Ewan A. Macpherson*

*Corresponding author’s address: National Centre for Audiology, Western University, 1201 Western Rd, London, N6G 1H1, ON, Canada, ewan.macpherson@nca.uwo.ca

Our studies have quantified the salience and weighting of dynamic acoustic cues in sound localization via head rotation. Results support three key findings: 1) low-frequency interaural time-difference (ITD) is the dominant dynamic binaural difference cue; 2) dynamic cues dominate front/rear localization only when spectral cues are unavailable; and 3) the temporal dynamics of dynamic cue processing are particular to auditory-vestibular integration. ITD dominance is shown indirectly in findings that head movements are highly effective for localizing low-frequency targets but not narrow-band high-frequency targets. Direct evidence comes from manipulation of dynamic binaural cues in spherical-head simulations lacking spectral cues. If the stimulus provides access to dominant high-frequency spectral cues, location illusions involving head-coupled source motion fail. For low-frequency targets, localization performance improves with increasing head-turn angle, but decreases with increasing velocity such that performance depends primarily on stimulus duration; ~100 ms being required for accurate front/back localization. That duration threshold only applies in dynamic localization tasks, and not in auditory-only tasks involving similar stimuli. Correct spatial interpretation of dynamic acoustic cues appears to require vestibular information about head motion, thus the temporal threshold is likely a property of vestibular-auditory integration.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

A listener can take an active role in sound localization by means of moving the head. In that case, dynamic information is provided by the relationship between the motion of the head and the resulting changes in the interaural-difference cues. In particular, for a given head rotation, the direction of change of interaural time-difference (ITD) and interaural level-difference (ILD) for a stationary source in the front hemisphere is opposite to that for a source in the rear. Previous studies have clearly demonstrated the benefits to sound localization of relatively large head movements for both free-field and virtual auditory space stimuli (e.g. Wallach, 1940; Perrett and Noble, 1997; Wightman and Kistler, 1999; Iwaya et al., 2003). Such studies have typically not used methods that permitted experimenter control over the amount of dynamic information available to the listener, and therefore have not addressed the limits of salience of the dynamic cues for small head movements that might be more typical of natural behavior. In our laboratory we have employed an experimental paradigm that does offer this level of control, and we review here the results of completed and ongoing studies that address the relative weighting of dynamic ITD and ILD cues, the relative influence of spectral cues and dynamic interaural cues, and the temporal dynamics of cue integration in sound localization via head rotation.

HEAD-SWEEP METHOD

The experiments described here all used variants of a common stimulus presentation paradigm we refer to as the head-sweep method, illustrated in Fig. 1. Blocks of trials were run with the head fixed (facing 0-deg azimuth) or the head in motion at a low (25 or 50 deg/s), intermediate (100 and 200 deg/s), or high (400 deg/s) velocities. Each head-motion trial began with the listener's head turned 45 degrees to one side. The listener then began a head rotation at a practised velocity. Head orientation was tracked continuously. When the orientation entered a selected spatial window (widths between 2.6 and 40 degrees), the stimulus was gated on with a 5-ms raised-cosine ramp. When the head's orientation exited the window, the stimulus was gated off. The listener continued the head rotation to 45 degrees on the other side. Following each head-fixed or head-moving stimulus presentation, the listener indicated the apparent direction of the stimulus by turning the body and orienting with the head. A final reading of head orientation constituted the listener's response. The locations of the sound sources were independent of the spatial window, and were typically arranged in 22.5-, 30-, or 45-degree intervals spanning 360 degrees around the horizontal plane. Stimuli were presented over loudspeakers in a darkened, anechoic room or in virtual auditory space using individually measured head-related transfer functions (HRTFs) and real-time, head-tracked HRTF filtering. Data analysis typically involved computing the proportion of “correct” or “small-error” responses, which were defined as those falling within 30 degrees of the true target location. In some cases we instead computed the proportion of correct front/rear hemisphere responses. Wider windows allowed for more onset-to-offset cue change and longer stimulus durations, and were therefore expected to lead to more accurate localization.

