Audio Engineering Society

Convention Paper 6124

Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

Acoustic positioning and head tracking based on binaural signals

Miikka Tikander[1], Aki Härmä[1], and Matti Karjalainen[1]

[1] *Helsinki University of Technology, Espoo, 02150, Finland*

Correspondence should be addressed to Miikka Tikander (miikka.tikander@hut.fi)

ABSTRACT

Tracking a user's movement and orientation is essential for providing realistic mobile augmented reality audio (MARA) services. For mobile use the tracking system needs to be light-weight, wearable and wireless. Binaural microphones offer a convenient and practical solution for tracking user movement and orientation. These sensors can be easily integrated with portable headphones. In addition to tracking, the microphones also offer several possibilities to control the user's acoustic environment. This paper reviews the latest results in binaural head tracking with known anchor sources and also discusses the case where there are no known anchor (reference) sources available. Some transducer issues are also discussed.

1. INTRODUCTION

In *mobile* and *wearable augmented reality audio* (MARA, WARA) systems the acoustic environment around a user is enriched by adding virtual audio objects to the environment. Typically, the objective is to produce an impression that virtual sources cannot be discriminated from real sources around the user. One key factor in providing realistic augmented virtual services is the ability to track a user's position and orientation. This can be done with the use of a head- and motion-tracker.

There are already several accurate and robust position and motion trackers (electro-magnetic, inertial and mechanical) available but in most cases they suffer from poor mobility and short range. Some techniques, such as GPS, are very mobile but they lack the ability to track a user's orientation and they show problems in precision and robustness, particularly inside buildings. Also some acoustic head-trackers are available but most of them are based on a configuration where the user is wearing a user element emitting ultrasound and head-tracking is based on processing of signals from a microphone array placed in the environment. Sometimes this is called an *outside-in* system for tracking [2]. The downside of such a system is that the positioning information is processed at the receiver even though the position information is often needed at the user end. Also ultrasonic frequencies require specific hardware.

Recently there has been research on using binaural signals for estimating a user's position and orientation [3, 4]. In binaural head-tracking a user is wearing a pair of microphones that are capturing surrounding sounds. The positioning is performed in relation to surrounding reference sound sources (anchors). There are several benefits in using binaural signals for tracking. In most MARA applications the virtual sound objects are placed in the acoustic environment according to the user's own coordinates [1]. Because the binaural positioning is done in relation to subject's ears the positioning results can be directly utilized in MARA applications. As the subject moves the binaural microphone array follows the subject's movements accurately.

Technically, binaural positioning is very similar to acoustic [5] or binaural [6] source localization techniques where a static microphone array is locating static or moving sound sources in the environment whereas in (binaural) positioning the microphone array is moving and the sound sources are assumed to be static.

This paper reviews the recent research done on binaural head tracking in a case where there are known anchor sources in the environment. Also some system considerations are discussed and as a pointer to future study some thoughts on blind binaural positioning are given. This refers to a case where there are no known anchors available.

2. SYSTEM CONSIDERATION

The basic scenario in a binaural positioning situation is that a user is in some space that has some static sound sources in the environment. These sound sources can be anything that emit a sound with some detectable feature. One such sound source could be for example an air conditioning outlet on the wall. The sound sources could also be purposely placed in the environment and emitting deterministic signals, like loudspeakers emitting a known reference signal. Depending on configuration the positioning scenarios could be divided into the following three categories:

I) *Anchor locations, signals, and signaling times are known (synchronous tracking).* In this case the user's exact distance and orientation in relation to each anchor can be solved. However this requires that some sort of synchronization signal is transmitted to the user.

II) *Anchor locations and signals are known (asynchronous tracking).* This allows estimating the users orientation and movement in relation to each anchor. Synchronization is needed to be done online.

III) *No knowledge of surrounding sound sources.* In this scenario the positioning system needs to locate static sound sources with some detectable feature in the acoustic environment and then estimate the orientation in relations to these unknown anchors.

The positioning is done by detecting the sound arrival times and their differences. This way the orientation and relative distances can be solved. Typically this is done by analysis of cross-correlation function. The time delays of arriving sounds are derived from peak locations of the cross-correlation functions. When anchor signals are not known (case III in the above list) the cross-correlation must be computed between the two microphone signals. This case is very prone to any interfering sounds.

