azimuth (blue), localization gain increased with sound level and stayed close to 1 (i.e., accurate localization) at high levels. Horizontal brackets indicate significant decreases in the gain compared with the maximum gain. Although for cat 38, there was a significant decrease of horizontal gain at the highest sound level, the gain was actually closer to 1 and the performance was thus more accurate (i.e., there was less overshooting) at the highest sound level. Localization in elevation (red) showed profound negative level effect for all three cats, although the optimal sound level at which the highest gain was achieved (marked by the starting point of the brackets) varied with individual subjects. In contrast, when the phase of the same click train was randomized after a Fourier Transform, the localization pattern in elevation changed dramatically—the negative level effect was no longer observed at high sound levels (Fig. 2B), although there was statistical significance in the change of both horizontal and vertical gains for cat 36.

The localization gain measures, on average, how far the gaze movement was in the direction of the target localization; trial-by-trial variations were not reflected in this measurement. Figure 3 shows a different measurement, δ , which can be thought of as an inverse indicator of localization precision. A large δ is usually a result of large scatter in the gaze responses, and is unrelated to how close the responses are to the actual target locations. For both the click train (Fig. 3A) and the random-phase stimulus (Fig. 3B), δ decreased at high sound levels for both azimuthal and elevational measurements, although when a negative level effect was present for the gain, there was a slight increase of δ at very high sound levels (Fig. 3A). In general, although for the click train the localization gain was low at both very low and very high sound levels, the detection patterns were actually different. At low levels, the sound was either undetected or unlocalizable, and the responses were scattered (e.g., Fig. 1A, *leftmost*). At high levels, the azimuthal gaze positions were pinned to the vicinity of the azimuthal target positions, but the elevational movements were compressed, i.e., close to zero (e.g., Fig. 1A, *rightmost*).

The response latency also exhibited different patterns for the click train vs. the random-phase stimulus. There was a systematic decrease in the azimuthal latency for the random-phase stimulus (Fig. 4B, blue), but not as much in the azimuthal latency for the click train (Fig. 4A, blue) or the vertical latency in either case (Fig. 4A, B, red). Interestingly, the azimuthal latency for the click train remained relatively large and variable even at high sound levels

FIG. 3. Localization precision measured with δ for the click train (A) and the random-phase noise (B) as a function of target sound level in the same format as Fig. 2. The error bars are 95 % confidence intervals derived with the bootstrap method. A large δ is an indication of large scatter, and thus low precision.

FIG. 4. Response latency for the click train (A) and the random-phase noise (B) as a function of target sound level. The error bars are standard deviations.

where the azimuthal performance was high (Fig. 4B, blue). It appeared that, on some trials, the cat purposely delayed its responses so that its decision can be made based on multiple clicks (inter-click intervals were 110 ms). In general, the latency values obtained with these three cats (cats 36, 38, and 33) were comparable to what was measured in our previous behavioral studies.

To examine the effect of presenting multiple clicks on localization in both azimuth and elevation, the experiment was repeated in two cats using a single click (Fig. 5). For azimuthal localization, the accuracy was generally lower than the corresponding accuracy for click trains, and a higher sound level was required for the performance to reach a decent level (Fig. 5A, blue).

FIG. 5. Localization accuracy (A), precision (B), and latency (C) for the single click as a function of target sound level. Note that the sound level covered a higher range (x-axis) than what was used for the click train and random-phase noise (Figs. 2, 3, and 4). The error bars in A and B are 95 % confidence intervals; the error bars in C are standard deviations. Each horizontal bracket indicates a significant decrease in the gain relative to the maximum ($p < 0.01$).

The elevational localization was even more affected, with the maximum performance being highly compromised even though a clear negative level effect was also observed (Fig. 5A, red; horizontal brackets). Similar to what was observed with the click train, the elevational δ was generally small (Fig. 5B, red), due to compressed elevational responses towards the horizontal plane. The response latency for targets in azimuth (Fig. 5C, blue) was indeed shorter and less variable than what was observed with the click train (Fig. 4A, blue) except at the lowest sound levels for cat 36 (Fig. 5C, left). Therefore, it was likely that when multiple clicks were presented, the cats purposely waited on some trials to gain more samples before a decision was made.

