

Binaural Modeling from an Evolving-Habitat Perspective



Jonas Braasch

Abstract Functional binaural models have been used since the mid-20th century to simulate laboratory experiments. The goal of this chapter is to extend the capabilities of a cross-correlation model so it can demonstrate human listening in complex scenarios found in nature and human-built environments. A ray-tracing model is introduced that simulates a number of environments for this study. This chapter discusses how the auditory system is used to read and understand the environment and how tasks that require binaural hearing may have evolved throughout human history. As use cases, sound localization in a forest is examined, as well as the binaural analysis of spatially diffuse and rectangular rooms. The model is also used to simulate binaural hearing during a walk-through a simulated office-suite environment.

1 Introduction

The goal of this chapter is to examine binaural models from an evolving-habitat perspective. While the evolution of the auditory system has been studied extensively from a phylogenetic perspective to establish knowledge of how the auditory system developed anatomically over time, the auditory system's ability to adapt to changing habitats over tens of thousands of years has been hardly investigated. Since it is impossible to travel back in time, the topic cannot be studied directly. This chapter describes an attempt of an initial study examining this by simulating different environments with a ray tracing model and using an extended binaural model for an auditory-specific analysis. When studying how the auditory system can adapt to different habitats, one must keep in mind that the anatomical changes of the auditory system took place over millions of years. While the structure of the auditory system continues to change over time, these changes occur at a much slower pace than most sociological changes. Thus, it can be assumed that our auditory system is basically

J. Braasch (✉)
School of Architecture, Rensselaer Polytechnic Institute,
Troy, NY 12180, USA
e-mail: braasj@rpi.edu

© Springer Nature Switzerland AG 2020
J. Blauert and J. Braasch (eds.), *The Technology of Binaural Understanding*,
Modern Acoustics and Signal Processing,
https://doi.org/10.1007/978-3-030-00386-9_10

structured the same way that it was during the beginning of modern civilization, which started about 50,000 years ago (Peck 1994). Since the neurological structure of the brain is very flexible, mammals and other organisms can easily adapt to new environments and situations—especially during the early post-natal phase (Peck 1995). This flexibility allows us to adjust to new sonic environments. Humans can, for example, comprehend and appreciate classical music in modern concert halls using an auditory system that primarily developed in natural habitats.

Traditionally, binaural models have been designed to simulate laboratory scenarios, for example, to predict the lateral position of a binaural stimulus presented over headphones. In this chapter, it is attempted to extend this knowledge for better understanding and predicting binaural-hearing tasks in natural environments and other complex situations that arose as civilization evolved and the built environment changed. Also, a bridge will be created to the robotic community, which has its own distinct way of designing sound-sensing systems. Experts in robotics often attempt to solve tasks in complex environments, for example, acoustically navigating systems, but without the desire to understand how these tasks are accomplished in biological systems. In the context of this chapter, understanding will be defined as *the ability to make judgments from perceived information*. In some cases, the understanding consists of the ability to accurately decode the intended meaning sent by a communication partner, for example, a conversation partner, or the ability to interpret unintended cues—such as the sounds of an approaching predator. In any case, understanding allows us to infer something from the received acoustic signals, and these signals then become information.

Evolutionary biologists agree that a biological organism needs to be successful in this behavioral complex of four tasks to survive as a species, namely, (i), *feeding*, (ii), *fleeing*, (iii), *fighting*, and (iv), *flirting (reproduction)*¹—compare, for instance, Graham (2014). Spatial awareness is essential to success in all these goals—to find food, avoid predators and to communicate with tribe members for various reasons ranging from cooperation to mating. As a starting point for the binaural analysis, the need for spatial acoustic communication and sensing will now be examined in the view of the main four survival tasks mentioned above.

1.1 Feeding

The early *Homo Sapiens* survived mainly as hunters and gatherers. Unlike other vertebrates, such as barn owls or bats, who find their prey acoustically, humans localize prey or gather objects using vision as their primary sense. Consequently, the acoustic-localization performance does not need to be as accurate as is the case for acoustical hunters, who must target their prey precisely. Most likely, acoustic communication between tribe members played a big role when hunting animals, for example, when engaging in an attack. Studies have found evidence that the early homo

¹Also known as the four F's: feeding, fleeing, fighting, fornication.

sapiens lived in the plains and hunted large animals from a distance using spears and other long-distance weapons—Villa and Soriano (2010). The ability to follow these hunting patterns was a direct result of the *Cognitive Revolution*—Wynn and Coolidge (2004, 2008), Coolidge and Wynn (2018). The rise of new cognitive abilities enabled homo sapiens to plan ahead, conduct better-coordinated group hunting and also to spatially navigate larger terrains—Baril (2012).

In contrast, *Homo Neanderthalensis* is believed to have been a hunter who killed animals in close combat, based on the type of spear perforations found in deer skeletons—Gaudzinski-Windheuser et al. (2018) and other evidence. In this context, it is noteworthy that homo sapiens has a voice box that is very different from that of *Homo Neanderthalensis* and other early human species. This results in a lower fundamental pitch (Fitch 2000). The need for lowering the voice could have resulted from the need to communicate acoustically over larger distances in the plain as the homo sapiens started to specialize in hunting animals from a distance. This would also explain why the frequency range is lower than it is the case of many other mammals. The first mammals were small nocturnal animals who presumably lived in densely vegetated areas—Gerkema et al. (2013). Their high-frequency hearing range is optimal for localizing potential predators at a close distance (Joris and Trussell 2018). However, high frequencies are not optimal for localizing sound sources from a larger distance because of air absorption (dissipation), which increases with frequency.

1.2 Fleeing

In contrast to when hunting prey, the angular localization accuracy is not that critical when fleeing from predators because one usually runs away from them. However, it is essential to detect the predators early on before they pose an imminent danger. Auditory cues always become predominant when visual cues are not available. This is the case when it is too dark to see, visual objects are occluded, or the acoustic sources are outside the visual field. The auditory sense also monitors the environment during sleep, and it can be shown that children are not yet disturbed by sounds at night (Busby et al. 1994). One explanation for this observation is that, from an evolutionary perspective, it is better for children at certain ages to have an undisturbed sleep and rely on their parents for monitoring than to monitor the environment themselves. A study on detecting fire alarms revealed that the sleep of children is often so deep that children do not wake up when the alarm is set off (Bruck 1999).

In discussions it is usually emphasized that the ability to detect signals to monitor predators is of particular relevance. However, an absence of sound can be equally important, because other animals will quiet down in a sector from where a predator is approaching. This might be one of the reasons for enjoying immersive sounds, namely, that they can serve as an inherent indicator that no predator is approaching.² While this hypothesis remains to be proven, several studies have shown that human subjects

²Personal communication with David Mountain, Boston University, April 5, 2013.

do not feel comfortable when performing tasks in extremely quiet environments (Volf 2012; DeLoach et al. 2015). Alternative strategies are pursued where fleeing is not an option. The females of the indigenous *BaYaka* group, for example, gather together and shouted group calls to make themselves appear as a large and well-coordinated group to scare away predators in lieu of fleeing (Knight and Lewis 2017, p. 442).

1.3 Fighting

Combat between humans is as nearly as old as the homo sapiens (Meller and Schefzik 2015; Ferrill 2018), and the remains of the oldest homicide victim found are 430,000 years old (Sala et al. 2015). The acoustic requirements for human combat against other humans or predators are very similar to the acoustic communication in hunting situations discussed above. Shouting calls are essential to coordinate attacks and warn others from counter attacks, demanding excellent spatial-hearing skills.

1.4 Reproduction

Many animals mostly rely on acoustic signaling and sound localization to find mating partners. Although many animals shout out mating calls over long distances, it is more likely human courting has always been an intimate social interaction, since humans always lived together in groups. It is widely believed that the fundamental pitch differences in human female and male voices have evolved to make each other more attractive to the opposite sex (Jones et al. 2010). In the context of spatial hearing, it is interesting though that homo sapiens engaged in artistic activities from early on that appear to address both erotic and musical desires. For example, in the Hohle-Fels cave, bone flutes were found next to venus figurines and phallus sculptures (Conard and Wolf 2014; Conard et al. 2009). Since then, the flute has been a typical courting instrument in indigenous cultures, see for example Conlon (2004). Adjacent to the inhabited part of the Hohle-Fels cave, where the bone flute was excavated, a much larger cavern exists with a reverberation time of about 2 s. It is not hard to imagine that our ancestors would have played the flutes in this larger cavern to enjoy the acoustics. At least, it is known that early humans were very aware of their acoustic environment. For instance, Reznikoff found that many prehistoric cave drawings were painted at places with dominant acoustic resonances (Reznikoff 2004/2005).

1.5 Modality and Bandwidth

In order to understand how our auditory system developed and was utilized, one needs to examine what cues and mechanisms are available to process these cues.

This way, situations can be determined in which acoustic cues supersede other cues. Of the five major senses, touch, taste, smell, vision, and audition, only the last three are useful to sense objects from a distance. Our olfactory sense is less developed than that of other species, including dogs. Human beings only have a directional sense of smell when they are moving, and even though it can be useful to detect the presence of a predator or food, this sense lacks the spatial precision of the auditory and visual systems. The early homo sapiens was primarily a diurnal hunter using the visual sense to hunt animals from a distance with spears and to collect food from plants. Most likely, the auditory sense had initially a support role until speech communication became increasingly important. The visual sense is limited to the binocular visual field and covers only about 214° in the horizontal plane (Rönne 1915). In contrast, the auditory sense is not spatially restricted, and it also helps us monitor our environment at night. Sound localization has been important to detect the direction of a predator quickly. It can also be assumed that it was important for our ancestors to localize each others' voice commands when tracking prey. Sound localization has also always been important in environments where vision is partially obstructed, for example when hunting deer in a dense forest (Gaudzinski-Windheuser et al. 2018; Roebroeks et al. 1992). Our first example deals with such a situation. A virtual walk-through in a forest is presented and analyzed in the next section.

2 Simulating Sound Localization in a Forest with Partially Obstructed Sight Using a Ray-Tracing Model

2.1 Introduction

To be able to better understand how the binaural system evolved over millions of years to perform robustly in complex environments, several scenarios were developed in which auditory cues are particularly important. Obviously, sound localization is always in demand when the source is out of sight. Aside from monitoring the environment at night, forests are a good test case because trees and other vegetation typically visually obstruct objects, and some of these objects might be looking for dinner. The forest simulation was set up using a ray-tracing simulation program, which is described in the next section. Circular boundaries are used as acoustic objects to simulate the acoustic behavior of tree trunks. The forest is simulated by randomly creating circles in an area of $100 \times 100 \text{ m}^2$. The diameters of the tree trunks are set by a stochastic process.

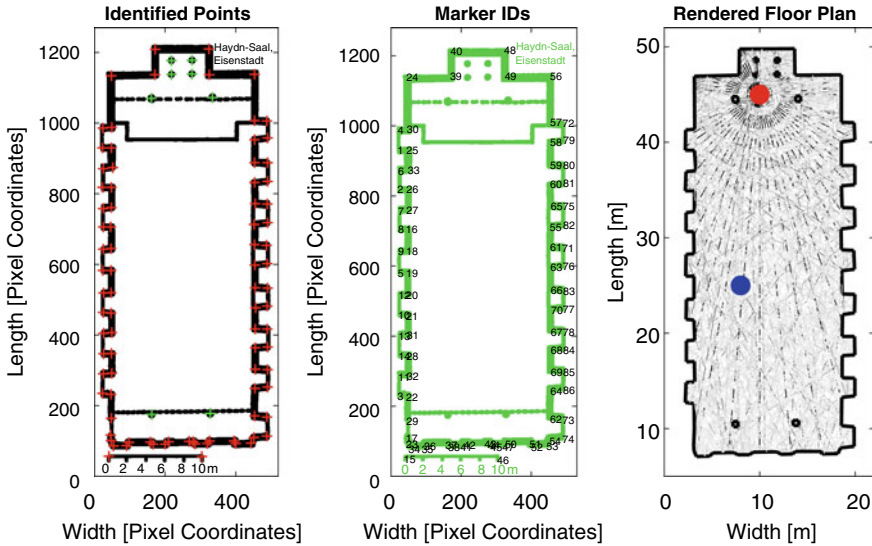


Fig. 1 Demonstration of the ray-tracing algorithm for the Haydn-Saal in Eisenstadt, Austria, where Joseph Haydn was active. **(Left)** Processed marked-up floor plan for visual inspection. **(Center)** Assignment of wall-corner identification numbers. **(Right)** Geometric model with sound source (red dot), receiver (blue dot), and calculated rays. The original floor plan was obtained from Meyer (1978, p. 147)

2.2 Creating Geometric Models

The method presented here is confined to simulations in the horizontal plane (two-dimensional rendering method) to allow fast calculations. This enables the simulation of complete walk-throughs using a batch process.³ The ray-tracing implementation was programmed in Matlab following common practice—details in Vorländer (1989), Lehnert and Blauert (1992a, b), Blauert et al. (2000). Additional features were added where needed, for example, an algorithm to create circular boundaries to simulate tree trunks and a method to generate models from floor plans rapidly.

