Evaluation of Binaural Synthesis by Minimum Audible Angles

Florian Völk, Armas Gottschalk, Hugo Fastl

AG Technische Akustik, MMK, Technische Universität München, 80333 München, Deutschland, florian.voelk@mytum.de

Introduction

Binaural synthesis (BS) is widely used as a tool aiming at eliciting specific auditory perceptions by means of headphones, ideally exactly as elicited by a real or hypothetical reference scene. In this contribution, an evaluation method is proposed, addressing the synthesis quality by comparing the minimum audible angles (MAAs) measured in the synthesized versus the corresponding real situation. The evaluation method is discussed by means of the underlying assumptions and conceptual possibilities and limitations. On that basis, an exemplary evaluation of the grid resolution required for transparent BS is presented with regard to implementation factors.

Theoretical Aspects

Dynamic BS is usually implemented based on adaptive signal processing realized by partitioned convolution of an audio signal with dynamically varying binaural impulse response pairs (BIRPs) as discussed by Völk (2011). The variation is controlled based on a geometrical scene description, materials properties, and the position and orientation of the listener's head. Regardless whether the BIRPs are acquired by measurement (data-based) or simulation (model-based), a spatially continuous scene adaption is not possible since in current implementations, the audio signal is processed block-wise. Thus, a discrete spatial grid has to be introduced. If the scenario is static for objects other than the listener, all possible listening situations are at a specific instant of time unambiguously identified by the listener position and orientation (the head is assumed fixed with regard to the body). For dynamically varying acoustic conditions, the positions and orientations of all objects within the scene have to be taken into account. Every single static situation is described by as many BIRPs as there are sound sources. For one or multiple independent sound sources in an anechoic listening environment, the relative angles and distances between sources and listener alone are sufficient to describe the situation completely. For that reason, the virtual sources for a BS of an anechoic listening situation with independent sources can be positioned with the desired orientation at the correct distance under an angle relative to the listener's head. Consequently, if for M_{s} different sources M_{d} distances, M_{so} source orientations, and M_{ho} head orientations are to be considered,

$$M_{\mathrm{ir,ae}} = M_{\mathrm{s}} \cdot M_{\mathrm{d}} \cdot M_{\mathrm{so}} \cdot M_{\mathrm{ho}} \qquad (1)$$

sets of BIRPs are necessary. In case the application scenario allows for different numbers of distances or orientations per source, the effort can be reduced accordingly.

The BIRP update for positioning relative to the head is done by transforming the source location to the current head related coordinates and selecting the BIRP matching closest the resulting angle relative to the head, the intended source orientation, and distance (*head related binaural synthesis*). The spatial resolution is given by the head orientations and distances where BIRPs are available, and so are the possible source positions.

In reverberant listening environments, the absolute positions of the listener and the sources within the scenario have to be taken into account. If for each of M_{s} different sources, M_{spo} source, and M_{hpo} head positions, the same number of head and source orientations are to be simulated, the number of BIRPs required is given by

$$M_{\mathrm{ir}} = M_{\mathrm{s}} \cdot M_{\mathrm{spo}} \cdot M_{\mathrm{hpo}} \cdot M_{\mathrm{so}} \cdot M_{\mathrm{ho}}. \qquad (2)$$

The current BIRP is then selected by transforming the current head orientation and position to world coordinates (*room related binaural synthesis*). The synthesis resolution is given by the head orientations and positions BIRPs are available for, while in contrast to head related BS the possible virtual source positions are given by the source positions where BIRPs are available. A reduction in complexity is possible by synthesizing only virtual sources in the horizontal plane or limitation to one distance, one sound source (one radiation pattern), and one source orientation. For reverberant environments, the number of head and source positions is usually limited, too. Head related BS using BIRPs measured in reverberant environments results in an unrealistic situation. However, the implementation effort is reduced compared to room related BS of the same reference situation.

To allow for high resolution BS, dynamic memory allocation is used here, based on pre-caching BIRPs from the hard drive to the main memory for a region R symmetric around the currently active BIRP. The idea of the pre-caching is that BIRPs adjacent to the currently used pair are available in the main memory accessible with little delay, while pairs at the boundaries of R are updated by pointer-arithmetic with the aim of keeping R symmetric around the currently used pair. The BIRPs at angular positions between the center and the boundaries of R act as a buffer, allowing the pre-caching to be delayed compared to the BIRP access for the convolution.

