Measurement of Time-Variant Binaural Room Impulse Responses for Data-Based Synthesis of Dynamic Auditory Scenes

Nara Hahn and Sascha Spors
Institute of Communications Engineering, University of Rostock, Germany
Email: {nara.hahn, sascha.spors}@uni-rostock.de

Introduction

In binaural synthesis, the ear signals are generated by filtering the source signal with a pair of binaural room impulse responses (BRIRs) that characterize the acoustic transmissions from a sound source to the outer ears [1]. Simulated BRIRs are used in model-based binaural synthesis, whereas measured BRIRs are used in data-based binaural synthesis. In the latter approach, the BRIRs are usually measured in static experimental setups. For the auralization of dynamic scenes, such static BRIRs have to be interpolated or extrapolated which is not straightforward under non-freefield conditions. In this study, an alternative approach for the data-based synthesis of dynamic scenes is presented. The dynamic acoustic system is excited by a test signal, and the response is captured by a dummy head, as shown in Fig. 1. By employing a recently proposed system identification method, the BRIRs are computed for every time instant. Thus, a massive number of BRIRs are obtained that describe the acoustic characteristics of the dynamic system.

Measurement of Time-variant BRIRs

In a dynamic acoustic scene, the acoustic transmission path from a sound source to each ear can be considered as a linear time-varying system,

$$y_{L,R}[n] = \sum_{k=0}^{N_h-1} x[n-k]h_{L,R}[k,n] \tag{1}$$

where $x[n]$ is the source signal, $h_{L,R}[k,n]$ the time-varying BRIRs, and $y_{L,R}[n]$ the corresponding ear signal. It is assumed that the system has a finite-length impulse response N_h, and that the measured signal is free of noise. Hereafter, the subscript L and R are omitted, as the derivation is identical for both ears.

To excite the system in a continuous manner, i.e., sample-by-sample, we employ a specific type of periodic signal $\tilde{x}[n] = \tilde{x}[n+N_x]$ that is commonly referred to as *perfect sequence* [2]. The periodic auto-correlation function of a perfect sequence is an impulse train,

$$\varphi_{\tilde{x}\tilde{x}}[m] = \sum_{n=0}^{N_x-1} \tilde{x}[n+m]\tilde{x}[n] = \sum_{\mu\in\mathbb{Z}} \delta[m-\mu N_x], \tag{2}$$

where $\delta[m]$ is the unit impulse function. The period N_x has to be longer than N_h such that the current BRIR does not overlap with the BRIR from the preceding time instant [3], and as short as possible for better system tracking. In the following derivation, a perfect sequence with the shortest period is considered, $N_x = N_h \equiv N$.

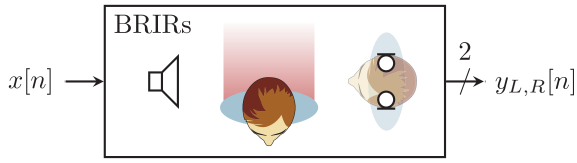


Figure 1: Data-based binaural synthesis of a dynamic scene.

Equation (2) shows that a perfect sequence is orthogonal to any m-shifted version of itself except for $m \bmod N = 0$. This agrees with the geometrical interpretation of perfect sequences provided in [4]. We can thus form an orthogonal basis set for $\mathbb{R}^N$,

$$\{\psi_m[n] = \tilde{x}[-(n-m)], m = 0\ldots N-1\}$$

where each basis sequence is time-shifted and also time-reversed. The reason for time-reversal becomes clear in the following.

Using these basis functions, BRIRs can be decomposed into orthogonal components,

$$h[k,n] = \sum_{m=0}^{N-1} a_m[n]\psi_m[k] = \sum_{m=0}^{N-1} a_m[n]\tilde{x}[m-k], \tag{3}$$

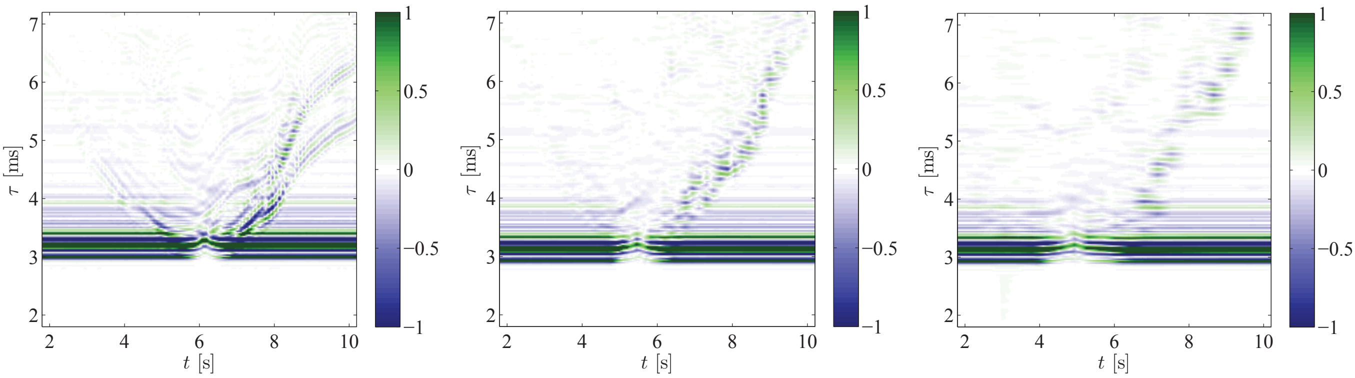
where $a_m[n]$ is the orthogonal expansion coefficient. Now, the output of the system is

