PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Binaural Modeling for Complex Environments

Jonas BRAASCH[(1)], Jens BLAUERT[(2)]

[(1)]Rensselaer Polytechnic Institute, USA, braasj@rpi.edu
[(2)]Ruhr-University Bochum, Germany, jens.blauert@rub.de

Abstract
Functional binaural models have been used since the 20th century to simulate basic listening scenarios. The objective of this paper is to apply the capabilities of a cross-correlation model that processes reflections and utilizes head movements to demonstrate human listening in complex scenarios. For this study, a ray tracing model is introduced to simulate an office environment. This article highlights how the auditory system reads and understands spatial properties of a complex environment. As a case of application, the model is used to simulate binaural listening during a walk in a simulated office suite environment.
Keywords: Binaural hearing, precedence effect, head movements, binaural activity maps

1 INTRODUCTION

The human auditory system is remarkably robust with regard to extracting acoustic information from the environment. Its ability to localize sound sources in reverberant spaces is often taken as an example for highlighting this fact. It is well known that the auditory system operates in reverberant environments by means of suppressing the directional information of the sound source's reflections. This effect is called *localization dominance*. However, localization dominance is only observed when the reflections arrive within a certain time span after the primary sound, this interval is called inter-stimulus interval (ISI). When the ISI for a single discrete reflection is shorter than 1 ms, the position of the auditory event is determined by both the primary source and its reflections. This effect is referred to as *summing localization*. When the time interval is larger than the so-called echo threshold – typically between 4.5 ms and 80 ms, depending on the featured signals (e.g., signal duration and signal bandwidth; Blauert & Cobben, 1978) – the reflections are perceived as one or more separate auditory events (echoes). The term "precedence effect" is the overarching term for all three effects: summing localization, localization dominance, and the perceptual disintegration of direct sound and echo. Extended reviews on the precedence effect were written by [1], [12], [2].

In most investigations on the precedence effect, a single reflection and the direct sound source were used as stimuli. In this case, the direct sound source is often referred to as the *lead*, while the reflection is termed the *lag*. In early investigations, natural sounds were employed, e.g., [10], [14]), while in most recent investigations click pairs or trains of clicks are utilized. In 1986, Lindemann proposed a model for simulating the effect of localization dominance using contralateral inhibition [11]. Later, the current authors developed a model algorithm, based on Lindemann's previous work, to simulate the precedence effect for ongoing sound sources [5]. The model analyzed ITDs and ILDs by using separate processors – in contrary to the original Lindemann model.

While most Precedence effect models aim to suppress the information on room reflections, a model that was recently developed separates the information related to the direct sound and reflections [4]. This way it is possible to make assumptions about the sonic environment the source is presented in. The Binaurally Integrated Cross-correlation/Auto-correlation Mechanism (BICAM) uses the following approach to establish a Binaural Activity Map that contains both information on the direct sound and specular reflections. In this paper, we describe how the BICAM model was used within the EU project TWO!EARS (www.twoears.eu) and how the model was designed to operate in a rescue scenario under realistic complex, reverberant conditions. The model is a component of a larger system that is described in [7]. Some of the challenges were to localize a sound source correctly in the front or rear hemisphere and to localize an occluded sound source. We describe the model

performance for an office suite, a scenario we developed for the TWO!EARS project using a ray tracing algorithm. The BICAM model is used to explore this office suite virtually by rendering the direct and reflected sound sources with head-related transfer functions (HRTFs). In the next section, we will summarize the model algorithms used for this study and continue with the analysis of the sound presented in the virtual office suite in the following section.