In summary, negative level effects were observed with the click train and the single click, but not with the random-phase stimulus, even though the click train and the random-phase stimulus had the same broad and flat frequency spectra.

A higher sound level can benefit localization of speech in elevation

Figure 6 shows the temporal waveforms (*top*) and frequency spectra (*bottom*) of the two speech segments. It is clear that most of the speech energy is below the high-frequency region that is critical for localization in elevation with our cats (i.e., the mid/high-frequency notches lie between 7 and 15 kHz, as shown by the HRTFs measured in Gai et al. 2013).

Figure 7 shows the localization performance of two cats for speech segment #1 (“Her shoes”, 680 ms). Cat 38 generally required a higher range of sound levels to achieve comparable performance as cat 39, which agreed with cat 38’s click performance described above (Fig. 2). Other than that, both cats showed similar patterns. First, their ability to localize speech in azimuth was better than the ability to localize speech in elevation, except for very high sound levels. For cat 39, at the highest sound level (0-dB attenuation), the performances in azimuth and in elevation were essentially the same in terms of localization accuracy and precision (Fig. 7, left column). For the other cat, the localization in elevation was never as good as in azimuth, especially when precision was considered (Fig. 7B, right). For both cats, there existed certain sound levels for which the azimuthal gain was about 0.9, whereas the elevational gain was below 0.3.

Second, elevational localization increased significantly with sound level, and no sign of any negative level effect was observed for either cat (Fig. 7A, red). We are confident in making this conclusion because, when there was the negative level effect, such as with the click trains (Fig. 2A), the optimal sound level usually occurred over a small dynamic range (e.g., from 55- to 35-dB attenuation for cat 33, and even

FIG. 6. Sound waveforms (*top*) and the corresponding spectral-temporal modulations (*bottom*). Speech segment #1, “Her shoes,” was 680-ms long. Speech segment #2, “A boy fell from” was 1-s long. The second speech was further tested with only the first 500 ms (“A boy”) to examine the duration effect on localization.

smaller for cat 36 and cat 38), before the performance quickly deteriorated. In contrast, cat 39’s localization gain with speech #1 remained high over a large range (45 dB). It is unlikely that a louder sound will trigger a negative level effect. For cat 38, at 0-dB attenuation, the elevational gain was still notably lower than its horizontal gain. Therefore, we added a louder condition by manually increasing the gain of the TDT Stereo Power Amplifier by 9 dB, which was the highest value allowed by the system. Although it may not be characterized as “uncomfortably loud,” when the experimenters listened to the sound, it was clear that normal speech is unlikely to be louder than this. Nevertheless, this increased sound level (referred to as -9 dB attenuation) only improved her elevational gain (Fig. 7A, *right*, red).

Last, their response latencies increased systematically with decreased sound level (Fig. 7C). It should be noted that cat 39 was a somewhat special subject. As was observed in our other experiments using different paradigms (not shown), this cat seems to have lower hearing thresholds and considerably longer response latencies, especially when the task is difficult. Since our lab started using the search-coil technique many years ago, we have been able to apply the 600-ms criterion (see “METHODS”) to determine whether a saccade is valid, i.e., if a saccade happens more than 600 ms after the stimulus onset, even though the cat may receive a reward for correct gaze movements, the saccade is not included in the analysis. This criterion seems to work for all of our subjects except cat 39, who frequently shows correct movements significantly later than 600 ms. Therefore, we included saccadic movements up to 1,500–2,000 ms for this subject.

Figure 8 shows examples of cat 39’s final-gaze positions for speech #1 at a low (*left*) and a high (*right*) sound levels. At the high level (0-dB attenuation), her final gaze positions clustered around the true target locations, resulting in high gains and small

FIG. 7. Localization accuracy (A), precision (B), and latency (C) for speech #1 (680 ms, “Her shoes”) as a function of target sound level. For both cats, the data at each sound level had $N > 76$ trials across all locations. The error bars in A and B are 95 % confidence intervals; the error bars in C are standard deviations.

