Full-Sphere Binaural Direction-of-Arrival Estimation Incorporating Head Rotation Information

Erik Fleischhauer and Peter Jax
Institute of Communication Systems (IKS), RWTH Aachen University, Germany
{fleischhauer, jax}@iks.rwth-aachen.de

***Abstract*—We previously proposed movement-invariant beamforming (MIBF) as a binaural direction-of-arrival (DoA) estimator that incorporates head rotation information along the horizontal axis. We showed that including yaw head rotation information helps to resolve the front-back confusion problem and leads to a more robust DoA estimate during head movements. In this contribution, we extend MIBF by head rotation information along all three axes. We analyze how head rotations resolve cone-of-confusion (CoC) ambiguities through beampattern analysis and experiments on real data. We show that yaw rotations are most helpful to resolve front-back ambiguities and roll rotations are most helpful to resolve up-down ambiguities.**

***Index Terms*—beamforming, beampattern, binaural, cone-of-confusion, direction-of-arrival, head rotation**

I. Introduction

Binaural direction-of-arrival (DoA) estimation is required in applications like speech enhancement for hearing aids [1] or immersive playback using binaural cue adaptation [2], [3]. One challenge is that a human's ability to rapidly move their head affects binaural signal properties. This can be addressed by measuring and incorporating head rotation information into localization algorithms.

If binaural DoA estimation is applied with only two microphones, the cone-of-confusion (CoC) problem arises. This states that without additional information, directions of sources on a cone around the interaural axis cannot be distinguished reliably using only the differences between the signals of two non-moving microphones. This implies that it is difficult to estimate the elevation angle or to determine whether a source is in front of or behind the listener (front-back confusion) [4, Sec. 2.1]. Humans overcome this problem by the use of visual cues [5], monaural cues [6] and dynamic cues [7]. Dynamic cues describe the changes in the binaural cues due to head rotations. Yaw rotations are particularly effective to resolve front-back ambiguities [5]. In [8] it was shown that roll rotations help in estimating the elevation angle. Pitch rotations do not appear to be as effective in resolving ambiguities [9].

In [10] a spherical harmonics based DoA estimation algorithm for robots was proposed that includes motion information. It was shown that motion generally helps to improve the DoA estimation performance. There are approaches that incorporate yaw head movements for binaural horizontal plane localization, such as a Bayesian inference-based approach in [11], deep learning-based approaches in [12], [13] or our proposed movement-invariant beamforming (MIBF) approach in [14]. They showed that yaw head movements can improve the DoA estimation performance and help to reduce the front-back confusion problem. Further, in [15] a Bayesian inference-based approach was proposed that incorporates head movements along all three axes for full-sphere localization. They analyze with numerical simulations how different head rotations helps to reduce the CoC problem.

In this contribution, we extend the MIBF by incorporating head rotation information along all three axes and we analyze the CoC problem for different head movements. In [16] it was shown how the movement of an array grid avoids spatial aliasing by analyzing beampatterns averaged over the movement. We apply this method by analyzing the MIBF's beampatterns for different head rotations and we show how head rotations help in binaural localization. Further, we evaluate our algorithm on real binaural recordings containing head movements. To investigate the CoC effect on the localization performance we propose a cone-of-confusion distance (CoCD).

The paper is structured as follows: Section II describes the binaural signal model and the extension of MIBF for rotations along all three axes. Section III analyzes the ability of the extended MIBF to exploit head rotation information for the resolution of the CoC problem using beampattern analysis. The experimental evaluation of the proposed method on real recordings is discussed in Section IV. The paper concludes in Section V.

II. Binaural DoA Estimation

To analyze the role of head rotations in DoA estimation, we consider the single-source case and model the direction of the source as a position on the unit sphere. Head rotations are modeled using quaternion calculus (see e.g. [17]). First, a signal model for binaural recordings with dynamic head motion is presented. Subsequently, MIBF is extended to include rotation information around all three axes.

A. Binaural Signal Model

Fig. 1 illustrates the used coordinate system. A distinction is made between the head-related coordinate system and the world coordinate system. The x-axis of the head-related coordinate system always corresponds to the listener's line of sight and the y-axis to the listener's interaural axis. Since only head rotations are considered, the origin of the world

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 509806277

coordinate system is identical to the origin of the head-related coordinate system, but is unaffected by head movements. We describe head rotations using roll-pitch-yaw angles where *roll* denotes rotations around the x-axis, *pitch* denotes rotations around the y-axis and *yaw* denotes rotations around the z-axis.

