Binaural Sound Source Localisation and Tracking using a Dynamic Spherical Head Model

Christopher Schymura, Dorothea Kolossa

Institute of Communication Acoustics
Ruhr-Universität Bochum, Germany
{christopher.schymura,dorothea.kolossa}
@rub.de

Fiete Winter, Sascha Spors

Institute of Communications Engineering
University of Rostock, Germany
{fiete.winter, sascha.spors}
@uni-rostock.de

Abstract

This paper introduces a binaural model for the localisation and tracking of a moving sound source's azimuth in the horizontal plane. The model uses a nonlinear state space representation of the sound source dynamics including the current position of the listener's head. The state is estimated via an unscented Kalman Filter by comparing the interaural level and time differences of the binaural signal with semi-analytically derived localisation cues from a spherical head model. The localisation performance of the model is evaluated in combination with two different head movement approaches based on open- and closed-loop control strategies. The results show that adaptive strategies outperform non-adaptive ones and are able to compensate systematic deviations between the spherical head model and human heads.

Index Terms: Binaural localisation, head movements, machine hearing

1. Introduction

The human auditory system exploits the acoustic properties of the outer ear including torso, head and pinna for localising sound sources. Modelling the shape of the human head as a rigid sphere is a simple geometric approach to approximate the influence of the outer ear on the Head Related Transfer Functions (HRTFs). It has been shown in the past [1], that this approach provides insights in to how far the human head contributes to the localisation capabilities of the auditory system. Brungart and Rabinowitz [2] showed that the dependency of measured Interaural Level Differences (ILDs) and Interaural Time Differences (ITDs) on the distance of nearby sound sources can be reasonably approximated with a spherical head model. While the influence of the head shape is well covered, the effects of the pinna are however completely ignored by the model. This leads to significant deviations from measured HRTFs at high frequencies [3, p.100]. The model provides an analytic connection between the position of the sound source and the ILD and ITD. Contrary to other approaches [4, 5], no prior supervised training on measured HRTF datasets is necessary to establish this connection.

The model proposed in this study uses an Unscented Kalman Filter (UKF) [6] to infer the position of the sound source from measured ITDs and ILDs. Similar approaches have already been introduced in previous works, either using Kalman filtering techniques [7, 8] or particle filters [9]. Additionally, the effects of translatory movements and head rotations of the listener on localisation performance using a particle filter have been investigated in [10]. The results conform with the work of Wallach [11], indicating that rotational head movements improve azimuth localisation by resolving front-back ambiguities, which are likely to occur if sound sources are positioned within the cone of confusion [12]. A probabilistic framework with similar capabilities was introduced in [13], though only step-by-step head rotations were considered in this work. Extensions of the model [13] include the evaluation of different head-rotation strategies [14] and the increase of robustness in reverberant and noisy conditions [15].

This paper introduces a binaural model that is capable of conducting continuous head movements and analyses in how far the insufficiencies of the spherical head approximation can be compensated by considering acoustic scene dynamics and adaptive head movement strategies.

2. Binaural Model

The model introduced in this work is represented by a generic nonlinear dynamical system

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k, u_k) + \boldsymbol{v}_k \tag{1}$$

$$\boldsymbol{y}_k = \boldsymbol{g}(\boldsymbol{x}_k) + \boldsymbol{w}_k \, , \tag{2}$$

where $\boldsymbol{x}_k$ and $\boldsymbol{y}_k$ denote the hidden state and the observation vectors at time frame k. The control input u_k is used to steer the head towards the desired orientation. $\boldsymbol{f}(\cdot)$ and $\boldsymbol{g}(\cdot)$ are nonlinear functions describing the model dynamics and the observations generated by the spherical head model, respectively. $\boldsymbol{v}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$ and $\boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$ are zero-mean, Gaussian distributed noise vectors, with covariance matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$. Throughout the course of this paper, it is assumed that both $\boldsymbol{Q}$ and $\boldsymbol{R}$ are either known in advance or can be estimated accordingly.