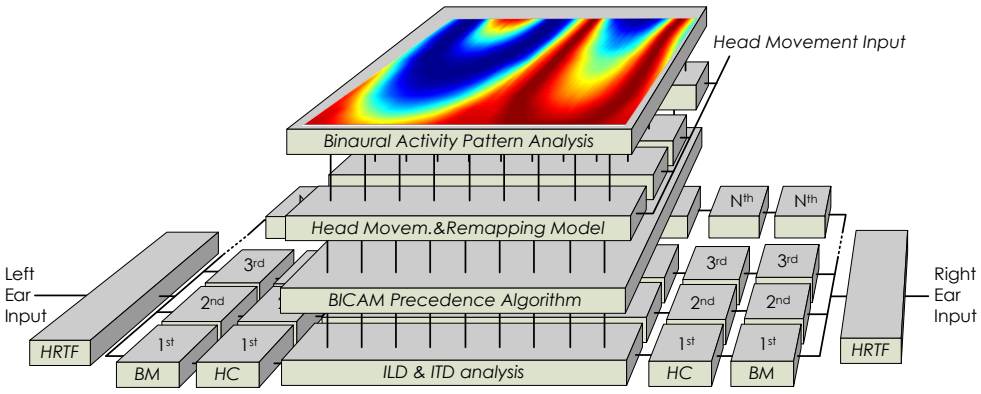

Figure 1. Binaural Model Architecture to analyze sounds in reverberant environments including stages to simulate the auditory periphery, the BICAM algorithm to process room reflections, a head-movement compensation algorithm to compute binaural activity maps. The auditory periphery consists of simulated basilar membranes (BM) with a gammatone filter banks, and a half-wave rectifier to simulate the hair-cell (HC) behavior.

2 BINAURAL MODEL

2.1 BICAM MODEL

The Binaurally Integrated Cross-correlation/Auto-correlation Mechanism (BICAM) uses a binaural signal to robustly localize a sound source in the presence of multiple reflections [4]. The model also extracts the delays (compared to the direct sound source) and lateral positions of each of the distinct reflections. The core algorithm resembles a dual-layer spatiotemporal filter to separate auditory features for the direct and reverberant segments of a windowed signal to localize the direct sound source. Using head-related transfer functions to spatialize the sound sources, the model can accurately localize a signal in the presence of two or more early side reflections and late reverberation.

At the first stage, the model separates the incoming binaural signal into auditory bands. Next, the model performs a set of auto-/cross-correlation analyses of the left and right ear signals within each auditory band. A 2nd-layer cross-correlation algorithm is then performed on top of the combined autocorrelation/cross-correlation algorithm. The underlying goal was to develop a method that incorporates the causality of the direct sound and its reflections. The conventional cross-correlation method does not reflect the temporal order of the incoming direct sound/reflections, whereas the human auditory system takes this into account, as demonstrated by the precedence effect. The second-layer crosscorrelation analysis if performed over the autocorrelation signal, R_{xx}, in one-channel and the 1st-layer cross-correlation signal, R_{xy}, in the second channel. The model compares the side peaks of both frequency-integrated functions (autocorrelation function and cross-correlation function). The side peaks for the left and right channels are correlated with each other and, by aligning them in time, the temporal offset between both main peaks can be used to determine the interaural time difference (ITD). In this way, the ITD of the direct sound can be found.

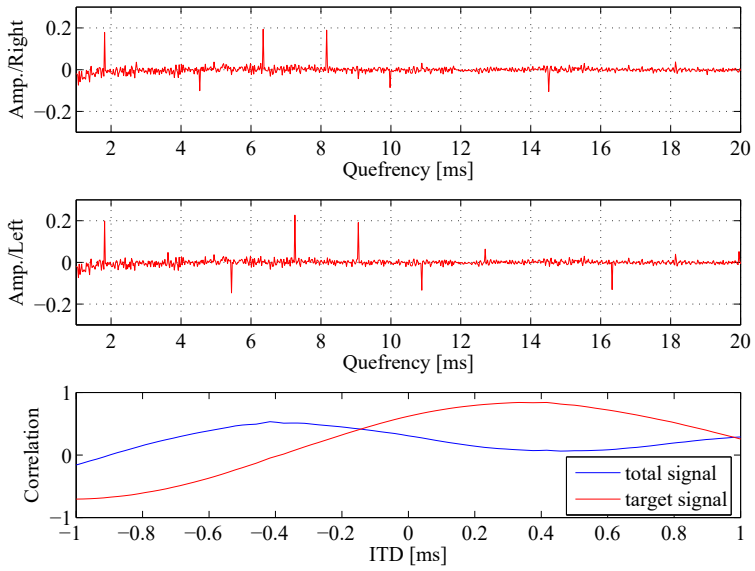

Figure 2. Cepstrum-based localization model. Top: Cepstrum for the left ear signal, Center: Cepstrum for the right ear signal, Bottom: Cross-correlation function for the direct signal (red) in comparison to the cross-correlation function for the total signal (blue)

