EFFICIENT RANGE EXTRAPOLATION OF HEAD-RELATED IMPULSE RESPONSES BY WAVE FIELD SYNTHESIS TECHNIQUES

Sascha Spors and Jens Ahrens

Quality and Usability Lab, Deutsche Telekom Laboratories, Technische Universität Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, Germany
Email: {Sascha.Spors,Jens.Ahrens}@telekom.de

ABSTRACT

Head-related impulse responses (HRIRs), which are measured from an acoustic source to the left and right ear, characterize the acoustic properties of the outer ear. Due to the involved measurement effort, HRIRs are typically available only for source positions on a circle or surface of a sphere. Range extrapolation techniques aim at calculating HRIRs with a different source distance. A number of techniques have been proposed in the past which are based on expansions of HRIRs in terms of surface spherical harmonics. This paper presents an alternative approach which bases on the interpretation of a BRIR dataset as virtual loudspeaker array which is driven by Wave Field Synthesis in order to achieve range extrapolation.

Index Terms— head-related transfer functions, range extrapolation, interpolation, Wave Field Synthesis

1. INTRODUCTION

1.1. Head-Related Impulse Responses (HRIRs)

The human auditory system exploits the acoustic characteristics of the outer ear in order to deduce spatial information [1]. The outer ear includes the pinnae, head and upper torso. The acoustic properties from the outer ear can be captured by measuring the impulse response from an acoustic source to a defined position in the ear canal of both ears. These functions are known as head-related impulse responses (HRIRs) or in the frequency domain as head-related transfer functions (HRTFs). They are typically measured in anechoic environments. HRIRs are frequently used for virtual auditory environments (VAEs), where virtual sound sources are created by filtering the signal of the desired source by the left/right ear HRIR, and reproducing the resulting signals via headphones. This reproduction method is often referred to as *binaural reproduction*. HRIRs depend in general on the position of the listener, the position of the acoustic source, the acoustic properties of the source and the environment. Hence for VAEs, a database of HRIRs is required that ideally covers all potential parameters.

1.2. Measurement of HRIRs

In order to limit the measurement effort, HRIRs are typically measured for a fixed head position and source positions having a fixed distance and varying angles with respect to the head. For simplicity, we consider a two dimensional scenario in the sequel were the sources are located in the horizontal plane. Figure 1 illustrates the considered geometry. The impulse responses from a source at position $\mathbf{x}_0 = R\ [\cos(\alpha_0)\ \ \sin(\alpha_0)]^T$ to the left and right ear is measured under free-field conditions. The left/right BRIR depends in general on the head orientation φ, the source distance R and source angle α_0. The HRIR is denoted by $h_{\{\mathrm{L,R}\}}(\alpha_0, R, \varphi, t)$ throughout this paper.

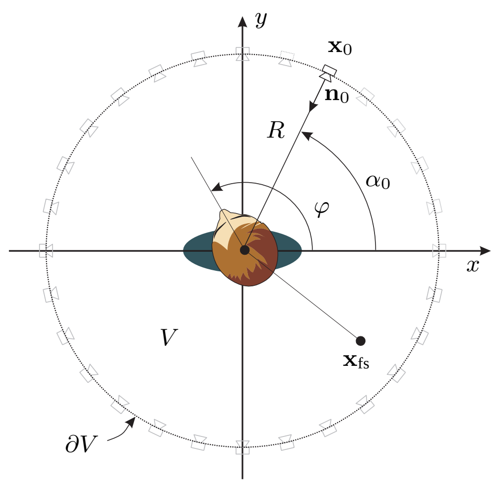


Fig. 1: Geometry used for measurement and range extrapolation of HRIR datasets.