FIG. 8. Scatter plots of the final gaze positions for speech #1 obtained with cat 39 at low (left) and high (right) levels. The large open symbols represent target locations, and the small filled symbols represent gaze positions. The gain and δ are indicated by the two values in brackets computed separately for horizontal (first value) and vertical (second) gaze movements.

δ values. At the low level (65-dB attenuation), the elevational gain was very low (0.23), and there was trial-by-trial variation ($\delta=8.7^\circ$). The localization accuracy in azimuth remained high (gain=0.84), but the variation across trials was very large ($\delta=17.5^\circ$). Note that this level (65-dB attenuation) was already 10-dB lower than the lowest tested for the other cat (cat 38), who could not localize in either azimuth or elevation even at 55-dB attenuation.

The second speech segment, “The boy fell from,” was 1,000-ms long. Both cats showed very similar performance (Fig. 9, solid lines) to what was observed with the first speech. We further tested cat 39 with only the first half of the speech, “The boy” (Fig. 9, left, dashed lines). At low sound levels, the accuracy, the precision, and the latencies all turned out to be lower (Fig. 9, left) when the duration was reduced. It is possible that the subject’s localization ability in both azimuth and elevation can benefit from combining information over hundreds of milliseconds, with either stationary HRTFs or dynamic HRTFs introduced by possible pinna movements (see “DISCUSSION”). Alternatively, because the fricative “f” provided more high-frequency energy (Fig. 6, right), the improved performance may be largely due to the added energy rather than temporal integration. To examine the role of temporal integration, a more strict approach would be to simply present the first half of the sentence twice consecutively and see whether or not the performance changes.

It is also possible that the introduction of the fricative “f” provided more high-frequency energy (Fig. 6, right), which led to the improvement of performance.

In summary, for speech sound, increasing the sound level seemed to always improve localization in elevation without the negative level effect.

DISCUSSION

The negative level effect

Hartmann and Rakerd (1993) first brought up the term, the negative level effect, to describe this

counterintuitive phenomenon observed with clicks but not with broadband noise. They hypothesized that this effect “arises because the tonotopic excitation pattern is broadened for intense impulsive sounds.” The fact that negative level effects occur for

FIG. 9. Localization accuracy (A), precision (B), and latency (C) for speech #2 (“A boy fell from”) as a function of target sound level. Solid lines represent performance obtained with the whole 1,000-ms speech, and dashed lines represent the first 500-ms of the speech.

short-duration but not long-duration sounds has been generally confirmed, with the exception of Vliegen and Van Opstal (2004), who also found some negative-level effect for their 100-ms stimulus.

Using an auditory nerve (AN) model that accurately describes many response features of the AN physiology (Zilany et al. 2009), our previous simulation study showed that the negative level effect is always predicted even for long broadband noise (Gai et al. 2013). Note that as a phenomenological model, the model successfully predicts long-term response of the AN fibers, even though no efferent circuit was explicitly simulated. Nevertheless, certain efferent effects may have been missed for the issue we are examining here. Different stimulus paradigms have also been explored to specifically examine the role of peripheral adaptation, which is unable to explain the absence of negative level effect for continuous and long-duration sound (please see Gai et al. 2013 for simulation details.) Note that at low sound levels, spectral notches may not be well encoded either, due to low audibility. Since we focus on the negative level effect that happens at high levels, we have not examined the low-level model performance explicitly.

The present behavioral study confirms the finding that the elevational localization of sound with temporally continuous energy does not suffer from the negative level effect, whereas elevational localization of impulsive sound decreases dramatically with increasing sound level. For both human (Brungart and Simpson 2008) and our cats, randomizing the phase of a click train, resulting in continuous temporal waveforms, diminished the negative level effect. Here, the term “continuous” does not mean that the sound has to be stationary, such as a white noise. Even for speech sound that is highly fluctuating over time, the auditory system is able to maintain its sensitivity to the elevational location at high sound levels. In fact, high sound levels only benefit elevational localization of speech sound.