$$\begin{aligned} y[n] &= \sum_{k=0}^{N-1} \tilde{x}[n-k]\left(\sum_{m=0}^{N-1} a_m[n]\tilde{x}[m-k]\right) && (4)\\ &= \sum_{m=0}^{N-1} a_m[n]\underbrace{\left(\sum_{k=0}^{N-1} \tilde{x}[m-k]\tilde{x}[n-k]\right)}_{=\delta[m-(n \bmod N)]} && (5)\\ &= a_\nu[n], && (6)\end{aligned}$$

where $\nu \equiv n \bmod N$. This states that the output $y[n]$ is the ν-th expansion coefficient of $h[k,n]$. This also shows that $h[k,n]$ cannot be fully determined, as only one coefficient is known at one time instant. The missing $N-1$ coefficients thus has to be estimated as proposed in [5]. In this study, we estimate $a_m[n]$ by linear interpolation,

$$\hat{a}_m[n] = \begin{cases} (1-\alpha)y[n-\nu+m] + \alpha y[n-\nu+N+m], & m=0\ldots\nu-1\\ y[n], & m=\nu\\ (1-\alpha)y[n-\nu+m] + \alpha y[n-\nu-N+m], & m=\nu+1\ldots N-1\end{cases}$$

where $\alpha = \frac{|m-\nu|}{N}$. This method is illustrated in Fig. 2.



(a) $N_x = 2^{12}$ (b) $N_x = 2^{13}$ (c) $N_x = 2^{14}$

Figure 3: BRIRs (left ear), $h(\tau, t) = h[k/f_s, n/f_s]$. Each vertical slice is one instantaneous BRIR. Only the earlier part ($2\text{ms} \leq \tau \leq 7\text{ms}$) of the BRIRs is shown.

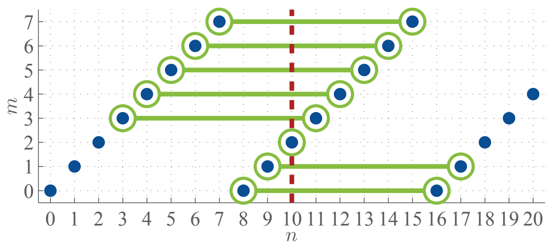


Figure 2: Example illustrating the presented method ($N = 8$ and $n = 10$). To compute $h[k, n]$, $a_m[n](m = 0 \ldots N-1)$ are estimated by linear interpolation. The filled circles ● shows the relation of $y[n]$ and $a_\nu[n]$. The green horizontal lines o—o connect the linear interpolation pairs.

Results and Discussions

By using the presented method, time-varying BRIRs were measured in a dynamic scene. A loudspeaker and a dummy head were placed in fixed positions. During the measurement, a person walked inbetween the loudspeaker and the dummy head, thereby obstructing the direct sound (See. Fig. 1). Perfect sequences with different periods $N_x = 2^{\{12,13,14\}}$ were used [6]. The sampling frequency of the audio signal was $f_s = 44.1$ kHz.

The results are shown in Fig. 3. Each vertical slice of the plots shows the BRIR coefficients at the corresponding moment. Due to the fixed positions of the source and the receiver, the direct sound ($\tau \approx 3$ ms) is almost constant with time. For the same reason, the direct components are identical for different N_x. When the direct transmission path is disturbed, however, the direct sound is attenuated and slightly time-shifted due to diffraction. The first wavefront is broken because of acoustic shadowing. Note that the movement of the obstructer was not accurately reproduced in the experiment, and thus, the obstruction occurs at different moments: (a) $t_{\text{ob}} \approx 6$ s, (b) $t_{\text{ob}} \approx 5.5$ s, and (c) $t_{\text{ob}} \approx 5$ s. Though, the deformed wavefronts look similar.

The reflections off the obstructer vary with time, as they depend on the position of the obstructer. Stronger reflections are captured when the obstructer enters the ipsilateral side, $t > t_{\text{ob}}$. Although direct comparison for different N_x is not possible, the dissimilarities are apparent. While individual reflections are well tracked in 3(a), the reflections are smeared in 3(b) and 3(c). It seems that, the upper limit of the changing rate of a system that can be tracked is determined by N_x. If this limit is exceeded, the system is identified such as it has a lower changing rate. This is also the case in some region in 3(a), where the time-variance is high, ($t \approx 8$ s, $\tau \approx 5$ ms). Such artifacts are attributed to the fact that the orthogonal expansion coefficients are undersampled, which have a sampling rate of $\frac{f_s}{N_x}$. Thus, it can be interpreted as aliasing artifacts.

Ear signals were generated by filtering source signals with the BRIRs (available for download at `http://spatialaudio.net/paper-dynamic-hrtfs/`). In informal listening, the presence of the moving obstructer was well perceived. No suspicious artifacts due to imperfect tracking, mentioned in the previous paragraph, were audible. This suggests that the presented method can be used for the data-based binaural synthesis of dynamic scenes. The perceptual properties of the presented approach are still subject to further investigation.

References

[1] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

[2] H. D. Lüke, *Korrelationssignale*. Springer, 1992.

[3] C. Antweiler and G. Enzner, "Perfect sequence LMS for rapid acquisition of continuous-azimuth head related impulse responses," in *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2009.

[4] C. Antweiler and M. Dörbecker, "Perfect sequence excitation of the NLMS algorithm and its application to acoustic echo control," in *Annales des Telecommunications*, vol. 49, p. 386–397, Springer, 1994.

[5] N. Hahn and S. Spors, "Identification of dynamic acoustic systems by orthogonal expansion of time-variant impulse responses," in *International Symposium on Communications, Control and Signal Processing (ISCCSP)*, 2014 (*accepted*).

[6] Y. Suzuki, F. Asano, H.-Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *The Journal of the Acoustical Society of America*, vol. 97, no. 2, p. 1119–1123, 1995.