It is obvious that the measurement of HRIR datasets becomes a complex task for a densely sampled space of source positions. Therefore, most of the currently available datasets consider only one measurement distance with a limited, sometimes not constant, angular resolution. Typical distances are in the range of 1.5 to 3 meters with an angular resolution of 5 to 10 degrees. It is generally assumed that further increasing the distance does not change the characteristics of the HRIRs essentially [1]. However, it is known that HRIRs change significantly for nearby sources [2]. Such HRIRs are typically termed as near-field HRIRs. Note, the term near-field is not used in a strict physical sense in this context. Near-field HRIRs are very interesting for VAEs. However, since the measurement is complex and the resulting datasets may become large, a number of techniques have been proposed the compute near-field HRIRs from HRIRs measured at a fixed distance. These will be reviewed briefly in the following subsection.

1.3. Range Extrapolation of HRIRs

A lot of work has been performed in the area of angular interpolation of HRIRs. We will not review these techniques since such interpolation techniques are not in the scope of this paper. We focus on the computation of near-field HRIRs. Due to the close relation of the problem to sound field extrapolation such methods have been termed as range extrapolation of HRIRs. We will exemplarily review two different approaches in this field.

In [3, 4] two techniques are introduced which are based upon the expansion of HRIRs into surface spherical harmonics. These form a orthogonal basis that can be used for angular interpolation and range extrapolation. However, the expansion into spherical harmonics is only valid for a region without sources/scatterers. Hence it can not be applied straightforwardly to measured HRIRs due to the scattering of the head/torso. A solution to this problem has been found by considering the reciprocity of the acoustic wave equation. Here it is assumed that the transfer paths from the ears to the source positions are measured. The expansion is then performed with respect to the source positions. However, extrapolating the HRIRs is only possible if no scattering objects are located within the extrapolation region. The application of these techniques to the computation of near-field HRIRs is therefore limited to distances not including reflections from the upper torso.

Another class of techniques uses the concept of virtual loudspeaker arrays. The HRIR measurements characterize the acoustic paths from a circular/spherical distribution of loudspeakers to the ears. The distribution of sources can be interpreted as a virtual loudspeaker array. When this array is driven with appropriate signals to synthesize a desired virtual source, the HRIRs from the virtual source to the ears is synthesized. In [5, 6] two techniques have been introduced which are based on higher-order Ambisonics (HOA) for the computation of the driving signals. The synthesis of virtual sources at closer distances than the loudspeaker array may be subject to instabilities when no countermeasures are taken.

Both approaches are based upon expansions into spherical harmonics. A general problem is that such expansions may become numerically unstable for sparsely sampled or unequally sampled datasets, and their numerical complexity. In this paper we also follow the concept of a virtual loudspeaker array. However, we use WFS for the computation of the driving signals. This has a number of benefits as will be illustrated in the following sections.

2. WAVE FIELD SYNTHESIS

Sound field synthesis (SFS) techniques aim at synthesizing the sound field $S(\mathbf{x},\omega)$ of a virtual source accurately inside a potentially large listening area. WFS is a well established technique in this context.

2.1. Basic Principles

The concept of WFS [7] has initially been developed for linear distributions of secondary sources and has later on been extended to arbitrarily shaped convex distributions. WFS is based on the Kirchhoff-Helmholtz integral, which states that SFS can be realized by a distribution of secondary monopole and dipole sources located on the boundary ∂V of the listening area V which are driven by the directional gradient and the pressure of the sound field of the virtual source $S(\mathbf{x},\omega)$, respectively.

WFS is based on a number of reasonable approximations in order to realize this principle by a continuous distribution of secondary monopole sources. Using point sources for the synthesis of a sound field in a plane constitutes a so called 2.5-dimensional problem. The resulting properties are somewhere between two and three dimensional synthesis [8].