Two autocorrelation functions, one for the left and the other one for the right ear, are aligned such the left and right main peaks match the ITD of the direct sound. The method of how to extract the ITD for the direct signal is described in detail in [4]. An alternative way to calculate the interaural time difference (ITD) for the direct signal is by estimating deconvolution filters from the cepstra of the left and right signals of a binaural signal. Figure 2 depicts this approach for a speech signal with a direct signal (0.4-ms ITD) and two reflections (−0.4-ms ITD each, delays of 7 and 9 ms with respect to the direct signal, each reflection has the same amplitude as the direct signal). The top graph shows the cepstrum for the left signal, the center graph the cepstrum for the right signal. The cepstra are used as deconvolution filters for the left and right signals before the cross-correlation function is obtained. The bottom graph shows that this method allows us to calculate the lateral position of the target sound correctly (red curve). The blue curve shows the cross-correlation function for the total signal which has its maximum near −0.4-ms because the combined amplitude of the reflections is twice as high as the amplitude of the direct sound.

Once the two autocorrelation functions are aligned against each other using the ITD of the direct signal, a running cross-correlation function is computed between the right side peaks of both autocorrelation functions including the realigned center peaks. This running cross-correlation function becomes a binaural activity map when plotted as a 3D plot The binaural activity maps show the zones of energy concentration as a function of delay and sideness – the latter in form of an interaural time difference. The direct signal is located at a delay near zero, early reflections appear at higher delays. Characteristic for the BICAM model is that the algorithm ignores diffuse reverberation. However, it can be assumed that human listeners have problems, too, representing the temporal order of diffuse reflections – see [13].

2.2 HEAD-MOVEMENT ALGORITHM

The BICAM model can be combined with a head-movement algorithm to compute binaural activity maps. The head-movement algorithm, which is described in detail in [3] can be used to extend current cross-correlation models, including the BICAM approach, to compensate for head movements. The algorithm tracks sound

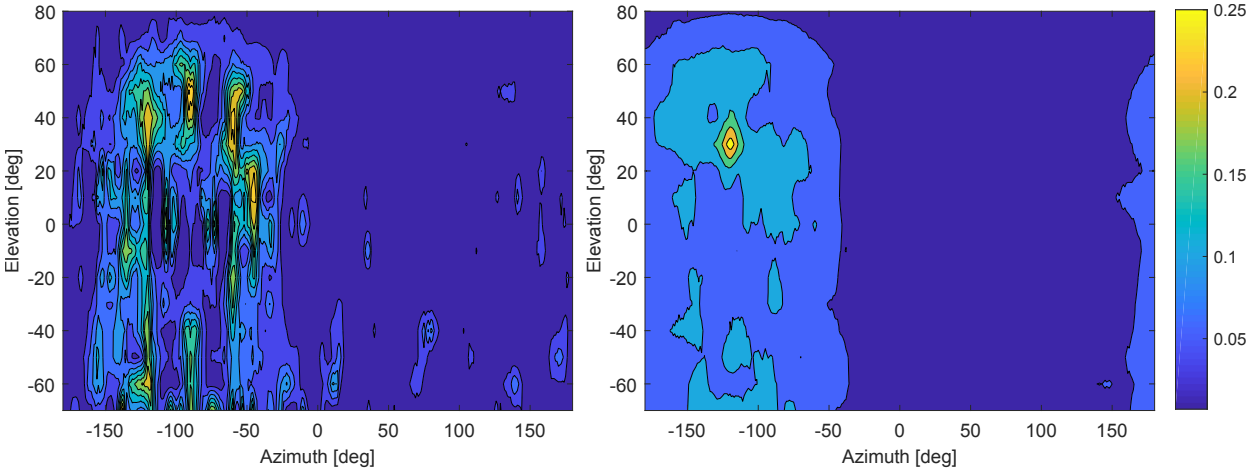

Figure 3. Azimuth/elevation maps for the head-movment model analysis for a sound source located at -120° azimuth/30° elevation. The left graph shows the results for the starting position before head movement, the right graphs shows the results after head movement with head movement compensation.

sources in the head-related coordinate system (HRCS) as well as in the room-related coordinate system (RRCS). It is also aware of the current head position within the room. The sounds are positioned in space using an HRTF catalog at 1° azimuthal resolution. The position of the sound source is determined through the inter-aural cross-correlation (IACC) functions across several auditory bands, which are mapped to functions of azimuth and superposed. The maxima of the cross-correlation functions determine the position of the sound source, but unfortunately, usually, two peaks occur – one at or near the correct location and the second one at the front/back reversed position. When the model is programmed to virtually turn its head, the degree-based cross-correlation functions are shifted with the current head angle to match the RRCS. During this procedure, the IACC peak for the correct hemisphere will prevail if integrated over time for the duration of the head movement, whereas the front/back reversed peak will average out.

