Localization performance correlates with binaural fusion for interaurally mismatched vocoded speech

Deepa Suneel, Hannah Staisloff, Corey S. Shayman, Julia Stelmach, and Justin M. Aronoff[a]

Speech and Hearing Science Department, University of Illinois at Urbana-Champaign,
901 South 6th Street, Champaign, Illinois 61820, USA
dsuneel93@gmail.com, staislo2@illinois.edu, shayman@ohsu.edu, stelmac2@illinois.edu,
jaronoff@illinois.edu

Abstract: Bilateral cochlear implant users often have difficulty fusing sounds from the two ears into a single percept. However, measuring fusion can be difficult, particularly with cochlear implant users who may have no reference for a fully fused percept. As a first step to address this, this study examined how localization performance of normal hearing subjects relates to binaural fusion. The stimuli were vocoded speech tokens with various interaural mismatches. The results reveal that the percentage of stimuli perceived as fused was correlated with localization performance, suggesting that changes in localization performance can serve as an indicator for binaural fusion changes.

1. Introduction

Bilateral cochlear implants (CIs) can improve speech understanding in noisy environments as well as improve sound localization (Dunn *et al.*, 2010; van Hoesel, 2004; Litovsky *et al.*, 2012). Patients with bilateral CIs receive some of the benefits that come with binaural hearing, but they often still have difficulty with binaural fusion, i.e., fusing sounds from the two ears together into a unitary percept (Kan *et al.*, 2013). However, reliably measuring binaural fusion can be challenging. The concept of binaural fusion is often difficult to explain to participants, resulting in participant-dependent differences that may not reflect differences in binaural fusion. Additionally, some CI users may not perceive fused sounds with their processors (Fitzgerald *et al.*, 2015) and thus not have a reference for what a fused percept sounds like, increasing the potential unreliability of the task for that population.

Unlike fusion, localization judgments are something that participants regularly do in their daily life. Additionally, localization performance and binaural fusion may be affected by similar factors. For example, Goupell *et al.* (2013) and Kan *et al.* (2013) found that large interaural mismatches resulted in both reduced lateralization (which relies on the same binaural cues as localization) and reduced binaural fusion. As such, although localization likely does not depend solely on binaural fusion, it may be possible to use changes in localization performance as an indicator of changes in binaural fusion. However, this can only be done if changes in localization performance are correlated with changes in binaural fusion. The purpose of this study was to determine whether this is the case.

To examine the relationship between changes in binaural fusion and changes in localization ability, a vocoder simulation was used to manipulate interaural mismatches, which have been shown to affect both binaural fusion and localization abilities (Goupell *et al.*, 2013; Kan *et al.*, 2013).

2. Methods

2.1 Participants

Eight normal hearing listeners, 4 female and 4 male, ages ranging from 22 to 23 years old participated in this experiment. All participants had pure tone thresholds $\leq 25\,dB$ hearing level from 0.25 to 8 kHz. Thresholds did not differ by more than 15 dB between the left and right ear.

2.2 Equipment

The stimuli were presented using an Edirol UA-25 external soundcard and delivered over Sennheiser HDA 200 headphones in a double-walled sound attenuating booth.

---

[a]Author to whom correspondence should be addressed.

The left and right headphones were separately calibrated with the SoundCheck 12.0 software using an artificial ear, microphone, and preamplifier (Brüel and Kjær type 4153, 4192, and 2669, respectively).

2.3 Vocoder

The stimuli were vocoded by first high-pass filtering them at 1200 Hz with a 6 dB per octave roll-off to add pre-emphasis. Next, eight bandpass filters were used for each ear to simulate an eight-channel cochlear implant, with a frequency range between 200 Hz and 9 kHz using a 4th order Butterworth filter with forward filtering. This corresponded to a center frequency spanning 21 mm along the cochlea. These filters were designed to sample frequency ranges that were equally spaced along the cochlea based on the equation by Greenwood (1990). The envelope of each band was extracted by half-wave rectification followed by low pass filtering at 160 Hz using a 4th order Butterworth filter. These parameters were chosen because they are similar to what are typically found in cochlear implants (e.g., Zeng, 2004). The envelopes for each channel were then convolved with narrowband noise.