Apart from the spread of excitation (or so-called tonotopic broadening), firing-rate saturation at the periphery is also likely to cause the negative level effect (Vliegen and Van Opstal 2004). Specifically, when the sound is loud, the AN firing rate at the notch frequency is high, whereas the firing rates at other non-notch frequencies do not increase much due to firing-rate saturation. This will inevitably flatten the population response. These issues may be reduced at higher auditory centers to counteract the negative level effect. For example, Blackburn and Sachs (1989) showed responses of a transient chopper in the ventral cochlear nucleus to long-duration

tones (their Fig. 12). The firing rate saturated at high levels for early responses (1 to 10 ms), but not for late responses (1 to 100 ms). Recordings in monkey primary auditory cortex showed that neurons’ late responses may even have different frequency tuning than what is obtained during the early part of responses (Fishman and Steinschneider 2009).

Meanwhile, higher center neurons may have narrow frequency tuning as sound level increases, possibly due to lateral inhibition (Hartmann et al. 2010). The type O units in the IC indeed show relatively constant response areas with sound level (Davis et al. 2003). It has also been shown that neurons in the auditory cortex are level tolerant when encoding horizontal and vertical sound locations (Zhou and Wang 2012). Other physiological studies have demonstrated that certain neurons in the inferior colliculus and the auditory cortex maintain their spatial sensitivity at high sound levels (Irvine and Gago 1990; Miller and Recanzone 2009), whereas the sensitivity of neurons in the lateral superior olive decreases with sound level (Tsai et al. 2010), although these studies only examined sound localization in azimuth.

Because cats have mobile ears (Tollin et al. 2009), pinna movements may play a role in counteracting the negative level effect for long-duration stimulus by acquiring multiple samples of the sound (Thurlow and Runge 1967). Pinna latencies measured with head-free cats can be as short as 25 ms for targets varying in elevation (Tollin et al. 2010). Since humans, who normally have immobile ears, have shown similar patterns in most of the studies that examined the negative level effect, it is possible that, when the sound is very long (such as 500 ms), human listeners can move their head to create multiple samples of the HRTFs in a way similar to pinna movements to counteract the negative level effect. Since the pinnae can start moving much sooner than head does, there might be certain durations for which cats show less negative level effect than humans. Indeed, our cats showed no negative level effect for the 164-ms sound in Gai et al. (2013), whereas humans showed some effect for 100-ms sound (Vliegen and Van Opstal 2004). Future studies that carefully examine those sensitive values for both cats and humans, or immobilize the pinna during sound localization for cats may provide more insights on this issue.

Localization of speech sound

Normal-hearing humans’ localization ability for speech sound has been tested using both free-field and virtual-space techniques, but more often in

FIG. 10. Comparison of the localization gains in elevation for the single click (red), the click train (green), and the random phase noise (magenta) for cat 36. The differences in performance for the three types of stimuli suggest possible mechanisms related to the negative level effect.

azimuth than in elevation (Begault and Wenzel 1993; Ricard and Meirs 1994; Gilkey and Anderson 1995; Best et al. 2005). The front/back confusion measured with speech sound is generally larger than with clicks or noise, though with one exception (Ricard and Meirs 1994). In addition, the free-field study by Gilkey and Anderson (1995) found larger localization errors with speech than with clicks in elevation, except for one subject whose error with clicks was also big. Two possible explanations were proposed to account for this observation, namely the lack of high-frequency energy contained in speech and the variability of the spectral envelope (Gilkey and Anderson 1995).

Our localization study with cats demonstrated that the elevational performance for speech sound was largely dependent on the sound level. At least for one cat (cat 39), when the speech sound was loud enough, the elevational localization can be almost as good as the azimuthal localization in terms of accuracy or precision (Figs. 7 and 9, left). This observation indicated that the variability of the spectral envelope contained in speech sound is unlikely a major factor, which agrees with our previous simulation result that a flat spectrum of the sound source, such as that of a white noise, is not critical for accurate localization in elevation (Gai et al. 2013).

Temporal integration vs. “multiple looks”

Based on localization performance using frequency-modulated sweeps, Hofman and van Opstal (1998) hypothesized that the elevation localization

system measures spectra on a short time scale of about 5 ms and conducts consecutive short-term analysis of the ongoing information to form “multiple looks.” They also showed that when increasing the inter-burst intervals of trains of 3-ms noise bursts, the performance kept decreasing. Note that the multiple samples can be related to changing HRTFs when the sound lasts long enough for head or pinna to move. The localization study in head-unrestrained cats by May and Huang (1996) examined localization accuracy for paired sound with long inter-stimulus intervals (3 to 5 s) coming from the same location. They showed that head movements towards the sound source during the interval greatly increased the localization accuracy of the second sound.