Coordinates of acoustic boundaries can be assigned to the ray-tracing algorithm in three different ways, (i) line segments with start and end points representing walls, (ii) squared pillars with the center coordinates and the pillar width, (iii) circles represented by center and radius coordinates. Figure 1 shows an example of the ray-tracing software for a concert hall in Eisenstadt, Austria, that was recreated from a floor plan. The software can work with annotated floor plans. For this purpose floor plans are marked up within a standard bitmap editor (GIMP, Photoshop, etc.) using red dots for room corners, green dots for squared pillars, and blue dots for circles—see Fig. 1, left graph. In the next step, the program plots annotated points on top of

³Originally, the ray-tracing method was implemented to create auralizations for the horizontal array of 128-channel loudspeakers at Rensselaer's CRAIVE-Lab.

the map providing a unique index for each annotated point—see Fig. 1, center graph. The user then creates a list of how the red points connect to walls. The scale of the floor plan needs to be annotated, and the original dimensions need to be handed over to the program as well (e.g., 1 m and 10m for two scale points respectively). The program then transforms these data points to an editable list of geometrical objects that can be extended by the user, for example, by adding identifiers for wall materials.

The Ray-Tracing Algorithm

The program sends out rays from a user-specified source position. The program computes the rays for equidistant azimuths covering the full 360° angular range. Intersection points are computed for each ray and boundary object as shown in Fig. 2, left graph. For each ray, the closest boundary intersection is determined. At the intersections, the reflection angle is calculated using Snell’s law, which predicts that the reflected angle measured from the normal of a plane surface equals the incoming angle, that is, $\cos(\alpha_o) = \cos(\alpha_i)$. Consequently, the next-order ray is sent out into the new direction until the maximum order (e.g., the number of reflections) as specified by the user is reached. The outgoing rays are stored as a sequence of ray elements containing the intersection points and the boundary-material identifiers. Since the initial angles are stored with the rays, a source-specific directivity pattern can be simulated after all rays have been traced.

Creating a Binaural Room Impulse Response

Next, the rays are collected by a receiver, which can be located anywhere in the rendered room. For this purpose, a virtual circle with an adjustable diameter is posi-

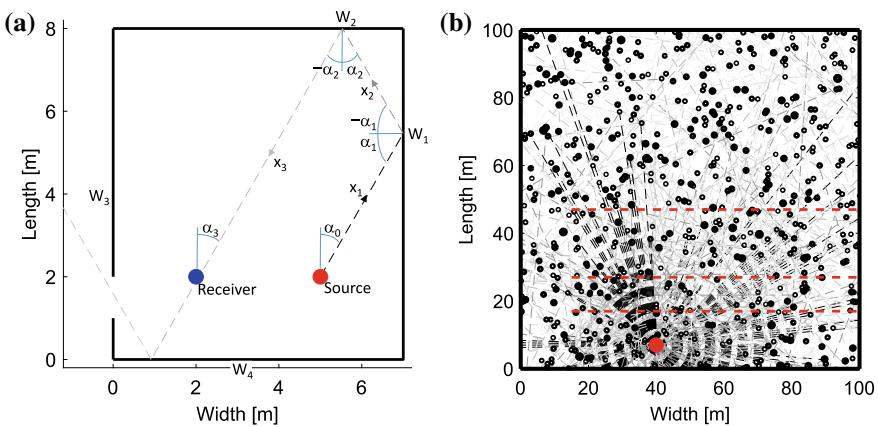


Fig. 2 (Left) Diagram to illustrate the ray-tracing method. (Right) Schematic of the simulated forest environment (top view)

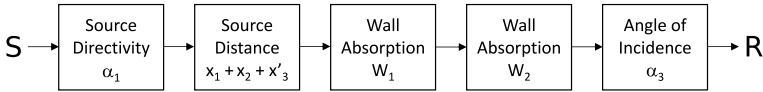


Fig. 3 Ray-tracing signal flow

tioned at the receiver location. Then, the algorithm calculates which ray elements intersect the circle, and for all positive cases, the total ray distance between sound source and receiver is calculated. All listed values are stored together with the final angle of incidence, that is, the angle that the ray was initially sent out, the reflection order, and the sequence of identifiers of the walls that the ray has hit on its way to the receiver. Based on these data, the impulse response is computed. The direct sound and the reflection are computed as delta peak at the delay that corresponds to the path length that the ray has traveled from the source to the receiver. In addition, each impulse is transformed in the following way—see also Fig. 3.

1. The magnitude of the ray is reduced based on the inverse-square law.
2. The high frequencies are filtered out based on dissipation effects in the air.
3. The absorption coefficients of the walls and other boundaries are simulated using a cascaded Finite Impulse Response (FIR) filter. These material-specific filters are chosen from a publicly available database (DIN 1968). The number of cascaded filters matches the order of the reflection.
4. In the final step, the incoming direct sound and the reflections are selected by their close proximity passing the receiver position. These are then filtered with *Head Related Transfer Functions* (HRTFs) that correspond to the closest available HRTF measurement for the direction of incidence. At this point, the room impulse response is transformed into a stereo signal. An overlap-method ensures that the delayed reflections can partially overlap.

Simulation of Late Reverberation

Late diffuse reverberation is computed in addition to the early reflections that are generated by the ray-tracing model. Since the late reverberation tail is formed by a stochastic process with an underlying Gaussian distribution, the fine structure of the simulated reverberation tail is constructed from a Gaussian noise sample. The duration of the Gaussian noise sample is adjusted to twice the value of the maximum reverberation time. Next, the noise sample is processed through a filter bank with nine adjacent octave-wide bandpass filters. An exponentially decaying time window, y_k , adjusted to the frequency-specific reverberation time, is calculated for each octave band, k :

$$y_k = e^{\frac{-t \cdot 20 \cdot \log(10)}{T_k \cdot 60}}. \quad (1)$$

with the reverberation time, T_k , in the k th frequency band, and the time, t , in seconds. Afterward, the total exponentially-decaying noise signal, x_t , is reassembled by summing up, sample-by-sample, the octave-filtered noise signals, multiplied with the exponentially decaying time window:

$$x_t = \sum_{k=1}^K x_k \cdot y_k. \quad (2)$$

This process is repeated for each channel, two for a binaural signal in each case, using independent Gaussian noise samples for each channel while keeping all other parameters constant. The frequency-specific reverberation times, T , is calculated using the Eyring formula, which is based on a three-dimensional room model that takes the room volume and the effective absorptive surface area into account, namely,

$$T_{60} = 0.161 \cdot V / (A + 4m \cdot V) \text{ s}, \quad (3)$$

whereby, A is total effective absorption, defined as the sum of all surface elements, S_k , multiplied with their specific absorption coefficients, α_k :

$$A = \left(\sum_{k=1}^K \alpha_k \cdot S_k \right). \quad (4)$$

In order to estimate the room volume, V , the area of the floor plan is calculated and then multiplied by the average room height, which has to be provided to the program. The formula for this calculation is:

$$V = A_F \cdot h. \quad (5)$$

The total effective absorption is calculated from the wall elements in the ray-tracing model, each multiplied with the average height. The frequency-specific absorption coefficients are determined with the values stored in the DIN database (DIN 1968) via the wall material identifiers. A linear onset ramp is calculated to gradually blend in the late reverberation tail with the direct sound and early reflections. The starting and end points of the ramps can be adjusted by the user.

Two methods are available to calculate the direct-to-reverberant energy ratio. The first method estimates the critical distance, the distance from the sound source at which the sound-pressure levels of the direct sound matches the sound-pressure level of the reverberant field. For an omnidirectional sound source, the critical distance can be calculated using the following equation (Kuttruff 2000, p.317),

$$r_c = 0.057 \cdot \sqrt{\frac{\gamma V}{T}}, \quad (6)$$

with the volume, V , the reverberation time, T , and the directivity coefficient, γ . In the subsequent calculation, omnidirectional sound sources are assumed to have a directivity coefficient that equals to one.

In the next step, the impulse response is calculated at a receiver position at the critical distance. The overall energy of the impulse response, E_T , is the sum of the direct sound energy, E_D , the early reflection energy, E_E , as well as the late reverberant energy, E_L , as follows,

$$E_T = E_D + E_E + E_L. \quad (7)$$

At the critical distance, the following condition has to be met for an omnidirectional source and receiver pair,

$$E_D = E_E + E_L. \quad (8)$$

Consequently, the energy of the late reverberation has to be adjusted to

$$E_L = E_D - E_E, \quad (9)$$

$$\sum_{t=0}^{2RT} p_L^2 = \sum_{t=0}^{2RT} p_D^2 - \sum_{t=0}^{2RT} p_E^2, \quad (10)$$

with the sound pressure, p , which is, of course, proportional to the digital signal amplitude.

In the second method, the exponentially-decaying amplitude of the reverberation tail is fitted to the exponentially-decaying amplitudes of the reflection pattern. For this purpose, both signals are logarithmized so that the decaying impulse response can be fitted by a linear-regression curve. The amplitudes of the decaying slope are then matched and a cross-fade method is used to blend out the early reflections while gradually blending in the late reverberation.

2.3 The Forest Walk-Through

Coming back to the forest simulation example, the environment is depicted in Fig. 2b. The red dot shows the sound source, which is located at the coordinate 40 m/7 m (x/y coordinates in meters). In cases where two circles overlapped, one of the circles was removed since two tree stems cannot occupy the same space. The absorption coefficient was set to 5% based on measurement results for tree barks (Reethof et al. 1977). Diffuse reverberation that results from leaves and other objects was added. The reverberation time was adjusted to 1.6 s and the reverberation ratio, adjusted to the interaural coherence, was 0.4 at a source-to-receiver distance of 40 m. Both values were chosen based on forest-acoustics measurements by Sakai et al. (1998). The rays are depicted through dashed lines that become lighter with increasing order. Three walk pathways were computed at different y-coordinates that were held constant for each condition, that is, 17 m, 27 m, and 47 m, labeled as 10-m, 20-m and

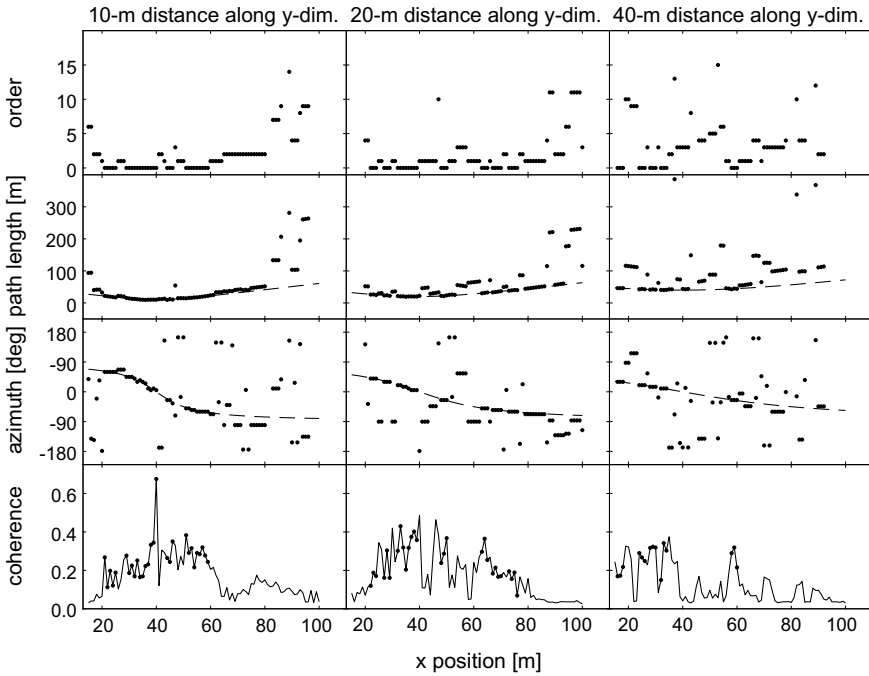


Fig. 4 Results of cue analysis of a simulated forest walk-through. The **left panels** show the results for the pathway that is 10m behind the listener in the y-direction. The **center panels** show the results for the 20-m condition, and the **right panels** for the 40-m condition. For each condition, the order of the first arriving reflection is shown in the top graph as a function of the x-position. An order of zero indicates that the direct signal passes to the listener position and is not obstructed by the tree trunks. The **panels 2nd from the top** show the actual path lengths of the first wavefront from the source to the receiver—indicated by the **dots**. The **dashed line** shows the direct distance between the sound and the listener. The **panels 3rd from top** show the angles of incidence of the first-arriving wavefront indicated by the **dots**. The **dashed lines** show the actual azimuths between the sound source and the listener. The **bottom graphs** depict the coherence indicated by the **solid lines**. All coherence values that correspond to cases where the direct signal was not obstructed are emphasized by additional **dots**

40-m conditions—referring to the distance in the y-dimension between source and receiver. Each walkway covers the distance from 15 to 100m along the x-coordinate. Binaural impulse responses were computed in 1 m increments along the x-axis and then analyzed.