If the source signal is known (cases I and II in the above list) the correlation can be calculated between the known source signal and the microphone signals. This

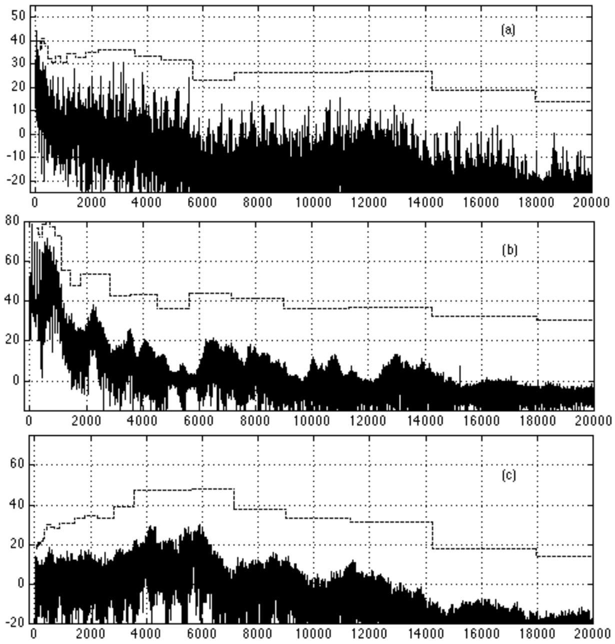

Fig. 1: Spectra of noises interfering with anchor sounds (linear frequency scale and no frequency weighting used for easy physical interpretation). (a) Background noise spectrum in a typical office room, consisting of ventilation noise (lowest frequencies) and computer fan noise. (b) Vowel /a:/ spectrum pronounced normally by a male subject, measured 1cm from ear canal entrance. (c) Fricative /s/ in vowel /a:/ context, same conditions as in case (b). Vertical axis: calibrated sound pressure level; horizontal axis: frequency in Hz. Dotted lines for 1/3-octave spectra. Signals were recorded through a B&K 2238 sound level meter.

way uncorrelated background noise affects less the cross-correlation and this increases the robustness of localization considerably. In a synchronous case two anchors are needed to solve an exact position in a half plane divided by the anchors. With three anchors, not placed on the same line, the position can be solved anywhere in the space covered by the anchors. In an asynchronous case, a change in position in a half plane divided by the anchors can be solved with two anchors, and with three anchors an exact position can be solved.

2.1. Anchor signals

One problem with multiple anchors is how to separate each anchor signal from recorded binaural signals. One way to do this is to divide the reference signal into sub-

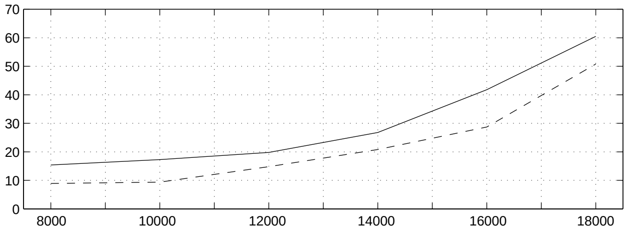

Fig. 2: JND threshold of band-passed anchor sound (Δf = 2 kHz) in dB as a function of carrier frequency (—) for bandpass noise and (- - -) bandpass-filtered impulse train (repetition period 100 ms). Average of two young subjects measured in an anechoic chamber.

bands that do not overlap in the frequency domain. At the receiver the recorded signal can be separated into subbands by using the same filters that were used to divide the reference signal for multiple anchors. After this the orientation and relative movement (or distance in synchronous case) in relation to each anchor can be solved. The more anchors there are the more robust estimation of a user location and orientation can be estimated.

Any kind of signal, e.g. music, noise, signals with modulated data, could be used as a reference signal. Ideally the reference signal would be hidden completely under the existing background noise. Here background noise means any sounds except the reference signal itself. This way the user would never hear the reference signals. One way to achieve this is to use some sort of intelligent anchors that can equalize the reference signal below the masked threshold according to current background noise level. This should be done adaptively so that the reference signal would continuously have maximum S/N but still be masked by the background noise. The problem in this scenario is that the reference signal must be hidden from all the people in the surrounding space and estimating the required masking curves might be close to impossible. In any case the reference signal must be designed with human hearing in mind. It is known that the human auditory sensitivity decreases at lower and higher frequencies. One practical way, when there is no information on background noise, is to weight the reference signal with the inverse of the human auditory sensitivity curve.

The most distinct causes of interfering background noise are the reflections from surrounding surfaces and objects, static background noise (e.g. computer fans and air conditioning), and subject's own voice. There is one common factor to all of these noise sources. They all have a distinct low-pass behavior. Fig. 1a shows spectrum of a typical office room background noise. Figs. 1b and 1c show the spectra of a wovel /a:/ and fricative /s/ recorded at the user's ear. The main trend in those spectra is that they decay toward higher frequencies. Though fricative sounds have some energy at higher frequencies as well. Also, subject's own voice is typically very loud compared to surrounding sounds and some voice activity detections methods must be used to minimize the interference.