Figure 10 superimposes the elevational gain for the random-phase stimulus (magenta), the click train (green), and the single click (red). The fact that the maximum gain obtained with a click train (green) was higher than the maximum gain with a single click (red) confirms the hypothesis that multiple looks of the spectral information can improve the elevational localization. Note that this is also the case for azimuthal localization, as is shown by comparing Figures 2A and 5A (blue). Yet despite the multiple looks provided by the click train, its elevational performance was never as good as the performance with the random-phase stimulus (purple), especially at high sound levels. This supports Hofman and van Opstal’s hypothesis that the elevational system is unable to integrate spectral information over a long interval (here the inter-click intervals were 110 ms).

In summary, focusing on a specific phenomenon, the negative level effect, the present study demonstrates that the temporal features of a sound can greatly affect the localization of targets in elevation, which is presumably dominated by spectral cues. The localization performance with speech sound increases monotonically with sound level, which is more likely affected by the amount of energy contained in the relevant frequency range in the speech, rather than the spectral flatness.

ACKNOWLEDGMENTS

We thank J Sekulski and Y Yan for helping with computer programming, and A Jones for animal care and training. This work was supported by NIH DC-07177.

Conflict of Interest There is no conflict of interest in this study.

REFERENCES

- BEGAULT DR, WENZEL EM (1993) Headphone localization of speech. *Hum Factors* 35:361–376
- BEST V, CARLILE S, JIN C, VAN SCHAIK A (2005) The role of high frequencies in speech localization. *J Acoust Soc Am* 118:353–363
- BLACKBURN CC, SACHS MB (1989) Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. *J Neurophysiol* 62:1303–1329
- BRONKHORST AW (1995) Localization of real and virtual sound sources. *J Acoust Soc Am* 98:2542–2553
- BRUNGART DS, SIMPSON BD (2008) Effects of temporal fine structure on the localization of broadband sounds: potential implications for the design of spatial audio displays. *Proceedings of the 14th International Conference on Auditory Display*, Paper No. ICAD08-1-08-5
- CARLILE S, DELANEY S (1999) The localization of spectrally restricted sounds by human listeners. *Hear Res* 128:175–189
- DAVIS KA, RAMACHANDRAN R, MAY BJ (2003) Auditory processing of spectral cues for sound localization in the inferior colliculus. *J Assoc Res Otolaryngol* 4:148–163
- EFRON B, TIBSHIRANI R (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* 1:54–77
- FISHMAN YI, STEINSCHNEIDER M (2009) Temporally dynamic frequency tuning of population responses in monkey primary auditory cortex. *Hear Res* 254:64–76
- GAI Y, RUHLAND J, YIN TC, TOLLIN D (2013) Behavioral and modeling studies of sound localization in cats: effects of stimulus level and duration. *J Neurophysiol* 110:607–620
- GARDNER MB, GARDNER RS (1973) Problem of localization in the median plane: effect of pinnae cavity occlusion. *J Acoust Soc Am* 53:400–408
- GILKEY RH, ANDERSON TR (1995) The accuracy of absolute localization judgments for speech stimuli. *J Vestib Res* 5:487–497
- HARTMANN WM, RAKERD B (1993) Auditory spectral discrimination and the localization of clicks in the sagittal plane. *J Acoust Soc Am* 94:2083–2092
- HARTMANN WM, BEST V, LEUNG J, CARLILE S (2010) Phase effects on the perceived elevation of complex tones. *J Acoust Soc Am* 127:3060–3072
- HEBRANK J, WRIGHT D (1974) Spectral cues used in the localization of sound sources on the median plane. *J Acoust Soc Am* 56:1829–1834
- HOFMAN PM, VAN OPSTAL AJ (1998) Spectro-temporal factors in two-dimensional human sound localization. *J Acoust Soc Am* 103:2634–2648