![Figure 1: Head-sweep presentation paradigm.](image_url)
WEIGHTING OF DYNAMIC ITD AND ILD IN LOCALIZATION VIA HEAD ROTATION

Free-field dynamic localization of wideband and low- and high-frequency narrowband noise targets

In a free-field experiment, we measured static and dynamic localization performance with bursts of wideband (0.5-16 kHz), or low- (0.5-1 kHz) or high- (6-6.5 kHz) frequency narrow-band noise. The narrow-band noises did not carry accurate spectral cues for front/rear location, and the listener therefore was forced to rely on dynamic ITD and ILD to localize accurately. Bursts presented in head-fixed conditions were 200 ms in duration.

Target-response plots for one typical listener (S203) in selected conditions (head-fixed and 50-deg/s motion) are shown in Fig. 2. Each column corresponds to one spatial window width. With the head fixed ("width" = 0, leftmost column), wideband stimuli were localized accurately, but all of this listener’s responses to low- and high-frequency noise fell in the front hemisphere, producing many back-to-front reversals. In head-motion conditions: responses to wideband noise were somewhat more scattered, but generally accurate; performance for low-frequency noise improved with increasing spatial window width; similar improvement was not observed for high-frequency noise. That latter result is remarkable; the fact that responses lie near either the positive or negative diagonal shows that the listener can certainly use the available interaural cues to report accurately the left/right component of the target location, but the listener could not use the head-motion-related change in those cues to disambiguate front from rear for the high-frequency targets.

![Figure 2: Selected target/response scatter plots for one typical subject. Stimuli were bursts of wideband, low-frequency narrowband, or high-frequency narrowband noise.](image)

Mean performance for four listeners is plotted in Fig. 3 as a function of spatial window width for each stimulus type and head-movement condition. For the wideband noise (left panel), increasing head velocity had a small negative effect on performance. For the low-frequency noise (middle panel), performance for a given spatial window decreased with increasing head velocity. For the low-frequency noise, performance improved monotonically with window width. Significant improvement was not observed for the high-frequency noise (right panel) except for the largest spatial window (40 degrees) and the lowest head velocity (50 deg/s). Because the primary available interaural cue for the low-frequency stimulus was ITD, and that for the high-frequency stimulus was ILD, these data suggest that, coupled with head-movement information, low-frequency dynamic ITD is a very salient cue for source location whereas high-frequency dynamic ILD is not.
Virtual auditory space dynamic localization with spherical-head HRTFs

A concern with the high-frequency narrow-band stimuli yielding the poor performance shown in Fig. 2 and Fig. 3 is that although they do not provide access to accurate spectral cues, they do in fact carry rather potent spectral cues of their own that could generate percepts of subject-dependent phantom locations (e.g. Middlebrooks, 1992; Morimoto and Aokata, 1984; Blauert, 1969/70). Simply replacing those stimuli with high-pass noise is problematic because such stimuli do provide access to accurate spectral cues and therefore do not force the listener to rely on dynamic cues.

In an attempt to overcome this problem, we are conducting ongoing studies using virtual auditory space presentation of stimuli generated with sphere-related transfer functions (SRTFs) lacking any front/rear spectral cues. These are computed using the method of Duda and Martens (1998) with the sphere radius selected to provide a close match between the low-frequency ITDs in each listener’s HRTFs and SRTFs. Figure 4 shows mean localization performance (four listeners) for wideband (0.5–16 kHz), low-frequency (0.5–1 kHz), and high-pass (4-16 kHz) noise targets as a function of spatial window width and transfer-function type (HRTF or SRTF). The head-sweep velocity was 50 deg/s.

As expected, static performance with HRTFs (open black circles) was good for the wideband and high-pass noises and poor for the low-pass noise. With SRTFs, static performance was poor for all target types due to the lack of front/rear spectral cues. While head-movement benefit was observed for each of the three target bandwidths with SRTFs, the benefit was lowest for the high-pass stimulus, for which only dynamic ILD was available for front/rear localization and highest for the low-pass stimulus, for which ITD would have been the available dynamic cue. The fact that the head movement benefit under SRTFs was less for the wideband stimulus than for the low-pass stimulus is perhaps related to poor externalization of the SRTF-filtered stimuli, but this remains to be assessed directly.
To further investigate the relative weighting of dynamic ITD and ILD cues, we are conducting a second study using SRTFs in which one or other of those binaural difference cues is “frozen” so that it does not change during head motion. For example, given a target at 30 degrees azimuth, the ITD is frozen by creating a new set of SRTFs in which each location has the normally computed spherical spectral cues (and thus normal spherical ILD), but in which the ITD for the 30-degree location was applied to every location in the set. Conversely the ILD can be frozen by applying the 30-degree left- and right-ear spectra to every location while allowing the ITD to vary normally. Figure 5 shows the mean asymptotic performance (data for 20- and 40-degree spatial window conditions pooled) for four listeners under normal and frozen-cue SRTFs with wideband targets. Paired-sample t-tests indicated that performance was significantly poorer when only dynamic ILD was available (ITD frozen) than when only dynamic ITD was available (ILD frozen), again suggesting that ITD is the primary dynamic localization cue. This parallels the finding that low-frequency ITD is the dominant static cue for source lateral angle (Macpherson and Middlebrooks, 2002; Wightman and Kistler, 1992).

