Sound source localization is a multisystem process

William A. Yost1,1*, M. Torben Pastore1 and Michael F. Dorman2
1Spatial Hearing Laboratory, Speech and Hearing Science, Arizona State University, PO Box 870102, Tempe, Arizona, 85287, USA
2Cochlear Implant Laboratory, Speech and Hearing Science, Arizona State University, PO Box 870102, Tempe, Arizona, 85287, USA

Abstract: A review of data published or presented by the authors from two populations of subjects (normal hearing listeners and patients fit with cochlear implants, CIs) involving research on sound source localization when listeners move is provided. The overall theme of the review is that sound source localization requires an integration of auditory-spatial and head-position cues and is, therefore, a multisystem process. Research with normal hearing listeners includes that related to the Wallach Azimuth Illusion, and additional aspects of sound source localization perception when listeners and sound sources rotate. Research with CI patients involves investigations of sound source localization performance by patients fit with a single CI, bilateral CIs, a CI and a hearing aid (bimodal patients), and single-sided deaf patients with one normal functioning ear and the other ear fit with a CI. Past research involving CI patients who were stationary and more recent data based on CI patients’ use of head rotation to localize sound sources is summarized.

Keywords: Sound source localization, Auditory-spatial cues, Head-position cues, Cochlear implants

PACS number: 43.66.QP [doi:10.1250/ast.41.113]

1. INTRODUCTION

This article provides a review of two related research topics: Sound Source Localization as a Multisystem Process and Sound Source Localization by Cochlear Implant (CI) Patients. The review focuses on research published or presented by the authors, but includes other published research. The overarching theme is that studies of sound source localization, when sound sources and listeners move, reveal that localizing sound sources in an actual sound field requires an integration of (head-centric) auditory-spatial cues and cues providing information about the position of the head in the sound field (world-centric head position). Because cues indicating world-centric head position are provided by a variety of neural systems, sound source localization is a multisystem process. CI patients can take advantage of head rotation to avoid front-back reversals (FBRs) when locating sound sources. Some of the literature that supports the proposal that sound source localization is a multisystem process will be reviewed, as will a literature indicating the conditions in which CI patients can localize sound sources and when they can benefit from head rotations.

2. SOUND SOURCE LOCALIZATION: A MULTISYSTEM PROCESS

Wallach [1] was probably the first to propose that sound source localization requires an interaction of auditory-spatial and head-position cues. Yost et al. [2] made a similar assertion. Figure 1 shows how the world-centric (w) angle of a stationary loudspeaker presenting sound (S), $\theta_{Sw}$, could be determined based on the sum (see Yost et al. [2]) of the head-centric (h) sound source angle, $\theta_{Sh}$, which is estimated entirely using auditory spatial cues, and the world-centric head (H) position angle, $\theta_{Hw}$ (H or S refer to the head or sound source, w or h to either a world-centric or head-centric reference system, and r to a reverse angle). $\theta_{Hw}$ could be estimated by any number of neural systems such as vision. Note that $\theta_{Sw}$ cannot be determined as a sole function of auditory-spatial cues, nor can it be determined by other neural system cues indicating head position alone. The geometric explanation shown in Fig. 1 is not the only way to explain the integration of auditory-spatial and head-position cues [1] nor is it a formal model. Such models would require connecting auditory-spatial cues to head-centric angle, head-position cues to head angle, and specifying a method of integrating the cues [3].

If the auditory spatial cues are interaural time (ITD) or level (ILD) differences, then two head-centric angles
produce the same ITD and (probably) ILD (Fig. 1A). For any sound source location there is a corresponding head-centric angle $\theta_{Sh}$, but this angle also specifies a front-back reversed location, $\theta_{Shr}$, at $180^\circ - \theta_{Sh}$. This leads to the possibility of an FBR, where the head-centric angle is $\theta_{Sh}$ rather than $\theta_{Shr}$. If there are auditory-spatial cues other than ITDs and/or ILDs (e.g., spatial cues derived from spectral complexities resulting from the head-related transfer function, HRTF), then these cues provide listeners enough information to localize sound sources to the correct front-back location. Thus, spectral complexity (especially for high frequencies) can help resolve FBRs.