The synthesized sound field $P(\mathbf{x},\omega)$ for the considered geometry (see Fig. 1) is given by [8]

$$P(\mathbf{x},\omega) = \int_0^{2\pi} D(\alpha_0,\omega)\,\frac{1}{4\pi}\,\frac{e^{-j\frac{\omega}{c}|\mathbf{x}-\mathbf{x}_0|}}{|\mathbf{x}-\mathbf{x}_0|}\,R\,d\alpha_0\ , \qquad (1)$$

where $\mathbf{x_0} = R[\cos(\alpha_0)\ \sin(\alpha_0)]^T$ denotes a position on the boundary ∂V, $D(\alpha_0,\omega)$ the driving signal for a secondary source at $\mathbf{x_0}$. The driving signal is given by [8]

$$D(\alpha_0,\omega) = 2R\,a(\alpha_0)\,c_{2.5\text{D}}(\alpha_0,\omega)\,\frac{\partial}{\partial\mathbf{n}_0}S(\alpha_0,R,\omega)\ , \qquad (2)$$

where $a(\alpha_0)$ denotes a window function, $c_{2.5\text{D}}(\alpha_0,\omega)$ a correction for 2.5-dimensional synthesis and $\frac{\partial}{\partial\mathbf{n}_0}$ the directional gradient evaluated at position $\mathbf{x}_0$ with the normal vector $\mathbf{n}_0(\alpha_0) = -[\cos(\alpha_0)\ \sin(\alpha_0)]^T$.

The window function $a(\alpha_0)$ takes care that only those secondary sources are active where the local propagation direction of the virtual source at the position α_0 has a positive component in direction of the normal vector $\mathbf{n}_0$ of the secondary source. The correction $c_{2.5\text{D}}(\alpha_0,\omega)$ depends in general on the desired virtual sound field and the geometry of the secondary source distribution. Note, that (1) together with (2) can be interpreted as a high-frequency approximation of the sound field extrapolation problem.

2.2. Focused Sources

The goal is to create the illusion of an acoustic source that is situated in front of the loudspeaker array. Since the secondary sources emit a sound field that travels towards the listener, one can only expect that the desired sound field of a focused source is correct if the focus point is located in between the active secondary sources and the listener. In the context of SFS, this is a well known limitation of focused sources [9].

The driving function for a focused source can be derived by modeling an acoustic sink at the focus point together with a sensible selection of the active secondary sources. It reads [9]

$$D_{\text{fs}}(\alpha_0,\omega) = a_{\text{fs}}(\alpha_0)\sqrt{-j\frac{\omega}{c}}\hat{S}_{\text{fs}}(\omega)R\sqrt{\frac{2}{\pi}}\frac{(\mathbf{x}_0-\mathbf{x}_{\text{fs}})^T\mathbf{n}_0}{|\mathbf{x}_0-\mathbf{x}_{\text{fs}}|^{3/2}}e^{j\frac{\omega}{c}|\mathbf{x}_0-\mathbf{x}_{\text{fs}}|}\ , \qquad (3)$$

where $\mathbf{x}_{\text{fs}} = r_{\text{fs}}[\cos(\alpha_{\text{fs}})\ \sin(\alpha_{\text{fs}})]^T$ with $r < R$ denotes the position of the focused source and $\mathbf{n}_{\text{fs}}$ its nominal orientation. The window function a_{fs} is given by

$$a_{\text{fs}}(\alpha_0) = \begin{cases} 1 & \text{, if } \langle \mathbf{x}_{\text{fs}}-\mathbf{x}_0, \mathbf{n}_{\text{fs}}\rangle > 0, \\ 0 & \text{, otherwise.} \end{cases} \qquad (4)$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product. Inverse Fourier transformation of the driving signal (3) yields

$$d_{\text{fs}}(\alpha_0,t) = a_{\text{fs}}(\alpha_0)s(t) * h(t) * w(\alpha_0)\,\delta(t+\tau_0(\alpha_0))\ , \qquad (5)$$

