# Biomimetic direction of arrival estimation for resolving front-back confusions in hearing aids

**Alan W. Archer-Boyd**
*Institute of Communication Acoustics, Ruhr-Universität Bochum, Building ID 2/231, 44780 Bochum, Germany*
*alan.archer-boyd@rub.de*

**William M. Whitmer and W. Owen Brimijoin**
*Medical Research Council/Chief Scientist Office Institute of Hearing Research–Scottish Section, New Lister Building, Glasgow Royal Infirmary, 16 Alexandra Parade, Glasgow G31 2ER, United Kingdom*
*bill@ihr.gla.ac.uk, owen@ihr.gla.ac.uk*

**John J. Soraghan**
*Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G11XQ, United Kingdom*
*j.soraghan@strath.ac.uk*

**Abstract:** Sound sources at the same angle in front or behind a two-microphone array (e.g., bilateral hearing aids) produce the same time delay and two estimates for the direction of arrival: A front-back confusion. The auditory system can resolve this issue using head movements. To resolve front-back confusion for hearing-aid algorithms, head movement was measured using an inertial sensor. Successive time-delay estimates between the microphones are shifted clockwise and counterclockwise by the head movement between estimates and aggregated in two histograms. The histogram with the largest peak after multiple estimates predicted the correct hemifield for the source, eliminating the front-back confusions.


[QJF]
**Date Received:** January 16, 2015 **Date Accepted:** March 26, 2015

## 1. Introduction

Many hearing-aid algorithms require reliable direction-of-arrival (DOA) estimates in order to enhance or suppress sound sources. Determining DOAs using a bilateral microphone array comprising a microphone at each ear faces a particular problem: Sound sources at the same angle in front or behind the array will produce the same DOA, resulting in a front-back confusion. Human listeners take into account head movements in judging sound-source location,[1] and further, use these movements to resolve front-back confusions.[2] Here, we present a proof-of-concept for a novel bilaterally communicating microphone system that mimics human behavior by measuring head movements to computationally resolve front-back confusions. We previously used bilateral microphone input, generalized cross-correlation (GCC), and head-movement information from micro-electromechanical systems (MEMS) to provide robust front-hemifield accuracy during head movement.[3] The new system extends this technique using head-movement-corrected DOA histograms to allow much more reliable sound-source localizations across hemifields, increasing the potential for signal enhancement. By using MEMS to disambiguate DOA estimates, it also avoids the need for a larger microphone array or more complex localization techniques such as comparing the microphone signals to a library of head-related transfer functions.[4]

The development and miniaturization of MEMS such as accelerometers and gyroscopes in recent years has made positional information available for use in signal/noise-localization algorithms in hearing aids. These low-power devices sense acceleration due to

 

gravity (accelerometer) and angular velocity (gyroscope) in three-dimensional space and so can provide information about a hearing-aid user's head movements. The incorporation of MEMS devices has previously been used to select hearing-aid programs by classifying a limited number of listening situations, using features extracted from long-term recordings of eye, head, and body movements in addition to acoustic information.[5] Besides our previous study limited to the front hemifield, there has been no application of MEMS to DOA estimations for hearing-aid algorithms.

### *1.1 Time-delay estimates for direction of arrival detection*

The time-delay estimate (TDE) between the signals arriving at two spatially separated microphones (e.g., bilaterally connected hearing aids) can be used to estimate the DOA of a sound source for hearing-aid algorithms. The GCC algorithm[6] for TDE is commonly used in many audio applications (e.g., using multiple microphones to record live sound)[7] and has previously been investigated for hearing aids on stationary heads.[8] The simplest configuration of a single active source in an anechoic space recorded by two spatially separated microphones can be described by

$$x_1[n] = s[n - \tau_1], \tag{1}$$

$$x_2[n] = s[n - \tau_2], \tag{2}$$

where $x_1[n]$ and $x_2[n]$ are the microphone signals, $s[n]$ is the source signal, and $n$ is the sample time step. The time delay of the source signal between the microphones ($\tau_s$) is the difference between $\tau_2$ and $\tau_1$.