Surface absorption typically increases toward higher frequencies as well. For this reason the reverberation and reflections from surfaces interfere most at lower frequencies. In this light, using low frequency reference signals is not favorable even though, due to lower auditory sensitivity, signals could be played louder.

2.2. High-frequency anchors

There are many benefits at using high frequencies as reference signals in positioning. As seen in Figs. 1a-c, static background noise and human speech is concentrated on lower frequencies and there is not much energy at higher frequencies. Also, typically absorption increases toward higher frequencies so reflections from surrounding surfaces are less interfering with high frequency sounds. In addition, human auditory sensitivity decreases at higher frequencies (see Fig. 2) and this allows playing the anchors at higher sound pressure levels.

The shadowing effect also increases toward high frequencies. For this reason all obstacles, including the subject's own head, start shadowing the anchor signals. When using high frequency reference signals more anchors are needed, compared to wide band references, to get a robust estimate of position.

The most straight forward way of using high frequencies would be just to apply high-pass filtered reference signals in a similar way as described earlier with wide band signals. With multiple anchors the occupied frequency band could be divided in sub-bands for each anchor. The down-side is that available bandwidth is very narrow even with few anchors.

2.3. Coherently modulated anchor signals

One way of utilizing high frequencies was introduced in [4]. In this method a low-passed (ca. 1-2 kHz) reference signal is used to modulate different carrier frequencies. Each carrier frequency corresponds to a different anchor.

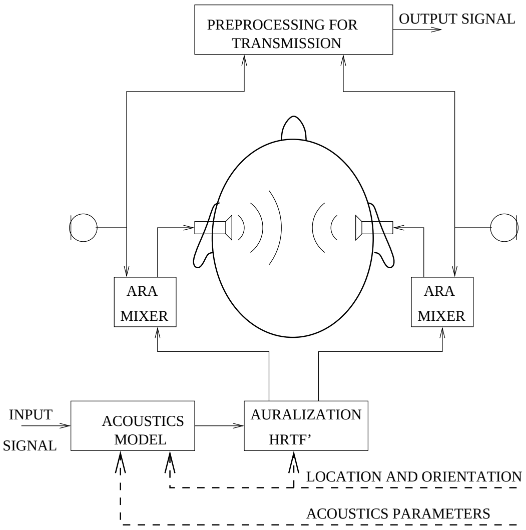

Fig. 3: Mobile augmented reality audio system based on a specific headset and signal processing architecture.

Each anchor is synchronously playing the same reference signal modulated to a different frequency region. By demodulating the recorded signal with an appropriate carrier frequency, each anchor signal can be decoded.

There are two main benefits in using coherently modulated anchor signals: Coherence between anchor signals and computational savings. Because the same reference signal is synchronously emitted by the anchors the arrival time differences between anchors can be computed, even if the reference signal is not known. This can be done by computing the cross-correlation between two demodulated anchor signals. Also, because the reference signal is a low-passed signal the calculations can be done at a lower sampling frequency to save computational load.

2.4. Headsets

Nowadays people are used to wearing different kinds of headphones or earplugs for longer periods of time while listening to music, hands-free headsets or hearing aids. The headphones offer a practical place to integrate the MARA microphones with. One big advantage is that the binaural microphone array follows the user's head and ear movements accurately and thus provides appropriate coordinates for virtual audio services.

For a positioning point of view the type of headphones is not too critical. Binaural cues related to pinnae are

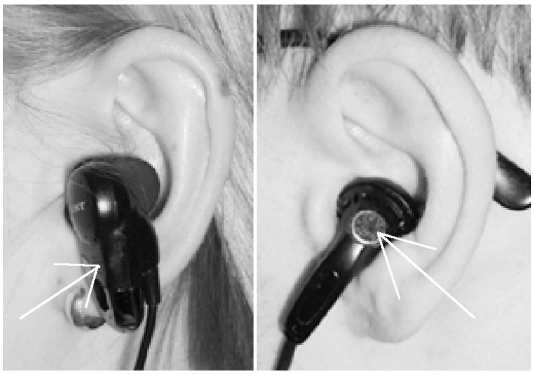

Fig. 4: Two different types of headsets used in MARA applications. The arrows point at the microphones.

easily affected by the headset but these cues are too delicate to be used in positioning anyway. On the other hand, the more important cues *interaural level difference* (ILD) and *interaural time delay* (ITD) are fairly robust to changes in headset design.