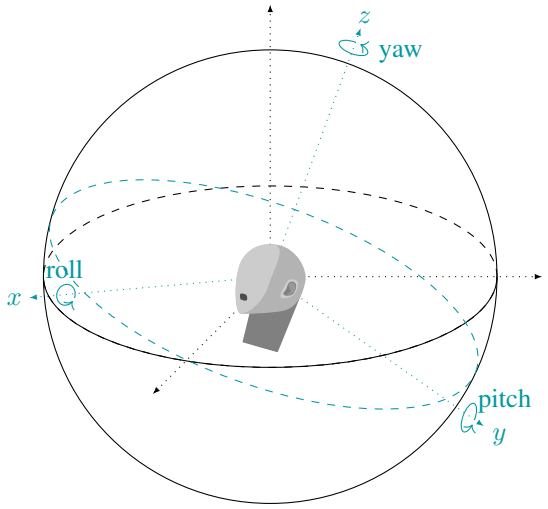


Fig. 1: World coordinate system shown in black and rotated head-related coordinate system shown in turquoise.

The beamformer's steering direction is modeled as a coordinate on a unit sphere at $\mathbf{d} = (0, x, y, z)^{\mathrm{T}} \in \mathbb{H}$, $\|\mathbf{d}\| = 1$ in the world coordinate system, where $\mathbb{H}$ is the set of quaternions, T is the transpose operator and $(x, y, z)^{\mathrm{T}} \in \mathbb{R}^3$ are cartesian coordinates. The beamformer's steering direction in the head-related coordinate system can be described by $\mathbf{d}'_\lambda$, where λ is the time index. The relation between two coordinates can be described by the rotation quaternion $\mathbf{q}_\lambda \in \mathbb{H}$ with

$$\mathbf{d}'_\lambda = \mathbf{q}_\lambda \otimes \mathbf{d} \otimes \mathbf{q}_\lambda^* \in \mathbb{H}, \tag{1}$$

where $\otimes$ denotes the quaternion product and $(\cdot)^*$ denotes the quaternion conjugate. In this contribution we assume that head rotation information $\mathbf{q}_\lambda$ is given.

A binaural recording is linked to the head-related coordinate system and can be described as a vector $\mathbf{x}_{\lambda,\mu} = (x_{0,\lambda,\mu}, \ldots, x_{M-1,\lambda,\mu})^{\mathrm{T}}$ of complex-valued short-time Fourier transform (STFT) coefficients, with μ as the frequency bin index and M as the number of microphones. The underlying signal model of a single source recording is given by

$$\mathbf{x}_{\lambda,\mu} = \mathbf{v}_\mu\left(\mathbf{d}'_\lambda\right) \cdot s_{\lambda,\mu} + \mathbf{n}_{\lambda,\mu}, \tag{2}$$

where $s_{\lambda,\mu}$ represents the source signal and $\mathbf{v}_\mu\left(\mathbf{d}'_\lambda\right)$ represents the acoustic transfer functions (ATFs) between a source at direction $\mathbf{d}'_\lambda$ in the head-related coordinate system and the M microphones. The vector $\mathbf{n}_{\lambda,\mu}$ models additive noise components, representing, for example, reverberation and diffuse noise.

B. Extension of the MIBF DoA Estimator

In [14] we proposed movement-invariant beamforming (MIBF) for binaural localization in the horizontal plane, utilizing only yaw head rotation information. The underlying idea is to estimate the averaged beamformer output power in the world coordinate system and thus invariant to head rotations. This leads to the advantage that CoC ambiguities in the averaged beamformer output power reduce and a more robust estimate, which is less affected by CoC ambiguities, is obtained. In this contribution we extend MIBF to incorporate head rotation information along all three axes. The beamformer output power steered into the direction $\mathbf{d}$ in world coordinates is estimated utilizing first-order recursive smoothing with

$$\hat{\Lambda}_{\lambda,\mu}\left(\mathbf{d}\right) = \gamma \hat{\Lambda}_{\lambda-1,\mu}\left(\mathbf{d}\right) + \left(1-\gamma\right) E_{\lambda,\mu}\left(\mathbf{d}\right), \tag{3}$$