The estimation of TDE using GCC is a frequency-domain technique for calculating $\tau_s$, defined as

$$\Psi_{\mathrm{GCC}}[n] = F^{-1}\{X_1^*[k] \cdot X_2[k]\}, \tag{3}$$

where $F^{-1}$ is the inverse fast Fourier transform, $X_1[k]$ and $X_2[k]$, $k = 0, \ldots, N-1$ are the frequency domain representations of the microphone signals $x_1[n]$ and $x_2[n]$, respectively, * is the complex conjugate and $N$ is the analysis window size in samples. $\tau_s$ is equal to the maximum peak in the GCC function ($\Psi_{\mathrm{GCC}}[n]$). The GCC can be made more robust to noise and reverberation by applying a phase transform (PHAT), setting all frequency magnitudes equal to 1,

$$\psi_{\mathrm{GCC\text{–}PHAT}}[n] = F^{-1}\left\{\frac{X_1^*[k] \cdot X_2[k]}{|X_1^*[k] \cdot X_2[k]|}\right\}. \tag{4}$$

For DOA estimation, each estimate of $\tau_s$ is converted to its equivalent angle. This can be done using the Woodworth model,[9,10] or as in this case, using previous empirical measurements of $\tau_s$ at known angles using a dummy head.[3] These estimates are then aggregated in a DOA histogram.

### *1.2 Gyroscopically compensated DOA (GC-DOA) estimation*

Techniques such as GCC-PHAT are shown to have degraded short-term performance in the presence of multiple signals and noise.[11] Aggregating DOA estimates over time and analyzing the resulting peaks of the aggregate histogram can produce more robust results and increase the number of sources that can be located. If the head moves during the aggregation period, however, estimates will be spread across the angle of head movement.

With the addition of a head-mounted gyroscope, however, it is possible to shift and update the aggregate histogram built up during each measurement frame in order to compensate for any head movement.[3] In this method, the histogram of previous DOA estimates were shifted by the difference in head angle between the previous and current estimate. This compensation produces a peak in the DOA histogram at the end of a measurement frame, corresponding to the position of the source relative to the current position of the head. The resulting gyroscopically compensated DOA

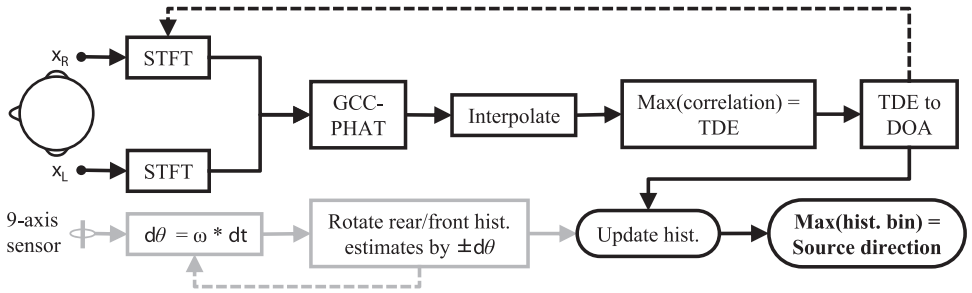


Fig. 1. Diagram of the B-DOA system. $\omega$ is the angular velocity of the head and $dt$ is measurement time of the analysis window (40 ms). For a full description, see text.

(GC-DOA) implementation was described in more detail in our previous paper.[3] This technique was found to be successful for robust, *in situ* DOA estimates of up to four speech sources but could not distinguish between front vs back sources.

## 2. Implementation of biomimetic direction of arrival (B-DOA) estimation

Figure 1 shows a diagram of the biomimetic direction of arrival (B-DOA) system. The system uses two microphones as inputs that are mounted on either side of the head and a head-mounted nine-axis sensor: A combined three-axis accelerometer, three-axis gyroscope, and three-axis magnetometer. This nine-axis sensor improved head-tracking accuracy in comparison to a gyroscope alone, as using a calibrated magnetometer greatly reduces or eliminates measurement drift over time.[12] Audio inputs $x_L$ and $x_R$ (left and right microphones, respectively; see top of Fig. 1) were recorded at 44.1 kHz, which was then downsampled to 16 kHz to approximate the sample rate of a modern hearing aid. A Hann window was applied to each 40 ms $x_L$ and $x_R$ segment. No overlap between windowed segments was used so each sample from the nine-axis sensor was synchronized with a single segment of audio. The short-time Fourier transforms (STFT) were taken of each audio segment, resulting in a DOA analysis rate of $25\,s^{-1}$. The GCC-PHAT algorithm was then applied to each pair of audio segments. Assuming the size of the head to be approximately 16 cm in diameter, and the sampling frequency to be 16 kHz, the maximum achievable delay (i.e., output of the GCC-PHAT) at 90° (i.e., the maximum distance between the microphones) was 10 samples, resulting in a range of ±10 samples and a resolution of 9°. The output of the GCC-PHAT was interpolated by a factor of 8:1 using a polyphase filter (the *resample* command in MATLAB). Experimental results showed that angles between 80° and 90° resulted in no change in time-delay estimation (TDE) after interpolation. This interpolation produced a resolution of 1°/sample from 0° to 80°. After interpolation, the delay of the largest correlation peak in the interpolated IFFT was selected as the estimate of time delay ($\tau_s$) between the microphones, and then converted to angle in degrees to produce a DOA estimate. This estimate was then placed in the corresponding histogram bins of two histograms and the process repeated for the next analysis window.