Wallach [1] indicated how a head rotation could also disambiguate an FBR yielding the perception of the sound source at angle $\theta_{Sw}$. Figure 1B indicates a rotation of the head, and the resulting change in the angles ($\theta_{Sw}', \theta_{Sh}', \theta_{Swr}', \theta_{Shr}'$, where $\theta'$ indicates an estimate of an angle after a change due to sound source and/or listener rotation). Wallach [1] proposed a "selective principal of rest" to suggest that the perceived location of a sound source when other sources do not move is that which does not change location (remains "at rest") as the head moves. Thus, because $\theta_{Sw}'$ does not change as the head moves in Fig. 1B compared to Fig. 1A, $\theta_{Sw}'$, and not, $\theta_{Swr}'$, is the predicted angle of the perceived sound source (i.e., the sound source would be at loudspeaker location #4, not loudspeaker location #14, in Fig. 1B).

In Fig. 1C, both the sound source and the listener rotate such that the sound source rotates at twice the angle of the listener (compare Fig. 1A to Fig. 1C). Wallach [1] suggested this "2-1" rotation scenario to investigate the interaction of auditory-spatial and head-position cues. In this "2-1" rotation scenario, the angle, $\theta_{Swr}'$, does not change as the head rotates (the reverse sound source remains at loudspeaker #12 in Fig. 1C compared to Fig. 1A); but $\theta_{Sw}'$ does rotate (actual sound moves from loudspeaker #4 to #6). Using Wallach’s [1] “selective principal of rest,” this leads to what has been called the Wallach Azimuth Illusion [4]; listeners perceive a stationary sound source even though the actual sound source is rotating [1,4–8]. As the head and sound rotate, the illusory, stationary location of the sound source is at $\theta_{Swr}$, the reversed world-centric location of the actual sound source. In other words, if the sound and listener start rotating when the listener faces the sound source, the illusionary perception is of a stationary world-centric sound source location behind the listener. The robust Wallach Azimuth Illusion is strong evidence in support of Wallach’s [1] proposal that sound source localization requires an interaction of auditory-spatial and head-position cues, and is consistent with Yost et al.’s [2] argument that sound source localization is a multisystem process.

In a study (eight listeners) of the Wallach Azimuth Illusion, Pastore and Yost [7] rotated listeners clockwise at constant velocity in a computer-controlled rotating chair at $45^\circ$/s, and presented filtered noise bursts, rotating the sounds clockwise from one loudspeaker to another around a 24-loudspeaker azimuth array (Fig. 1) at $90^\circ$/s, consistent with the “2-1” rotation scenario. Five filtered, 200-ms noise bursts (shaped with $20\text{ms} \cos^2$ rise/fall times and presented at 65 dBA) were presented: low-frequency (center frequency, CF, of 250 Hz), two-octave wide noise
The data confirm the predictions described above for eyes-open conditions (Panels A, B) with LF noise stimuli: nearly all listeners perceived a stationary sound source at the “front-back reversed” location. The eyes-open data also confirm the predictions described above for the HF, 2-O and BB filtered noises: listeners perceived a rotating sound source. The data for the HF, 1/10-O eyes-open conditions show that some listeners indicated that the perceived sound source was rotating while others indicated that it was stationary. The differences between listeners were correlated with the percent of FBRs they demon-
strated. For those conditions with the highest percent of FBRs, listeners indicated a stationary sound source. For those conditions with the lowest percent of FBRs, listeners indicated a rotating sound source, somewhat consistent with the predictions provided above.

When listeners’ eyes were closed, the data of Fig. 2 suggest that for seven of the eight listeners the predictions described above were accurate in that these seven listeners usually indicated that all of the filtered noises rotated clockwise around their head. Counterclockwise responses were recorded only 12 times and for only three listeners, all for the two LF filtered-noise conditions. The eighth listener indicated that the sound image was stationary (90% of the time) in the eyes-closed condition, although this listener could not reliably indicate which loudspeaker presented the sounds.