- INOUE J (2001) Effects of stimulus intensity on sound localization in the horizontal and upper-hemispheric median plane. *J UOEH* 23:127–138
- IRVINE DR, GAGO G (1990) Binaural interaction in high-frequency neurons in inferior colliculus of the cat: effects of variations in sound pressure level on sensitivity to interaural intensity differences. *J Neurophysiol* 63:570–591
- KING RB, OLDFIELD SR (1997) The impact of signal bandwidth on auditory localization: implications for the design of three-dimensional auditory displays. *Hum Factors* 39:287–295
- MACPHERSON EA, MIDDLEBROOKS JC (2000) Localization of brief sounds: effects of level and background noise. *J Acoust Soc Am* 108:1834–1849
- MARTIN RL, WEBSTER WR (1987) The auditory spatial acuity of the domestic cat in the interaural horizontal and median vertical planes. *Hear Res* 30:239–252
- MAY BJ, HUANG AY (1996) Sound orientation behavior in cats. I. Localization of broadband noise. *J Acoust Soc Am* 100:1059–1069
- MIDDLEBROOKS JC (1992) Narrow-band sound localization related to external ear acoustics. *J Acoust Soc Am* 92:2607–2624
- MILLER LM, REGANZONE GH (2009) Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proc Natl Acad Sci U S A* 106:5931–5935
- MOORE JM, TOLLIN DJ, YIN TC (2008) Can measures of sound localization acuity be related to the precision of absolute location estimates? *Hear Res* 238:94–109
- MUSICANT AD, CHAN JC, HIND JE (1990) Direction-dependent spectral properties of cat external ear: new data and cross-species comparisons. *J Acoust Soc Am* 87:757–781
- POPULIN LC, YIN TCT (1998) Behavioral studies of sound localization in the cat. *J Neurosci* 18:2147–2160
- RICARD GL, MEIRS SL (1994) Intelligibility and localization of speech from virtual directions. *Hum Factors* 36:120–128
- RICE JJ, MAY BJ, SPIROU GA, YOUNG ED (1992) Pinna-based spectral cues for sound localization in cat. *Hear Res* 58:132–152
- ROFFLER SK, BUTLER RA (1968) Factors that influence the localization of sound in the vertical plane. *J Acoust Soc Am* 43:1255–1259
- SHAW EA (1974) Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *J Acoust Soc Am* 56:1848–1861
- THURLOW WR, RUNGE PS (1967) Effect of induced head movements on localization of direction of sounds. *J Acoust Soc Am* 42:480–488
- TOLLIN DJ, YIN TC (2003) Spectral cues explain illusory elevation effects with stereo sounds in cats. *J Neurophysiol* 90:525–530
- TOLLIN DJ, YIN TC (2009) Sound localization: neural mechanisms. In: Squire LR (ed). *Encyclopedia of neuroscience*. Elsevier, New York, pp 137–144
- TOLLIN DJ, POPULIN LC, MOORE JM, RUHLAND JL, YIN TC (2005) Sound-localization performance in the cat: the effect of restraining the head. *J Neurophysiol* 93:1223–1234
- TOLLIN DJ, RUHLAND JL, YIN TC (2009) The vestibulo-auricular reflex. *J Neurophysiol* 101:1258–1266
- TOLLIN DJ, McCLAIN EM, YIN TC (2010) Short-latency, goal-directed movements of the pinnae to sounds that produce auditory spatial illusions. *J Neurophysiol* 103:446–457
- TOLLIN DJ, RUHLAND JL, YIN TC (2013) The role of spectral composition of sounds on the localization of sound sources by cats. *J Neurophysiol* 109:1658–1668
- TSAI JJ, KOKA K, TOLLIN DJ (2010) Varying overall sound intensity to the two ears impacts interaural level difference discrimination thresholds by single neurons in the lateral superior olive. *J Neurophysiol* 103:875–886
- VLIEN J, VAN OPSTAL AJ (2004) The influence of duration and level on human sound localization. *J Acoust Soc Am* 115:1705–1713
- ZHOU Y, WANG X (2012) Level dependence of spatial processing in the primate auditory cortex. *J Neurophysiol* 108:810–826
- ZILANY MS, BRUCE IC, NELSON PC, CARNEY LH (2009) A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. *J Acoust Soc Am* 126:2390–2412