During each analysis window, the rotational velocity information measured by the head-mounted nine-axis sensor was used to determine the angle through which the head had rotated since the previous analysis window (see bottom of Fig. 1). Two histograms of DOA estimates were created: One was shifted counterclockwise by the measured head-rotation angle ($+d\theta$), being the correct shift for a source in the front hemifield (front histogram), while the other was shifted clockwise by the measured head-rotation angle ($-d\theta$), being the correct shift for a source in the rear hemifield (rear histogram). The current DOA estimate was added to the histogram unchanged. One measurement frame consisted of a 1-s aggregate of 25 analysis windows, in order for the head-movement compensation to have a measurable effect on the histograms. Each measurement frame was discrete from the next. Shifting the histogram in the correct direction (counterclockwise for the front and clockwise for the rear hemifield) to compensate for head movement

produced a strong peak at one DOA in the measurement frame, whereas shifting the histogram in the wrong direction produced no peak during continuous head movement and a widened histogram distribution. By shifting the histogram in both directions, the histogram with the largest peak at the end of a measurement frame corresponded to the strongest source active during the measurement frame and the correct hemifield selection.

## 3. Experimental tests

In the experimental testing of the proposed method a single loudspeaker (JBL Control 1 Pro, JBL, Northridge, CA) was placed 1.5 m directly in front (0°) or behind (180°) the participant at a height of 1.2 m. The participant's head moved through approximately 60°, from −30° to +30°. The signal was composed of concatenated, same male-talker sentences from the IEEE York corpus[13] and presented at 65 dB in a 6.5 m × 5 m × 3 m room with an $RT_{30}$ of 0.35 s.

Two in-ear microphones (Sound Professionals MS-TFB-2, The Sound Professionals, Hainesport, NJ) were placed on top of the participant's pinnae to simulate the position of behind-the-ear hearing-aid microphones. A Zoom H4n was used as the microphone preamp and analog-to-digital converter. A nine-axis sensor (Sparkfun SEN-10724, SparkFun Electronics, Boulder, CO) was calibrated using the RAZOR AHRS software (P. Bartz, Berlin, Germany).[12] Head-position data were collected from it using an Arduino Uno microcontroller (Arduino, Torino, Italy). Both audio and motion data were recorded by the same computer. To maintain synchronization of the audio and motion recordings, the most recent reading in the motion buffer of the nine-axis sensor was recorded at the end of each analysis window (every 40 ms). The nine-axis sensor ran at a sampling rate of 50 Hz to maintain a reading in the motion buffer at all times and resulted in a maximum asynchrony of 20 ms between the audio and motion data. The maximum head-turn velocity during recording was 36.6° $s^{-1}$, producing a maximum variability between the measured and actual head position of 0.7°. This would be sufficient to shift DOA estimates into a histogram bin adjacent to the correct one.

The nine-axis sensor required several seconds at start-up to obtain enough information to calculate its initial position, therefore the first 5 s of each recording were discarded. Recordings were made for 15 s, with the first 5 s discarded to allow for sensor self-calibration, resulting in 10 s of data (i.e., ten measurement frames) from each recording. After start-up, the system requires only 1 s of recording (one frame) to determine the position of a source in the front or rear hemifield. The participant was instructed to make smooth and steady head movements back and forth between two markers placed on a facing wall at ±30° (0° being straight ahead) throughout each recording.