2.1. Model dynamics

The model dynamics (1) are represented by the 3-dimensional state vector

$$\boldsymbol{x}_k = [\phi_k \; \dot{\phi}_k \; \psi_k]^T \tag{3}$$

including the source position ϕ_k, the angular source velocity $\dot{\phi}_k$ and the head orientation ψ_k (see Fig. 1). The process equations of the former two can be described by

$$\phi_{k+1} = \phi_k + T\dot{\phi}_k + v_{\phi,k}, \quad v_{\phi,k} \sim \mathcal{N}(0, \sigma_\phi^2) \tag{4}$$

$$\dot{\phi}_{k+1} = \dot{\phi}_k + v_{\dot{\phi},k}, \quad v_{\dot{\phi},k} \sim \mathcal{N}(0, \sigma_{\dot{\phi}}^2), \tag{5}$$

where T denotes the frame length of the Kalman filter in seconds. σ_ϕ^2 and $\sigma_{\dot{\phi}}^2$ are the variances of the noise terms. The

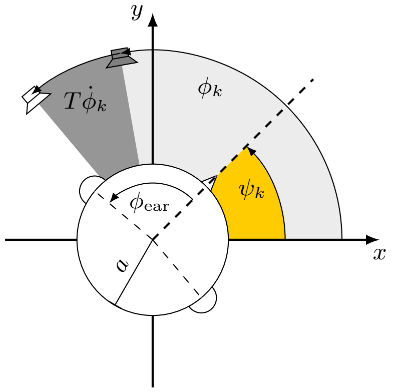


Figure 1: For this work a right-hand coordinate system is used. The parameter ϑ_{ear} measuring the angle between the ears' position and the z-axis is omitted for convenience.

process equation of the look direction is represented as

$$\psi_{k+1} = \text{sat}\Big(\psi_k + T\dot{\psi}_{\max}\,\text{sat}(u_k)\Big) + v_{\psi,k},\; v_{\psi,k} \sim \mathcal{N}(0, \sigma_\psi^2), \tag{6}$$

where $\dot{\psi}_{\max}$ is the maximum angular velocity for the head rotation in radians per second, which is assumed to be constant. In order to model physical constraints of the maximum head displacement and restricted control inputs, two saturation functions $\text{sat}(x) = \min(|x|, x_{\max}) \cdot \text{sgn}(x)$ are introduced in Eq. (6). Hence, the sets of possible angular head positions and control input values that the system can handle are defined by

$$\mathcal{H} = \Big\{\psi_k \in \mathbb{R}^1 \,\Big|\, |\psi_k| \leq \psi_{\max}\Big\}, \tag{7}$$

$$\mathcal{U} = \Big\{u_k \in \mathbb{R}^1 \,\Big|\, |u_k| \leq u_{\max}\Big\}, \tag{8}$$

where $\psi_{\max}$ is the maximum rotational angle and $u_{\max}$ is the system input limit. Furthermore, it is necessary to restrict the possible initial values for the look direction of the head within the range $\psi_0 \in [-\psi_{\max}, \psi_{\max}]$ on the frontal hemisphere.

By assuming uncorrelated disturbances affecting the state variables, the corresponding process noise covariance matrix can be expressed by

$$\boldsymbol{Q} = \begin{pmatrix} \sigma_\phi^2 & 0 & 0 \\ 0 & \sigma_{\dot{\phi}}^2 & 0 \\ 0 & 0 & \sigma_\psi^2 \end{pmatrix}. \tag{9}$$

It is worth noting, that the absolute angular position of the source and the look direction of the head are circular variables that lie within the range $(-\pi, \pi]$. This would imply that discontinuities are present in the state space, which would degrade the estimation results when using a generic UKF. To circumvent this issue, both quantities are treated as 2π-periodic variables that are unbounded in $\mathbb{R}$, which can be handled by the spherical head model that will be presented in Section 2.4.

2.2. Integration of head-movement strategies

The proposed system allows the integration of continuous head movements through the control input u_k. A positive control input of $u_k = u_{\max}$ triggers a clockwise head rotation with maximum angular velocity. Similarly, negative values induce a counter-clockwise head rotation.

In this work, the influence of continuous head rotations on the localisation performance is investigated. Two alternative strategies are proposed, relying on a purely feedforward and an adaptive feedback paradigm.