Using this approach it is also possible to estimate the elevation of a sound source. To achieve this, the functions that describe the azimuth-degree vs. ITD of maximum peak relationships need to be extended to 2-dimensional maps, where the azimuth-degree vs. ITD of maximum peak relationships are contained for individual elevation bands as shown in Fig. 3. For each elevation angle, an individual function to establish the relationship between azimuth and ITD at cross-correlation peak positions is computed. The method is identical to the one described in [3] but computed at multiple elevation angles in steps of 10 deg. Now, when the model moves its head virtually, the ITDs change differently over angle – a form of motion parallax. The greatest ITD changes occur in the median plane, where ITDs of up to 1 ms can be found for angles of $\pm 90^\circ$. The smallest ITD changes of zero are found at the poles of $\pm 90^\circ$ elevation. Because of this effect, only the peaks at the actual sound-source elevation remain constant. All other peaks wander with head movement and average out – see Fig. 3, right. Here, only the actual sound-source location prevails, whereas, in the non-movement case, all solutions on the cone of confusion are indicated.

3 OFFICE SUITE WALK-THROUGH SIMULATION

3.1 RAYTRACING SIMULATION

To test the binaural model in a real-world scenario, a ray-tracing model was implemented to generate binaural impulse responses for the model analysis. A geometrical model was defined as shown in Fig. 5, based on

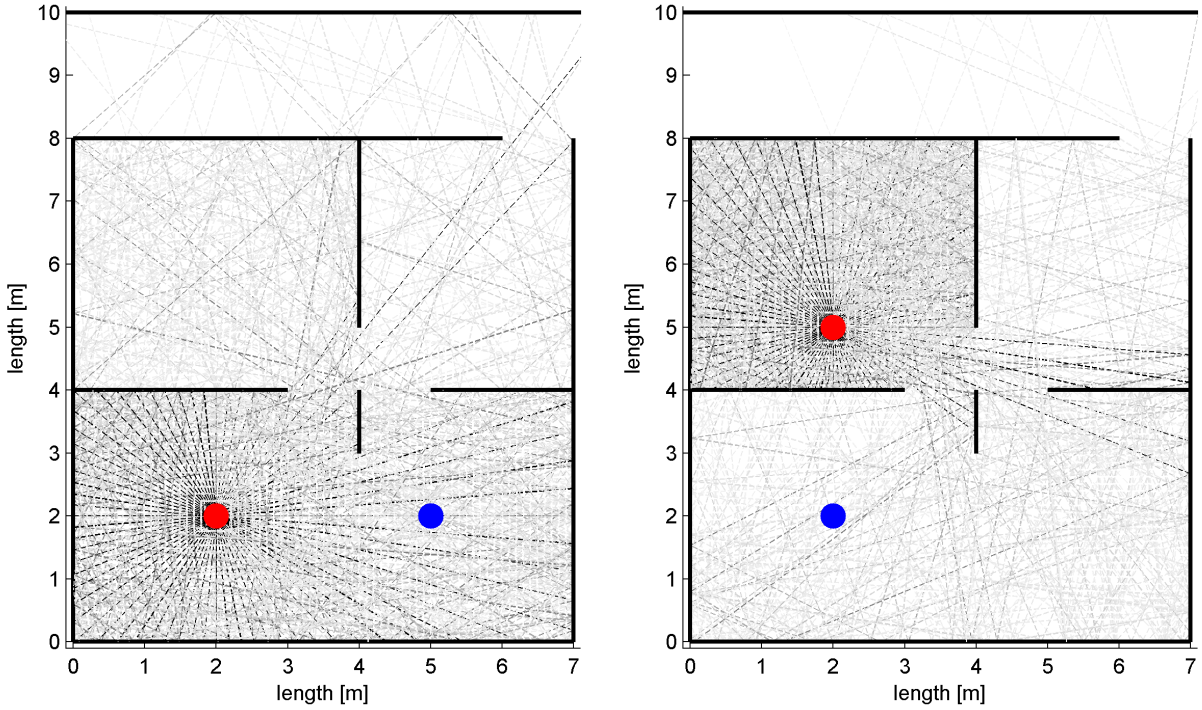

Figure 4. Ray Tracing Simulation in a computer-generated office suite, left: Scenario 1 with a non-included sound source, right: Scenario 2 with an occluded sound source. Sound sources are depicted as a red dot, binaural receivers as a blue dot. The gray level of the rays lighten with decreasing distance and amplitude.

