Toronto, Canada
International Symposium on Room Acoustics
2013 June 9-11

Recording and analysis of head movements, interaural level and time differences in rooms and real-world listening scenarios

Alan W. Boyd (alan@strath.ac.uk)
Centre for excellence in Signal and Image Processing (CeSIP)
Department of Electronic and Electrical Engineering
University of Strathclyde
204 George Street
Glasgow
G1 1XW

William M. Whitmer (bill@ihr.gla.ac.uk)
Michael A. Akeroyd (maa@ihr.gla.ac.uk)
MRC Institute of Hearing Research Scottish Section
Glasgow Royal Infirmary
16 Alexandra Parade
Glasgow
United Kingdom
G31 2ER

ABSTRACT

The science of how we use interaural differences to localise sounds has been studied for over a century and in many ways is well understood. But in many of these psychophysical experiments listeners are required to keep their head still, as head movements cause changes in interaural level and time differences (ILD and ITD respectively). But a fixed head is unrealistic. Here we report an analysis of the actual ILDs and ITDs that occur as people naturally move and relate them to gyroscope measurements of the actual motion.

We used recordings of binaural signals in a number of rooms and listening scenarios (home, office, busy street etc). The listener's head movements were also recorded in synchrony with the audio, using a micro-electromechanical gyroscope. We calculated the instantaneous ILD and ITDs and analysed them over time and frequency, comparing them with measurements of head movements.

The results showed that instantaneous ITDs were widely distributed across time and frequency in some multi-source environments while ILDs were less widely distributed. The type of listening environment affected head motion. These findings suggest a complex interaction between interaural cues, egocentric head movement and the identification of sound sources in real-world listening situations.

1 INTRODUCTION

In many real-world situations, the auditory environment we experience can change in a number of ways. There may be multiple sound sources active at the same time, resulting in multiple

auditory cues.[1] The sources may have their own motion relative to the listener, causing changes in the interaural level and time differences (ILDs and ITDs respectively) experienced by the listener, in addition to subtle changes in the filtering of sound due to the head-related transfer function, HRTF.[2] The movements of the listener, in particular their head movements, produce similar changes in the auditory cues. However, these do not result in a perception of source movement, due to vestibular feedback to the auditory system.[2] The majority of auditory localization research has used controlled, artificial stimuli, often with the participant's head fixed in place, or using headphone presentation. However, head movements during laboratory localization tasks have been recorded,[3] in addition to head movements while evaluating other attributes of a sound[4] and the ILD and ITD cues available to listeners in rooms.[5] Motion strategies for binaural localization have been modelled for artificial listeners[6] and head motion has been analysed and synthesised for talkers.[7] Head movements have also been considered for inclusion in auditory displays[8] and improved localization when using them.[9]

The current study used a synchronised binaural signal and head-motion recording system to capture the binaural audio signals that listeners are exposed to and their head movements while listening and moving freely in a number of real-world situations. A comparison made between the most prominent and reliable ITD and ILD cues and head motion to ascertain whether head motion could improve source detection and tracking.

2 RECORDING SYSTEM AND ANALYSIS METHODS

The recording system combined binaural in-ear microphones with sensors to determine head orientation. In-ear microphones (Sound Professionals MS TFB-2) were fixated just below the concha of the listener. Signals were recorded at 16-bit, 48-kHz sampling rate. The head movements were recorded using a micro-electromechanical system (MEMS) comprising an accelerometer, gyroscope and magnetometer. The MEMS was connected to an Arduino Uno controller running the Razor attitude and head rotation sensor (AHRS) firmware developed by Peter Bartz at TU-Berlin.[10] Head movements were recorded at a sampling rate of 50 Hz. Synchronization of the two independent systems (audio/MEMS) was achieved by recording one 960-sample audio frame for the most recent MEMS sample. Synchronization was verified by attaching the binaural microphones to the MEMS device and tapping the combined unit. The systems were synchronised within one MEMS sample (i.e., 0.02 s).