2.2.1. Static head position

If the control input is set to $u_k = 0 \,\forall\, k$, the look direction of the head will remain in its initial position. The static head position will serve as a baseline for all conducted experiments described in the following section.

2.2.2. Periodic scanning

The Periodic Scanning (PS) strategy is based on a feedforward controller

$$u_k = \sin\Big(2\pi k \frac{T}{T_{\text{p}}}\Big), \tag{10}$$

where T_{p} denotes the duration of one scan cycle in seconds. This triggers the head to perform a bidirectional periodic rotation around the initial look direction.

2.2.3. Smooth Posterior Mean

In addition to the previously introduced PS strategy, a second approach using a closed-loop feedback controller

$$u_k = \left[1 - \frac{1}{1 + |\phi_k - \psi_k|}\right] \cdot \text{sgn}\Big(\phi_k - \psi_k\Big) \tag{11}$$

is introduced and investigated. This approach is called the Smooth Posterior Mean (SPM) strategy, because the controller (11) steers the head on a smooth trajectory towards the posterior mean of the source position ϕ_k for each update of the UKF.

2.3. Binaural Front-End

The ILD and the ITD are the two main auditory cues for localising sound sources in the horizontal plane. The auditory front-end proposed by May et al. [4] is used to estimate the ILDs and the ITDs of a discrete binaural signal $\boldsymbol{s} = [\boldsymbol{s}_{\text{L}}\ \boldsymbol{s}_{\text{R}}]^T$. The subscripts L and R denote the signals corresponding to the left and the right ear, respectively. Both ear signals are sampled with a rate of $f_s = 1/T_s = 44.1$kHz. Each ear signal is then decomposed into $M = 32$ auditory channels using a phase compensated gammatone filterbank. The channel center frequencies f_c are equally distributed on the equivalent rectangular bandwidth (ERB) scale between 80 Hz and 5kHz. Half-wave rectification is applied to each frequency channel in order to extract the envelope. The ILDs and ITDs are then estimated for each frequency channel independently using non-overlapping, rectangularly windowed time frames with a length of 2048 samples ($T \approx 46.4$ms). The output of the binaural front-end is a vector

$$\begin{aligned}\boldsymbol{b}(\boldsymbol{s}_{k,\text{L}}, \boldsymbol{s}_{k,\text{R}}) = \Big[&\tau_1\Big(\boldsymbol{s}_{k,\text{L}},\, \boldsymbol{s}_{k,\text{R}}\Big), \ldots, \tau_M\Big(\boldsymbol{s}_{k,\text{L}},\, \boldsymbol{s}_{k,\text{R}}\Big), \\ &\delta_1\Big(\boldsymbol{s}_{k,\text{L}},\, \boldsymbol{s}_{k,\text{R}}\Big), \ldots, \delta_M\Big(\boldsymbol{s}_{k,\text{L}},\, \boldsymbol{s}_{k,\text{R}}\Big)\Big]^T\end{aligned} \tag{12}$$

containing the ITDs $\tau_i(\cdot)$ and the ILDs $\delta_i(\cdot)$ for each frequency channel i estimated for the k-th time frame. A generic implementation of the auditory front end used in this study is publicly available at [16].

2.4. Spherical Head Model

The nonlinear mapping function $\boldsymbol{g}(\boldsymbol{x}_k)$ introduced in Eq. (2) establishes a semi-analytical connection between the state vector (3) and the expected ILDs and ITDs. This is achieved by using an analytically derived binaural impulse response $\boldsymbol{r}(\boldsymbol{x}_k)$, which solely depends on the apparent sound source azimuth, as an input to the auditory front-end described in 2.3. This allows to express the nonlinear function in the measurement model as

$$\boldsymbol{g}(\boldsymbol{x}_k) = \boldsymbol{b}\left(\boldsymbol{r}_{\mathrm{L}}(\boldsymbol{x}_k),\, \boldsymbol{r}_{\mathrm{R}}(\boldsymbol{x}_k)\right), \tag{13}$$

where $\boldsymbol{b}$ is given in Eq. (12). The full measurement model is derived by inserting Eq. (13) into the generic measurement equation (2). In this paper, the disturbances affecting the measurements are assumed to be uncorrelated. Hence, the corresponding covariance matrix is defined as

$$\boldsymbol{R} = \begin{pmatrix} \sigma_\tau^2 \boldsymbol{I}_M & \boldsymbol{0}_M \\ \boldsymbol{0}_M & \sigma_\delta^2 \boldsymbol{I}_M \end{pmatrix}, \tag{14}$$

where $\boldsymbol{I}_M$ is the $M \times M$ identity matrix and $\boldsymbol{0}_M$ is a matrix of the same dimension containing only zeros. The variances of the noise affecting the ITDs and ILDs are denoted as σ_τ^2 and σ_δ^2.