where $E_{\lambda,\mu}\left(\mathbf{d}\right)$ represents the instantaneous output power from direction $\mathbf{d}$, and $0 < \gamma < 1$ denotes a smoothing constant. The smoothing constant can be calculated as a function of an exponential time constant τ, using $\gamma = \exp\left(-T_\lambda / \left(f_s \cdot \tau\right)\right)$ [18], where T_λ represents the STFT frame-advance parameter and f_s denotes the sampling rate. To calculate the beamformer output power, a set Ω_H of direction-dependent ATF estimates is needed. In the notation $\hat{\mathbf{v}}\left[\mathbf{d}\right]$, square brackets indicate the nearest ATFs to the direction $\mathbf{d}$ within the set Ω_H. The instantaneous output power for the direction $\mathbf{d}$ in the world coordinates is given by

$$E_{\lambda,\mu}\left(\mathbf{d}\right) = \frac{\left|\hat{\mathbf{v}}_\mu^{\mathrm{H}}\left[\mathbf{q}_\lambda \otimes \mathbf{d} \otimes \mathbf{q}_\lambda^*\right] \mathbf{x}_{\lambda,\mu}\right|^2}{\left\|\hat{\mathbf{v}}_\mu\left[\mathbf{q}_\lambda \otimes \mathbf{d} \otimes \mathbf{q}_\lambda^*\right]\right\|^2}, \tag{4}$$

where H denotes the hermitian transpose. Since the recording $\mathbf{x}_{\lambda,\mu}$ is always tied to the listener's coordinate system, we need to choose an ATF from Ω_H which compensates for head movement, using Eq. (1). The online broadband DoA estimate in world coordinates is determined as the direction that maximizes the beamformer output power

$$\hat{\mathbf{d}}_\lambda = \underset{\mathbf{d} \in \Omega_S}{\operatorname{argmax}} \sum_{\mu}^{N_{\mathrm{DFT}}} \hat{\Lambda}_{\lambda,\mu}\left(\mathbf{d}\right), \tag{5}$$

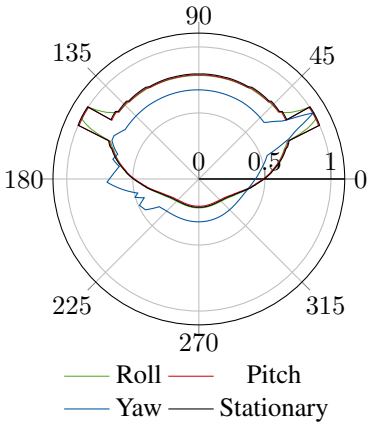
where N_{DFT} is the discrete Fourier transform (DFT) size. The search space Ω_S contains directions in world coordinates for which the MIBF searches for the position estimate. In [14] this was restricted to the horizontal plane. In Sec. IV we vary the search space for different configurations.

III. Beampattern Analysis

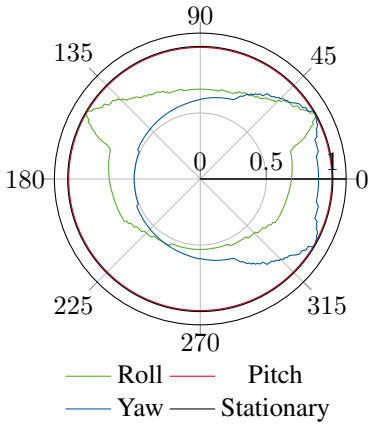
We want to investigate MIBF's beampattern to analyze the role of head rotations in resolving CoC ambiguities. The beampattern can be defined for a stationary array grid [19, Sec. 15.4]. In [16] the beampattern of a moving array grid was calculated by averaging over the rotation. In this contribution we determine the broadband rotating beampattern by

$$G\left(\mathbf{d};\mathbf{u}\right) = \frac{1}{N_{\mathrm{DFT}}} \sum_{\mu}^{N_{\mathrm{DFT}}} \frac{1}{T} \sum_{\lambda}^{T} \frac{\left|\hat{\mathbf{v}}_\mu^{\mathrm{H}}\left[\mathbf{d}'_\lambda\right] \mathbf{v}_\mu\left[\mathbf{u}'_\lambda\right]\right|^2}{\left\|\hat{\mathbf{v}}_\mu\left[\mathbf{d}'_\lambda\right]\right\|^2}, \tag{6}$$