## 4. Results

Figure 2 shows a 10-s example of a recording with the head moving between −32° and 35°. The source is in front of the participant (0°). The head movement as measured by the head tracker is shown in the top row. The second row shows the DOA histograms with no shift applied to them. The third row shows the clockwise-shifted DOA histograms for ten 1-s measurement frames (25 analysis windows in each frame). The fourth row shows the counterclockwise-shifted DOA histograms for the same measurement frames. The shift (clockwise or counterclockwise) that produces the largest peak in each measurement frame is used as a prediction of the source hemifield: Larger peaks in the counterclockwise-shifted histogram predict a source in the front hemifield, and larger peaks in the clockwise-shifted histogram predict a source in the rear hemifield. Large peaks in the histograms with no shift represent periods of little or no head movement. Though these may be similar to one of the shifted histograms, they provide no information on the source hemifield. In every measurement frame in Fig. 2, the counterclockwise-shifted histogram (fourth row) produced a higher peak than the corresponding clockwise-shifted histogram. Therefore in each measurement frame, the difference in the histograms correctly placed the signal in the front hemifield. Table 1 shows the peak angles in the counterclockwise-shifted (CCW) histograms and the angle obtained from the head tracker at the end of each measurement frame. The absolute localization error was the absolute difference between the two values and

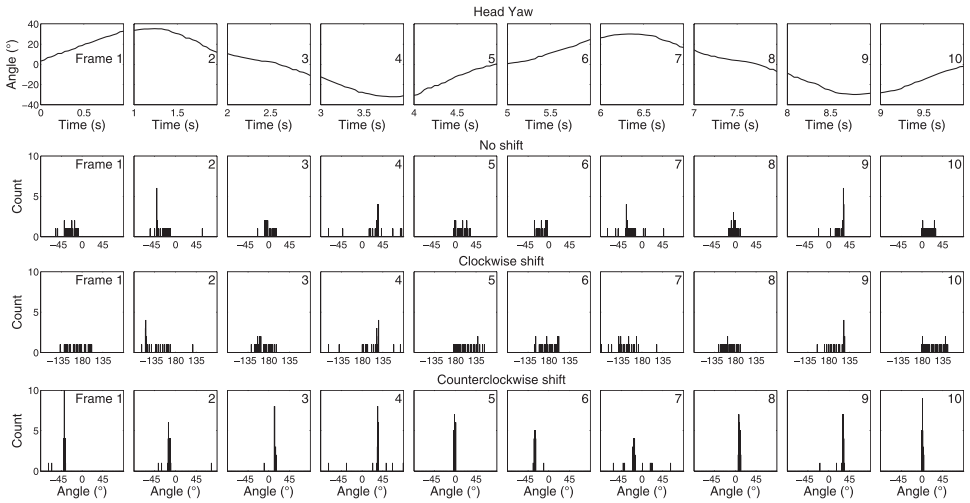


Fig. 2. Tracking results for a source in the front hemifield (0°). Head yaw during each measurement frame is shown in the top row. Second row shows the histograms with no shift applied. Third and fourth rows show the clockwise-shifted and counterclockwise-shifted histograms, respectively, of DOA estimates for each measurement frame of 1 s duration.

gives a measure of the source localization accuracy. The absolute localization error varied between 0.7° and 6.2°.

Figure 3 shows a 10-s example of a recording with the head moving between −34° and 35°. The source is at the rear of the participant (180°). The head movement and shifted DOA histograms are shown as in Fig. 2. The selection rules are applied as in Fig. 2. In the histograms with no shift, large peaks represent periods of little or no head movement. In every measurement frame in Fig. 3, the clockwise-shifted histogram produces a higher peak than the corresponding counterclockwise-shifted histogram. Therefore in each measurement frame, the source was correctly placed in the rear hemifield. Table 1 shows the peak angles in the clockwise-shifted (CW) histograms and the angle obtained from the head tracker at the end of each measurement frame. The absolute localization error varies between 0.6° and 11.2°, a larger range than the results for the previous front-hemifield example.