The audio recordings were analysed using time windows of 20 and 100 ms with no overlap. The ITDs were calculated using fourth-order Butterworth bandpass filtered audio from 500 to 1600Hz. The lower cut-off eliminated low-frequency building noise; the higher cut-off approximates the upper limit of useful ITDs. The ITD for each time window (τ_s) was calculated as the peak in the generalized cross-correlation (GCC):

$$\psi_{GCC}[n] = F^{-1} X_R^*[k] \cdot X_L[k] \tag{1}$$

$$\tau_s = \arg\max_n \psi_{GCC}[n] \tag{2}$$

for frequencies k = [0, ... , N-1], where N is the analysis window size, F^{-1} is the inverse Fourier Transform, X_L and X_R are the frequency domain representation of the left and right microphone signals respectively and * is the complex conjugate. By oversampling each window by a factor of 2, the ITD resolution was 10.4 µs.

The ILDs were calculated using fourth-order Butterworth bandpass-filtered audio from 2 to 4 kHz. A small frequency range was chosen to avoid ILD frequency dependence. The difference in power was calculated between the left and right audio signals in each frame and converted to decibels.

The head movement was calculated using the output from the accelerometer, gyroscope and magnetometer. This information was combined using the direction cosine matrix method.[11] The AHRS measurement system was placed on the head to read 0° pitch and rotation for the listener (first author) in a relaxed standing position. Yaw was recorded relative to magnetic North and converted so that 0° was equivalent to on-axis (mid-saggital plane), as determined by the magnetometer calibration.

3 RECORDINGS

Recording environments were chosen to cover a wide range of everyday situations and listening scenarios. Each recording (excluding the initial test) was 10 minutes long.

Initial tests: ITD and ILD analysis was performed using two male, continuous speech signals from the IEE York corpus in an acoustically dampened room without moving the listener's head.

One-to-one conversation: One male speaker was recorded by a naturally-interacting listener, both seated in an office measuring approximately 2.5 x 2.5 x 2.7 m. Reverberation time (RT_{60}) was estimated to be < 0.5 s. The talker was 1 metre from the listener.

Two-to-one conversation: Two male speakers were recorded by a naturally-interacting listener, all seated in a small laboratory measuring approximately 4 x 3 x 2.7 m, with an adjoining area measuring 4 x 2 x 2.7 m. RT_{60} was estimated to be < 1 s. The talkers were 2 m from the listener.

Watching television: The audio output of a television in a living room was recorded from 2.5 m away by a seated, naturally-interacting listener. The room measured approximately 5 x 3.5 x 3 m. RT_{60} was estimated to be ~1 s. The television was 2.5 m from the listener

Hospital foyer: A large, busy hospital foyer was recorded by a listener walking through it. The foyer consisted of a large rectangular space, approximately 15 x 10 x 4 m at its largest, narrowing to a long hallway approximately 4 m wide at one end, with multiple perpendicular hallways attached to this hallway.

4 RESULTS

The results of recording 10 seconds of two continuous talkers keeping the head still are shown in Figure 1. The ITD histogram shows clear peaks at 310 and -360 μs. The ILD histogram shows peaks at -9 and 4 dB (ILD σ = 5.96 dB). Both ITDs and ILDs vary several times a second in the time domain.