sound-reflecting walls, a source (red dot), and a receiver (blue dot). A set of rays is sent out from the sound source in every direction within the horizontal plane at equiangular distances of 5°. Each ray is then traced, and every time a ray meets a wall it is reflected back using Snell's law considering that the outgoing angle equals the incoming angle. The ray is traced until the 20th reflection occurs unless the ray exits the geometrical model. At every reflection, the sound level is attenuated by 2 dB across frequency to simulate acoustic wall absorption. The sound intensity is also attenuated over distance based on the inverse square law, assuming the sound source to be of omnidirectional character. The collection of rays is shown in Fig. 4 as gray lines such that the rays become lighter in color with distance and decreasing sound pressure.

All rays are then collected at the receiver position assuming a spatial window 0.6 m wide. Each calculated ray is tested if it intersects the spatial window at the receiver position. For each intersecting ray, it is then calculated how far it traveled from the source position to the receiver position, at which azimuth angle it arrives at the receiver position and how much times it had been reflected (reflection order). Based on these data a binaural room impulse response is calculated in which a left/right HRTF pair is inserted at the correct delay and a head orientation-based direction-of-arrival angle. In each case the direct-sound source and the reflections were spatialized using head-related transfer functions (HRTFs) from the MIT KEMAR database [9]. Each HRTF pair is calibrated to the amplitude the ray should have based on distance traveled and the number of wall reflections. In addition, a late reverberation tail is generated at a constant level (assuming a statistically evenly distributed diffuse reverberation field) using an exponentially decaying Gaussian noise burst adjusted to a reverberation time of 0.7 seconds. At the position shown in the left graph of Figure 4, the diffuse reverberation level was about -10 dB lower than the combined level of the direct sound and the early reflections.