Interaural mismatches were created by shifting the carrier filters for the apical end of the right simulated array to cover an approximately 20.8, 19.9, 18.4, 16.8, or 15.3 mm extent of the cochlea (see the top panel of Fig. 1), corresponding to an average mismatch of 0, 0.5, 1.2, 2, and 2.8 mm. This approach resulted in high frequency carrier filters that were largely interaurally matched alongside low frequency carrier filters that contained interaural mismatches. These conditions were chosen as ones that would sample a large range of fusion from nearly complete fusion to virtually no fusion based on the results from Aronoff *et al.* (2015). Additionally, the range of interaural mismatches approximately spanned the range of pitch mismatches found with clinical maps in Aronoff *et al.* (2016). All channels were summed for each ear and the waveforms were combined to create a single two-channel audio signal.

2.4 Fusion task

The subjects were seated in front of a monitor within a double-walled sound attenuating booth. The stimuli were based on those from Aronoff *et al.* (2015) and consisted of vocoded versions of the words *yam* and *pad* spoken by a male speaker, presented at 65 dB(A). Two fusion measures were used, referred to here as the fused/unfused measure and the punctate/diffuse measure. For the fused/unfused measure, the following question was presented on the screen for each participant: "Do you hear the same sound at both ears or a different sound at each ear?" Participants had two buttons to
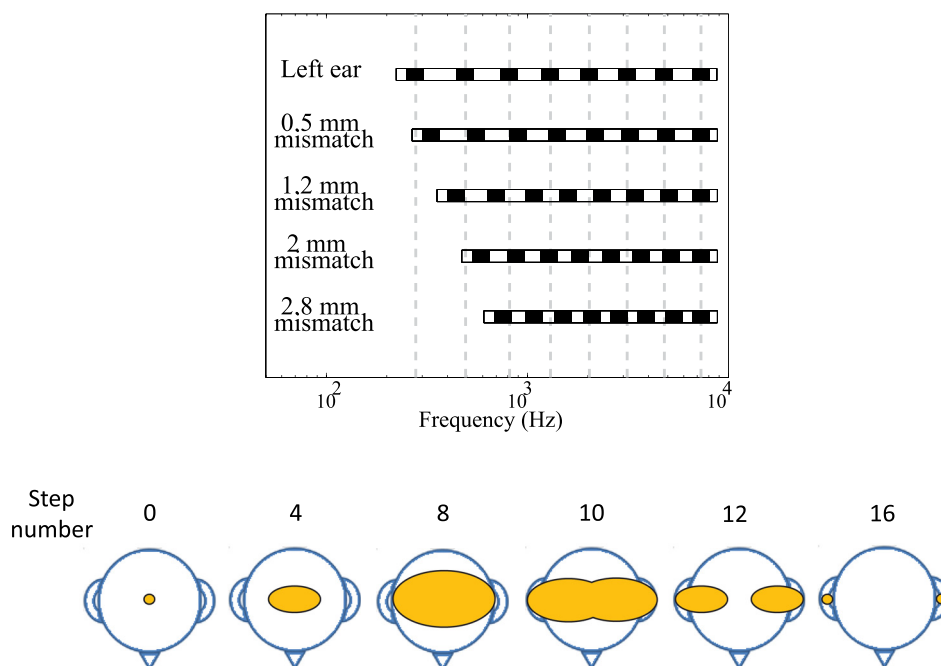

Fig. 1. (Color online) Experimental procedures. The top panel shows the center of the carrier filters across conditions. The analysis filters were always the same as the left ear carrier filters. The bottom panel shows examples from the punctate/diffuse measure. The numbers indicate the step number used for analysis, the heads indicate what the participant saw for different amounts of dial movements.

select from, one labeled "same" and one labeled "different." The next trial was initiated once the participant selected a response. There were 60 trials per condition for each of the five conditions. Trials were grouped into three blocks of 100 trials each. All conditions were tested in all blocks.