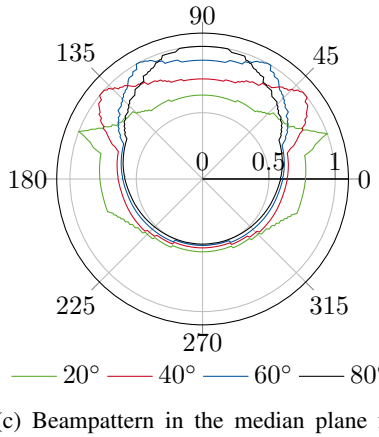
where $\mathbf{d}$ denotes the steering direction of the beamformer, $\mathbf{u}$ denotes the position of the source in the world coordinate system, $\hat{\mathbf{v}}$ is the given ATF estimate and $\mathbf{v}$ is the actual



(a) Beampattern in the horizontal plane for a source at position 30° azimuth.



(b) Beampattern in the median plane for a source at position 30° elevation.



(c) Beampattern in the median plane for different source positions along the median plane and for a roll rotation of 30°.

Fig. 2: Normalized beampatterns in the world coordinate Systems for varying source positions, rotations and planes. All head rotations start in the same position and rotate 30° around the given axis.

ATF. The head-related positions $\mathbf{d}'$ and $\mathbf{u}'$ are calculated with Eq. (1).

For the simulations depicted in Figure 2, we use the spherical head-related transfer function (HRTF) model from [20] with a resolution of 1° for both the estimated and actual ATF. We investigate the stationary case, where the head remains in a fixed position, and ideal head rotations, which are modeled as a linear movement of 30° along a given axis in steps of 1° per time index. Note that natural head rotations include rotations along all three axes (as it can be seen in Table I). The sampling rate was set to $f_\mathrm{s} = 48\,\mathrm{kHz}$ and the DFT size was set to $N_\mathrm{DFT} = 512$.

Figure 2a depicts the beampattern in the horizontal plane for a source at position 30° azimuth. In the absence of head rotations, the beampattern exhibits two peaks at azimuths 30° and 150°, indicating front-back confusion. A yaw movement results in the disappearance of the phantom peak at 150° and thus eliminates the front back-confusion. A similar effect of yaw rotations has been found in psychoacoustic experiments with humans [7]. Further, the maximum at 30° is narrower than in the stationary case, which may indicate a more robust and precise localization performance. The beampattern of the pitch rotation is comparable to that observed in the stationary case. This can be attributed to the fact that the direction-dependent HRTF according to the spherical head model stays the same over the course of the rotation. Similarly, the roll rotation does not resolve the front-back confusion, as both peaks remain.

Figure 2b illustrates the beampattern for the median plane, with the source positioned at 30° elevation. It can be observed that the beampattern in the stationary case and for the pitch rotation forms a unit circle, which indicates a CoC. Consequently, elevation angles cannot be determined at all in the stationary case. However, the roll and the yaw rotations enable an estimation. Pure roll rotations still present a front-back confusion problem, as do pure yaw rotations, which still exhibit an up-down confusion problem. These findings are align with those of [8] which indicate that roll rotations, in particular, can assist in estimating the elevation angle.

Figure 2c displays the beampattern in the median plane for a roll rotation by 30° and varying source positions along the median plane. It can be observed that the peak at the source position is more pronounced for source positions in close proximity to the horizontal plane. This may indicate that detecting the DoA near to the horizontal plane is more feasible than in the polar region. This assumption requires further investigation.

IV. Evaluation

This section presents an evaluation of the performance of the extended MIBF for different head rotations and different localization planes, based on real binaural recordings from humans. First, the procedure used to collect the recordings is described. Subsequently, the evaluation setup is outlined. Finally, the results of the experiment are presented and discussed.

A. Data Acquisition

The recordings were conducted in the IKS|Lab of the Institute of Communication Systems (IKS) at RWTH Aachen University, in a ITU-R BS.1116-3 compliant room with a low reverberation time ($T_{60} \approx 0.25\,\mathrm{s}$). The participants were seated in a chair and speech signals from the VCTK corpus [21] were played back at a sampling rate of $48\,\mathrm{kHz}$ by Neumann KH 120 loudspeakers positioned around the participant's head. The signals were played from ten different positions, with seven in the horizontal plane and four in the median plane. The six participants were equipped with B&K 4101-B binaural microphones, and their head rotation data was recorded with a HTC Vive tracker mounted on the participant's head. The participants were instructed to perform a specific one-sided rotation around either the roll, pitch or yaw axes, or to maintain their head stationary while the sound signal was played. The head rotation speed was left unrestricted. At the outset of each

trial, the participant was instructed to rotate their head to the reference position ($0°$ azimuth, $0°$ elevation). For each head movement, 120 binaural recordings and the corresponding head rotations were recorded, resulting in a total of 480 recorded signals with an average duration of 6.55 s.