## 5. Discussion

The results show that a sound source can be robustly identified to be in the front or rear hemifield relative to a participant by combining a bilateral array comprising one

Table 1. Head angle at the end of each frame, angle of histogram peak for counterclockwise-shifted (CCW) histograms (source at 0°, front hemifield) and clockwise-shifted (CW) histograms (source at 180°, rear hemifield) and absolute localization error (abs. loc. err. = |head angle – angle of histogram peak|).

| | Frame | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Source at 0° (front) | | | | | | | | | | |
| Head angle (°) | 32.8 | 11.9 | −11.1 | −31.1 | 0.4 | 24.5 | 16.7 | −7.2 | −28.7 | −2.0 |
| CCW hist. peak (°) | −39.0 | −15.0 | 13.0 | 34.0 | −2.0 | −29.5 | −16.0 | 8.0 | 32.0 | 1.0 |
| Abs. loc. err. (°) | 6.2 | 3.1 | 1.9 | 2.9 | 1.6 | 5.0 | 0.7 | 0.8 | 3.3 | 1.0 |
| Source at 180° (rear) | | | | | | | | | | |
| Head angle (°) | 14.5 | −15.0 | −32.8 | −0.3 | 31.6 | −1.3 | −31.4 | −9.0 | 27.6 | 12.5 |
| CW hist. peak (°) | 9.0 | −22.0 | −44.0 | −6.0 | 31.0 | −8.0 | −42.0 | −18.0 | 27.0 | 10.0 |
| Abs. loc. err. (°) | 5.5 | 8.8 | 11.2 | 5.7 | 0.6 | 6.7 | 10.6 | 9.0 | 0.6 | 2.5 |

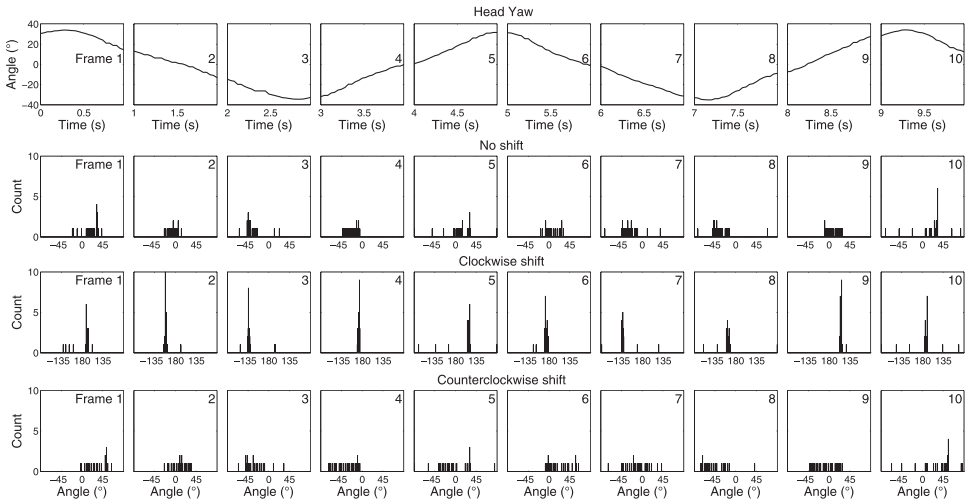


Fig. 3. Same as Fig. 2 except for a source in the rear hemifield (180°).

microphone at each ear, the head movements of the participant, and selecting the (shifted) histogram with the largest peak. Showing both clockwise and counterclockwise rotations of the head demonstrates that the system can determine the position of a source for both rotational directions without additional *a priori* knowledge.

The counterclockwise-shifted histogram peaks for a source in the front hemifield were generally higher and narrower than the clockwise-shifted histogram peaks for a source in the rear hemifield. Absolute localization errors at larger head angles are smaller for counterclockwise-shifted peaks than for the clockwise-shifted peaks. For example, counterclockwise-shifted frame four reports a head angle of −31.1° and an absolute localization error of 2.9°, whereas clockwise-shifted frame reports a similar head angle of 31.4° and an absolute localization error of 10.6°. Therefore, the counterclockwise-shifted peaks are seen to be more accurate. The relationship between time-delay and direction of arrival was derived by empirical measurement for the front hemifield only.[3] This relationship may not be the same in the rear hemifield, resulting in the reduction in accuracy at larger angles. Further work is required to determine the source of this inaccuracy, perhaps by extending the empirical model or applying an improved Woodworth model that takes account of variations such as ear position and source distance.[10] The maximum temporal offset between the nine-axis sensor and recorded audio could also be reduced by increasing the sample rate of the sensor, which would reduce the number of incorrectly shifted estimates and increase the height of the histogram peak. Increasing the height of the histogram peak would make the technique more robust to the increased number of inaccurate estimates that can be caused by diffuse noise and longer reverberation times. In addition, the system is only able to determine the correct hemifield of a source if the head moves during a measurement frame. The minimum angle of head movement required during a frame for robust hemifield detection has not yet been investigated.