For the punctate/diffuse measure, participants indicated what they heard, ranging from a punctate sound to a diffuse sound to separate sounds at each ear. Participants saw an image of a head with a small oval in the center. By turning a dial, they were able to make the oval, representing the number and size of the auditory images in their head, larger or smaller (see the bottom panel of Fig. 1). Initially, clockwise dial turns increased the size of the oval and counter-clockwise dial turns decreased the size of the oval. Once the oval was as large as the head, continuing to turn the dial clockwise resulted in the oval splitting into two ovals, one at each ear, with the two ovals decreasing in size as the dial was moved further clockwise. The dial creates discrete steps and the number of steps (i.e., $0 =$ small oval in the center of the head; $8 =$ large oval in the center of the head; $16 =$ small ovals at each ear; see Fig. 1) was used as an ordinal scale of fusion. Responses could range anywhere from 0 to 16 steps. There were 60 trials per condition for each of the five conditions. Trials were grouped into three blocks of 100 trials each. All conditions were tested in all blocks.

2.5 Localization task

For the localization task, the stimuli from the fusion task were processed by a head related transfer function (HRTF) to create virtual locations. These HRTFs were generated by recording from a microphone in a Zwislocki coupler on a Knowles Electronics Manikin for Acoustic Research (KEMAR) and were the HRTFs labeled "acoustic hearing" in Aronoff *et al.* (2012). The vocoded stimuli were convolved with the HRTF. To reduce the time requirements for the localization task, only the *yam* stimuli were used. The same mismatch conditions from the fusion task were used (i.e., an average mismatch of 0, 0.5, 1.2, 2, and 2.8 mm).

Testing largely followed the procedures in Aronoff *et al.* (2012). Subjects were asked to locate the stimulus presented from one of twelve virtual locations in the rear field, where the sensitivity to auditory spatial cues is most critical. The locations were spaced 15° apart, ranging from 97.5° to 262.5°. The locations were numbered from 1 to 12, with number 1 located at 97.5° (right) and number 12 located at 262.5° (left). Subjects were provided with a screen that showed each location relative to a drawing of a head. The subject's task was to identify the location from which the stimulus originated, clicking on the button located at the appropriate position.

Prior to testing in each condition, subjects were familiarized with the stimulus locations by listening to the stimulus presented once at each of the twelve locations, starting from the rightmost location and ending at the leftmost location. The location of each stimulus was indicated by having the corresponding location briefly turn yellow. After familiarization, subjects were presented with a practice test that included each location presented once in a random order. There was no minimum score required to move from the practice to the test session. No feedback was provided during the practice or test sessions. For the practice and test sessions, the target was presented once at a given location prior to the subject indicating the perceived location. Each location was presented six times, randomly ordered. The final score was determined by calculating the RMS localization error, in degrees, based on all 72 responses.

3. Data analysis

Because there is dependency across conditions and dependency across tasks (i.e., the same people completed all tasks and conditions), but not across subjects, and dependent variables are being compared, it was not valid to use a standard regression, a standard correlation, or a mixed effects regression analysis. To handle the complex dependencies, a bootstrap analysis was conducted whereby participants were randomly chosen with replacement and the entire dataset for that participant was used as part of the bootstrap distribution. This process was repeated until the number of participants in the bootstrap distribution was the same as the number in the original dataset. Because sampling was done with replacement, some participants were included multiple times and some not at all for a given bootstrap distribution. This meant that the bootstrap distributions maintained the same dependencies as the original dataset.

To determine the relationship between localization ability and the two fusion metrics, two analyses were conducted. A Pearson correlation was conducted comparing the localization and fused/unfused measure using the bootstrap distribution. Because the punctate/diffuse measure is ordinal data, a Spearman correlation was conducted

comparing the localization and punctate/diffuse measure using the same bootstrap distribution. The r values from the correlations were recorded.

This process of resampling with replacement and conducting Pearson and Spearman correlations was repeated 599 times and r values from each analysis were rank ordered. The range of the 572 (i.e., 95%) central r values for each analysis indicated the 95% confidence intervals for the correlation. If these confidence intervals included zero, it would indicate that there was not a significant correlation.