where $*$ denotes convolution, $\delta(\cdot)$ the Dirac delta function and $s(t)$ the input signal of the virtual source. The impulse response $h(t)$ denotes the inverse Fourier transformation of $\sqrt{-j\omega/c}$. All frequency independent weights of the driving function are collected in $w(\alpha_0)$. The shift in the Dirac delta function is given as $\tau_0(\mathbf{x}_0) = |\mathbf{x}_0-\mathbf{x}_{\text{fs}}|/c$. This essentially constitutes an anticipation of the source signal. Typically, a pre-delay is introduced to ensure causality in practical implementations. Hence, the driving function can be computed efficiently by (i) filtering the signal of the virtual source $s(t)$ with the filter $h(t)$ and (ii) weighting/delaying this prefiltered signal. No numerical issues are present for arbitrary source positions and geometries of the secondary source distribution.

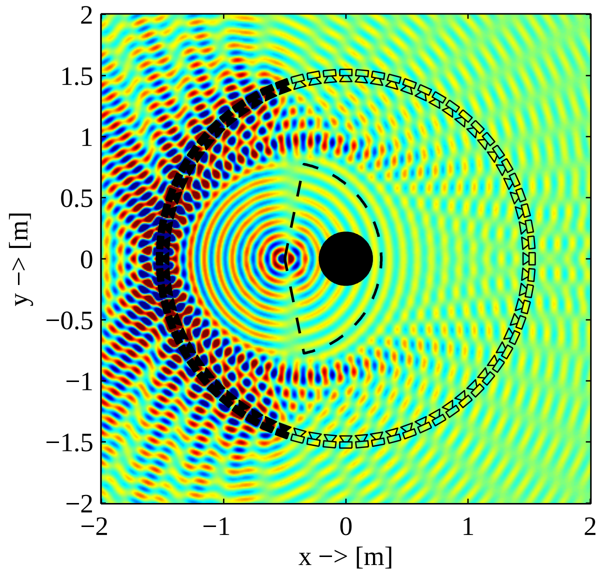


Fig. 2: Synthesis of a focused source with WFS using a circular loudspeaker array ($N = 72$, $R = 1.5$ m, $f_{\text{fs}} = 3$ kHz, $\mathbf{x}_{\text{fs}} = [-0.5\ 0]^T$ m, $\mathbf{n}_{\text{fs}} = [1\ 0]^T$). The active loudspeakers are filled.

2.3. Spatial Sampling

A limited number of loudspeakers is used in practice to realize the secondary source distribution. This constitutes a spatial sampling of the driving function at the spatially discrete loudspeaker positions. The resulting spatial sampling artifacts are well investigated for non-focused [10] and focused sources [9]. With non-focused sources it is such that, above a given frequency termed as spatial aliasing frequency f_{al}, artifacts are apparent anywhere in the listening area. The appearance of spatial sampling artifacts for focused sources depends on the distance to the focus point. In comparison to non-focused sources, focused sources allow the accurate synthesis for much higher frequencies in the vicinity of the focus point.
This is illustrated with an example. Figure 2 shows the synthesized sound field for a focused source using WFS. For the simulated situation and non-focused sources sampling artifacts would be present for frequencies above $f_{\text{al}} \approx 1.5$ kHz [10]. However, it is obvious from Fig. 2 that almost no spatial sampling artifacts are present in a circular region with a radius of about 0.75 m around the focus point for $f_{\text{fs}} = 3$ kHz.
The size of the region without major artifacts can be estimated by considering the spatial dimensionality that can be synthesized by a limited number of secondary sources [11]. The radius r_{al} of the region with negligible sampling artifacts can be estimated as

$$r_{\text{al}} < \frac{N'_{\text{a}} c}{\pi e f} \ , \tag{6}$$

where $N'_{\text{a}} = 2N_a + 1$ with N_a denoting the number of active loudspeakers. Note, that the effective listening area is further limited by truncation [9]. The resulting listening area, as well as, the typical head size is indicated in Fig. 2 by the dashed line and the filled circle, respectively.