2.4.1. Computation of Binaural Impulse Responses

The spherical head model [2, 17] is defined by three parameters, namely the head radius a and the angle pair $(\phi_{\mathrm{ear}}, \vartheta_{\mathrm{ear}})$ describing the position of both ears on the sphere (see Fig. 1). The time-frequency spectrum of the impulse response for the left ear $\boldsymbol{r}_{\mathrm{L}}(\phi_k)$ is given as its Fourier transform

$$\begin{aligned} R_{\mathrm{L}}(\boldsymbol{x}_k, \omega) = \frac{c}{4\pi\omega a^2} \sum_{\nu=0}^{\infty} \frac{h_\nu\left(\frac{\omega}{c}d\right)}{h'_\nu\left(\frac{\omega}{c}a\right)} \times \\ (2\nu+1)\, L_\nu\Big(\sin(\vartheta_{\mathrm{ear}}) \cos\left(\phi_k - \psi_k - \phi_{\mathrm{L}}\right)\Big), \end{aligned} \tag{15}$$

where $\phi_{\mathrm{L}} = -\phi_{\mathrm{R}} = \phi_{\mathrm{ear}}$. The impulse response for the right ear $\boldsymbol{r}_{\mathrm{R}}(\phi_k)$ can be generated accordingly. The spherical Hankel function of second kind and νth-order [18, sec. 10.1.1] is denoted by $h_\nu(\cdot)$, while $L_\nu(\cdot)$ symbolizes the νth-degree Legendre polynomial [18, sec. 8.6.18]. The sound source position is described in polar coordinates by the distance d and the current apparent azimuth angle $(\phi_k - \psi_k)$. The speed of sound is denoted by c. The impulse responses for both ears are generated for each time frame k and truncated to 2048 samples each.

2.4.2. Implementation details

Two practical aspects have to be considered for the implementation of the spherical head model: First, the series involved in (15) has to be truncated sensibly with respect to accuracy and computational efficiency. According to [19, (43)], the series can be truncated at

$$N = \left\lceil \frac{\pi \mathrm{e}}{2c} a\omega \right\rceil + \max\left(0, \left\lceil \ln\left(\frac{0.67848}{\epsilon}\right) \right\rceil\right), \tag{16}$$

where $\lceil \cdot \rceil$ and ϵ denote the ceiling operator and the upper bound of the truncation error. Secondly, Eq. (15) has to be evaluated on a regular grid of frequencies f in order to perform an inverse discrete Fourier transform, whose result is the discrete binaural impulse response $\boldsymbol{r}_{\{\mathrm{L,R}\}}(\phi_k)$. However, numerical instabilities are likely to occur, when dividing the two spherical Hankel functions (especially for high orders). A numerically stable approach uses a cascade of first- and second-order Infinite Impulse Response (IIR) filters, whose accumulated impulse response coincides with $\boldsymbol{r}_{\{\mathrm{L,R}\}}(\phi_k)$. The coefficients of these filters are derived from Eq. (15) using digital filter design methods. For a detailed description of this approach, the reader is referred to [20, 21].

Table 1: The table shows the parameters introduced in this publication and the respective values used for the evaluation of the binaural model.