The data for the eyes-closed condition suggest that the listening condition was a head-centric reference system for most listeners because there appeared to be little, or no, world-centric head-position information when the naïve listeners were slowly rotated at constant velocity with their eyes closed [2]. Had there been significant head-position information it is likely it could have been integrated with the auditory-spatial cues to generate the Wallach Azimuth Illusion of a stationary sound source, rather than the rotating sound that was reported by most listeners most of the time. The interview with the eighth listener following the experiment revealed that providing somewhat different instructions as how to respond in the eyes-open versus eyes-closed conditions might confuse some listeners. Thus, the data for the eyes-closed condition should be viewed with caution, especially concerning conditions that substantially reduce world-centric head-position cues.

The results of the Wallach Azimuth Illusion experiment described above are in good agreement with the results of the Yost et al. [2] study. That is, world-centric localization requires, as Wallach [1] and Yost et al. [2] stated, an integration of auditory-spatial and head-position cues. Without head-position cues, sound source localization is based on a head-centric reference system and listeners cannot locate the source of sound within the environment in which the sound source exists. Yost et al. [2] used three experimental paradigms to show that head-position cues are integrated with auditory spatial cues to determine the measures of sound source localization perception obtained in these experiments. In all conditions, listeners rotated at 45°/s at constant velocity, accelerated or decelerated at 9°/s², or remained stationary. A broadband (125 Hz to 8,000 Hz) filtered noise burst rotated from loudspeaker to loudspeaker around the 24-loudspeaker array (see Fig. 1) or remained stationary. There is no output indicating listener rotation recorded by the semicircular canals when rotation is at constant velocity, but there is vestibular output when rotation is accelerating and decelerating — see Lackner and DiZio [11] for an explanation of the vestibular system as an accelerometer. Thus, varying the form of rotation allows for control of vestibular information about listener rotation. The results of Yost et al. [2] clearly showed that constant-velocity rotation when listeners’ eyes were closed did not provide head-position information, as
is consistent with the results described above for Pastore and Yost [7].

The Yost et al. [2] results were mixed regarding whether or not listeners, when their eyes were closed and they were accelerating or decelerating, made sound source localization judgments in a world- or head-centric reference system. Yost et al. [2] suggested that some problems with the way in which listeners were rotated in their study might confound interpretations of the role of acceleration and deceleration versus constant velocity rotation. Thus, it is still not clear how the form of rotation effects head-position cues and how it might reveal information about vestibular function in determining head-position cues.

However, when the eyes are open, a growing literature partially reviewed above strongly supports Wallach’s [1] proposal that locating sound sources in the real world requires an integration of auditory-spatial and head-position cues. One of the many additional ways to explore this proposal is to investigate how impairment in processing auditory-spatial cues and/or head-position cues affect sound source localization perception. In the next section, we provide a brief summary of research completed by the authors involving sound source localization with hearing-impaired listeners fitted with Cochlear Implants (CI).

3. SOUND SOURCE LOCALIZATION BY COCHLEAR IMPLANT (CI) PATIENTS

Normal hearing and CI listeners indicated which of thirteen loudspeakers, spaced 15 deg. apart in the front hemifield, presented a noise burst (loudspeakers #19 to #7 in Fig. 1). Stimuli were only presented from loudspeakers #20 to #6, in order to avoid edge effects [12]. The stimuli were one of three filtered noise bursts (200-ms duration, 20-ms \( \cos^2 \) rise/fall times, 65 dBA): 2-Octave wide bandpass noise with \( CF = 250 \) Hz (LF), 2-Octave wideband noise with \( CF = 4,000 \) Hz (HF), and a 125-Hz to 8,000-Hz filtered broadband noise (BB). These filter conditions were chosen in order to implicate ITD processing for the LF noise, ILD processing for the HF noise, and both ITD and ILD processing for the BB noise.

Figure 3 displays data (mean, range, and standard deviation) from several published experiments [13–20] conducted by the authors regarding sound source localization accuracy measured as root-mean-square (rms) error in degrees (azimuth plane). Panel 3A are data for bilateral CI patients presented with the three filtered noises. Figure 3B shows rms error as a function of device configuration: One CI (patients have severe hearing loss in both ears and only one ear has been fit with a CI), Bilateral CIs; BiModal (CI fit to one ear and a hearing aid to the other); and single-sided deaf (SSD, CI fit to one ear and normal unaided hearing in the other ear). Additional details about these subject populations and their devices can be found in the literature cited above. In addition, normal-hearing mean rms error in degrees and ± one-standard deviation are included as the bar labeled “NH.” The 95% confidence interval for Chance performance (rms error degrees) in the localization identification task is indicated by the upper, gray-shaded horizontal bar. Numbers at the top Fig. 3 indicate the number of listeners (“n”) for each CI configuration. NH data represent 95 listeners [21–24].