3. EFFICIENT COMPUTATION OF NEAR-FIELD HRIRS

The following section illustrates the application of WFS to the efficient computation of near-field HRIRs. The basic scheme is introduced, as well as, some practical aspects that have to be considered for implementation of the proposed technique.

3.1. Basic Scheme

The synthesis equation (1) of WFS states that a superposition of the sound field of appropriately driven secondary sources allows the synthesis of almost any desired virtual sound field. The terms besides the driving function in the integral of (1), characterize the sound propagation of a monopole source under free-field conditions. The sound propagation from one source position to both ears is captured in the respective BRIRs. An entire BRIR dataset consequently captures the propagation paths from a distribution of sources to both ears. The concept of the proposed technique for range extrapolation of HRIRs is to synthesize a focused source using the measured points as virtual loudspeaker array driven by WFS.
Using the above outlined concept the computation of near-field HRIRs using WFS techniques is straightforward. BRIR datasets are typically measured on a finite number of positions with fixed distance to the listener (see Fig. 1). Hence, spatial sampling of (1) is required. For simplicity it is assumed that the BRIR dataset has been taken at N equiangular positions $\alpha_{0,n}$. The extension to non-equiangular sampling is straightforward. Introducing the driving function $d_{\text{fs}}(\alpha_{0,n}, t)$ of a focused source into the spatially discrete time-domain counterpart of (1) yields the range extrapolated BRIRs for the left/right ear

$$h_{\{\text{L,R}\}}(\alpha, r, \varphi, t) = \sum_{n=1}^{N} d_{\text{fs}}(\alpha_{0,n}, t) * h_{\{\text{L,R}\}}(\alpha_{0,n}, R, \varphi, t) \ , \tag{7}$$

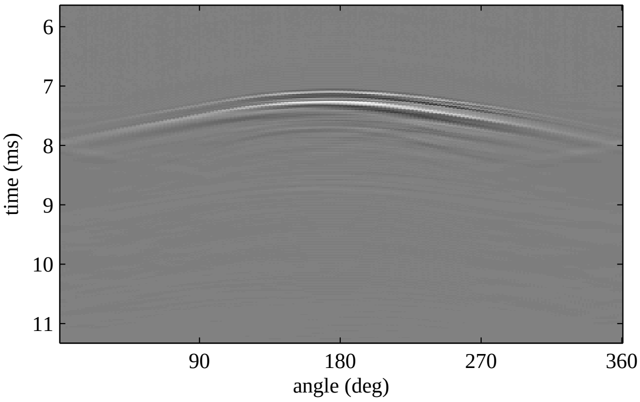
with $\alpha_{\text{fs}} = \alpha$ and $r_{\text{fs}} = r$. Introducing the driving function for a focused source (5) with $\hat{s}(t) = \delta(t)$ into (7) allows to deduce a very efficient realization

$$h_{\{\text{L,R}\}}(\alpha, r, \varphi, t) = \\ h(t) * \sum_{n=1}^{N} a_{\text{fs}}(\alpha_{0,n}) w(\alpha_{0,n}) h_{\{\text{L,R}\}}(\alpha_{0,n}, R, \varphi, t + \tau_{0,n}) \ . \tag{8}$$