TABLE I: Root mean square deviation of the roll, pitch and yaw angles averaged over 120 recordings.

Instructed Motion	Measured Motion		
	roll [°]	pitch [°]	yaw [°]
stationary	0.24	0.35	0.24
roll	22.83	3.56	4.13
pitch	2.57	18.19	1.78
yaw	2.14	1.73	29.7

Table I presents the measured root mean square rotation along the specific axes for each instructed motion averaged over 120 signals. Since natural head movements are given, it can be observed that in addition to the instructed rotation, rotations around the other axes occurs. Furthermore, the amount of the different head rotations differ depending on the axis of rotation.

B. Evaluation Setup

The binaural recordings were transformed into the STFT domain using a DFT length of $N_{\text{DFT}} = 512$ samples, and an overlapping window procedure with a frame advance of $T_\lambda = 256$ and a Hann analysis window [22]. The time constant was chosen to $\tau = 300\,\text{ms}$. The ATFs representing Ω_H were generated using the spherical HRTF model from [20], which only depends on the absolute angle, i.e. the angle between the acoustic source and the ear. The absolute angle resolution of the spherical HRTFs in Ω_H is $1°$. Consequently, there is a mismatch between the recorded participants' HRTFs and those used for DoA estimation.

As performance metric the great-circle distance (GCD) is used which is defined by

$$\Delta\sigma_{\text{GCD}}\left(\hat{\mathbf{d}}_\lambda, \mathbf{d}_\lambda\right) = \arccos\left(\hat{\mathbf{d}}_\lambda^{\text{T}}\mathbf{d}_\lambda\right), \tag{7}$$

where $\hat{\mathbf{d}}$ is the estimated direction using Eq.(5) and $\mathbf{d}$ represents the actual source direction. The GCD does not distinguish whether the estimation errors are caused by the CoC problem or by a generally poor estimation performance. Therefore, we propose the cone-of-confusion distance (CoCD), which measures the closest distance between the estimated position in the head-related coordinate system and the CoC which is spanned by the actual position $\mathbf{d}'$. If the GCD is significantly larger than the CoCD, then the estimator is obviously affected by the CoC problem. The CoCD can be calculated with

$$\Delta\sigma_{\text{CoCD}}\left(\hat{\mathbf{d}}_\lambda, \mathbf{d}_\lambda\right) = \left|\arccos\left(\mathbf{e}_y^{\text{T}}\hat{\mathbf{d}}'_\lambda\right) - \arccos\left(\mathbf{e}_y^{\text{T}}\mathbf{d}'_\lambda\right)\right|, \tag{8}$$

where $\mathbf{e}_y = [0, 0, 1, 0]^{\text{T}}$ is the quaternion y-axis unit vector, and $\mathbf{d}$ and $\hat{\mathbf{d}}$ are transformed into the head-related coordinate system using Eq. (1). Since both ears lies on the y-axis by definition, the intersection between the CoC and the surface of the unit sphere is a circle around the y-axis at the position of the y-value of $\mathbf{d}'$. The mean absolute error can be calculated for the GCD and the CoCD with

$$\overline{\Delta\sigma} = \frac{1}{T}\sum_{\lambda}^{T}\Delta\sigma\left(\hat{\mathbf{d}}_\lambda, \mathbf{d}_\lambda\right). \tag{9}$$

Moreover, the accuracy can be calculated for both metrics with

$$Acc = \frac{100\%}{T}\sum_{\lambda}^{T}\mathbb{1}\left(\left|\Delta\sigma\left(\hat{\mathbf{d}}_\lambda, \mathbf{d}_\lambda\right)\right| \leq \sigma_T\right), \tag{10}$$

where $\mathbb{1}(\cdot)$ is the indicator function and σ_T is a threshold value. In this contribution $\sigma_T = 10°$ is chosen.