For MARA applications the design of headsets is much more crucial. The basic scenario is that the sound captured by the microphones is routed directly to the headphones and this way the user is hearing a binaural recording of the surrounding acoustic environment, the *pseudo-acoustic environment*. For this reason the headset should disturb the outer ear acoustics as little as possible and the microphones should be placed as close to the ear canal entrances as possible. If the microphones are positioned too far from the ear canal entrances or the headphones are big (thus changing outer ear geometry) the natural localization accuracy may suffer [8]. Fig. 3 illustrates a schematical view of the MARA system used in [1]. The *augmented reality audio mixer* (ARA-mixer) takes care of mixing the pseudo-acoustic environment and virtual sound objects to the user.

Fig. 4 shows two different types of headsets used in [1, 3, 4]. The headset on the left has microphones slightly outside and below the ear canal entrance and the headset on the right has microphones close to the ear canal. For binaural positioning and MARA applications the main requirements for the headset microphones are low self-induced noise and omni-directionality. For practicality a small size and high sensitivity are beneficial. Low self-induced noise is important when tracking anchor sources that are played at very low sound pressure levels. Omni-directionality of the microphones ensures that anchors can be detected from all directions and the user can hear normally in all directions in the pseudo-acoustic environment.

3. BINAURAL POSITIONING

Typically the time arrival differences are calculated by analyzing the cross-correlation function, either between the signals recorded at both ears or between the reference signal emitted by an anchor and the recorded signals. In the following the basic idea of estimating the lateral angle and change in distance in relation to one anchor is given.

3.1. Estimating the orientation

Let us assume there are *N* static anchors in the system (see Fig. 5). Each anchor is emitting (looping) a known reference sound sample and all the samples are divided into *N* non-overlapping sub-bands with filters $P_i(f)$, where $i = 1, \ldots, N$. The orientation can be estimated by detecting the *interaural time difference* (ITD) of an arriving reference signal. In the following the procedure of calculating the ITD for one anchor is shown. For the other anchors the calculation is identical.

The reference signal emitted by the *i*th anchor is given by

$$X_{\mathrm{ref},i}(f) = X_{\mathrm{ref}}(f)P_i(f), \tag{1}$$

where $X_{ref}(f)$ is the Fourier transform of the original known reference signal and $P_i(f)$ is the filter for dividing the signal in sub-bands.

The cross-correlation between the known reference and the recorded signal is given by

$$R_{\mathrm{l,ref},i} = \int_{\tau-T}^{\tau} \Phi_i(f) S_{\mathrm{l,ref},i}(f) e^{j2\pi f\tau} df, \tag{2}$$

where

$$S_{\mathrm{l,ref},i}(f) = E\{X_{\mathrm{l}}(f)X_{\mathrm{ref},i}(f)^*\} \tag{3}$$

is a sampled cross-spectrum between the recorded left channel and the reference signal. The weighting function $\Phi_i(f)$ is given by

$$\Phi_i(f) = \frac{P_i(f)}{|S_{\mathrm{l}}(f)S_{\mathrm{ref},i}(f)|^\gamma}, \tag{4}$$

where $P_i(f)$ is a frequency mask for the *i*th source and is the same as used for filtering the anchor signal. γ is

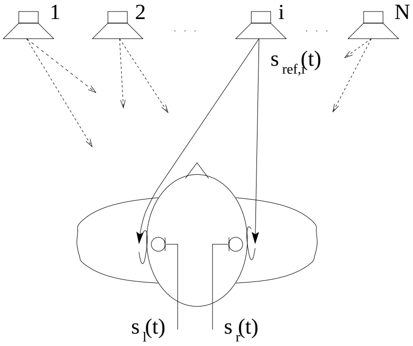

Fig. 5: *Schematical view of a system with N static sound sources.*

a parameter for changing the amount of magnitude normalization (GCC ↔ PHAT) [9]. In the same manner, the cross-correlation $R_{\mathrm{r,ref},i}$ is calculated for the right channel as well. Now the estimate for the ITD_i is given by the distance of the maximum values of the $R_{\mathrm{r,ref},i}$ and $R_{\mathrm{l,ref},i}$.

$$\mathrm{ITD}_i = \mathrm{maxarg}(R_{\mathrm{l,ref},i}) - \mathrm{maxarg}(R_{\mathrm{r,ref},i}). \quad (5)$$

When the ITD is known the lateral angle of the user relative to the *i*:th anchor can be solved by assuming an ITD model, such as delay $d_{i,\mathrm{ITD}} = R_{\mathrm{head}}\{\phi_i + \sin(\phi_i)\}$, where R_{head} is the radius of subject's head. The lateral angle for the rest of the anchors is solved the same way.

3.2. Estimating the distance

In a synchronous case where the timing of anchor signals is known the distance to each anchor is just the traveling time from an anchor to a microphone turned into distance. In an asynchronous case the exact distance cannot be determined from recorded signals because there is no information on when the signal was emitted. So, the initial value of the time delay estimate D just sets a reference point where to compare the upcoming delays with.