Parameter Description	Symbol	Value	Sec.	Ref.
Head rotation limit	$\psi_{\max}$	90°	2.1	
Max. rotation speed	$\dot{\psi}_{\max}$	$45^\circ\,\mathrm{s}^{-1}$	2.1	
Control input limit	$u_{\max}$	1	2.1	
Azimuth variance	σ_ϕ^2	0.25	2.1	
Velocity variance	$\sigma_{\dot{\phi}}^2$	0.01	2.1	
Look dir. variance	$\sigma_{\dot{\psi}}^2$	10^{-8}	2.1	
Scan cycle time	T_{p}	1 s	2.2	
No. of channels	M	32	2.3	
Head radius	a	8.5 cm	2.4	[17]
Ear's azimuth angle	ϕ_{ear}	93.60°	2.4	[17]
Ear's polar angle	ϑ_{ear}	110.67°	2.4	[17]
Source distance	d	3 m	2.4	
Speed of sound	c	$343\,\mathrm{ms}^{-1}$	2.4	
Truncation error	ϵ	10^{-3}	2.4	
ITD noise variance	σ_τ^2	0.01	2.4	
ILD noise variance	σ_δ^2	1	2.4	

3. Evaluation

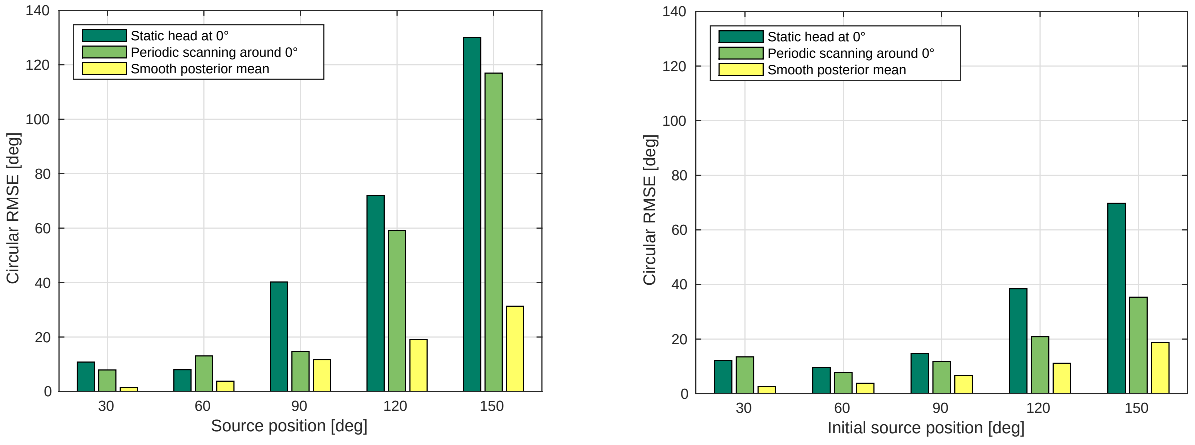
3.1. Evaluation scenarios

The proposed binaural model was evaluated in two single-source localisation scenarios. In the first scenario, a static sound source was positioned at five different target azimuth angles: 30°, 60°, 90°, 120° and 150°, covering the whole range of the cone of confusion [12]. Since the localisation task was not restricted to the frontal plane, the binaural model had to deal with potential front-back ambiguities. The second evaluation scenario contained a dynamic scene, where the initial source positions were chosen identically to the first scenario. Additionally, the scene involved a counter-clockwise, uniform circular movement of the source by 180° over the total simulation time. No additional prior knowledge was provided to the model in either scenario.

3.2. Experimental setup

For the scenarios described in Sec. 3.1, the ear signals are generated by convolving the source signal with a suitable HRTF. Based on the apparent source azimuth $(\phi_k - \psi_k)$ the HRTF is reselected for each signal frame k. Scene dynamics are simulated by cross-fading the HRTFs of subsequent frames. The binaural simulation tool used for this purpose is publicly available [22]. For the experiments, parts of the anechoic HRTF dataset released by Wierstorf et. al [23] are used. The HRTFs are measured with a Knowles Electronics Manikin for Acoustic Research (KEMAR), where the sound source is situated on a circle of 3m radius in the horizontal plane with an azimuthal resolution of 1°.