The analysis results are shown in Fig. 4. The three columns show the results for the different distances along the y-coordinate. The top row shows the order of the first arriving wave. For the 10-m condition, the unobstructed direct signal (0th order) arrives at the listener position in 37% of the cases—see top-left graph of Fig. 4. In those cases, where the receiver is further away from the sound source based on the x-axis position (x-position > 60 m), the direct line of sight is obstructed in all cases and the first ray that reaches the receiver is typically on the order of two or higher.

For $x > 80$ m, the effective path length becomes much greater than in the other cases with values of 100 m and beyond—see graph second from the top in the left panel of Fig. 4. This also greatly affects the azimuth of the first arriving wavefront, which is not necessarily the direct signal—see graph second from the bottom in the left panel of Fig. 4. With a few exceptions, the first wave front arrives from the azimuth direction of the sound source or an angle close by if the receiver is located at an x -position between 20 and 60 m. Outside this range, the azimuth values differ greatly from the actual sound-source angle.

Next, the interaural coherence is investigated. The interaural coherence estimates how similar the left and the right ear signals are in the time domain after they have been amplitude and time aligned. It is a measure of how reverberant the sound field is at the listener position—knowing that the presence of reverberation decorrelates both signals, thus making them more dissimilar. The interaural coherence can be calculated as the absolute maximum of the normalized cross-correlation function, which is defined as

$$R_{l,r}(n, m) = \frac{\sum_{n=n_0}^N x_l(n - m) \cdot x_r(n)}{\sqrt{\sum_{n=n_0}^N x_l^2(n - m) \cdot \sum_{n=n_0}^N x_r^2(n)}}, \quad (11)$$

with the time n , the internal delay, m , the left input signal, x_l , and the right input signal, x_r .

The interaural coherence for the 10-m condition is shown in the bottom-left graph of Fig. 4, solid line. It is noteworthy that the interaural coherence becomes noticeably larger with the absolute distance from the source. Therefore the interaural coherence is generally smaller in the 20-m–60-m x -position range than for the outside positions. All values that correspond to cases that include the direct sound are emphasized through dots.

The 20-m condition is shown in the center column of Fig. 4. The relative number of x -positions, where the direct signal is not obstructed on the pathway to the listener position is slightly lower than in the 10-m condition—33% versus 37%. The average coherence, 0.17, is the same as found for the 10-m condition. Also, the coherence is noticeably higher for most cases, where the direct signal reaches the listener's ears—as indicated by the dots.

The 40-m condition is shown in the right column of Fig. 4. Here, the direct path is often obstructed and the direct signal reaches the listener only in 17% of the cases. Consequently, only a few azimuth values indicate the correct sound-source position—see graph second from the bottom in the right panel of Fig. 4. The coherence values are lower than for the other two conditions with an average of 0.13. However, also in this case, the relative coherence values are higher when the direct signal is not obstructed on its way to the listener.

In general, it can be concluded that in a dense forest environment the direct signal is often obstructed, making it both acoustically and visually challenging to localize

salient objects. By moving through the environment, the receiver can find locations where the sound source arrives directly without obstructions. These positions are usually characterized by coherence values that are higher than those found for obstructed sound sources. The forest scenario is a good example of where hearing and vision can work together to locate sound sources quickly because the sound source becomes visible once it is no longer obstructed by objects.

3 Understanding the Fundamental Sound of Caves to Concert Halls Using a Precedence-Effect Model

The next section investigates how binaural models can be used to extract room-acoustic features from a running signal and compare the results to the first-known type of concert venue—the cave. At the core of the human ability to extract information from sound sources in reverberant spaces are auditory mechanisms related to the *precedence effect* (Blauert 1997; Litovsky et al. 1999). The precedence effect, formerly also called the *law of the first wave front*, describes the ability of the auditory system to suppress information about secondary sound sources that are reflected from walls and other objects. This enables the auditory system to localize the actual position of a sound source by making the localization cues pertinent to the direct-signal component available. This is a non-trivial task for the auditory system since the direct signal and the reflected signal parts overlap in time and frequency. The primary cues to localize a sound source are *Interaural (arrival) Time Differences* (ITDs) and *Interaural Level Differences* (ILDs). ITDs occur because the path lengths between a sound source and both ears differ depending on the incoming azimuth angle. The cross-correlation algorithm, (11), is an adequate algorithm to simulate the processes in the auditory system when extracting ITD cues. The lateral position of the cross-correlation peak as a function of the internal delay, m , is used to determine the ITD. ILDs occur because of shadowing effects of the head toward the contralateral ear. For more details on ILDs, see, for instance, Breebaart et al. (2001), Braasch (2003, 2005).

The third type of spatial cues are called “*monaural cues*”. Monaural cues are direction-dependent, pinna-induced spectral modifications that require only one ear for analysis—Blauert (1969/1970, 1997), Zakarauskas and Cynader (1993). These cues are especially important for judging the elevation and front/back orientation of a sound source. Yet, it is shown in this chapter that these dimensions can also be handled by ITD-based algorithms if head movements are considered—compare Pastore et al. (2020, this volume). The focus of the chapter will, however, continue to focus on sound-source localization and information extraction in reverberant environments. Further, regards survival, the auditory system’s ability to segregate sound sources is relevant as well—for details see, for instance, Bodden (1993), Roman et al. (2003), Roman et al. (2006), Deshpande and Braasch (2017), Mi et al. (2017).

When conducting room analyses, it should be kept in mind that modern concert halls have not been around until very recently in the scheme of human history. It will now be discussed how the auditory system can extract room-acoustic features by using a system that does not have at its disposal auditory experience of millions of years to adapt to rectangular-shaped rooms.

The first-known musical instrument at all is a 40,000-year-old rim flute made from a vulture bone. Yet, in the context of the current paper is not so much the instrument itself that is important but rather the cave where it was found. This is the *Hohle-Fels* cave near Schelklingen, Germany—Conard et al. (2009). It is hard to imagine that this instrument has not been played in the cave. In 2016, the current author had the opportunity to visit the cave and to record some impulse-responses in it. This allowed him to estimate the cave's mid-frequency reverberation time, which came out as about 2 s—Braasch (2019). It is remarkable that the reverberation time of the *Hohle-Fels* cave is in the range of modern classical concert halls. For example, the Haydn-Saal in Eisenstadt, as was shown in Fig. 1, also has a reverberation time of about 2 s in the mid-frequency range when the hall is occupied—Meyer (1978), p.147. However, what is important in the context of this paper, is the following. Despite the similarities in reverberation times of the *Hohle-Fels* cave and a typical concert hall, there is a fundamental acoustic difference between them. Concert halls are typically rectangularly shaped or have at least large plane surfaces, while the surface of a cave is very irregular. The latter leads to a very diffuse echogram while the concert hall has a few very distinct reflections. While reverberation chambers for technical acoustic measurements are often kept diffuse, there are very few music facilities that build on this diffuseness notion. The most distinct two in the world are probably Studio C at Blackbird Studios in Nashville, TN, (Bonzai 2018), and the Studios 1 and 2 at Rensselaer Polytechnic Institute's Experimental Media and Performing Arts Center (EMPAC) in Troy, NY. Studio C was conceived and designed by George Massenburg, and its walls are treated with 40 ton of long wood beams similar to the absorptive wedges in anechoic chambers but with an irregular pattern and being sound-reflective. While the sound of the studio is reverberant, this unique design avoids spectral colorations imposed by comb filtering. Concurrently, a lot of spatial properties that would commonly originate from the pattern of specular reflections are not present in this studio. An anecdote illustrates the acoustic features of this space. According to George Massenburg, a session was booked with a blind pianist. When the musician entered the studio, he walked around the music stand and the piano with the help of his cane but then, without the usual direct reflections of planar surfaces, walked straight into a wall that he did not perceive to be there. Studio 1 and 2 at EMPAC were conceived by EMPAC director Johannes Goebel with the diffuse acoustics of a forest opening in mind.

3.1 Precedence-Effect Model

Estimating the ITD of the Direct Signal

The precedence-effect model as reported here, has the task of analyzing the reverberant conditions. It is on the *Binaurally Integrated Cross-correlation Auto-correlation Mechanism (BICAM)*—Braasch (2016). Modifications were made to the original algorithm to calculate more accurate binaural-activity maps. Unlike traditional precedence-effect models that suppress the energy or spatial information of early reflections, the BICAM algorithm separates the auditory cues for the direct signal and from the early reflections but does not remove or suppress the latter. This is important for this section because the aural quality of the room that the sound source is presented in can thus be evaluated. The model separates auditory features for the direct signals and early reflections from a running signal using a dual-layer spatiotemporal filter. Figure 5 shows the architecture of the model. The incoming signal ascends from bottom to top. The model separates the incoming binaural signal into auditory bands at the initial stage—as shown in the bottom row of boxes. Then, the model performs a set of auto-/cross-correlation analyzes within all auditory bands as depicted in boxes, labeled “AC” and “CC”, that are shown in the 2nd row from the bottom. During this process, the following autocorrelation/crosscorrelation sequences are calculated from the left and right ear signals, x and y —depicted as Steps 1 and 2 in Fig. 6,

$$R_{xx}(m) = E[x_{n+m}x_n^*] \quad (12)$$

$$R_{yx}(m) = E[y_{n+m}x_n^*] \quad (14)$$

$$R_{xy}(m) = E[x_{n+m}y_n^*] \quad (13)$$

$$R_{yy}(m) = E[y_{n+m}y_n^*], \quad (15)$$

with the cross-correlation sequence, R , the expected-value operator, $E\{\dots\}$. The regular, non-normalized cross correlation is defined as follows:

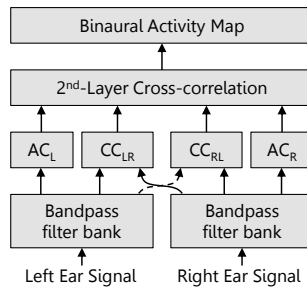


Fig. 5 System architecture and signal flow of the BICAM model—(AC)...autocorrelation, (CC)...cross correlation

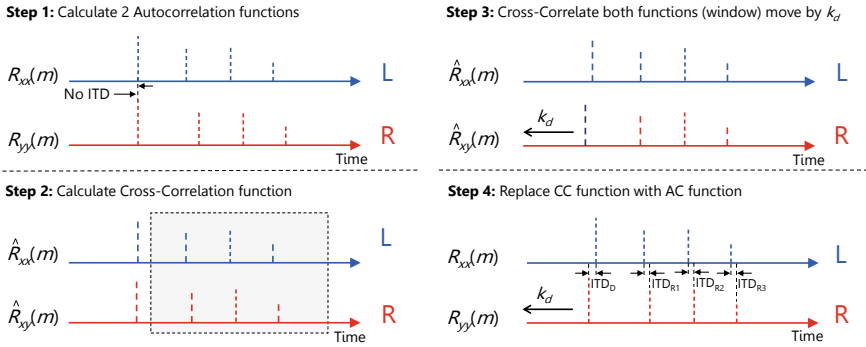


Fig. 6 Autocorrelation/cross-correlation procedures that are performed using the BICAM architecture to estimate the binaural room impulse response. The variable $R_{x,x}$ is the autocorrelation for the left channel, $R_{y,y}$ denotes the same for the right channel. $R_{x,y}$ represents the cross-correlation function between the left and the right channels. The hats over the variable R indicates that only the right side is considered. The **gray window** in Step 2, is used to compare the left-ear channel, L , to the right-ear channel, R

$$R_{i,j}(n, m) = \sum_{n=n_0}^N i(n - m) \cdot j(n), \tag{16}$$

with, the time, n , the internal delay, m , and the input signals i, j . The variable n_0 is the start time of the analysis window and N the end time. The left and right input signals are assigned to the variables i and j . For the case $i = j$, R denotes the autocorrelation. For the BICAM model, the range of the internal delays, $-M$ to M , needs to exceed the duration of the reflection pattern of interest. Otherwise, the impulse response is not shown in its entire duration. Alternatively, $\pm M$ can also be set to just show the early part of an impulse response. The variable n typically ranges from the beginning of the signal, $n = 0$, to the end of the signal, N . The calculation can be performed as a running analysis over shorter, overlapping time segments.