When the distance between an anchor and the microphones changes, as the user moves, the maximum values in the cross-correlation responses move accordingly. Now this information can be used to estimate the user movement. The chance in distance relative to the *i*th anchor is the average movement of the maximum values of the cross-correlation responses. The change in distance d_i in samples to the *i*th anchor can be estimated with

$$d_i = \frac{1}{2}(\mathrm{maxarg}(R_{\mathrm{l,ref},i}) + \mathrm{maxarg}(R_{\mathrm{r,ref},i})) - D, \quad (6)$$

where D is the initial value of d_i.

When the distances and the angles to all of the anchors are estimated, solving the user's position and orientation is a standard geometrical problem and is not discussed here in more detail. The more there are anchors the more robust estimate can be obtained.

3.3. Coherently modulated anchor signals

With coherently modulated anchor signals a low-passed reference signal is used to modulate different carrier frequencies. Synchronously, each anchor emits the reference signal with different carrier frequency. Each anchor signal can be resolved from the recorded signal by demodulation.

Assume that there is a low-passed reference signal $x_{\mathrm{ref}}(t)$ and N static anchors as shown in Fig. 5. Now, for the *i*th anchor the reference signal is formed by modulating a carrier signal $\cos(2\pi f_i t)$ with the reference signal. Thus the signal played by the *i*th anchor is

$$x_{i,\mathrm{ref}}(t) = x_{\mathrm{ref}}(t)\cos(2\pi f_i t). \quad (7)$$

The same procedure is repeated to obtain a modulated reference signal for each anchor, see Fig. 6. Care must be taken that the modulated reference signals do not overlap in the frequency domain. This means that the bandwidth of the reference signal must be smaller than half of the difference of the carrier signal frequencies:

$$BW_{\mathrm{ref}} < \frac{|f_i - f_{i-1}|}{2}. \quad (8)$$

To resolve the reference signal emitted by the ith anchor the recorded signal must the demodulated with the corresponding carrier frequency. Because the room absorption is different at different frequencies the sidelobes of the recorded signal are not necessary mirror images. For this reason a complex valued demodulation by $e^{-j\omega_i t}$ and further cross-correlation processing is needed. After demodulation the signal could also be low-pass filtered and decimated to save computational load.

When the reference signals for each anchor are resolved the orientation and distance can be calculated as described above for base-band reference signals. In a case where the reference signal is not known but the carrier frequencies are known the arrival time differences between anchors can be computed by calculating the cross-correlation between two recorded anchor signals.

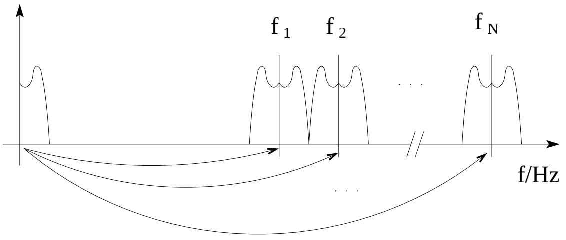

Fig. 6: *A low-passed reference signal modulated with N different carrier frequencies.*

3.4. Synchronization

The reference signal samples used should be long enough compared to the dimensions of the surrounding space. The minimum length of the reference signal is determined by the longest distance between anchors. This way the relative delays between anchors yield a single correlation peak and reliable estimates.

In the system introduced in [3], the synchronization is done as follows. Assume that the reference signal $s_{\mathrm{ref},i}(t)$ played by the *i*th anchor is *n* samples long (See Fig. 5). When the tracking system is started the first *n* samples are buffered. Then a cross-correlation between the *i*th reference signal $s_{\mathrm{ref},i}(t)$ and the other channel of the recorded signal $s_{\mathrm{r}}(t)$ or $s_{\mathrm{l}}(t)$ is calculated. Then the reference signal in the memory is circularly shifted to synchronize the system with the recorded signal. After synchronization the recorded frame should match a frame from the reference signal. The same synchronization is repeated for each anchor separately.

As the user moves the synchronization degrades and eventually when the recorded frame and the frame from the reference signal do not correlate enough the system needs to be re-synchronized. This is done the same way as the initial synchronizing. The on-line synchronization can be performed in the background while running the system without interrupting the tracking. The online synchronizing needs only to be done for the anchors that are running out of synchronization. This reduces the computational load compared to the initial synchronization.

4. COMPARISON OF DIFFERENT ANCHOR SIGNAL METHODS

In a case where anchor signals are not known the cross-correlation method locates the loudest sound signals in the environment. If the characteristics of background noise is known this can be taken into account for example by filtering the recorded signal so that the most interfering sounds are left out in the cross-correlation calculations. Unfortunately in real situations only very general trends of background noise can be assumed. When using known anchor signals the system's robustness to surrounding interfering sound is increased considerably.