4. Results

The correlation between localization performance and the percentage of stimuli perceived as fused was analyzed. The results indicated that there was a significant inverse correlation between localization performance and performance on the fused/unfused task (95% confidence interval for $r = -0.6$ to -0.8; median value $= -0.7$; note that increased fusion leads to higher numbers on the fused/unfused task but better localization is indicated by lower numbers). The correlation between localization performance and the diffuseness of the percept was also analyzed. There was also a significant relationship between localization performance and performance on the punctate/diffuse task (95% confidence interval for $r = 0.1$ to 0.6; median value $= 0.3$; note that more punctate sounds result in lower numbers and better localization also leads to lower numbers; see Fig. 2).

5. Discussion

Localization RMS errors varied with the degree of interaural mismatch. However, even in the condition with no mismatch, the localization error averaged over listeners was 31°, which is slightly larger than errors usually found in CI listeners. For example, Nopp *et al.* (2004), Aronoff *et al.* (2010), and Majdak *et al.* (2011) found RMS errors of 29°, 27°, and 23°, respectively. The better localization performance found in CI listeners may originate from their better adaptation to the listening situation. NH subjects typically listen to stimuli without any interaural mismatch. Also the NH subjects in the current study did not have training in listening via a CI simulation, which reduces the availability of cues required for accurate sound localization.

Similarly, both the likelihood of a fused percept and the diffuseness of the percept varied with the degree of interaural mismatch. It is difficult to directly compare fusion across studies since different studies use different measures of fusion. However, subjects were slightly more sensitive to increases in interaural mismatch than predicted based on the linear fit in Aronoff *et al.* (2015) of similarly vocoded and mismatched stimuli (26% per mm mismatch in Aronoff *et al.*, 2015, compared to 33% in the current study). This may reflect subject-specific differences in ITD sensitivity, consistent with the variability in performance across subjects in the current study (see Fig. 2).

Comparing fusion and localization results indicated that fusion correlated with localization performance. The correlation was strongest with the fused/unfused task, accounting for nearly 50% of the variance. In contrast, the correlation with the punctate/diffuse task accounted for less than 10% of the variance, suggesting that localization is less affected by the diffuseness of a percept. Additionally, despite results from the fused/unfused task indicating that unfused percepts were common, particularly for the 2 and 2.8 mm mismatch conditions, only a few participants perceived the stimuli as split into a left and right ear percept in the punctate/diffuse task (a score of 9 or
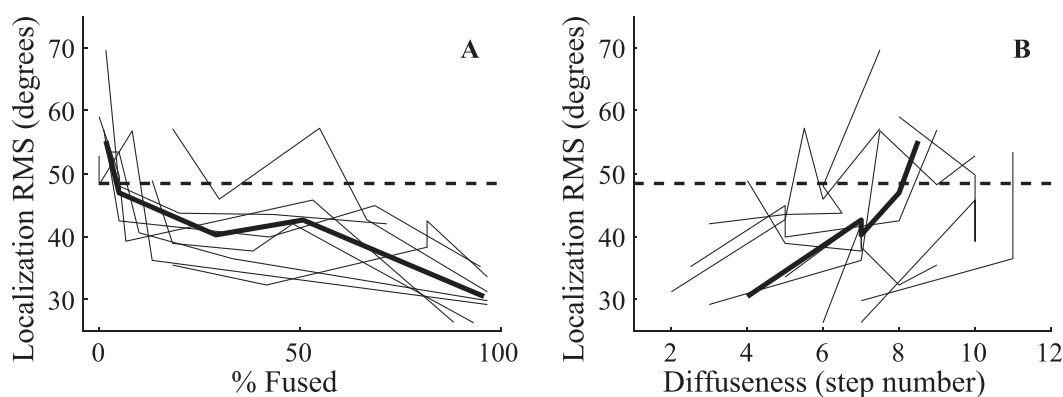
Fig. 2. Results showing that localization performance is correlated with fusion. Light lines indicate individual performance, dark lines indicate median values. The dashed line indicates chance as determined in Aronoff *et al.* (2010) using a Monte Carlo simulation.

more). This may suggest that the measure of binaural fusion depends on the task, and the interpretation of the percentage of fused responses is complicated.