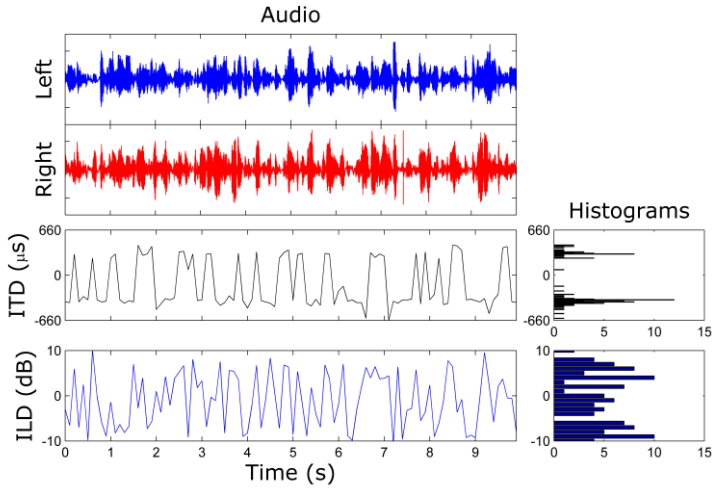

Figure 1: Analysis of 10 seconds of 2 continuous talkers at ±45° azimuth, using a 100 ms window size. The top two plots display the in-ear microphone recordings at each ear on a linear scale. The bottom two plots display ITD and ILD, respectively, as a function of time. The histograms to the right display the frequency of a given ITD (top right panel) or ILD (bottom right) over the plotted measurement period.

The analysis for 30 seconds of a live talker and listener using 20-ms time windows is shown in Figure 2. The ITD histogram peaks at -180 μs and the ILD at -1 dB, with a standard deviation (σ) in the ILDs of 2.7 dB. The area marked by the black rectangle highlights a rapid 55° head movement and the corresponding change in ILD. The change in the ITD due to this head movement cannot be observed.

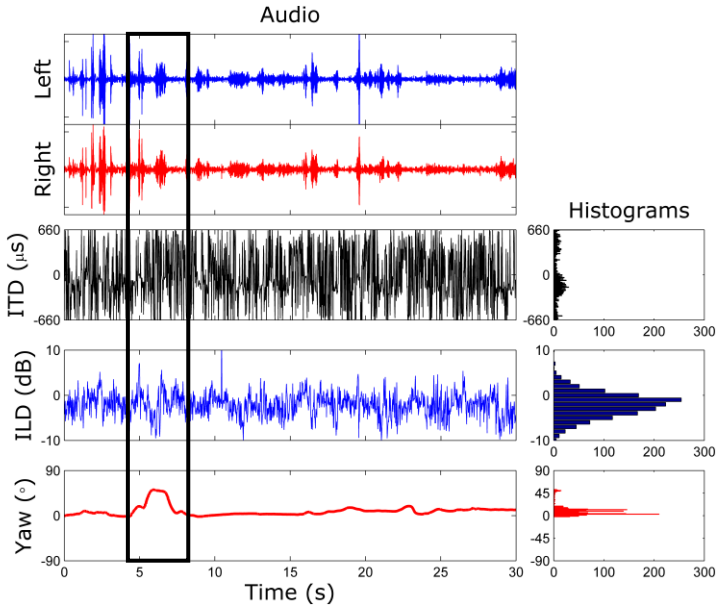

Figure 2: Analysis of 30 seconds of one live talker, seated -15° **(**listener's left), using a 20-ms window duration. The top two plots display the recordings at each ear on a linear scale. The lower plots display ITD, ILD and yaw (head orientation angle), respectively, as a function of time. The histograms to the right display the frequency of a given ITD, ILD or yaw, respectively, over the plotted measurement period. The black rectangle shows audio/MEMS output interaction.

An analysis of the same recording using 100-ms time windows is shown in Figure 3. The ITD histogram peak is -220 μs and for ILD is -1 dB (ILD σ = 2.3). The ITD variation is reduced and it is possible to observe a change in ITD due to head movement in the time domain, in addition to the ILD change. This finding indicates that 100 ms is a better time window than 20 ms to determine sources based on either ITDs or ILDs for this particular analysis. 100-ms time windows are used for the remaining analyses.