As the articles cited above indicate, and as is evident in Fig. 3, there is enormous variability in sound source localization accuracy across patients in and across the various CI populations. The majority of bilateral CI
patients localize the LF filtered noises with far less accuracy (high rms error in degrees) than the HF or BB filtered noises, suggesting that CI users are not as able to process ITDs as well as they are ILDs for sound source localization. Similar conclusions have been reached by several other studies [25,26]. This is most likely because acquiring ITD information requires an interaural neural comparison of a sound’s temporal fine structure (TFS). This is probably not possible for bilateral CI users, as the devices at the two ears are not synchronized, CI devices process a sound’s envelope and not its TFS, and CI electrodes cannot deliver sounds to the cochlear apex where low-frequency TFS is processed.

NH listeners have statistically the same rms error for all three filtered noises [21,22] implying that for sounds with bandwidths of two octaves or more sound source localization accuracy is the same for ITD and/or ILD processing [21,22]. On average, CI patients have similar rms error in degrees for the HF and BB filtered noises, suggesting that as long as CI patients have access to ILD cues, they can localize the sources of sound in the front azimuthal hemifield, but rarely with the same accuracy as NH listeners. CI configuration also makes a difference in sound source localization accuracy, as the references cited above describe.

Spectral cues are not available to assist sound source localization at low frequencies, and CI patients are poor at localizing low-frequency sound sources, so we asked: “Can bilateral CI patients’ process spectral cues for the high-frequency sound sources many of them can localize, and what role might head rotation play in their ability to localize high-frequency sound sources?” Microphones used in many CI processors are placed behind-the-ear (BTE), not allowing the sound to be affected by the pinna and, therefore, probably not allowing these CI patients an ability to take advantage of pinna-based (HRTF) spectral cues. Because many of these bilateral CI patients (see Fig. 3) can localize high-frequency sound sources on the azimuth plane, one might infer that these CI patients use ILD cues because there are few, if any, spectral cues available to them. If so, they might be prone to FBRs for high-frequency sounds, whereas NH listeners are not (i.e., NH listeners have access to spectral cues).

In a study of the role of head rotation in sound source localization, normal hearing and CI listeners were asked to localize six sound sources spaced an equal distance apart (60°) around the azimuth circle (the six loudspeaker locations marked with dark squares shown in the middle inset of Fig. 4 represent the six loudspeakers used in the experiment). The stimulus was a 3-second duration, high-frequency (HF) filtered noise (2,000–8,000 Hz). The four panels (A, B, C, and D) of Fig. 4 display confusion matrices (i.e., the loudspeaker location reported by listeners as a function of the location of the actual loudspeaker presenting a sound). In the top row listeners (Fig. 4A: Normal Hearing listeners and Fig. 4B: CI patients) kept their heads still. Figures 4C and 4D show confusion matrices for the same listeners.

**Fig. 4** Confusion matrices for normal hearing (NH, 7 listeners) and Cochlear Implant (CI, 7 listeners) listeners of reported loudspeaker locations as a function of presented loudspeaker locations. The six loudspeaker locations shown in inset. Black bubbles for correct actual sound source responses, gray bubbles for FBRs, open bubbles for errors. A: NH listeners, no head movement; B: CI users, no head movement; C: NH listeners, head movement; D: CI users, head movement.
whose data are depicted in the top row, when the listeners were instructed to move their heads in order to help them localize the sounds. The instructions emphasized that head movements must be limited to within ±30° rotation and that listeners should move only their head, not their body or shoulders. The confusion matrices are organized so that dark bubbles on the positive diagonal represent correct sound source localization responses, grey bubbles on the negative diagonal represent FBRs, and bubbles in any other cell represent sound source localization errors. The area of the bubble is proportional to the total number of responses across listeners and repetitions (12 repetitions per loudspeaker per listener).