An important caveat is that, unlike the NH subjects in the current study, CI users have considerable deficits in terms of interaural time difference (ITD) sensitivity when using their clinical processors (Aronoff *et al.*, 2010; Seeber and Fastl, 2008). The availability of ITD cues was minimally altered in this experiment, particularly since the HRTFs were applied after vocoding, better preserving fine structure ITD cues. The degree to which the correlations found in this study will extend to cochlear implant patients may depend on the role of ITD cues in these correlations and requires further study.

In conclusion, the results of the study indicate that binaural fusion can be correlated with localization performance. However, the degree of the correlation depends on the task used. Within a listener, a degradation in localization performance can serve as an indicator for less binaural fusion when listening to interaurally mismatched vocoded sounds.

Acknowledgments

References and links

Aronoff, J. M., Freed, D. J., Fisher, L. M., Pal, I., and Soli, S. D. (**2012**). "Cochlear implant patients' localization using interaural level differences exceeds that of untrained normal hearing listeners," J. Acoust. Soc. Am. **131**, EL382–EL387.

Aronoff, J. M., Padilla, M., Stelmach, J., and Landsberger, D. M. (**2016**). "Clinically paired electrodes are often not perceived as pitch matched," Trends Hear. **20**, 1–9.

Aronoff, J. M., Shayman, C., Prasad, A., Suneel, D., and Stelmach, J. (**2015**). "Unilateral spectral and temporal compression reduces binaural fusion for normal hearing listeners with cochlear implant simulations," Hear. Res. **320**, 24–29.

Aronoff, J. M., Yoon, Y., Freed, D. J., Vermiglio, A. J., Pal, I., and Soli, S. D. (**2010**). "The use of interaural time and level difference cues by bilateral cochlear implant users," J. Acoust. Soc. Am. **127**, EL87–EL92.

Dunn, C. C., Noble, W., Tyler, R. S., Kordus, M., Gantz, B. J., and Ji, H. (**2010**). "Bilateral and unilateral cochlear implant users compared on speech perception in noise," Ear Hear. **31**, 296–298.

Fitzgerald, M. B., Kan, A., and Goupell, M. J. (**2015**). "Bilateral loudness balancing and distorted spatial perception in recipients of bilateral cochlear implants," Ear Hear. **36**, e225–e236.

Goupell, M. J., Stoelb, C., Kan, A., and Litovsky, R. Y. (**2013**). "Effect of mismatched place-of-stimulation on the salience of binaural cues in conditions that simulate bilateral cochlear-implant listening," J. Acoust. Soc. Am. **133**, 2272–2287.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species–29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Kan, A., Stoelb, C., Litovsky, R. Y., and Goupell, M. J. (**2013**). "Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users," J. Acoust. Soc. Am. **134**, 2923–2936.

Litovsky, R. Y., Goupell, M. J., Godar, S., Grieco-Calub, T., Jones, G. L., Garadat, S. N., Agrawal, S., Kan, A., Todd, A., Hess, C., and Misurelli, S. (**2012**). "Studies on bilateral cochlear implants at the University of Wisconsin's Binaural Hearing and Speech Laboratory," J. Am. Acad. Audiol. **23**, 476–494.

Majdak, P., Goupell, M. J., and Laback, B. (**2011**). "Two-dimensional localization of virtual sound sources in cochlear-implant listeners," Ear Hear. **32**, 198–208.

Nopp, P., Schleich, P., and D'Haese, P. (**2004**). "Sound localization in bilateral users of MED-EL COMBI 40/40+ cochlear implants," Ear Hear. **25**, 205–214.

Seeber, B. U., and Fastl, H. (**2008**). "Localization cues with bilateral cochlear implants," J. Acoust. Soc. Am. **123**, 1030–1042.

van Hoesel, R. J. (**2004**). "Exploring the benefits of bilateral cochlear implants," Audiol. Neurootol. **9**, 234–246.

Zeng, F. G. (**2004**). "Trends in cochlear implants," Trends Amplif. **8**, 1–34.