In the next step, which is typically not found in traditional cross-correlation models (Sayers and Cherry 1957; Blauert and Cobben 1978; Stern and Colburn 1978), a cross-correlation algorithm is performed on top of the combined autocorrelation/cross-correlation algorithm as shown in the second top box in Fig. 5 and also in Step 3 of Fig. 6. The goal of this procedure was to develop a method that incorporates the causality of the direct sound and its reflections, which is not provided by conventional cross-correlation models. Using the second-layer cross-correlation analysis over the autocorrelation signal (e.g., $R_{x,x}$) in one-channel and the cross-correlation signal (e.g., $R_{x,y}$) in the second channel, the spatial information in the direct signal and in the individual reflections can be segregated.

A key to the function of the model is a comparison of the right side peaks of both functions (autocorrelation function and cross-correlation function) as shown in the gray box in Step 2 of Fig. 6. These side peaks are correlated to each other by windowing out the direct peaks and the left side of the (auto-)correlation func-

tions. The temporal offset between both main peaks can be obtained by aligning the side peaks in time to determine the interaural time difference (ITD) of the direct sound. The alignment of the side peaks is accomplished by cross-correlating the two autocorrelation/cross-correlation functions—(12) and (15)—over the segments of both functions that contain the side peaks for positive internal delay values, m , (gray areas in Step 2 of Fig. 5). It is important to zero out all remaining segments so that the main peaks and the side peaks for negative m values cannot affect the alignment of the positive side peaks. Mathematically, this operation can be stated as

$$\hat{R}_{ij} = R_{ij} | \forall m > w \wedge \hat{R}_{ij} \stackrel{!}{=} 0 | \forall -M \leq m \leq w. \quad (17)$$

In the next step, the variables, i and j , are substituted with the left and right ear signals, x and y , to compute the following four functions, \hat{R}_{xx} , \hat{R}_{xy} , \hat{R}_{yx} , and \hat{R}_{yy} . The variable w is the length of the window to remove the main peak. The method works if the cross terms (correlations between the reflections) are within certain limits.

Using these functions, the 2nd-layer cross-correlation is calculated. The ITD for the direct signal, $k_{\bar{d}}$, can then be computed from the product of the 2nd-layer cross-correlation terms—see Step 3 in Fig. 5:

$$k_{\bar{d}} = \max_m \arg \left\{ \sqrt{|R_{\hat{R}_{xy}\hat{R}_{xx}} \cdot R_{\hat{R}_{yy}\hat{R}_{yx}}|} \right\}. \quad (18)$$

The solution for $k_{\bar{d}}$ represents the lateral position of the direct signal. In the next step, this solution is used to further expand the algorithm to derive a binaural-activity map that also contains information about the locations and delays of individual early reflections—see top box in Fig. 5.

Binaural-Activity-Map Calculation

A binaural activity map is a three-dimensional plot of a binaural room impulse response that depicts the temporal course of the reflections on the x-axis, the spatial positions of the reflections on the y-axis and the amplitude of the reflections on the z-axis—see Braasch (2005) for more information. In order to create the binaural activity map, the ITD of the direct signal, $k_{\bar{d}}$, is used to shift one of the two autocorrelation functions, R_{xx} or R_{yy} . The latter two functions are, in some form, a representation the early reflection patterns for the left and right channels—see Step 4 in Fig. 5. The respective equations are

$$\check{R}_{xx}(m) = R_{xx}(m), \quad (19)$$

$$\check{R}_{yy}(m) = R_{yy}(m - k_{\bar{d}}). \quad (20)$$

A series of cross-correlation functions is calculated over moving segments of the time aligned autocorrelation functions, \check{R}_{xx} and \check{R}_{yy} , for positive time values in order to estimate the delays, ITDs and relative amplitudes of the reflections.

3.2 Acoustical Analysis

In order to demonstrate the effects of the two opposite sound environments, two idealized environments were created, one with mostly diffuse reflections and the other one with mostly specular reflections, using the ray-tracing model that was introduced in Sect. 2.2. In the first case, simulating the cave, the impulse response for the diffuse reverberation of the *Hohle-Fels cave* was simulated using decaying Gaussian-noise burst roughly matching the RT of the cave, namely, 2 s, and an initial time-delay gap of 12 ms. The direct sound source was simulated using a delta peak convolved with an HRTF pair corresponding to 0° azimuth and 0° elevation. All HRTF catalogs used for this chapter have been measured at the Institute of Communication Acoustics of the Ruhr-University Bochum, Germany—for details see Braasch and Hartung (2002). The direct-to-reverberant-energy ratio between the direct signal and the late reverberation was set to 3 dB. An anechoic male voice sample of 12 s duration was used as the sound signal for all examples in this section. For the simulation, the auto- and cross-correlation terms, \hat{R}_{xx} , \hat{R}_{xy} , \hat{R}_{yx} , and \hat{R}_{yy} of the BICAM algorithm were employed. The values for Eq. (17) are calculated in separate auditory bands using the same gammatone-filter bank (Patterson et al. 1995) with 15 auditory bands from 100 to 1600 Hz. The beginning of the window w in (17) was set to 100 samples (2268 μ s). The length of the window equaled to 40 ms. The ITD, $k_{\bar{d}}$, for the direct signal was then estimated from the frequency-superposed 2nd-layer cross-correlation functions according to (18).

Figure 7a shows the binaural impulse response extracted from the running signal using the BICAM model. One can clearly see that both the left and the right channels, shown as blue and red lines, depict the direct sound but not the exponentially decaying reverberation tail. Only some residual noise that results from the autocorrelation process can be found due to the limited duration of the source signal. The binaural-activity map that was computed using the estimated binaural impulse response is shown in Fig. 8a. Based on the discussed features of the impulse response, it comes as no surprise that the binaural-activity map only shows a single peak for the direct signal but no trace of the reflections. With the exceptions of a few artifacts at the late end of the binaural activity map, the outcome is very similar to the binaural-activity map computed for an anechoic environment but an identical direct sound source—see Fig. 7b and Fig. 8b. The results support the StudioC anecdote and are also in line with a study by Teret et al. (2017) that demonstrates that listeners have no temporal representation of a Gaussian reverberation tail independent of the sound stimuli that are convolved with that reverberation tail.

For the concert-hall example, the impulse response was composed of the same direct signal and its reverberant tail as has already been used to simulate the cave. In addition, four specular reflections were simulated at the following locations and delays,

Azimuth	Delay	Reflection coefficient
-45°	16 ms	0.7
+45°	19 ms	0.7
-60°	22 ms	0.5
+60°	25 ms	0.4

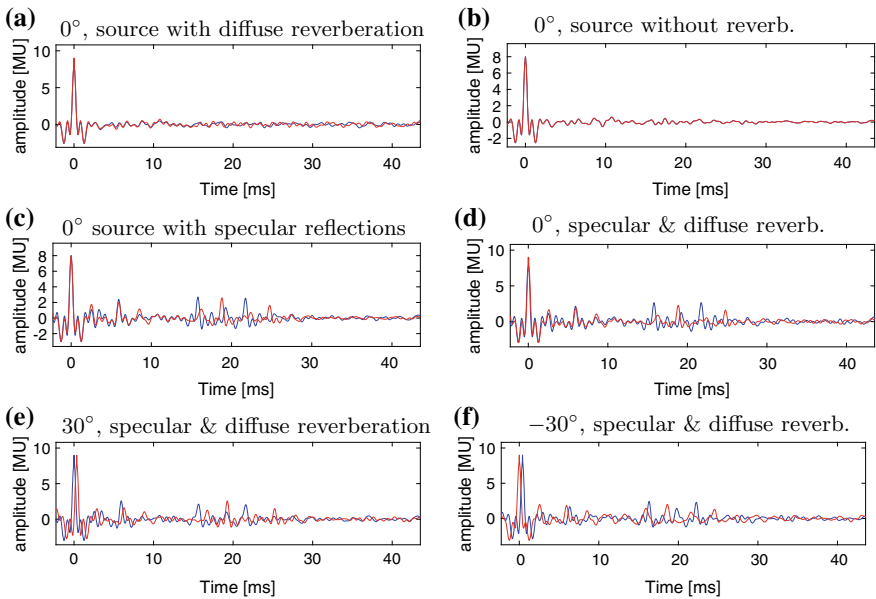


Fig. 7 Binaural room impulse responses, estimated from running signals

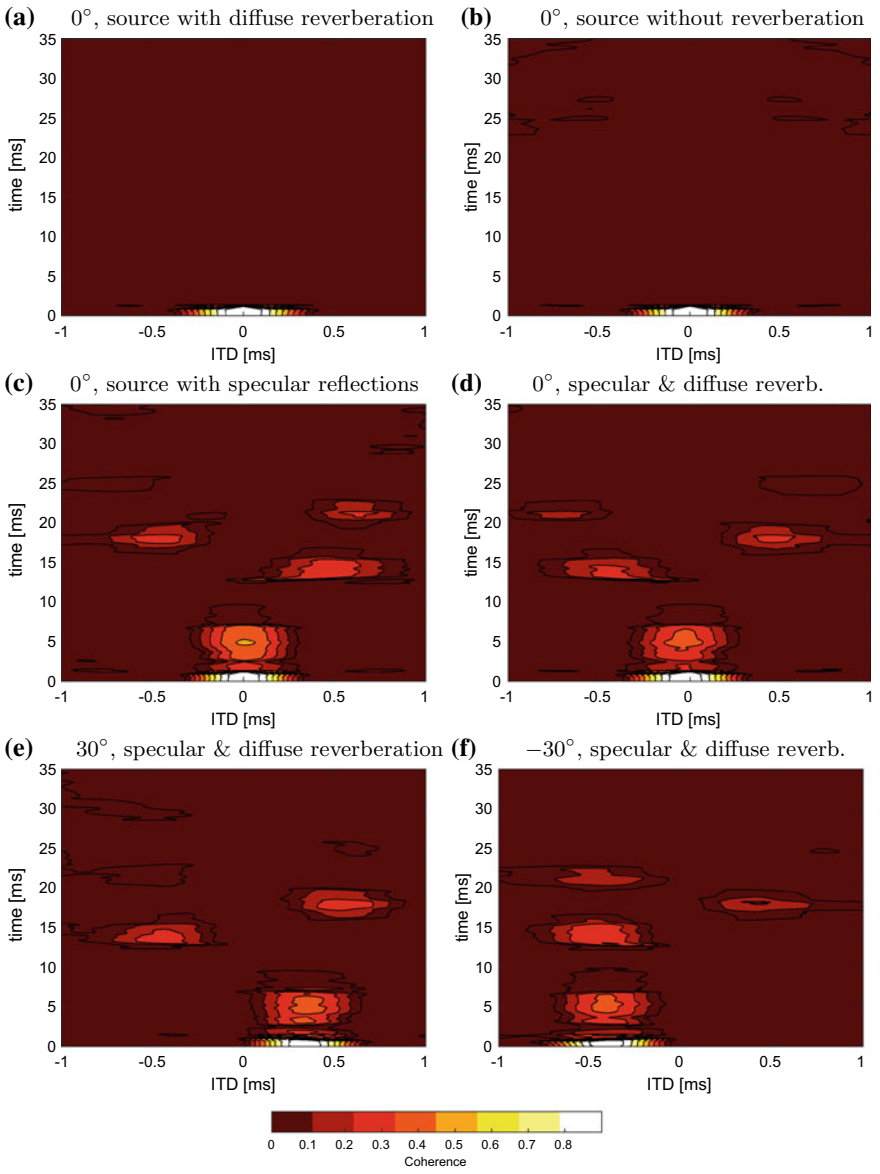


Fig. 8 Binaural-activity-map results of BICAM-model analyses for different conditions, including diffuse and specular reflections as indicated in the individual graphs above