C. Experiment

We want to evaluate how different head movements affect the MIBF's DoA estimation performance in the horizontal and the median plane and on the full-sphere localization task. For horizontal localization, the search space Ω_s contains positions on the horizontal plane with azimuth values between $-180°$ and $179°$ with a $1°$ resolution. For frontal median plane localization Ω_s contains positions in the median plane with elevation values between $-40°$ and $90°$ with $1°$ resolution. Consequently, it is assumed that the source is in front of the listener. For full-sphere localization Ω_s contains direction from an equal-angle grid with $1°$ resolution and a minimum elevation angle of $-40°$.

1) Horizontal Plane: Table II depicts the outcomes of the horizontal plane localization experiment. The stationary case has a high GCD while the CoCD is relatively low. This means that the main part of the GCD can be explained by localization issues due to the CoC problem. The relatively small CoCD lies for all rotations in the same order of magnitude indicating that the DoA estimator is for all kind of head movements generally working and the discrepancy to the GCD shows how strong DoA estimation is affected by the front-back confusion problem for the different head movements. In line with the beampattern analysis presented in Fig. 2a and [14], yaw head rotations help to reduce front-back confusion. The enhanced performance for the instructed roll and pitch rotations in comparison to the stationary case may be attributed to the fact that, as evidenced in Table I, the actual head rotation also encompasses a yaw component, with the roll movement having a greater yaw component than the pitch rotation.

TABLE II: DoA estimation performance along the horizontal plane averaged over 84 recordings.

Motion	GCD		CoCD	
	$\overline{\Delta\sigma}$ [°]	Acc [%]	$\overline{\Delta\sigma}$ [°]	Acc [%]
yaw	13.1	79.2	6.3	85.9
roll	31.2	53.6	5.7	85.5
pitch	38.2	51.9	6.3	82.5
stationary	47.1	41.2	6.4	80.4

2) Frontal median plane: Tab. III shows the results of the frontal median plane localization experiment. As the CoC lies near to the localization plane or, in the stationary case, on the localization plane, the CoCD is low or close to zero.

As anticipated in Fig. 2b, the roll rotation followed by the yaw rotation yields the best results for estimating elevation angles, which generally shows that the elevation angle can be estimated with the help of head movements. Examining only at the estimated elevation angle, up-down confusion is possible, but not front-back confusion. Therefore, in the scenario where we only consider frontal median plane localization the roll performance is not limited by ambiguities, whereas the yaw performance is. As expected, no localization is possible in the stationary case.

TABLE III: DoA estimation performance along the frontal median plane averaged over 48 recordings.

	GCD		CoCD	
Motion	$\overline{\Delta\sigma}$ [°]	*Acc* [%]	$\overline{\Delta\sigma}$ [°]	*Acc* [%]
roll	12.2	62.8	2.4	97.0
yaw	31.1	29.3	3.8	92.4
pitch	46.7	16.8	1.7	99.9
stationary	57.9	7.5	0.1	100.0

3) Full-sphere: Tab. IV shows the localization performance on the full-sphere. Note that the dataset only contains sound sources in the horizontal and median planes and is therefore not balanced across directions. Again a small and similar CoCD can be seen for all motions indicating that localization errors mainly comes from the CoC problem. It can be seen that the yaw and roll rotations performs best, but there is a significant difference to the best performing results of the previous two experiments. This is in line with the results of the beampattern analysis, which shows that no rotation could resolve all ambiguities. More complex rotations may be needed with MIBF to achieve better performance in full-sphere localization.

TABLE IV: DoA estimation performance for full-sphere localization averaged over 120 recordings.

	GCD		CoCD	
Motion	$\overline{\Delta\sigma}$ [°]	*Acc* [%]	$\overline{\Delta\sigma}$ [°]	*Acc* [%]
yaw	30.9	26.4	6.9	82.5
roll	44.3	24.2	7.0	79.0
pitch	57.3	10.7	5.9	85.1
stationary	59.7	9.3	5.2	84.7

V. Conclusion

In this contribution we extend MIBF by incorporating rotation information along all three axes. Using beampattern analysis, we show theoretically that head rotations are useful for resolving CoC ambiguities. In particular, yaw motions are suitable for resolving front-back ambiguities and roll motions are suitable for resolving up-down ambiguities. With experiments on real recordings, we confirm the results of the beampattern analysis and show that head movements improve the localization accuracy.

References

[1] Mehdi Zohourian, Gerald Enzner, and Rainer Martin, "Binaural speaker localization integrated into an adaptive beamformer for hearing aids," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 515–528, Mar. 2018.