The speed and extent of the head movements performed during the recording are within the bounds that could be expected during a conversation. The constant oscillatory movement is not natural localization behavior, but serves as a proof of concept. Frame by frame, it can be seen that though the starting point, change in angle and thus the speed of head movement changes, the system correctly predicts the position of the source in the front or rear hemifield. In the real world, the minimum required angle would depend upon the sample rate and accuracy of the nine-axis sensor, and the audio sample rate and upsampling method used for the GCC-PHAT.

This approach may work if there were a number of sources in the same hemifield. That question was partially explored in a previous paper.[3] In that paper, up to four active sources were accurately localized during a head movement using a similar system. The previous system had two major differences to the current one. First, only a counterclockwise histogram was produced previously, as the sources were assumed to be in the front hemifield. Second, the length of measurement frame was previously four seconds (100 estimates), making the update speed of the previous system four times slower than the current system. While we see no theoretical reason for the system to fail with multiple sources, this has not yet been explicitly tested.

## 6. Conclusion

A biomimetic system for resolving front-back confusions in DOA estimation was designed and tested. The system utilized head motion to differentially rotate GCC-PHAT estimates of DOA against head motion over time. Using simple peak size comparisons between congruent histograms, it was shown that the correct hemifield was selected for a single source. This system extends the robust measurement space of a two-microphone, time-delay estimation technique from 180° to 360° without the need for a larger array or more complex localization techniques such as comparing the microphone signals to a library of head-related transfer functions.[4]

## Acknowledgments

The Scottish Section of the IHR was supported by intramural funding from the Medical Research Council (Grant No. U135097131) and the Chief Scientist Office of the Scottish Government.

## References and links

[1]H. Wallach, "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol. **27**(4), 339–368 (1940).
[2]W. O. Brimijoin and M. A. Akeroyd, "The role of head movements and signal spectrum in an auditory front/back illusion," Iperception **3**(3), 179–182 (2012).
[3]A. W. Boyd, W. M. Whitmer, W. O. Brimijoin, and M. A. Akeroyd, "Improved estimation of direction of arrival of sound sources for hearing aids using gyroscopic information," Proc. Meet. Acoust. **19**, 030046 (2013).
[4]T. Usagawa, A. Saho, K. Imamura, and Y. Chisaki, "A solution of front-back confusion within binaural processing by an estimation method of sound source direction on sagittal coordinate," in *IEEE Region 10 Conference* (November 21–24, 2011), pp. 1–4.
[5]B. Tessendorf, M. Debevc, P. Derleth, M. Feilner, F. Gravenhorst, D. Roggen, T. Stiefmeier, and G. Tröster, "Design of a multimodal hearing system," Comp. Sci. Info. Sys. J. **10**(1), 483–501 (2013).
[6]C. H. Knapp and G. C. Carter, "Generalized correlation method for estimation of time delay," IEEE Trans. Acoust. Speech Signal Process. **24**, 320–327 (1976).
[7]A. Clifford and J. Reiss, "Calculating time delays of multiple active sources in live sound," in *Audio Engineering Society 129th Convention* (November 4, 2010), Paper 8157.
[8]T. Rohdenburg, S. Goetze, V. Hohmann, K.-D. Kammeyer, and B. Kollmeier, "Combined source tracking and noise reduction for application in hearing aids," in *Proceedings of the 8th ITG-Fachtagung Sprachkommunikation*, Aachen, Germany (2008).
[9]R. S. Woodworth, *Experimental Psychology* (Holt, New York, 1938), pp. 520–523.
[10]N. L. Aaronson and W. M. Hartmann, "Testing, correcting, and extending the Woodworth model for interaural time difference," J. Acoust. Soc. Am. **135**(2), 817–823 (2014).
[11]W. R. Aichner, H. Buchner, S. Wehr, and W. Kellermann, "Robustness of acoustic multiple-source localization in adverse environments," in *Proceedings of the 7th ITG-Fachtagung Sprachkommunikation*, Kiel, Germany (2006).
[12]P. Bartz, "Razor attitude and head rotation sensor," Quality and Usability Lab, TU-Berlin. https://github.com/ptrbrtz/razor-9dof-ahrs (Last viewed April 13, 2015).
[13]P. C. Stacey and A. Q. Summerfield, "Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech," J. Acoust. Soc. Am. **121**(5), 2923–2935 (2007).