The extracted impulse responses are shown in Fig. 7c,—impulse response direct sound and specular reflections only—and in Fig. 7d,—impulse response with specular reflections and diffuse reverberation tail. The corresponding binaural-activity maps are shown in Fig. 8c, d. In both cases, the map correctly identifies the lateral positions and the delays of the direct sound and of most of the early reflections. In order to demonstrate the ability of the model to provide an independent but joint analysis of the direct sound and the early reflections, only the lateral position of the direct sound source was moved while the lateral positions and delays of the early reflections were maintained. Figure 8e shows the same binaural-activity map as Fig. 8d but for a direct signal that has been moved laterally to $+30^\circ$. The result shows that the binaural-activity map indicates the new lateral position correctly while maintaining, in principle, the positions of the side peaks that indicate the delays and lateral positions of the reflections. The positions of the side peaks are also maintained when the direct-sound source is moved to -30° —see Fig. 8f. Figure 7e, f depict the binaural room impulse responses that were extracted from the running signal for the two conditions with a lateralized direct-sound source. In comparison with the laterally centered direct-sound source, shown in Fig. 8d, one can see that the later parts of the binaural room impulse responses are very similar while the onset delays between left and right channels, shown in blue and red, are clearly visible.

Before concluding this section, the fundamental differences of the BICAM algorithm when processing specular reflections and diffuse reflections should be discussed. In this context it is worth noting that the binaural-activity map for the condition with laterally centered direct signal and early specular reflections, as shown in Fig. 8c, does not change much when a late, diffuse reverberation tail is added—see Fig. 8d. Both maps are very similar indeed despite the fact that the stimulus of Fig. 8d contains a diffuse reverberation tail in addition to the early specular reflections. The similarity in both maps re-emphasizes that the BICAM method is “blind” toward diffuse reverberation tails because these do not produce a distinct autocorrelation map. Obviously, human listeners are aware of the presence of a late reverberant field, otherwise acoustical designs like the Blackbird’s StudioC and the EMPAC’s Studios 1 and 2 would not have a meaningful purpose. It has to be kept in mind, though, that the proposed BICAM model is a localization model. Other types of psychoacoustic models, such as detection models, are needed to extract further room-acoustic features. Further standard methods estimate, for instance, interaural coherence and/or extract features of the (exponentially) decaying room impulse response from transients in the source signals, in particular, from impulses and abrupt stops.

While the interaural-coherence method is usually calculated from a measured room impulse response, it can also be calculated from a running signal. This was done for the forest walk-through—see (11). The drawback using the latter method is that the type of source signal will influence the interaural coherence, and the outcome is no longer solely based on the room parameters. However, also in real life, the perceived reverberance is highly influenced by the source signal employed—Teret et al. (2017). A further method is to estimate the reverberation time from the exponential-decay rate—see, for instance, Huang et al. (1999).

Among the three described methods, the binaural-activity-map analysis is the only method that allows for the extraction of information about the location (angle and distance) of reflective surfaces, for instance, of walls. Neither the interaural-coherence method nor the exponential-decay method provides these cues. Without them, the listener will not receive unambiguous information about the size of a room and the location of walls and of further sound-reflecting or obscuring obstacles. This is the reason why the blind pianist walked right into a wall in Blackbird's Studio C—the absence of salient cues.

4 Simulating an Office Walk-Through Using a Binaural Model Capable of Utilizing Head Movements

An acoustic walk-through a building is in many ways a modern-society version of the forest walk-through. Also in this case, the direct sight to an object can be obstructed and then one has to rely on the acoustic sense. However, obstructing objects have very different acoustic qualities. While the forest is a leaky reverberation chamber with diffusive character, office suites, and other small rooms are characterized by specular reflections that arrive shortly after the direct sound. From an evolutionary perspective, where anatomical changes occur over a span of several million years, rectangular caverns with flat walls have only been introduced recently during our early civilization and similar acoustic objects do not appear in nature. It is therefore important to understand how the auditory system is able to adapt to built rooms given that it has not specifically evolved to deal with such environments.

4.1 Head-Movement Algorithm

In order to simulate the office walk-through, an existing head-movement algorithm (Braasch et al. 2013) is added to the binaural model, such that the model can resolve back/front confusions and analyze the auditory scenes adequately. The model builds on a theory proposed by Wallach (1939).

Auditory Periphery

The model takes a step back from the elaborate BICAM mechanism and uses a traditional interaural-cross-correlation method as introduced by Sayers and Cherry (1957) to estimate ITDs—see (17). The basic model structure, shown in Fig. 9, is similar to the one proposed by Braasch (2002). The inputs signals are filtered with HRTFs from desired directions. Basilar-membrane and hair-cell behavior are simulated using a gammatone-filter bank with 36 bands and a simple half-wave rectifier at a sampling frequency of 48 kHz, as described by Patterson et al. (1995).

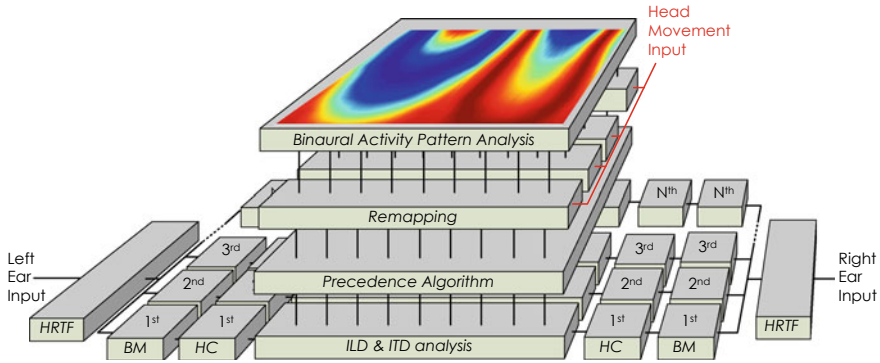


Fig. 9 General model structure of the binaural-localization model utilizing head rotations. **HRTF**...External-ear simulation/HRTF filtering. **BM**...Basilar membrane/bandpass filtering. **HC**...Hair cell/halfwave rectification **ITD & ILD analysis**...Interaural time difference-cue (ITD) extraction/interaural cross-correlation and interaural level-difference cue (ILD) analysis with EI-cells, precedence effect algorithm, remapping to azimuths with head-rotation compensation, binaural-activity-map analysis to the estimation the sound-source positions

Cross Correlation

After the half-wave rectification, the normalized interaural cross-correlation (11) is computed for each frequency band over a short time segment. Only the Frequency Bands 1 to 16 (23–1559 Hz) are analyzed, reflecting the inability of the human auditory system to resolve the temporal fine structure at high frequencies, as well as the fact that at low frequencies the interaural time differences in the fine structure are the dominant cues—provided that they are available at all (Wightman and Kistler 1992).

Remapping and Decision Device

Next, the cross-correlation functions will be remapped from interaural time differences to azimuth positions. This is important for the model to be able to predict the spatial position of the auditory event. In addition, this procedure helps to align the estimates for the individual frequency bands as one cannot expect that the interaural time differences are constant across frequency for a given angle of sound-source incidence. An HRTF catalog is analyzed to convert the cross-correlation function’s x-axis from interaural time differences to the azimuth. The HRTF catalog was measured at a resolution of 15° in the horizontal plane and then interpolated to 1° resolution using the spherical-spline method—see Hartung et al. (1999). After filtering the HRTFs with the gammatone-filter bank, the ITDs for each frequency band and angle are estimated using the interaural-cross-correlation (ICC) algorithm of (16). This frequency-dependent relationship between ITDs and azimuths are used to remap the output of the cross-correlation stage (ICC curves) from a basis of ITDs $m(\alpha, f_i)$, to

a basis of azimuth angles in every frequency band as follows:

$$m(\alpha, f_i) = g(\text{HRTF}_l, \text{HRTF}_r, f_i) \quad (21)$$

$$= g(\alpha, f_i), \quad (22)$$

with azimuth, α , elevation, $\delta = 0^\circ$, distance, $r = 2$ m, $\text{HRTF}_{l/r} = \text{HRTF}_{l/r}(\alpha, \delta, r)$, center frequency of bandpass filter, f_i .

Next, the ICC curves, $(R_{x,r}(m, f_i))$, are remapped to a basis of azimuths using a simple for-loop in Matlab using a step size of 1° :

```
for alpha=1:1:360
    R_rm(alpha, freq)=R(g(alpha, freq), freq);
end
```

Here, $R(m, \text{freq})$ is the original, frequency dependent, interaural-cross-correlation function with the internal delay, m . The function $g(\alpha, \text{freq})$ provides the measured m -value for each azimuth and frequency. Inserting this function as input, m , to R transforms the R -function into a function of the azimuth, using the specific Matlab syntax.

In the decision device, the average of the remapped ICC functions, $R_{\text{rm}}(\alpha, \text{freq})$, over the frequency bands 1–16 is calculated and divided by the number of frequency bands. The model estimates the sound sources at the positions of the local peaks of the averaged ICC function.

Figure 10 shows an example of a sound source in the horizontal plane with an azimuth of 30° for the eighth frequency band. The top-left graph shows the original ICC curve obtained using (16) as a function of ITD. The graph is rotated by 90° with the ICC on the x-axis and ITD on the y-axis to demonstrate the remapping procedure. The curve has only one peak at an ITD of 0.45 ms. The top-right graph depicts the relationship between ITD and azimuth for this frequency band. As mentioned previously, the data were obtained by analyzing HRTFs from a human subject. Now, this curve will be used to project every data point of the ICC-versus-ITD function to an ICC-versus-azimuth function, as shown for a few data points using the straight dotted and dashed-dotted lines. The bottom panel shows the remapped ICC function, which now contains two peaks, that is, one for the frontal hemisphere and one for the rear hemisphere. The two peaks fall together with the points where the cone-of-confusion hyperbolas intersect the horizontal plane for the ITD value of the maximum peak that is shown in the top-left panel.