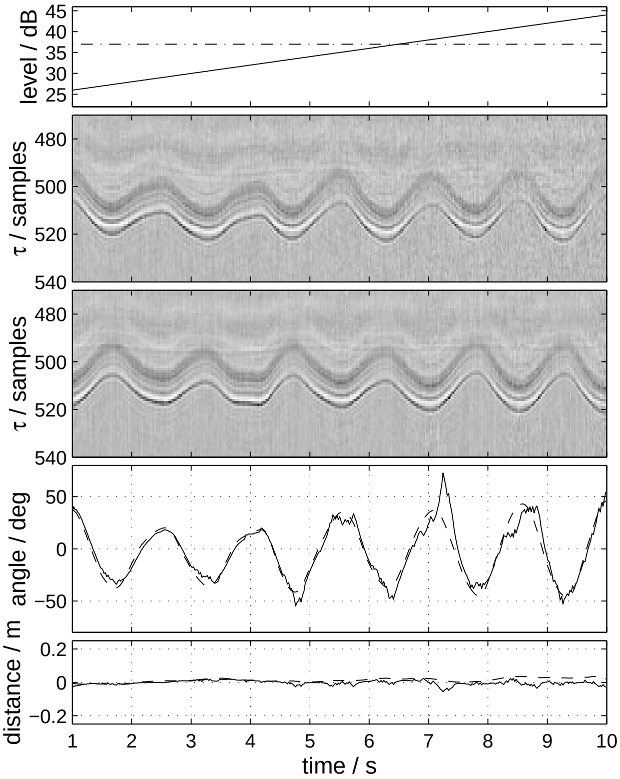

Fig. 7: *Top: Solid line is the interfering noise level and the straight line is the reference source level (37 dB, A-weighted). Middle and upper middle: Correlation responses at both recorded channels. Correlation function is coded by gray level. Lower middle and bottom: Estimated angle and distance. Solid line is the binaural tracking and the dashed line is the data from an electromagnetic tracker.*

Fig. 7 shows the results of a measurement performed in an anechoic chamber. In this case a known reference signal (0.74 second sample of white noise) was played by one loudspeaker 2.45 meters in front of a subject, who was turning his head from side to side. Interfering

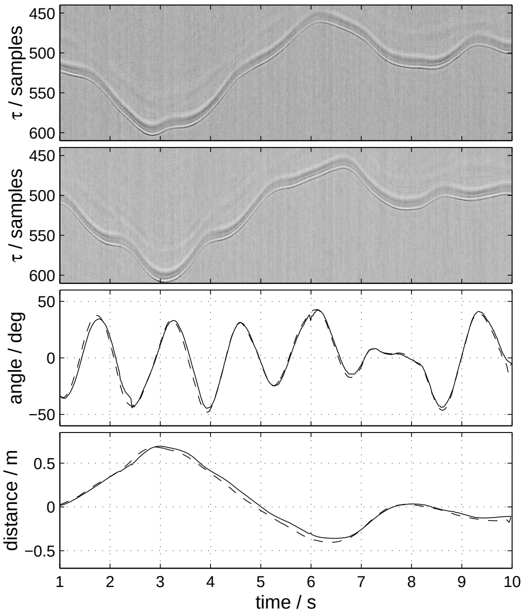

Fig. 8: *Measurement in a reverberant room. A-weighted background noise level was 40 dB and the A-weighted reference level was 53 dB. Top and upper middle: Correlation responses for both recorded channels. Lower middle and bottom: Estimated angle and distance. Solid line is the binaural tracking and the dashed line is the data from the electro-magnetic tracker.*

white noise that was increasing in volume was played by a loudspeaker. The top panel shows the A-weighted sound pressure levels of the anchor signal and interfering noise at the subject's position. Second and third panels (from the top) illustrate cross-correlation responses between the reference and the recorded signals. The responses are zoomed to show the vicinity of the maximum value of the correlation. The system performs well even when the S/N is below 0 dB. This suggests that the reference signal could be masked behind background noise or for example music and still be used for positioning. From the cross-correlation responses can be seen that at larger angles the head is shadowing the other microphone and the correlation decreases resulting in a decreased accuracy in the estimate of user's position. Especially when the interfering sound level is high and the reference level is low this effect is pronounced. The effect can be seen in the lower two panels where the binaural angle and distance estimates (solid line) are plotted. The dashed line is the data from an electro-magnetic positioning device.