The target source was speech signals taken from the GRID corpus [24]. The corpus consists of short utterances spoken by 34 native English speakers (18 male and 16 female speakers). The evaluation set was comprised of 100 randomly selected utterances from 5 male and 5 female speakers, where 10 utterances were taken per speaker. Speech pauses were excluded from the evaluation according to the corresponding alignments provided with the corpus. The individual utterances



(a) *Static scenario.*

(b) *Dynamic scenario.*

Figure 2: *Circular RMSE of the model using different head rotation strategies in both scenarios.*

were replicated to match a total duration of 5 seconds of speech. All experiments were conducted with identical model parameters listed in Tab. 1. State estimation was performed with a generic UKF [6], using the publicly available EKF/UKF Toolbox [25]. In all conducted experiments, the initial state was set to $\boldsymbol{x}_0 = [0\,0\,0]^T$, ensuring that the model is not provided with any prior knowledge about the source position and velocity. Localisation performance was measured for each utterance using the circular root mean square error (RMSE)

$$\mathrm{cRMSE} = \sqrt{\frac{1}{K}\sum_{k=1}^{K} \min_{l \in \mathbb{Z}} \left(\hat{\phi}_k - \phi_k + 2\pi l\right)^2}$$

where $\hat{\phi}_k$ denotes the estimated source position at frame k, ϕ_k is the corresponding ground truth and K is the total number of frames in one utterance.

3.3. Results and discussion

The achieved localisation performance for both scenarios is shown in Fig. 2. Each head rotation strategy was evaluated with all 100 utterances for 5 different (initial) positions of the target source. The average circular RMSEs depicted in Fig. 2 were computed as the mean over all utterance-level circular RMSEs for each experiment.

Fig. 2a shows localisation performance for the static scenario, indicating that controlled head movements yield improvements over open-loop controlled and static head positions. The SPM strategy outperforms both the static case baseline and the PS approach in all investigated source positions. The improvements are statistically significant according to a t-test conducted with $p < 0.01$. It can be seen that sources positioned at the rear of the listener decrease localisation performance for static and open-loop controlled head positions due to occurring front-back ambiguities. Furthermore, the baseline localisation error for a source that is positioned at 90° is still prominent, even though front-back ambiguities are not likely to occur in this case. The resulting error can be explained by systematic errors of the spherical head model itself, which produces ITDs and ILDs that do not match the corresponding measurements. The dynamic scenario yields comparable results to the static case, as depicted in Fig. 2b. The achieved localisation errors in all dynamic experiments are generally lower than for static sources. This can be explained by the fact that both the frontal and the rear hemisphere of the listener are covered by the moving sources for all investigated initial positions. This reduces the probability of occurring front-back ambiguities even if no head rotations are applied. However, the SPM movement strategy outperforms the PS and the baseline just as well as in the static scenario. These improvements are likewise statistically significant.

4. Conclusions and future work

In this study, a binaural model for localisation and tracking of sound sources, including continuous head rotations, was introduced. Experimental results have shown that closed-loop controlled head movements yield significant improvements in localisation performance over a pre-determined open-loop control strategy and no head movements in static and dynamic scenes. In particular, the investigated closed-loop strategy was able to reduce the effect of systematic errors of the underlying spherical head assumption of the model, in comparison with measured HRTFs from a KEMAR dummy head.

A next step for further investigations is the analysis of errors introduced by the spherical head model in order to improve robustness for localisation in different acoustic environments. The integration of sound distance as an additional state parameter is a second possibility for further extensions of the model. Additionally, the assumption of uncorrelated disturbances affecting the underlying state space model used in this study does not hold for practical applications. Therefore, it is necessary to further include and evaluate suitable techniques for the estimation of the corresponding covariance matrices describing the process and measurement noise. Due to the fact that the model proposed in this study is able to perform continuous head movements, it can also serve as a testbed for comparisons with human localisation performance assessed in listening tests.

5. Acknowledgements

This research has been supported by EU FET grant TWO!EARS, ICT-618075.

6. References

[1] R. O. Duda and W. L. Martens, “Range dependence of the response of a spherical head model,” *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3048–3058, 1998.

[2] D. S. Brungart and W. M. Rabinowitz, “Auditory localization of nearby sources. Head-related transfer functions,” *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1465–1479, 1999.

[3] D. S. Brungart, “Near-field auditory localization,” Ph.D. dissertation, Massachusetts Institute of Technology, 1998.