Integrating Head Rotation

In the following, it is assumed that the head rotates to the left while analyzing an incoming sound source from the front. Related to the head, the sound source will move toward the right. However, in the case that sound source was in the rear, the

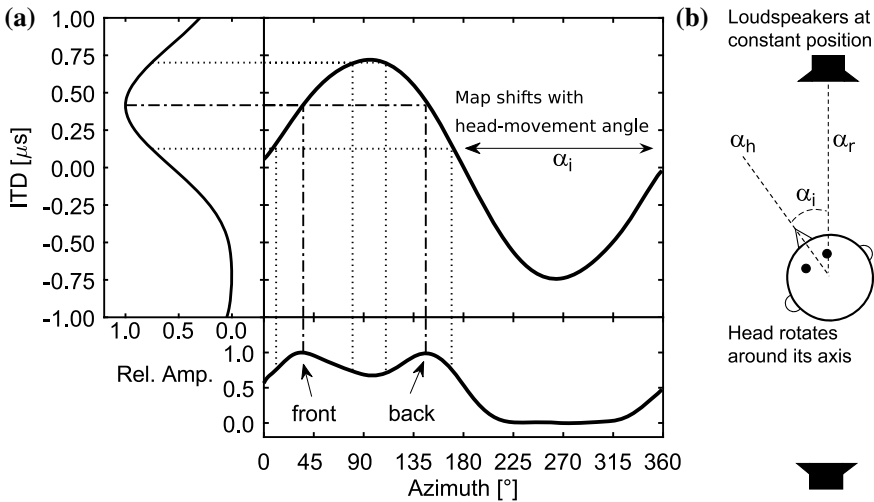


Fig. 10 (Left) Remapping of the cross-correlation function from ITD to azimuth, shown for the frequency band 8, centered at 434 Hz. The signal was presented at 30° azimuth, 0° elevation. **(Right)** Sketch to illustrate the front/back confusion problem. If an ongoing sound source is located in front of the listener who turns her head left, the sound source will move to the right from the perspective of the listener’s head. But if the sound source is located in the back, the sound source appears to move to the left for the same head rotation. The variable, α_r , denotes the azimuth in the room-related coordinate system, pointing here at 0°. The variable α_h is the azimuth in the head-related coordinate system, also pointing at 0° but for this coordinate system. The third angle, α_m , is the head-rotation angle, which indicates by how much the head is turned from the reference head orientation that coincides with the room-related-coordinate system

sound source had moved to the left. This phenomenon will now be used to distinguish between both options, that is, frontal and rear position. For this purpose, a different coordinate system is introduced, namely, the room-related coordinate system. The fact that human listeners maintain a good sense of the coordinates of a room as they move through it, motivates this approach. If a stationary head position is considered, the head-related coordinate system is fully sufficient. However, if the head rotates or moves, the description of stationary sound-source positions can become challenging because every sound source starts to move with alterations of the head position. An easy way to introduce the room-related coordinate system is to define a reference position and reference orientation of the human head, and then determine that the room-related coordinate system coincides with the head-related coordinate system for the chosen reference position—compare Pastore et al. (2020, this volume), for details on this topic, involving multimodal cues.

Consequently, the room- and head-related coordinate systems are identical if the head does not move. In this investigation, only head rotations within the horizontal plane are considered, and for this case, the difference between the head-related coordinate system and the room-related coordinate system can be expressed through the head-rotation angle α_i that converts the room-related azimuth α_r to the head-related

azimuth α_h —see Fig. 10, right graph. That is,

$$\alpha_r = \alpha_i + \alpha_h. \quad (23)$$

Given restricted head movement, the origin of both coordinate systems and the elevation angles are always identical. While the sound-source position changes relative to the head with head rotation, a static sound source will maintain its position in the room-related coordinate system. Using this approach, another coordinate transformation of the ICC function is executed in the model, namely, a transformation from of head-related to room-related azimuth. This can be accomplished by rotating the remapping function when the head is moving by $-\alpha_i$ to compensate for the head rotation.

If a physical binaural manikin were used—with a motorized head in connection with the binaural model—the HRTF would be automatically adjusted with the rotation of the manikin’s head. In the model discussed here, where the manikin or human head is simulated by means of HRTFs, the HRTFs have to be adjusted virtually. Also, at every moment in time the HRTFs have to correspond to the sound-source angle relative to the current head position. This can be achieved with the help of a running window function, where the sound source is convolved with the current HRTF pair. A Hanning window of 10 ms duration and a step size of 5 ms is used here for this purpose. The smooth edges of this window will cross-fade the signal allowing a smooth transition during the exchange of HRTFs. For each time segment, the model processes the following sequence:

1. First, it updates the current head-rotation angle, α_i
2. Then it calculates the current head-related azimuth angle, α_h , for each sound source located at its room-related azimuth, α_m
3. Next, the model selects the HRTF pair that correspond closest to α_h
4. Afterwards, it computes the normalized ICC, $R_{l,r}$, for each frequency as a function of the ITD
5. It converts the ICC function to a function of head-related azimuth, α_h , using the remapping function shown in Fig. 10.
6. Next, the model circular-shifts the remapping function based on the head-rotation angle by $-\alpha_i$ to transform the ICC curve into the room-related coordinate system
7. Then, it computes the mean ICC output across all frequency bands
8. It averages the ICC outputs over time
9. It estimates the position of the auditory event to be at the azimuth where the ICC peak has its maximum

The first example is based on a bandpass-filtered white-noise signal with a duration of 70 ms. The signal is positioned at -45° azimuth in the room-related coordinate system. At the beginning of the stimulus presentation, the head is oriented toward the front, $\alpha_h = 0^\circ$, and then rotates with constant angular velocity to the left until it reaches an angle of 30° at the time that stimulus is turned off. The ICC functions are integrated over the whole stimulus duration. Figure 11 shows the result of the simulation. The initial ICC-versus- α_r function the output of Step 6 for $\alpha_m = 0^\circ$ is

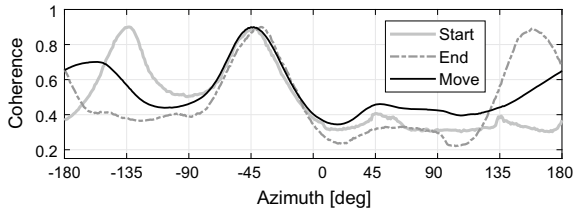


Fig. 11 Interaural-cross-correlation pattern for a sound source at -45° which is presented during a head rotation from $\alpha_m = 0^\circ$ to 45° . The **dashed line** shows the ICC curve for the initial time window, the **solid gray curve** for the last segment when the head is fully turned. Note that the ICC pattern was shifted in the opposite direction of the head rotation to maintain the true peak position at $\alpha_r = -45^\circ$. The **black curve** shows the time-averaged ICC curve for which the main ICC peak remains and the secondary ICC partly dissolves

depicted by the solid, light gray curve. Here clearly two peaks can be observed, one at $\alpha_{r=h} = -45^\circ$ and another one at $\alpha_{r=h} = -135^\circ$. At the end of the stimulus presentation shown as the dashed, dark gray curve $t = 70$ ms, $\alpha_m = 45^\circ$ —, only the position of the rear peak is preserved. This peak indicates the “true”, that is, the physical sound-source location, $\alpha_{r \neq h} = -45^\circ$, because the head rotation was compensated for by rotating the remapping function in opposite direction of the head movement.

However, in the case of a front peak, that is, the front/back confused position, the peak position was counter-compensated for and it rotates twice the value of the head-rotation angle, $\alpha_m = 30^\circ$. The new peak location is shifted by -60° to a new value of 165° . The time-averaged curve (the solid black line which shows the output of Step 7) demonstrates the model’s ability to robustly discriminate between front and rear angles. The secondary peak, the one representing the solution for a frontal sound source, is now smeared out across the azimuth because of the head rotation. Further, its peak height is reduced from 0.9 to 0.7 making it easy to discriminate between front and rear.

4.2 Analysis of an Office Walk-Through

In the next example, it is investigated how the combined head-movement and BICAM localization models can be applied to a real-world scenario, for example, to sound localization in an office suite. For this purpose, a ray-tracing model was implemented to generate binaural impulse responses for the binaural-model analysis. The left graph of Fig. 12 depicts the floor plan together with the trajectory of the walk-through. The encircled numbers indicate the positions of the binaural- analysis examples that as are discussed below in this section. All binaural room impulse responses for the simulations were rendered using the ray-tracing model that was discussed in Sect. 2.2. A geometrical model was defined as shown in Fig. 12b, namely, based on

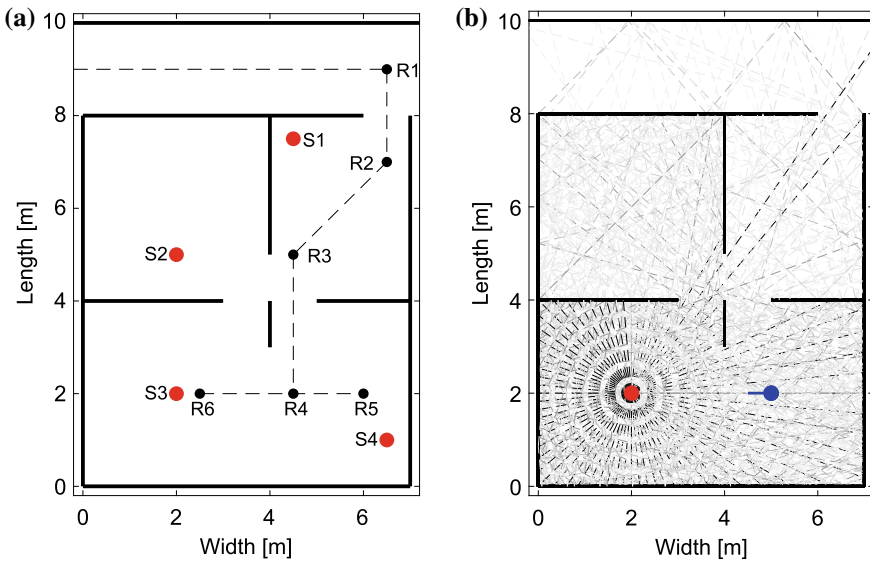


Fig. 12 **a** Diagram for the office walk-through with test positions for sources S1–S4, and receivers, R1–R6. **b** Ray-tracing simulation in a computer-generated office suite with a non-occluded sound source. The sound source is depicted as a **red dot**, the binaural receiver as a **blue dot**. The **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 at a resolution of 1 ray per 5° . Each ray is then traced, and every time a ray meets a wall it is reflected back using Snell's law, that is, 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 the acoustic absorption of the walls. 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. 12b as gray lines, such that the rays become lighter in color with distance and decreasing sound pressure.

All rays are finally collected at the receiver position, assuming a spatial window of 0.6 m width. Each calculated ray is tested for whether it intersects the spatial window at the receiver position. How far each ray traveled from the source position to the receiver is then calculated for each intersecting ray. Similarly, the azimuth of the arriving ray and the order of reflection for the incoming ray is determined. Based on these data, a binaural room impulse response is calculated in which a left/right HRTF pair is inserted at the correct delay, further, the head orientation-based direction-of-arrival angle of the respective ray. Each HRTF pair is calibrated to the amplitude that the ray should have, based on the distance traveled and the number of wall reflections that it has undergone. In addition, a late-reverberation tail is generated at a constant level by assuming a statistically-evenly distributed diffuse-reverberation field, using

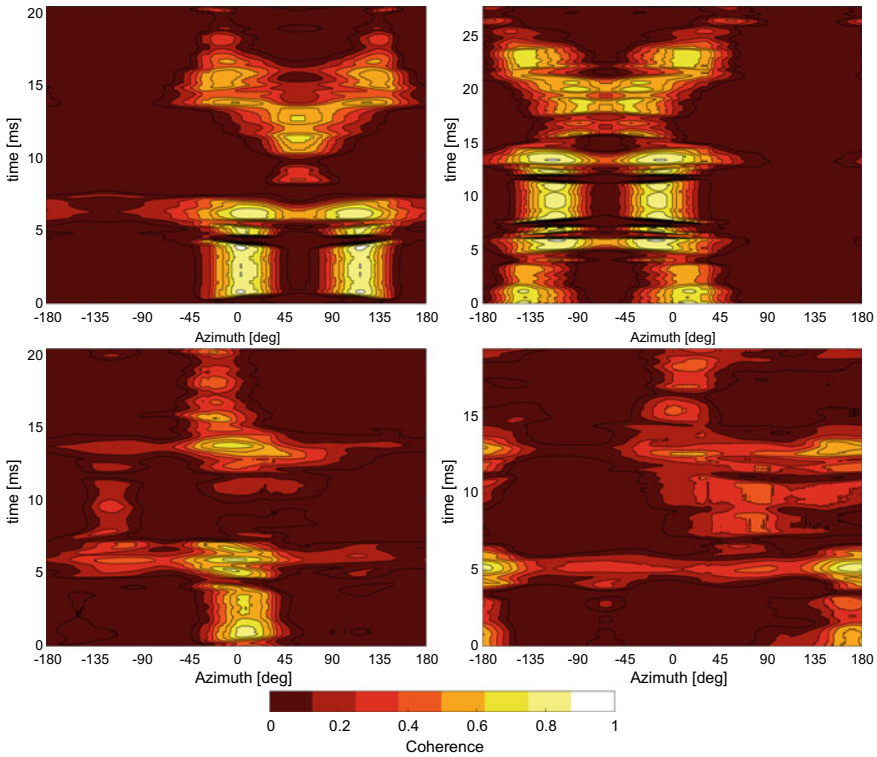


Fig. 13 Binaural-activity map results for the BICAM model analysis utilizing head movements. The **top-left graph** shows the results for Scenario 1 (Fig. 12) right, with the sound source pointing 30° left to the sound source (including a 30° head-movement compensation). The **top-right graph** shows the same condition but for the receiver pointing 30° to the right. The **bottom-left graph** shows the combined analysis for removal of front/back confusions for a receiver pointing into the direction of the sound source, 0°. The **bottom-right graph** shows the same condition as depicted in the **bottom-left graph** but for a receiver pointing away from the sound source, 180°

an exponentially decaying Gaussian noise burst adjusted to a reverberation time of 0.7 s. At the position shown in the right graph of Fig. 12, the diffuse reverberation level was about -10 dB lower than the combined level of the direct sound and the early reflections.