Proposed Evaluation Procedure

For it is the aim of BS to create the hearing sensations of a reference situation, a rather strict quality evaluation is the comparison of minimum audible differences in synthesis and reference situation. The lowest resolution for a specific BS system resulting in the reference scene MAA can be addressed by MAA measurements with systematically varying angular resolution. While the theory appears straight forward, the evaluation experiment has to be carefully adjusted to the BS situation, since the virtual source positioning is implemented differently in head and room related BS. In head related BS, the head orientations where BIRPs are available determine the synthesis resolution and possible virtual source directions. Therefore, the MAA for head related BS is limited to the angle between two adjacent head orientations where BIRPs are available. Room related BS allows for virtual source positioning independent of the synthesis resolution. For that reason, the resolution required can be addressed by means of MAA measurements without interference between the experimental method and a possibly limited synthesis resolution. Following Perrott and Pacheco (1989), a two-alternative forced choice 2-down/1-up method combined with Parameter Estimation by Sequential Testing for step size adaption is proposed for assessing the MAA. The subjects are asked to indicate where the second hearing sensation occurred in relation to the first (in a specified plane) by pressing one of two buttons, while the presentation sequence is chosen randomly. This procedure is repeated until both, the deviation between the last two minimum and the deviation between the last two maximum values remains smaller than a threshold dependent on the experimental setup. Since this procedure is converging towards the 70.7% point of the psychometric function, the MAA is defined here as the angular threshold where about 71% of all judgments of the relative positions of the sound sources are correct. As sound stimuli, uniform exciting noise (UEN) impulses (Fastl and Zwicker 2007, section 6) are proposed for containing equal intensity in all critical bands and thus being assumed to provide the listener with all spectral localization cues at the same perceptual weight. For the MAA measurement, two UEN impulses are successively presented to the subjects by two virtual sources positioned symmetrically to the left and right of the median plane. According to Blauert and Braasch (2007), impulse durations of more than 200 ms allow for orientating head movements by eliciting so-called dynamic localization cues. To provide dynamic localization cues, 700 ms impulse duration, 20 ms Gaussian gating, and 300 ms pause between the two stimuli is proposed. With regard to audio-visual interactions, the controlled visual stimulus darkness should be applied.

Verification and Discussion

To verify the evaluation procedure proposed here, the horizontal resolution required for a specific unrealistic room related dynamic BS is addressed by systematically varying the angular resolution. The synthesis is set up in the horizontal plane with blocked auditory canal artificial head (Neumann KU 80) BIRPs, average magnitude equalization, and Sennheiser HD 800 headphones selected based on Völk and Fastl (2011a) and Völk (2012). Additional parameters are the fixed virtual source distance 2 m, $M_\mathrm{s} = 2$, $M_\mathrm{spo} = 1$, $M_\mathrm{hpo} = 1$, $M_\mathrm{so} = 1$, and $M_\mathrm{ho} = [360 \ldots 3600]$.

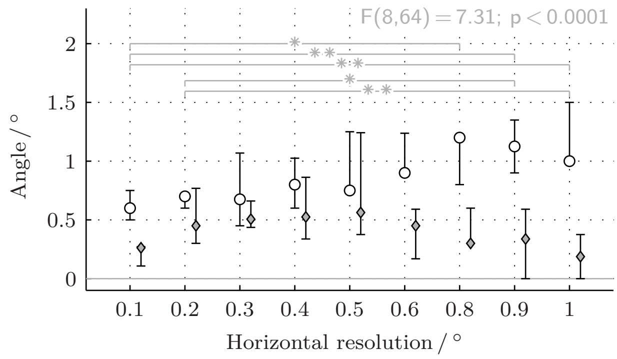

Figure 1: Quartiles of individual medians (open) and interquartile ranges (filled) of minimum audible angles for room related binaural synthesis in the horizontal plane (artificial head binaural impulse response set recorded in reverberant laboratory environment, average headphone equalization, nine normal hearing subjects, three repetitions per stimulus).

The results shown by figure 1 indicate about 0.5° resolution to be sufficient for the average MAA not to significantly decrease further with the grid resolution. This interpretation is additionally supported by one factorial analysis of variance, indicating a highly significant effect of the horizontal grid resolution, where post-hoc comparison reveals significant differences only between the resolutions below 0.3° and above 0.6°. Because of the unrealistic situation and the artificial head measurement, the results are not representative for arbitrary situations and BS with human head measurement. However, the similarity of the resulting angle to the MAA for broadband UEN with real sources indicates the correct order of magnitude and therefore the applicability of the procedure (Völk and Fastl 2011b). Considering individual BS and realistic reference scenes, it is important to take into account the procedural difficulties associated with high resolution measurements, especially on human subjects. In those cases, head movements during the measurements may exceed the grid resolution and thus corrupt the measurements.

This work has been supported in the context of Project FA 140/4 by Deutsche Forschungsgemeinschaft (DFG).

References

Blauert J., J. Braasch: *Chapter* Räumliches Hören. In *Handbuch der Audiotechnik*, ed. by S. Weinzierl. (Springer, 2007)

Fastl H., E. Zwicker: *Psychoacoustics – Facts and Models*. 3rd edition (Springer, Berlin Heidelberg, 2007)

Perrott D. R., S. Pacheco: Minimum audible angle thresholds for broadb. noise as a function of the delay between the onset of the lead and lag signal. J. Acoust. Soc. Am. **85**, 2669–2672 (1989)

Völk F.: System Theory of Binaural Synthesis. In *131st AES Convention* (2011) (Convention Paper 8568)

Völk F.: Headphone Selection for Binaural Synthesis with Blocked Auditory Canal Recording. In *132nd AES Convention* (2012)

Völk F., H. Fastl: Locating the Missing 6 dB by Loudness Calibration of Binaural Synthesis. In *131st AES Convention* (2011a)

Völk F., H. Fastl: Richtungsunterschiedsschwellen für ein zirkulares WFS-System . In *Fortschritte der Akustik, DAGA 2011*, 945–946 (Dt. Gesell. für Akustik e.V., Berlin, 2011b)