The results of the Pastore et al. [27] study indicate that when NH listeners did not move their heads they were almost 100% correct in accurately localizing the six sound sources for the HF filtered noise, and therefore there were almost no FBRs (Fig. 4A). The CI patients, on the other hand, had far fewer correct sound source localization responses with a large number of FBRs (41.9% FBRs) when they did not move their heads (Fig. 4B). There was essentially no change in the response pattern shown in the confusion matrix for the NH listeners when they moved their head (Fig. 4C). However, when CI patients rotated their heads, there was a substantial decrease in FBRs (to 6.7% FBRs) and a corresponding increase in correct responses (Fig. 4D). As the literature [9,10] indicates, NH listeners typically have very few FBRs for broadband, high-frequency sounds as NH listeners have access to HRTF spectral cues. Consequently, sound source localization accuracy for NH listeners does not change much, if at all, for such broadband, high-frequency stimuli when they rotate their heads.

In summary, the CI patients in this study with BTE microphones do not have access to HRTF spectral cues and CI patients, in general, have substantially reduced spectral resolution due to the filtering and electrode placement used by CI devices. Thus, if CI patients can localize high-frequency sound sources it is likely that they do so on the basis of ILDs (not spectral cues), which are likely to produce FBRs. Figure 4 shows that these CI patients do have substantial FBRs consistent with the argument made above. In addition, bilateral CI patients are able to reduce FBRs for high-frequency sounds by rotating their heads.

The data of Fig. 4 involving CI patients and their ability to use head rotations to help disambiguate FBRs is consistent with the overall theme of this review, that sound source localization is a multisystem process involving an integration of auditory-spatial and head-position cues. That is, as indicated in Fig. 4 head rotations can reduce FBRs because auditory-spatial cues are integrated with head-centric head-position cues, consistent with the arguments originally proposed by Wallach [1]. CI patients appear to integrate auditory-spatial cues (i.e., ILDs) with head-position cues to reduce FBRs, in ways similar to NH listeners.

ACKNOWLEDGMENTS

Research supported by a grant (to WAY) from the National Institute on Deafness and Other Communication Disorders (NIDCD), an NIDCD Post-Doctoral Fellowship (to MTP), and a grant from Facebook Reality Labs (to WAY and MTP); and by support from Advanced Bionics and MED EL Corporation (to MFD). We want to acknowledge the assistance of Dr. Yi Zhou, Kathryn Pulling, Sarah Natale, and Chris Montagne.

REFERENCES

[1] H. Wallach, “The role of head movements and vestibular and visual cues in sound localization,” J. Exp. Psychol., 27, 339–368 (1940).
[2] W. A. Yost, X. Zhong and A. Najam, “Judging sound rotation when listeners and sound rotate: Source sound localization is a multisensory process,” J. Acoust. Soc. Am., 38, 3293–3308 (2015).
[3] J. Braasch, S. Clapp, A. Parks and M. T. Pastore, “A binaural model that analyses aural spaces and stereophonic reproduction systems by utilizing head movements,” in The Technology of Binaural Listening, J. Blauert, Ed. (Springer-Verlag, Berlin-Heidelberg, 2013), pp. 201–224.
[4] W. O. Brimijoin and M. A. Akeryod, “The role of head movements and signal spectrum in an auditory front/back illusion,” Iperception, 3, 179–181 (2012).
[5] E. A. Macpherson, “Head motion, spectral cues, and Wallach’s ‘principle of least displacement’ in sound localization,” in Principles and Applications of Spatial Hearing, Y. Suzuki, D. Brungart and H. Kato, Eds. (World Scientific, Hong Kong, 2011), pp. 103–120.
[6] W. O. Brimijoin and M. A. Akeryod, “The effects of hearing impairment, age, and hearing aids on the use of self motion for determining front/back location,” J. Am. Acad. Audiol., 27, 588–600 (2016).
[7] M. T. Pastore and W. A. Yost, “Sound source localization as a multisensory/multisystem process: The Wallach Azimuth Illusion,” J. Acoust. Soc. Am., S141, 3635 (2017).
[8] M. T. Pastore and W. A. Yost, “Sound source localization as a multisensory process: Wallach Azimuth Illusion,” J. Acoust. Soc. Am., 146, 382–398 (2019).
[9] J. C. Middlebrooks and D. M. Green, “Sound localization by human listeners,” Annu. Rev. Psychol., 42, 135–159 (1991).
[10] J. Blauert, Spatial Hearing (The MIT Press, Cambridge, MA, 1997).
[11] J. R. Lackner and P. DiZio, “Vestibular, proprioceptive, and haptic contributions to spatial orientation,” Annu. Rev. Psychol., 56, 115–147 (2004).
[12] W. H. Hartmann, B. Rakerd and J. B. Gaalaas, “On the source-identification method,” J. Acoust. Soc. Am., 104, 3546–3557 (1998).
[13] M. F. Dorman, W. A. Yost, B. S. Wilson and R. H. Gifford, “Speech perception and sound localization by adults with bilateral cochlear implants, in Seminars in Hearing 32 (Thieme Medical Publishers, New York, 2011), pp. 73–89.
[14] M. F. Dorman, A. Spahr, L. Loiselle, T. Zhang, S. Cook, S. Fiebig and W. A. Yost, “Localization and speech understanding by a patient with bilateral cochlear implants and
bilateral hearing preservation," *Ear Hear.*, **34**, 245–248 (2012).