The results are then analyzed using the BICAM precedence-effect model (Braasch 2016) and a male-speech sample (Bang & Olufsen 1992). The BICAM algorithm was modified to transform the model’s ITD estimates into azimuths using a remapping function according to Braasch et al. (2013)—as shown in the binaural-activity map of Fig. 13 (top-left graph). The plot shows 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 right graph of Fig. 12. Note that the data are presented in a room-coordinate system that faces the sound source directly. As can be easily seen, each time slice shows

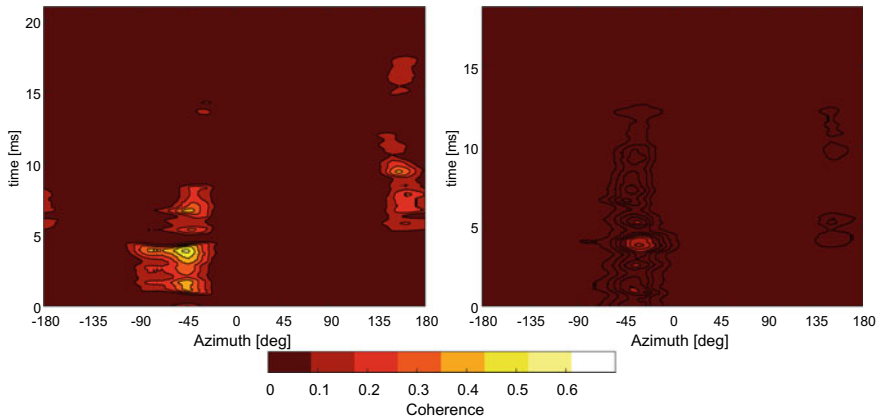


Fig. 14 Binaural-activity-map results for the BICAM model analysis utilizing head movements for an occluded direct sound source—simulating a scenario as depicted in Fig. 12a—with source position, S2, and receiver position, R6. The **left graph** shows the result for a single time interval, the **right graph** depicts the outcome for an average over 10 time intervals

two ambiguous peaks, namely, one for the front and one for the corresponding rear direction, a common problem that was discussed in detail in Braasch et al. (2013). 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. 13. This graph also displays the data in a room-related coordinate system. Now simply, the average is taken of the two binaural activity maps and the ambiguous front/back-confusion peaks average out—see the bottom-left graph in Fig. 13. To demonstrate the effectiveness of the head-movement algorithm, the same scenario was simulated again, but this time with the virtual head facing the rear at 180° with temporal head-movement shifts to 150° and 210° to resolve front/back directions. It should be noted that there are two main differences between the model presented here and the model of Braasch et al. (2013). 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. This renders to two advantages, that is, the direct-sound-source angle can be computed separately from the early reflections, which yields in a higher localization accuracy, and the algorithm can also estimate the front/back direction of the reflections. However, the new model cannot yet calculate front/back directions from a continuously turning head like it is the case for the Braasch et al. (2013) 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.

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 Fig. 12a (S2, R6). Figure 14 shows the binaural-activity maps for this case. In the left graph, the

binaural activity map was calculated from a single time interval. Here, a prominent peak is visible with a maximum correlation of 0.6 even though the direct signal was occluded. However, if the binaural-activity map is calculated as an average over 10 estimates, computed over 10 time intervals, the coherence drops to 0.2—see the right graph. The reason is that the prominent peak develops randomly at different positions for each of the ten computations. It should be noted that each segment by itself leads to a maximum coherence of one because the autocorrelation peaks always 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.

Following Wallach (1939, 1940), the head-movement model can also be used to estimate the elevation of sound sources by utilizing the fact that the ITD range is reduced with up- or downward elevation changes of the sound source from the horizontal plane—compare Pastore et al. (2020, this volume). Ideally, the ITD range is reduced monotonically with the elevation magnitude, minimizing to ITDs of zero at the -90° and 90° elevation poles.

In order to enable sound-source-elevation estimates, the head-movement model is slightly modified for processing different elevations from -70° to 80° in steps of 10° . For each elevation, a new set of frequency-dependent remapping functions is calculated according to (21). In principle, the model is an alternative implementation to an existing localization model by Parks (2014), which also draws from Wallach's ideas to estimate elevation angles. The results of the model simulation are shown in Fig. 15. Each horizontal color sequence corresponds to one elevation set as indicated on the y-axis. The sequence depicted for the elevation of 0° basically shows the same data as Fig. 11 but for different source and head-movement angles. The left graphs indicates the start position of the head-movement angle, the right graphs presents an averaged function over the head motion. The top row shows the results for a sound source at 30° azimuth and 60° elevation. At the beginning of the head-movement trajectory, the results are still ambiguous and the source could be at various elevations at 30° or 150° azimuth. After the head-movement, the model accurately locates the sound source at 30° azimuth and 60° elevation. Also in cases of a sound source located at 0° azimuth/ 0° elevation or one located at -40° azimuth/ -135° elevation, the actual sound source can be determined through head movement—see Fig. 15 center and bottom rows.

5 Conclusion

The goal of this chapter was to examine how auditory systems utilize binaural mechanisms to extract useful information from the environment to be able to read and understand a complex scene. Using idealized but complex simulated environments, it is tested how the auditory system can adapt to different scenarios. Most of the

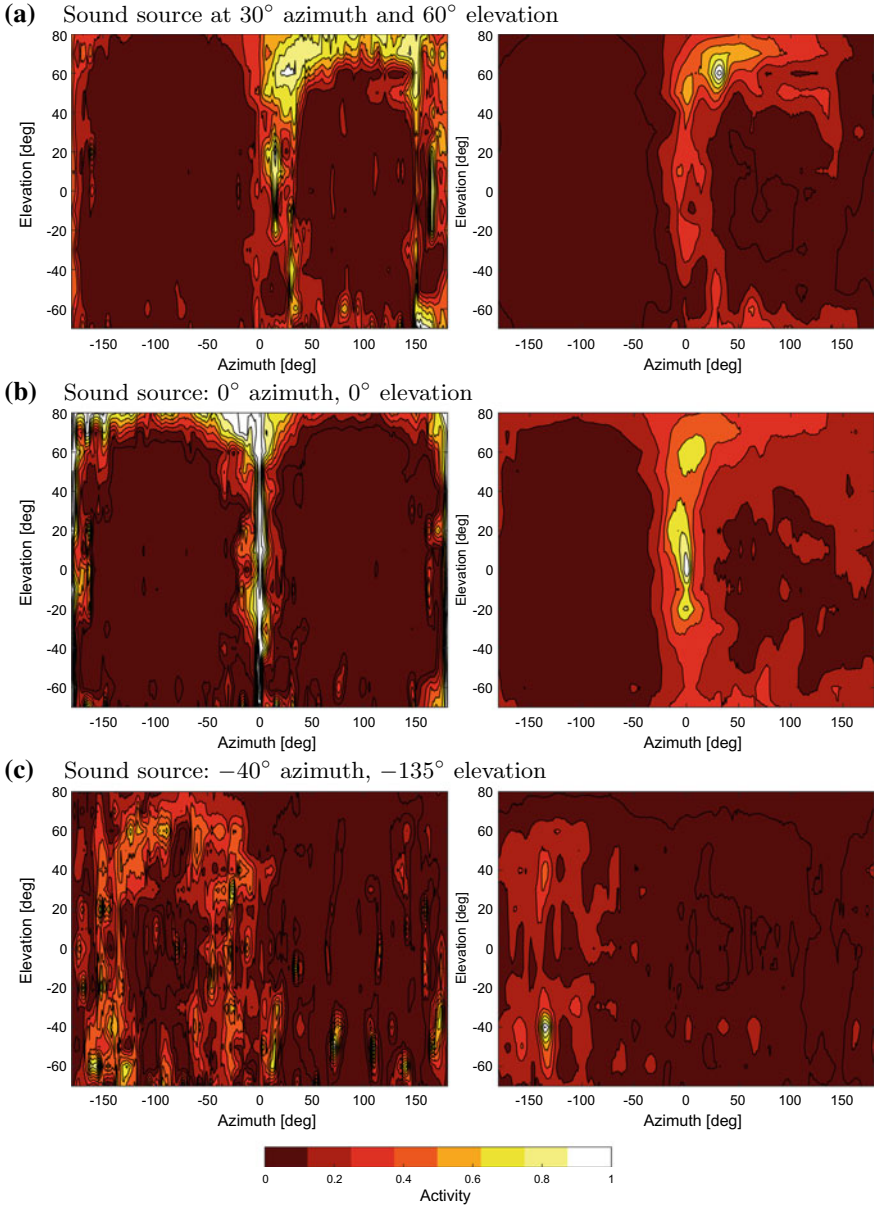


Fig. 15 Demonstration of the head-movement algorithm to estimate elevation. All **left graphs** depict the initial model performance before head movement, the **right graphs** the model performance with integrated head movement. In all cases the head rotated from 0° to 60° azimuth, maintaining the elevation at 0°. **Light areas** indicate a high likelihood of estimated source position, **dark brown areas** indicate a low probability of the source being present

auditory system's capabilities can be traced back to tasks that preceded human civilization, but a remaining mystery is how the auditory system can process specular reflections. One explanation is that our precedence mechanism is an evolutionary response to floor reflections and reflections from a cliff, where large plain surfaces can be found. Yet, it, is still amazing that this mechanism can handle sound perception in human-built rectangular rooms, which appeared very late during our evolutionary process. An alternative explanation is that the precedence effect largely falls out of the specific processing of the auditory system and is not necessarily the product of any precedence-effect specific mechanisms at all. For other tasks, the current demands are not that different from our pre-civilization experiences. Head movements, for example, can help to resolve front/back ambiguities for sound sources. A future goal is to extend the binaural analysis to actually measured environments and to support the findings with psychoacoustic experiments. The study presented here hopefully serves as an initial gateway to better understand how the binaural system reads the world under complex conditions.

Acknowledgements This material is based upon work supported by the National Science Foundation under Grant Nos. 1320059 and BCS-1539276 in alliance with the TwoEars! project (European Research Council, STREP, FP7-ICT-2013-C, No. 618075). The research was carried out within Rensselaer Polytechnic Institute's Cognitive and Immersive Systems Laboratory, a collaboration of RPI and IBM. Reiner Blumentritt provided access to the Hohle-Fels cave for the impulse-response measurement. M. Torben Pastore, William Yost, and Yi Zhou internally reviewed the manuscript. The author is appreciative for valuable comments and suggestions provided by two anonymous reviewers.

References

- Bang & Olufsen. 1992. Music for Archimedes. *CD B&O 101*.
- Baril, D. 2012. The success of homo sapiens may be due to spatial abilities. *PhysOrg*. May 09, 2012, <https://phys.org/news/2012-05-success-homo-sapiens-due-spatial.html> (last access: Dec. 15, 2019).
- Blauert, J. 1969/1970. Sound localization in the median plane. *Acta Acustica united with Acustica* 22: 205–213.
- Blauert, J. 1997. *Spatial Hearing*. Cambridge, MA: MIT Press.
- Blauert, J., and W. Cobben. 1978. Some consideration of binaural cross correlation analysis. *Acta Acustica united with Acustica* 39: 96–104.
- Blauert, J., H. Lehnert, J. Sahrhage, and H. Strauss. 2000. An interactive virtual-environment generator for psychoacoustic research. I: Architecture and implementation. *Acta Acustica United with Acustica* 86 (1): 94–102.
- Bodden, M. 1993. Modeling human sound source localization and the cocktail-party effect. *Acta acustica (Les Ullis)* 1: 43–55.
- Bonzai, M. 2018. George Massenburg builds a Blackbird room. *Digizine, Digidesign*. http://www2.digidesign.com/digizine/dz_main.cfm (last accessed: August 13, 2018).
- Braasch, J. 2002. Localization in the presence of a distracter and reverberation in the frontal horizontal plane: II. Model algorithms. *Acta Acustica United with Acustica* 88 (6): 956–969.