[4] T. May, S. van de Par, and A. Kohlrausch, “A Probabilistic Model for Robust Localization Based on a Binaural Auditory Front-End,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 1–13, Jan. 2011.

[5] H. Wierstorf, A. Raake, and S. Spors, “Binaural Assessment of Multi-channel Reproduction,” in *The Technology of Binaural Listening*, J. Blauert, Ed. Springer, 2013, pp. 255–278.

[6] S. J. Julier, “A New Extension of the Kalman Filter to Nonlinear Systems,” *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, vol. 3, 1997.

[7] A. Portello, P. Danes, and S. Argentieri, “Acoustic models and Kalman filtering strategies for active binaural sound localization,” in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, Sept 2011, pp. 137–142.

[8] J. Traa and P. Smaragdis, “A Wrapped Kalman Filter for Azimuthal Speaker Tracking,” *Signal Processing Letters, IEEE*, vol. 20, no. 12, pp. 1257–1260, Dec 2013.

[9] D. B. Ward, E. Lehmann, and R. Williamson, “Particle filtering algorithms for tracking an acoustic source in a reverberant environment,” *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 826–836, Nov 2003.

[10] Y.-C. Lu and M. Cooke, “Motion strategies for binaural localisation of speech sources in azimuth and distance by artificial listeners,” *Speech Commun.*, vol. 53, no. 5, pp. 622–642, May 2011.

[11] H. Wallach, “The role of head movement and vestibular and visual cues in sound localization,” p. 368, 1940.

[12] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*. Cambridge, Mass. MIT Press, 1997.

[13] C. Schymura, N. Ma, T. Walther, G. Brown, and D. Kolossa, “Binaural Sound Source Localisation using a Bayesian-network-based Blackboard System and Hypothesis-driven Feedback,” in *Proc. Forum Acusticum*, Kraków, Poland, 2014.

[14] T. May, N. Ma, and G. J. Brown, “Robust localisation of multiple speakers exploiting head movements and multi-conditional training of binaural cues,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2015.

[15] N. Ma, T. May, H. Wierstorf, and G. J. Brown, “A machine-hearing system exploiting head movements for binaural sound localisation in reverberant conditions,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2015.

[16] R. Decorsière, T. May, C. Kim, and H. Wierstorf, “Two!Ears Auditory Front-End 0.8,” 2015. [Online]. Available: http://dx.doi.org/10.5281/zenodo.13788

[17] V. R. Algazi, C. Avendano, and R. O. Duda, “Elevation localization and head-related transfer function analysis at low frequencies,” *J. Acoust. Soc. Am.*, vol. 109, no. 3, pp. 1110–1122, 2001.

[18] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Corporation, 1964, no. 55.

[19] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, “Intrinsic Limits of Dimensionality and Richness in Random Multipath Fields,” *Signal Processing, IEEE Transactions on*, vol. 55, no. 6, pp. 2542–2556, 2007.

[20] F. Zotter and M. Noisternig, “Near- and Far-Field Beamforming Using Spherical Loudspeaker Arrays,” in *3rd Congress of the Alps Adria Acoustics Association*, Graz, Austria, Sep. 2007.

[21] S. Spors, V. Kuscher, and J. Ahrens, “Efficient realization of model-based rendering for 2.5-dimensional near-field compensated higher order Ambisonics,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, 2011.

[22] F. Winter, H. Wierstorf, and T. May, “Two!Ears Binaural Simulator 0.8,” 2015. [Online]. Available: http://dx.doi.org/10.5281/zenodo.13982

[23] H. Wierstorf, M. Geier, and S. Spors, “A Free Database of Head Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances,” in *Proc. of 130th Aud. Eng. Soc. Conv.*, London, UK, 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16564

[24] M. Cooke, J. Barker, S. Cunningham, and X. Shao, “An audio-visual corpus for speech perception and automatic speech recognition,” *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2421–2424, 2006.

[25] J. Hartikainen and S. Särkkä, *Optimal filtering with Kalman filters and smoothers: A Manual for Matlab toolbox EKF/UKF*, P.O.Box 9203, FIN-02015 TKK, Espoo, Finland, Feb. 2008. [Online]. Available: http://www.lce.hut.fi/research/mm/ekfukf/