**Figure 5:** Effect of ITD or ILD cue-freezing on dynamic localization accuracy with spherical-head HRTFs.

**Relative Influence of Spectral Cues and Dynamic Interaural Difference Cues for Front/Rear Location**

Since both static spectral cues and dynamic interaural difference cues can provide information about the front/rear location of a sound source, it is of interest to determine the factors influencing the relative weighting of these cues. Wallach (1939, 1940) devised a mechanical signal switching apparatus that allowed the experimenter to alter the location of the active loudspeaker in a circular array based on listener head movement. In particular, by causing the active source to rotate in the same direction as the head, but at twice the angular velocity, a dynamic stimulus was created that had interaural difference cues corresponding to a stationary source in the opposite hemisphere, as illustrated in Fig. 6. It was reported that listeners consistently perceived the illusory stationary source location, and on this basis Wallach derived the ‘principle of least displacement’, which states that spectral cues are subordinated to the auditory system’s preferred stationary-source interpretation of the dynamic interaural cues. Macpherson (2011) reported, however, that in a virtual auditory space recreation of Wallach’s arrangement, listeners readily perceived wideband noise targets veridically moving in the correct hemisphere.

To examine more closely the effect of high-frequency spectral cues on the success of the Wallach illusion, we generated sound stimuli that varied between low-pass (0.5-1 kHz) and wideband (0.5-16 kHz) with a variety of attenuation slopes above 1 kHz, as shown in Fig. 7 (left). Participants localized these stimuli under both static-head and head-sweep conditions, and in the latter, the sources executed the double-head-speed rotation intended to elicit the Wallach illusion over a head-sweep spatial window of 40 degrees.

To determine the quality of the spectral cues carried by each stimulus spectrum, the mean proportion of correct front/rear hemisphere responses was computed over five listeners for the static-head listening condition. Figure 7 (right, blue bars) shows that front/rear localization was quite accurate (85% correct or...
higher) for all the stimuli except the low-pass (less than 60% correct), which lacked high frequencies entirely. Thus all but the low-pass stimulus seemed to have available spectral cues for front/rear localization. The same front/rear performance computation for the moving stimuli indicates the rate of failure of the Wallach illusion, and results are shown in Figure 7 (right, red bars). Only for the low-pass stimulus was the failure rate of the illusion low, indicating dominance of the static interpretation of the dynamic cues. For the other spectra, although the details of the response patterns varied among the five listeners, the illusion was not robust, and for most listeners the conflict between the spectral cues and dynamic cues appeared to cause uncertainty about the source location. Similar localization behavior has been reported by Kawaura et al. (1991) under conditions of spectral and dynamic cue mismatch in virtual auditory space.

This result, that dynamic cues dominate completely only when spectral cues are weak or absent, is consistent with two recent findings. Brimijoin and Akeroyd (2012) found the Wallach illusion to occur robustly in a free-field re-creation only when stimuli lacked high-frequency energy. Similarly, Martens et al. (2013) report success of the “Phantom Walker Illusion” only for stimuli dominated by low frequency energy. The Phantom Walker is analogous to the Wallach illusion, and manifests as a front/rear reversal of source location elicited by swapping the left and right ear signals by means of an electronic hearing instrument while the listener experiences small involuntary head movements associated with walking.