- Braasch, J. 2003. Localization in the presence of a distracter and reverberation in the frontal horizontal plane: III. The role of interaural level differences. *Acta Acustica United with Acustica* 89 (4): 674–692.
- Braasch, J. 2005. Modelling of binaural hearing. In *Communication Acoustics*, ed. J. Blauert, 75–108. Berlin: Springer
- Braasch, J. 2016. Binaurally integrated cross-correlation auto-correlation mechanism (BICAM). *The Journal of the Acoustical Society of America (Express Letter)* 140 (1), EL143–EL148.
- Braasch, J. 2019. *Hyper-specializing in Saxophone Using Acoustical Insight and Deep Listening Skills*. Berlin, Heidelberg: Springer.
- Braasch, J., and K. Hartung. 2002. Localization in the presence of a distracter and reverberation in the frontal horizontal plane. I. Psychoacoustical data. *Acta Acustica United with Acustica* 88 (6): 942–955.
- Braasch, J., S. Clapp, A. Parks, T. Pastore, and N. Xiang. 2013. A binaural model that analyses aural spaces and stereophonic reproduction systems by utilizing head movements. In *The Technology of Binaural Listening*, ed. J. Blauert, 201–223. Berlin, Heidelberg, New York: Springer and ASA Press.
- Breebaart, J., S. van de Par, and A. Kohlrausch. 2001. Binaural processing model based on contralateral inhibition. I. Model setup. *The Journal of the Acoustical Society of America* 110: 1074–1088.
- Bruck, D. 1999. Non-awakening in children in response to a smoke detector alarm. *Fire Safety Journal* 32 (4): 369–376.
- Busby, K.A., L. Mercier, and R. Pivik. 1994. Ontogenetic variations in auditory arousal threshold during sleep. *Psychophysiology* 31 (2): 182–188. <https://doi.org/10.1111/j.1469-8986.1994.tb01038.x>.
- Conard, N.J., and S. Wolf. 2014. *Der Hohle Fels bei Schelklingen: Ursprung für Kunst und Musik [The Hohle Fels near Schelklingen: Source of art and music]*. Schelklingen: Museumsgesellschaft Schelklingen.
- Conard, N.J., M. Malina, and S.C. Münzel. 2009. New flutes document the earliest musical tradition in Southwestern Germany. *Nature* 460 (7256): 737–740.
- Conlon, P. 2004. Reviewed work: Native America, Indian flute songs from Comanche land by Doc Tate Nevaquaya. *The World of Music* 46 (3): 208–210.
- Coolidge, F.L., and T. Wynn. 2018. *The Rise of Homo Sapiens: The Evolution of Modern Thinking*. Oxford: Oxford University Press.
- DeLoach, A.G., J.P. Carter, and J. Braasch. 2015. Tuning the cognitive environment: Sound masking with “natural” sounds in open-plan offices (A). *The Journal of the Acoustical Society of America* 137 (4): 2291.
- Deshpande, N., and J. Braasch 2017. Blind localization and segregation of two sources including a binaural head movement model. *The Journal of the Acoustical Society of America* 142 (1): EL113–EL117. <https://doi.org/10.1121/1.4986800>.
- DIN. 1968. Schallabsorptionsgrad-Tabelle (Table of absorption coefficients). In *German Standards (Deutsches Institut für Normung (DIN))*. Note: This table is no longer formally standardized.
- Ferrill, A. 2018. *The Origins of War: From the Stone Age to Alexander the Great*. Abingdon: Routledge.
- Fitch, W. 2000. The evolution of speech: A comparative review. *Trends in Cognitive Sciences* 4 (7): 258–267. [https://doi.org/10.1016/S1364-6613\(00\)01494-7](https://doi.org/10.1016/S1364-6613(00)01494-7).
- Gaudzinski-Windheuser, S., E.S. Noack, E. Pop, C. Herbst, J. Pflöging, J. Buchli, A. Jacob, F. Enzmann, L. Kindler, R. Iovita, M. Street, and W. Roebroeks. 2018. Evidence for close-range hunting by last interglacial Neanderthals. *Nature Ecology & Evolution* 1087.
- Gerkema, M.P., W.I.L. Davies, R.G. Foster, M. Menaker, and R.A. Hut. 2013. The nocturnal bottleneck and the evolution of activity patterns in mammals. *Proceedings of the Royal Society B: Biological Sciences* 280 (1765): 20130508. <https://doi.org/10.1098/rspb.2013.0508>.
- Graham, P.J. 2014. The function of perception. *Virtue Epistemology Naturalized*, 13–31. Berlin: Springer.

- Hartung, K., J. Braasch, and S.J. Sterbing. 1999. Comparison of different methods for the interpolation of head-related transfer functions. In *AES 16th International Conference on Spatial Sound Reproduction*, 319–329.
- Huang, J., N. Ohnishi, X. Guo, and N. Sugie. 1999. Echo avoidance in a computational model of the precedence effect. *Speech Communication* 27: 223–233.
- Jones, B.C., D.R. Feinberg, L.M. DeBruine, A.C. Little, and J. Vukovic. 2010. A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour* 79 (1): 57–62.
- Joris, P.X., and L.O. Trussell. 2018. The calyx of held: A hypothesis on the need for reliable timing in an intensity-difference encoder. *Neuron* 100 (3): 534–549.
- Knight, C., and J. Lewis. 2017. Wild voices: Mimicry, reversal, metaphor, and the emergence of language. *Current Anthropology* 58 (4): 435–453.
- Kuttruff, H. 2000. *Room Acoustics*. London, New York: Spon Press.
- Lehnert, H., and J. Blauert. 1992a. Aspects of auralization in binaural room simulation. In *Audio Engineering Society Convention 93*, Audio Engineering Society, Preprint No. 3390.
- Lehnert, H., and J. Blauert. 1992b. Principles of binaural room simulation. *Applied Acoustics* 36 (3–4): 259–291.
- Litovsky, R.Y., H.S. Colburn, W.A. Yost, and S.J. Guzman. 1999. The precedence effect. *The Journal of the Acoustical Society of America* 106: 1633–1654.
- Meller, H., and M. Schefzik. 2015. *Krieg: Eine archäologische Spurensuche [War: An archaeological search of traces]* (Landesamt Für Denkmalpflege und Archäologie, Landesmuseum für Vorgeschichte, Sachsen-Anhalt), Begleitband zur Sonderausstellung im Landesmuseum für Vorgeschichte, Halle an der Saale, Germany, 6 Nov 2015–22 May 2016.
- Meyer, J. 1978. Raumakustik und Orchesterklang in den Konzertsälen Joseph Haydns [room acoustics and orchestra sound in Joseph Haydn's concert halls]. *Acta Acustica united with Acustica* 41 (3): 145–162.
- Mi, J., M. Groll, and H.S. Colburn. 2017. Comparison of a target-equalization-cancellation approach and a localization approach to source separation. *The Journal of the Acoustical Society of America* 142 (5): 2933–2941. <https://doi.org/10.1121/1.5009763>.
- Parks, A.J. 2014. Listener orientation and spatial judgments of elevated auditory percepts. Ph.D. thesis, Rensselaer Polytechnic Institute.
- Pastore, M., Y. Zhou, and W.A. Yost. 2020. Cross-modal and cognitive processes in sound localization. In *The Technology of Binaural Understanding*, eds. Blauert, J. and J. Braasch, 315–350. Cham, Switzerland: Springer and ASA Press.
- Patterson, R.D., M.H. Allerhand, and C. Giguère. 1995. Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *The Journal of the Acoustical Society of America* 98 (4): 1890–1894.
- Peck, J.E. 1994. Development of hearing. Part I: Phylogeny. *Journal American Academy of Audiology* 5: 291.
- Peck, J.E. 1995. Development of hearing. Part III: Postnatal development. *Journal American Academy of Audiology* 6: 113.
- Reethof, G., O. McDaniel, and G. Heisler. 1977. Sound absorption characteristics of tree bark and forest floor. In *Proceedings of the Conference on Metropolitan Physical Environment*, eds. L.P. Heisler, M. Gordon and L.P. Herrington, 206–217. Gen. Tech. Rep. NE-25. Upper Darby, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station.
- Reznikoff, I. 2004/2005. On primitive elements of musical meaning. *The Journal of Music and Meaning* 3. www.musicandmeaning.net/issues/showArticle.php?artID=3.2 (last accessed: July 2017).
- Roebroeks, W., N.J. Conard, T. Van Kolfschoten, R. Dennell, R.C. Dunnell, C. Gamble, P. Graves, K. Jacobs, M. Otte, D. Roe, et al. 1992. Dense forests, cold steppes, and the palaeolithic settlement of Northern Europe - [and comments and replies]. *Current Anthropology* 33 (5): 551–586.
- Roman, N., D. Wang, and G.J. Brown. 2003. Speech segregation based on sound localization. *The Journal of the Acoustical Society of America* 114 (4): 2236–2252.

- Roman, N., S. Srinivasan, and D. Wang. 2006. Binaural segregation in multisource reverberant environments. *The Journal of the Acoustical Society of America* 120 (6): 4040–4051.
- Rönne, H. 1915. Zur Theorie und Technik der Bjerrumschen Gesichtsfelduntersuchung [On the theory and technique of Bjerrum's visual field examination]. *Archiv für Augenheilkunde* 78: 284–301.
- Sakai, H., S.-I. Sato, and Y. Ando. 1998. Orthogonal acoustical factors of sound fields in a forest compared with those in a concert hall. *The Journal of the Acoustical Society of America* 104 (3): 1491–1497.
- Sala, N., J.L. Arsuaga, A. Pantoja-Pérez, A. Pablos, I. Martínez, R.M. Quam, A. Gómez-Olivencia, J.M.B. de Castro, and E. Carbonell. 2015. Lethal interpersonal violence in the middle pleistocene. *PLoS one* 10 (5): e0126589.
- Sayers, B.M., and E.C. Cherry. 1957. Mechanism of binaural fusion in the hearing of speech. *The Journal of the Acoustical Society of America* 29: 973–987.
- Stern, R.M., and H.S. Colburn. 1978. Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position. *The Journal of the Acoustical Society of America* 64: 127–140.
- Teret, E., M.T. Pastore, and J. Braasch. 2017. The influence of signal type on perceived reverberance. *The Journal of the Acoustical Society of America* 141 (3): 1675–1682.
- Villa, P., and S. Soriano. 2010. Hunting weapons of Neanderthals and early modern humans in South Africa: Similarities and differences. *Journal of Anthropological Research* 66 (1): 5–38.
- Volf, M. 2012. Effects of soundscapes on attentional capacity. Master's thesis, Rensselaer Polytechnic Institute, Troy, NY.
- Vorländer, M. 1989. Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. *The Journal of the Acoustical Society of America* 86 (1): 172–178.
- Wallach, H. 1939. On sound localization. *The Journal of the Acoustical Society of America* 10 (4): 270–274.
- Wallach, H. 1940. The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology* 27 (4): 339–368.
- Wightman, F.L., and D.J. Kistler. 1992. The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America* 91: 1648–1661.
- Wynn, T., and F.L. Coolidge. 2004. The expert Neandertal mind. *Journal of human evolution* 46 (4): 467–487.
- Wynn, T., and F.L. Coolidge. 2008. A stoneage meeting of minds: Neandertals became extinct while Homo Sapiens prospered. A marked contrast in mental capacities may account for these different fates. *American Scientist* 96 (1): 44–51.
- Zakarauskas, P., and M.S. Cynader. 1993. A computational theory of spectral cue localization. *The Journal of the Acoustical Society of America* 94 (3): 1323–1331.