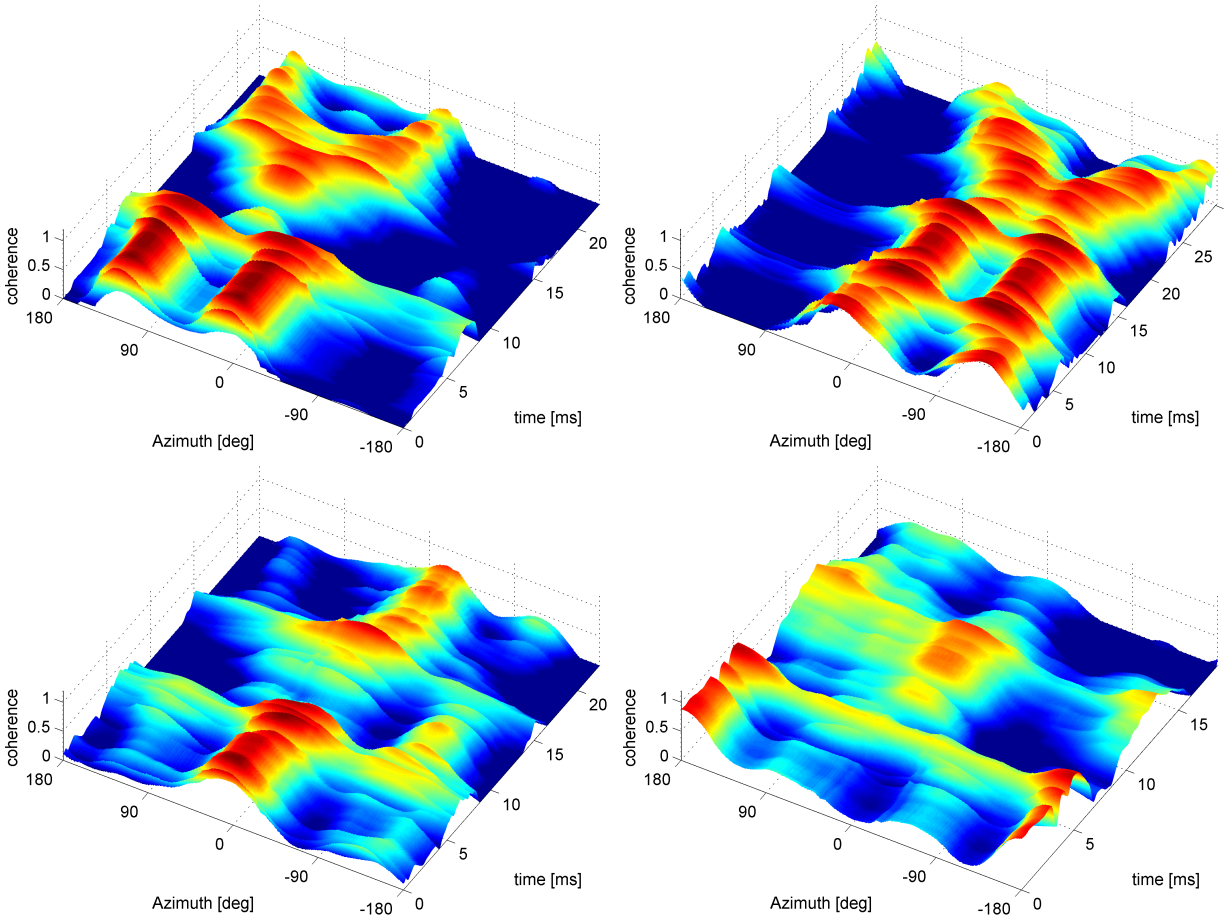

Figure 5. Binaural-activity-map results for the BICAM model analysis utilizing head movements. The top-left graph shows the results for Scenario 1 (Fig. 4) with the sound source pointing 30° left to the sound source (including 30° head-movement compensation), the top-right graph shows the same condition but for the receiver pointing 30° to the right. The bottom-left graph shows the combined analysis to remove front/back confusions for a receiver pointing in the direction of the sound source (0°). The bottom-right condition shows the same condition, but for a receiver pointing away from the sound source (180°).

3.2 BINAURAL ANALYSIS

The results are analyzed using the BICAM precedence effect model [4] and a male speech sample from the Archimedes CD [**?**]. The top-left graph Fig. 5 depicts the scenario in which the virtual head of the model is turned 30° away from the sound source based on the scenario shown in the left graph of Fig. 4. Note that the data is presented in a room coordinate system that faces the sound source directly. As can be easily seen, each time slice shows two ambiguous peaks. One for the front and one for the corresponding rear direction, a common problem that was discussed in detail in [6]. In order to resolve the ambiguous peaks, the virtual head of the model is shifted by 60° to the opposite side – see the top-right graph of Fig. 5, which also displays the data in a room coordinate system. Now we simply take the average of the two binaural activity map and the ambiguous front/back confusion peaks average out – see the bottom-left graph of the Fig. 5. To demonstrate the effectiveness of the algorithm, the same scenario was simulated again, but this time with the

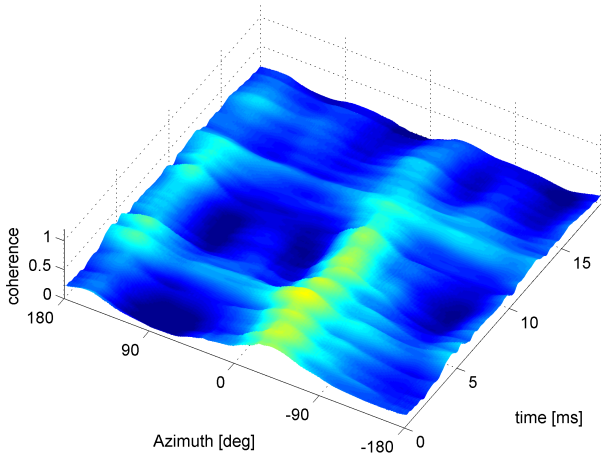

Figure 6. Binaural-activity-map results for the BICAM model analysis utilizing head movements for an occluded direct sound source – simulating Scenario 2 depicted in the right graph of Fig. 4.

virtual head facing the rear at 180° with temporal head-movement shifts to 150° and 210° to resolve front/back directions – see bottom-right graph of Fig. 5. It should be noted that there are two main differences between the model presented here and the model of [6]. Firstly, in the new model, the head-movement algorithm is now applied to the estimated binaural activity map and not to the binaural signal itself, which leads to two advantages: (i) the direct sound source angle can be computed separately from the early reflections which yields in higher localization accuracy, and (ii) the algorithm can also estimate the front/back direction of the reflections. Secondly, the new model cannot yet calculate front/back directions from a continuously turning head like it is the case for the [6] model. The reason for this is that the time alignment method for the two autocorrelation functions currently requires a stable head orientation. Therefore, the new model calculates the front/back directions based on two distinct head positions until a better solution is found for the time alignment method.