Fig. 8 illustrates the results for a measurement performed in a reverberant room (4.5 m x 10.5 m, T_{60} = 0.2 s). The same reference signal was used as in the above case in an anechoic chamber. During the measurement the subject walked back and forth in front of a loudspeaker, positioned 2.5 meters in front, while turning his head from side to side. The two upper panels show cross-correlation functions between the reference and the recorded signals in the two ears. The subject's movement can also been seen as changes in peak values of cross-correlation spectra. The two lower panels show estimated angle and distance in the proposed binaural tracking method (solid line) and in an electro-magnetic positioning device (dashed line).

To test the performance of sub-band and high-frequency anchors the same measurement data as in Fig. 8 (reverberant room) was used but now the calculation was done in sub-bands. The angle was estimated by using 2 kHz frequency band from the recorded signals. The mid-frequency was varied from 1 kHz to 19 kHz. Fig. 9 shows the estimated subject's orientation when using sub-band anchor signals. At the lowest frequencies the background noise and reflections from surrounding surfaces are the main causes of interference. At higher frequencies the head shadowing is the main cause of errors in the estimates. This can be seen as errors at larger angles. Though, at smaller angles the results are consistent even at higher frequencies. In the frequency range of 6-14 kHz the static background noise has already lowered considerably and the surface absorption has attenuated the reflections to some degree. In this frequency range the head is not yet shadowing the signals completely.

5. DISCUSSION

In this paper we have reviewed the latest research done in binaural head tracking. So far the research has concentrated on a case where the anchor signals are known. Different kinds of anchor signals have been studied and experimented with. The focus has been on designing anchor signals that are as imperceivable as possible and would allow using multiple anchors.

In favorable conditions the binaural positioning has been found useful and the method could be used in many practical MARA applications. Especially the ability to track a user's orientation is a very useful and important feature that is not offered in many other current mobile positioning technologies.