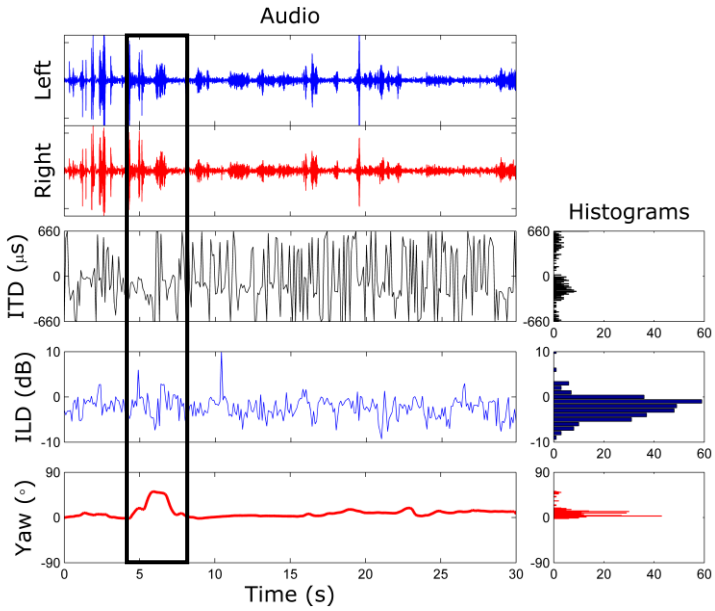

Figure 3: Same as Figure 2 but using a 100-ms window duration. The histograms (right panels) appear to show little change with window duration. In the time domain (left panels), a longer window length reduces the noise (fluctuations) in the measurements.

The analysis of a two talker conversation and listener scenario is shown in Figure 4. The ITD histogram shows peaks at 140, -140 and -480 μs. The ILD histogram peaks at 2 dB (ILD σ = 2.2 dB). The first rectangle highlights a shift of attention from one talker to the next, with an initial shift in ITD and ILDs, followed by a re-orientation of the head and a corresponding shift of the ITD and ILDs due to this head movement. The second rectangle highlights a simple head movement and corresponding shift in ITDs and ILDs.

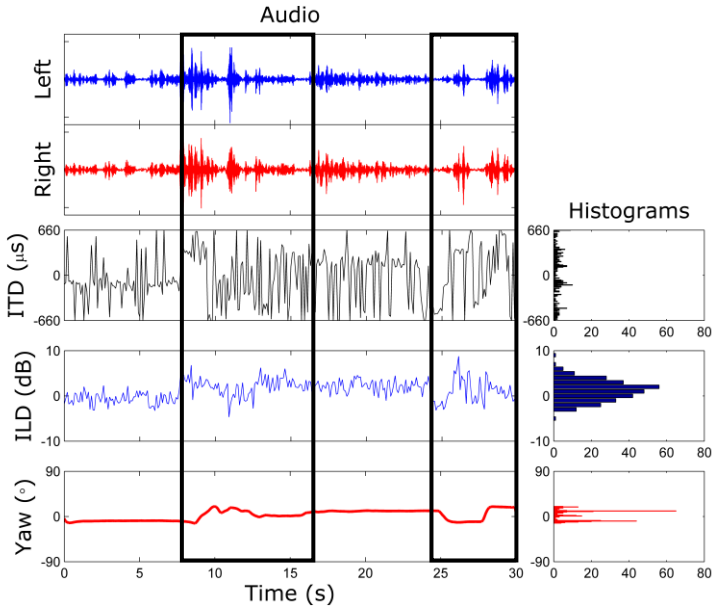

Figure 4: Analysis of 30 seconds of two talkers seated at ±25° azimuth. The top two plots display the recordings at each ear on a linear scale. The lower plots display ITD, ILD and yaw, respectively, as a function of time. The histograms to the right display the frequency of a given ITD, ILD or yaw, respectively, over the plotted period. The black rectangles show audio/MEMS output interactions.

The analysis of a listener watching television in a living room is shown in Figure 5. The ITD histogram displays a peak at 80 μs and the ILD histogram peaks at 3 dB (ILD σ = 1.33 dB).