**VESTIBULAR MEDIATION OF TEMPORAL DYNAMICS IN SOUND LOCALIZATION VIA HEAD ROTATION**

Faster head movements necessarily reduce stimulus duration for a given spatial window width. In Fig. 8, the mean performance data from Fig. 3 are replotted as a function of stimulus duration. The results for the
low-frequency noise (middle panel) are of primary interest. For a given duration, performance for the
low-frequency noise was very similar across head velocities despite spatial windows varying over a range of
almost an order of magnitude. The low-frequency results indicate that, regardless of head velocity, a stimulus
duration of approximately 100 ms is required for substantial head movement benefit (∼75% small errors).

![](image)

**Figure 8:** Effect of stimulus duration, head velocity and stimulus spectrum.

Initial results from an ongoing study have shown that active head movement is not required for correct
integration of dynamic interaural cues with motion of the head in space, and that performance in a passive
head movement condition is equally accurate (Kim and Macpherson, 2012). Figure 9 shows low-frequency
localization performance as function of stimulus duration for three listeners in the usual, head-sweep
condition (active, 25, 50, and 100 deg/s head velocity) and when being rotated passively in the same manner
on a swiveling chair. Stimuli were presented in virtual auditory space using individually measured HRTFs
and real-time head tracking and impulse response interpolation. The results indicate a ∼100-ms duration
requirement in both active and passive conditions, similar to that seen in Fig. 8. That suggests that neck
proprioception and efference copy are not necessary for this form of temporal integration, since information
from those modalities was greatly reduced in the passive rotation condition (Kim *et al.*, 2013).

![](image)

**Figure 9:** Localization performance under active or passive head-sweep conditions.

To determine whether these temporal dynamics are specific to vestibular/auditory integration, we
measured discrimination performance in an equivalent auditory-only task. For front/rear localization via
head rotation, a listener must determine whether the source moves left-to-right or right-to-left relative to the
head as the head rotates and also whether this is in the same direction as the head turn, indicating a
rear-hemisphere source, or in the opposite direction, indicating a front-hemisphere source. The auditory-only
task required listeners to discriminate the direction of motion of a 0.5–1-kHz band of noise presented over
headphones with monotonically increasing or decreasing ITD. The magnitude of the ITD change over the
duration of each stimulus was 25, 50, 100, 200, or 400 μs, and the rate of ITD change was ±250, ±500, ±1000,
or ±2000 μs/s. Since for an average-sized head, ITD naturally varies by ∼10 μs/deg across the midline, these
values approximate those that were produced by listener head rotation in head-sweep localization tasks. In
some blocks of trials, the ITD sweep was centered on 0 μs, and in others, the starting ITD was roved over a 500-μs range to prevent the start or end point being used as a cue for the direction of motion. 5-ms onset and offset ramps were applied to the left and right channels before applying the ITD.

Mean psychometric functions are shown in Fig. 10 for the same three listeners whose localization data are shown in Fig. 9. When performance is plotted as a function of ΔITD (top panels), the results show improved performance with greater cue change and a penalty for higher ITD velocity. ITD roving (right panels) somewhat decreased performance accuracy. When plotted as a function of stimulus duration, however, it is clear that the velocity penalty is too small to align the psychometric functions, and that, unlike in the active or passive head-sweep localization tasks, it is ΔITD rather than duration that is the better predictor of performance. Thus the duration-limited performance seen in passive or active dynamic localization seems to be specific to vestibular/auditory integration. Also suggestive of vestibular involvement is the similarity of the 100-ms integration window observed in dynamic localization to the 75-100 ms delay observed between the perception of head movement and sound onset (Raeder et al., 2012). Also, auditory-vestibular temporal binding windows are of similar duration and, as seen in our localization results, appear to be independent of rotational velocity (Chang et al., 2012).

**CONCLUSIONS**

The results presented lead to the conclusions that: low-frequency ITD is the primary dynamic interaural difference cue for front/rear location; dynamic interaural difference cues dominate localization judgments only when spectral cues are absent; and the ∼100-ms duration requirement for effective use of dynamic auditory cues is a property of vestibular-auditory integration and not of the auditory system alone.

**ACKNOWLEDGMENTS**

The author thanks Michael Barnett-Cowan for insights into vestibular psychophysics, David Grainger for technical support, and students and research assistants Alasdair Cumming, Sarah Gillespie, Devin Kerr, Janet Kim, Tran Nguyen, Robert Quelch, and Kristen Tonus for their contributions. This work was supported by funding from NSF (0717272, Perception, Action and Cognition Program), NSERC (Discovery Grant 386259), and Western University.
REFERENCES