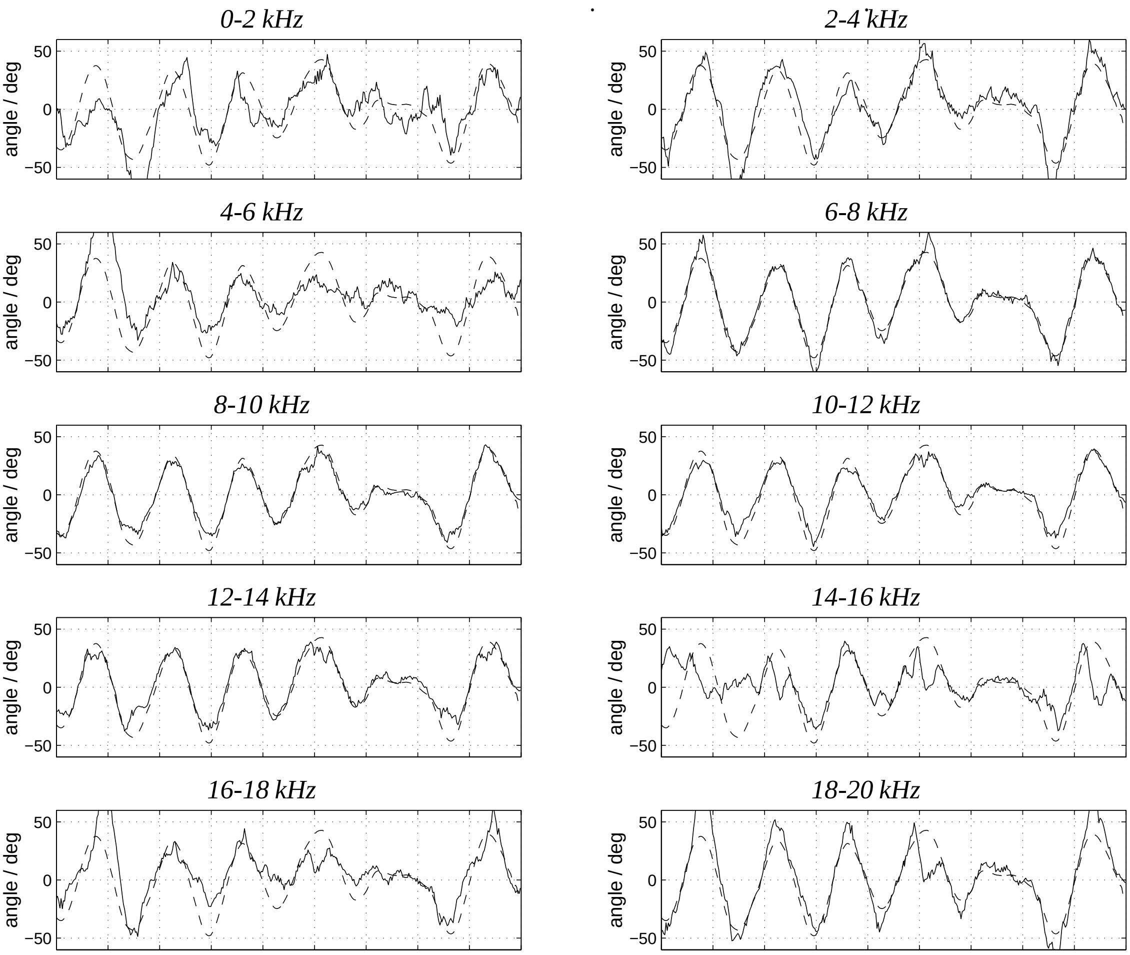

Fig. 9: *User orientation estimates when using 2 kHz frequency sub-bands by varying the sub-band mid-frequency. The measurement data is the same as in Fig. 8. Each plot is a 9 seconds take from a recorded data. Solid line is the binaural positioning estimate and the dashed line is data from an electro-magnetic motion tracker.*

The downside of binaural positioning, in a case where the reference signals are known, is the need of a fixed setup of loudspeakers or some other controllable sources. Though, there are numerous facilities that already have some sort fixed setup of loudspeakers that could be used as known anchors. Such facilities are available for example in stores, vehicles, and museums. The coordinates of the anchors (loudspeakers) could be known beforehand or the anchor coordinate data could be embedded in the reference signals and thus the anchor coordinates would be automatically available for each facility.

Furthermore, in the case where the reference signals are known, the reference signals must be generated and played with some controllable sound sources. Ideally, the users would never hear any anchor signals. This can be achieved by keeping the reference signals below the masked threshold of hearing. The usage of high frequency reference signals have been found useful because the hearing sensitivity and also interfering background noise decreases toward high frequencies.

For practical positioning multiple anchors are needed. This allows a 3D-positioning of a subject, robustness of positioning is increased and also wider areas can be covered. A wide-band reference signal can be divided to non-overlapping sub-bands to allow multiple anchors. Another way is to use a low-pass reference signal and apply it to modulate multiple high frequency carriers. This way the anchor signals can be transmitted efficiently and imperceivably to the user.

Ideally the tracking system would not depend on any external infrastructure. The system should automatically locate static sound sources in the environment and use them for tracking. In a typical acoustic environment we may identify many potential unknown anchors such as computers or air shafts in an office environment that could also be used for tracking the user's position. Because the user could be constantly moving the most problematic task is finding and defining static sources. Some first steps have been taken on this course of research but so far there are no usable solutions for blind binaural positioning.

6. ACKNOWLEDGMENT

This work has been carried out in collaboration between Helsinki University of Technology and Nokia Research Center. A. Härmä's work has been supported by the GETA graduate school and the Academy of Finland.

7. REFERENCES

[1] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, H. Nironen, and S. Vesa, “Techniques and Applications of Wearable Augmented Reality Audio”, in *Proc. 114th AES Convention*, preprint 5768, Amsterdam, March 22-25, 2003.

[2] K. Mayer, H. L. Applewhite, and F. A. Biocca, "A survey of position-trackers," *Presence: Teleoperators and virtual environments*, vol. 1, no. 2, pp. 173-200, 1992.

[3] M. Tikander, A. Härmä, and M. Karjalainen, “Binaural positioning system for wearable augmented reality audio”, in *Proc. of the IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoust. (WASPAA'03)*, New Paltz, New York, UxSA, October 2003.

[4] M. Karjalainen, M. Tikander, and A. Härmä, “Head-tracking and subject positioning using binaural headset microphones and common modulation anchor sources”, in *Proc. of IEEE Int. Conf. on Aoust., Speech, and Sig. Proc. (ICASSP'04)*, Montreal, Canada, 17-21 May, 2004.

[5] J. H. Dibiase, H. F. Silrverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. Ward, Eds., Springler-Verlag, 2001, ch. 7, pp. 131-154.

[6] N. Roman and D. Wang, "Binaural Tracking of multiple moving sources," in *Proc. IEEE Int. Conf. on Acoust., Speech, and Sig. Proc.*, Hong Kong, May 2003.

[7] S. Laugesen, K. B. Rasmussen, and T. Christiansen ”Design of a microphone array for headset”, in *Proceedings of IEEE Workshop on Appl. of Sign. Proc. to Audio and Acoust.*, October 19-22, 2003, New Paltz, New York, USA.

[8] D. S. Brungart, A. J. Kordik, C. S. Eades, and B. D. Simpson ”The effect of microphone placement on localization accuracy with electronic pass-through earplugs”, in *Proceedings of IEEE Workshop on Appl. of Sign. Proc. to Audio and Acoust.*, October 19-22, 2003, New Paltz, New York, USA.

[9] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech and Sig. Proc.*, vol. 24, pp. 320-327, August 1976.