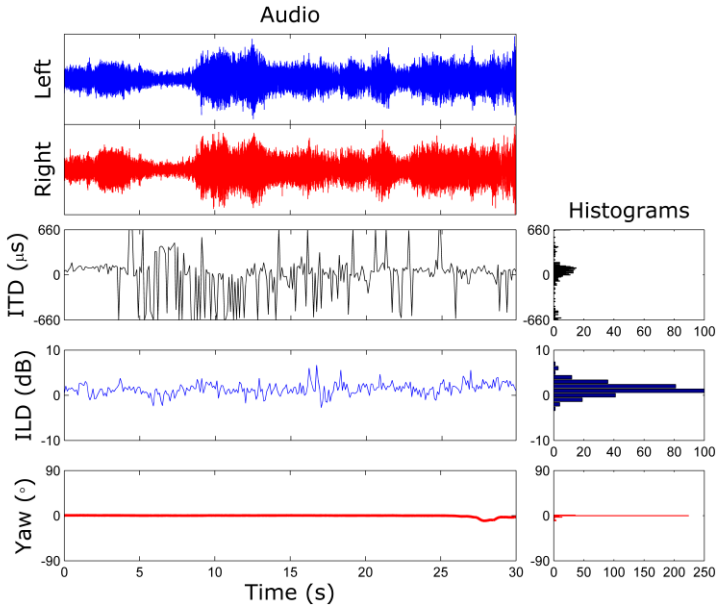

Figure 5: Analysis of 30 seconds of a seated listener watching a television + 5° (listener's right). The top two plots display the recordings at each ear on a linear scale. The lower plots display ITD, ILD and yaw, respectively, as a function of time. The histograms to the right display the frequency of a given ITD, ILD or yaw, respectively, over the plotted period.

The analysis of the hospital foyer recording is shown in Figure 6. The ITD histogram peaks at 160 and -30 μs, though these peaks are less well defined in relation to previous scenarios due to the level of background noise. The ILD histogram peaks at 1 dB (ILD σ = 1.33 dB). The rectangle highlights a change in ILD that was not caused by head movement.

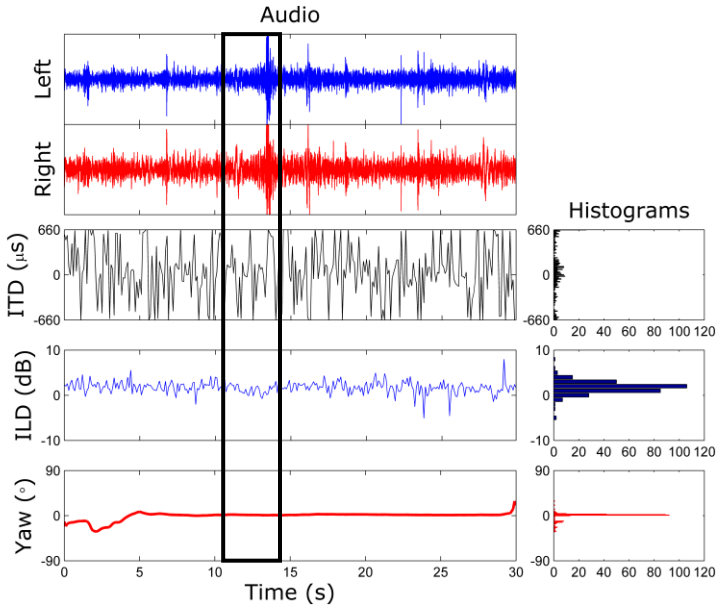

Figure 6: Analysis of 30 seconds of a listener walking through a busy hospital foyer. The top two plots display the in-ear microphone recordings at each ear on a linear scale. Each histogram displays the frequency of a given ITD (top) or ILD (middle) or head position (bottom) over the plotted measurement period. The black rectangles show audio/MEMS output interactions