3.3 CASE OF AN OCCLUDED SOUND SOURCE

The analysis is concluded by computing a scenario, in which the direct pathway between the source and the received is occluded by a wall as shown in the right graph of Fig. 4. Figure 6 shows the binaural activity map for this case. While there is a distinct peak visible, the maximum correlation of 0.6 is much lower than was the case for the first scenario which yielded a maximum correlation of 0.9. Note, that the binaural activity map was determined based on the average BICAM analysis of 10 segments. Each segment by itself leads to a maximum coherence of one because the autocorrelation peaks have a main peak of one. However, in the occluded case, the outcome of the analysis is heavily influenced by the diffuse reverberant signal component and the main peak averages out, since its lateral position moves from segment to segment. In the case of Scenario 1, the binaural activity map is stable from segment to segment and hardly influenced by the time-averaging method.

4 CONCLUSION & OUTLOOK

In this paper, we used an advanced binaural model to localize a sound source in a complex office environment. The model is robust to room reverberation and can distinguish front/back locations by means of head rotations. The algorithm is also able to localize sound sources vertically. In the future, we plan to extend the precedence effect algorithm so it can perform dynamic head movements rather than taking multiple looks. We also plan to

integrate an existing source-separation model that is based on an equalization/cancelation algorithm [8] into the model that was presented in this paper.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant Nos. IIS-1320059 and BCS-1539276, the European Research Council Project FP7-ICT-2013-C-#618075, www.twoears.eu), and the Cognitive and Immersive Systems Laboratory.

REFERENCES

[1] J. Blauert. *Spatial Hearing*. MIT Press, Cambridge, 1997.

[2] J. Blauert and J. Braasch. Acoustical communication: The precedence effect. In *Proceedings Forum Acusticum*, pages 15–19, Budapest, Hungary, 2005.

[3] J. Braasch. A precedence effect model to simulate localization dominance using an adaptive, stimulus parameter-based inhibition process. *J. Acoust. Soc. Am.*, 134(1):420—435, 2013.

[4] J. Braasch. Sound localization in the presence of multiple reflections using a binaurally integrated cross-correlation/auto-correlation mechanism. *J. Acoust. Soc. Am.*, 140(1):EL143–EL148, 2016.

[5] J. Braasch and J. Blauert. The precedence effect for noise bursts of different bandwidths. II. Comparison of model algorithms. *Acoust. Sci. & Tech.*, 24:293–303, 2003.

[6] J. Braasch, S. Clapp, P. T. Parks, A., and N. Xiang. *The Technology of Binaural Listening Application of Models of Binaural Listening Binaural Listening in Technology*, chapter A binaural model that analyses aural spaces and stereophonic reproduction systems by utilizing head movements, pages 201–223. Springer, Berlin, Heidelberg, New York, 2013.

[7] B. Cohen-L'hyver, C. R. Kim, H. Wierstorf, Y. Kashef, J. Blauert, J. Mohr, J. Braasch, N. Ma, S. Argentieri, T. Walther, G. Bustamante, P. Danès, T. Forgue, A. Podlubne, T. May, and Y. Guo. Fp7-ict-2013-c two!ears project 618075, Final integration-&-evaluation report, Deliverable D4.3. Technical report, European Research Council (ERC), 11 2016.

[8] N. Deshpande and J. Braasch. Blind localization and segregation of two sources including a binaural head movement model. *J. Acoust. Soc. Am.*, 142(1):EL113–EL117, 2017.

[9] W. Gardner and K. Martin. HRTF measurements of a KEMAR. *J. Acoust. Soc. Am.*, 97(6):3907–3908, 1995.

[10] H. Haas. Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache. *Acustica*, 1:49–58, 1951.

[11] W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front. *J. Acoust. Soc. Am.*, 80:1623–1630, 1986.

[12] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman. The precedence effect. *J. Acoust. Soc. Am.*, 106:1633–1654, 1999.

[13] E. Teret, M. T. Pastore, and J. Braasch. The influence of signal type on perceived reverberance. *J. Acoust. Soc. Am.*, 141(3):1675–1682, 2017.

[14] H. Wallach, E. B. Newman, and M. R. Rosenzweig. The precedence effect in sound localization. *Amer. J. Psychol.*, 57:315–336, 1949.