[2] Sebastian Nagel, *Interactive Reproduction of Binaurally Recorded Signals*, vol. 6 of *Aachen Series on Communication Systems*, Shaker Verlag, 2025.

[3] Erik Fleischhauer, Sebastian Nagel, Abisman Balachanthiran, and Peter Jax, "On the use of dereverberation algorithms in binaural cue adaptation," in *German Annual Conference on Acoustics*, Hannover, Germany, Apr. 2024, pp. 1600–1603.

[4] Jens Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, 1997.

[5] Hans Wallach, "The role of head movements and vestibular and visual cues in sound localization.," *Journal of Experimental Psychology*, vol. 27, no. 4, pp. 339–368, Oct. 1940.

[6] Frederic L. Wightman and Doris J. Kistler, "Monaural sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 101, no. 2, pp. 1050–1063, Feb. 1997.

[7] Frederic L. Wightman and Doris J. Kistler, "Resolution of front–back ambiguity in spatial hearing by listener and source movement," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2841–2853, May 1999.

[8] Jianliang Jiang, Bosun Xie, Haiming Mai, Lulu Liu, Kailing Yi, and Chengyun Zhang, "The role of dynamic cue in auditory vertical localisation," *Applied Acoustics*, vol. 146, pp. 398–408, Mar. 2019.

[9] Glen McLachlan, Piotr Majdak, Jonas Reijniers, Michael Mihocic, and Herbert Peremans, "Dynamic spectral cues do not affect human sound localization during small head movements," *Frontiers in Neuroscience*, vol. 17, Feb. 2023.

[10] Vladimir Tourbabin and Boaz Rafaely, "Direction of arrival estimation using microphone array processing for moving humanoid robots," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 2046–2058, Nov. 2015.

[11] Yan-Chen Lu and Martin Cooke, "Motion strategies for binaural localisation of speech sources in azimuth and distance by artificial listeners," *Speech Communication*, vol. 53, no. 5, pp. 622–642, May 2011.

[12] Ning Ma, Tobias May, and Guy J. Brown, "Exploiting deep neural networks and head movements for robust binaural localization of multiple sources in reverberant environments," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2444–2453, Dec. 2017.

[13] Daniel A. Krause, Guillermo García-Barrios, Archontis Politis, and Annamaria Mesaros, "Binaural sound source distance estimation and localization for a moving listener," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 996–1011, Dec. 2023.

[14] Erik Fleischhauer, Sebastian Nagel, and Peter Jax, "Binaural direction-of-arrival estimation incorporating head movement information," in *Proceedings of International Workshop on Acoustic Signal Enhancement*, Aalborg, Denmark, Sept. 2024, pp. 1–5.

[15] Glen McLachlan, Piotr Majdak, Jonas Reijniers, and Herbert Peremans, "Towards modelling active sound localisation based on bayesian inference in a static environment," *Acta Acustica*, vol. 5, pp. 45, Oct. 2021.

[16] Alfredo Cigada, Massimiliano Lurati, Francesco Ripamonti, and Marcello Vanali, "Moving microphone arrays to reduce spatial aliasing in the beamforming technique: Theoretical background and numerical investigation," *The Journal of the Acoustical Society of America*, vol. 124, no. 6, pp. 3648–3658, Dec. 2008.

[17] Arthur Cayley, "On the application of quaternions to the theory of rotation," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 33, no. 221, pp. 196–200, 1849.

[18] Jont B. Allen, David A. Berkley, and Jens Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *The Journal of the Acoustical Society of America*, vol. 62, no. 4, pp. 912–915, Oct. 1977.

[19] Peter Vary and Rainer Martin, *Digital Speech Transmission and Enhancement*, John Wiley & Sons, second edition, 2024.

[20] David Romblom and Hélène Bahu, "A revision and objective evaluation of the 1-pole 1-zero spherical head shadowing filter," in *Proceedings of AES International Conference on Audio for Virtual and Augmented Reality*, Redmond, WA, USA, Aug. 2018.

[21] Junichi Yamagishi, Christophe Veaux, and Kirsten MacDonald, "CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit (version 0.92)," Nov. 2019.

[22] Ronald Crochiere, "A weighted overlap-add method of short-time fourier analysis/synthesis," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 99–102, Feb. 1980.