5 DISCUSSION

The analysis of the two continuous talker scenario (figure 1) shows that sources can be reliably determined using cross-correlation for ITD estimates and short-duration ILD estimates in laboratory environments. For increasingly complex real-world scenarios, however, the increases in background noise, reverberation and sources result in a less accurate estimate of interaural cues over short durations (e.g., figures 4 & 6). The size of the temporal window is important for the reliability of single estimates, but multiple estimates over longer time-periods are less sensitive to temporal window size. ITD cues can produce accurate estimates of source direction over time, while ILDs display a Gaussian distribution around 0 dB in all natural scenarios over time. ILDs do indicate instantaneous changes in source direction. A change in interaural cues can be attributed to source movement or movement of the listener's head in real-world scenarios if head-movement information is available. The head movement information also shows that horizontal head movement is greatest when interacting with more than one talker and smallest when passively listening and visually fixated, such as when watching television.

6 CONCLUSIONS

The current study combined and synchronized head-motion recordings with binaural audio recordings to observe their interaction. This study found that head motion changes dependent on the type of listening and engagement with the listening scenario. In addition, head movement information can allow ITD and ILD changes to be attributed to the listener of the source's

movement to an extent. Future work will include taking longer time duration recordings and more listening situations (e.g., city streets, supermarkets) to produce more robust long-term results. A second AHRS may be used to measure the position of the body, allowing head movement to be accurately recorded while the listener is moving by subtracting the body-mounted measurements from the head-mounted measurements.

ACKNOWLEDGMENTS

A.W.B. was funded by a Ph.D. studentship from the Medical Research Council. The Scottish Section of the IHR is supported by intramural funding from the Medical Research Council (grant number U135097131) and the Chief Scientist Office of the Scottish Government.

REFERENCES

1 Barbara Shinn-Cunningham, “Acoustics and perception of sound in everyday environments,” Proc. 3rd Int. Workshop Spatial Media (2003)

2 Hans Wallach, “The role of head movements and vestibular and visual cues on head movements,” J. Exp. Psych. 27(4), 339-368 (1940)

3 Bernard Willard R. Thurlow, John W. Mangels, Philip S. Runge, “Head movements during localization,” J. Acoust. Soc. Am. 42(2), 489-493 (1967)

4 Tim Brookes, Chungeun Kim, Russell Mason, “Head movements when evaluating various attributes of sound,” Proceedings of the Audio Engineering Society Convention 122. (2007)

5 Donald Dirks, John P. Moncur, “Interaural Intensity and Time Differences in Anechoic and Reverberant Rooms,” J. Speech Hear. Res. 10, 177-185 (1967)

6 Yan-Chen Lu, Martin Cooke, “Motion strategies for binaural localisation of speech sources in azimuth and distance by artificial listeners,” Speech Comm. (2010)

7 Guillaume Gibert, Gérard Bailly, Denis Beautemps, Frédéric Elisei, Rémi Brun, “Analysis and synthesis of the three-dimensional movements of the head, face, and hand of a speaker using cued speech,” J. Acoust. Soc. Am. 118(2), 1144-1153 (2005)

8 Robert D. Sorkin, Frederic L. Wightman, Doris S. Kistler, Greg C. Elvers, “An Exploratory Study of the Use of Movement-Correlated Cues in an Auditory Head-up display,” Human Factors. 31(2), 161-166 (1989)

9 György Wersényi, “Effect of Emulated Head-Tracking for Reducing Localization Errors in Virtual Audio Simulation,” IEEE Trans. Audio Speech Lang. Proc. (2008)

10 Peter Bartz, “Razor attitude and head rotation sensor,” Quality and Usability Lab, TU-Berlin, https://dev.qu.tu-berlin.de/projects/sf-razor-9dof-ahrs (last visited: 14/03/2013)

11 Mark Euston, Paul Coote, Robert Mahony, Jonghyuk Kim, Tarek Hamel, “Complementary Filter for Attitude Estimation of a Fixed-Wing UAV,” Int. Conf. Intell. Robots Systems (2008)