[15] M. F. Dorman, L. Loiselle, W. A. Yost, J. Stohl, A. Spahr, C. Brown and S. Cook, “Interaural level differences and sound source localization for bilateral cochlear implant patients,” *Ear Hear.*, **35**, 633–640 (2014).

[16] L. Loiselle, M. F. Dorman, W. A. Yost and R. H. Gifford, “Sound source localization by hearing preservation patients with and without symmetric low-frequency acoustic hearing,” *Audiol. Neurotol.*, **19**, 234–256 (2014).

[17] D. M. Zeitler, M. F. Dorman, S. J. Natale, L. Loiselle, W. A. Yost and R. H. Gifford, “Sound source localization and speech understanding in complex listening environments by single-sided deaf listeners after cochlear implantation,” *Otol. Neurotol.*, **36**, 1467–1471 (2015).

[18] M. F. Dorman, D. Zeitler, S. Cook, L. Loiselle, W. A. Yost, G. B. Wanna and R. H. Gifford, “Interaural level difference cues determine sound source localization by single-sided deaf patients fit with a cochlear implant,” *Audiol. Neurotol.*, **20**, 183–188 (2015).

[19] M. F. Dorman, L. Loiselle, S. Cook, W. A. Yost and R. H. Gifford, “Sound source localization by normal hearing listeners, hearing-impaired listeners and cochlear implant listeners,” *Audiol. Neurotol.*, **21**, 127–131 (2016).

[20] L. Loiselle, M. F. Dorman, W. A. Yost, S. Cook and R. H. Gifford, “Using ILD and ITD cues for sound source localization and speech understanding in complex listening environment by listeners with bilateral and with hearing-preservation cochlear-implants,” *J. Speech Lang. Hear. Res.*, **59**, 810–818 (2016).

[21] W. A. Yost, L. Loiselle, M. F. Dorman, C. Brown and J. Burns, “Sound source localization of filtered noises by listeners with normal hearing: A statistical analysis,” *J. Acoust. Soc. Am.*, **133**, 2876–2882 (2013).

[22] W. A. Yost and X. Zhong, “Sound source localization identification accuracy: Bandwidth dependencies,” *J. Acoust. Soc. Am.*, **136**, 2737–2746 (2014).

[23] W. A. Yost, “Sound source localization accuracy: Level and duration dependencies,” *J. Acoust. Soc. Am.*, **140**, EL14–EL19 (2016).

[24] W. A. Yost, “Sound source localization accuracy: Envelope dependencies,” *J. Acoust. Soc. Am.*, **142**, 173–185 (2017).

[25] D. W. Grantham, D. H. Ashmead, T. A. Ricketts, R. F. Labadie and D. S. Haynes, “Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants,” *Ear Hear.*, **28**, 524–541 (2007).

[26] R. J. van Hoesel, “Exploring the benefits of bilateral cochlear implants,” *Audiol. Neurotol.*, **9**, 234–246 (2004).

[27] M. T. Pastore, S. J. Natale, W. A. Yost and M. F. Dorman, “Head movements allow listeners bilaterally implanted with cochlear implants to resolve front-back confusions,” *Ear Hear.*, **39**, 1224–1231 (2018).