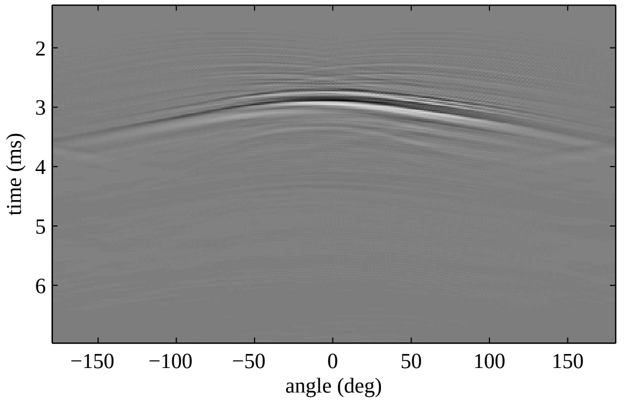
Hence, range extrapolation of HRIRs for $r < R$ can be performed by summing up weighted and delayed versions of the measured HRIRs and filtering the resulting signal by the pre-filter of WFS. The nominal direction of the focused source should be chosen as $\mathbf{n}_{\text{fs}} = -[\cos(\alpha)\ \ \sin(\alpha)]^T$. Note, the presented method implicitly also performs (angular) interpolation of BRIRs. A range extrapolated BRIR dataset can be computed by evaluating (8) for M focused source positions α_M on a circle with radius r.

3.2. Practical Aspects

A number of aspects have to be considered for the practical implementation of the proposed approach using (8).
The realization of the delay operation by quantization to the temporal sampling interval may result in audible artifacts. This may require to usage fractional delay filters. However, the perceptual impact of these artifacts has to be evaluated in the future. Ideally, the measured HRIR dataset should have been sampled densely enough to avoid spatial sampling artifacts. This might not be always the case and hence spatial aliasing artifacts may be present in the extrapolated



(a) measured HRIR dataset ($R = 2.5$ m)



(b) extrapolated HRIR dataset ($r = 0.5$ m)

Fig. 3: Measured and range extrapolated HRIR dataset for the left ear ($\alpha = 180^{\text{o}}$).

HRIRs. However, for typical HRIR datasets these artifacts will only be present for very high frequencies. The perceptual relevance of such artifacts is not clear at the current state. It is further well known that the 3dB per Octave pre-equalization filter of WFS should be flattened out above the spatial aliasing frequency [12]. The window function a_{fs} may produce truncation artifacts for specific positions of the focused source. This problem can be overcome by weighting the window function with a spatial window to smoothen its ends [8].

4. RESULTS

The proposed technique for range extrapolation of HRIRs has been implemented in MATLAB. The HRIR dataset that has been used for the experiment has been recorded by the FABIAN mannequin [13]. It consists of $N = 288$ equiangular measurements taken at a source distance of $R = 2.5$ m. Figure 3a shows the HRIRs for the left ear. A range extrapolated dataset consisting of $M = 288$ equiangular positions with $r = 0.5$ m has been computed. The resulting frequency up to which no sampling artifacts are present, as given by (6), is $f_{\text{al}} \approx 16$ kHz. The pre-equalization filter $h(t)$ has been designed accordingly. The extrapolated HRIRs for the same ear are shown in Fig. 3b. The results show that range extrapolation of HRIRs using the proposed technique works without major artifacts. Informal listening by the authors has further confirmed that the distance cues in the range extrapolated HRIRs are reconstructed correctly [1].

5. CONCLUSIONS

We have presented a novel method for range extrapolation of HRIRs. The underlying concept is the interpretation of a BRIR dataset as virtual loudspeaker array which is driven by WFS. Although, the method is not limited to the computation of near-field HRIRs, we have chosen this example due to its practical relevance. Note, also non-focused source models could be used, e.g. like point sources or plane waves to compute range extrapolated HRIR datasets for $r > R$. The method facilitates an efficient implementation by weighted delay lines and is numerically stable. Furthermore, non-equiangular sampled HRIR datasets can be used straighforwardly. Although we introduced the method for two-dimensional scenarios, its generalization to three dimensions is straightforward using three dimensional WFS [8]. A formal listening experiment will be carried out in the future to perceptually validate the method.

6. REFERENCES

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, 1996.

[2] T. Lentz, *Binaural Technology for Virtual Reality*, Ph.D. thesis, RWTH Aachen, 2007.

[3] R. Duraiswami, D.N. Zotkin, and N.A. Gumerov, “Interpolation and range extrapolation of HRTFs,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

[4] W. Zhang, T.D. Abhayapala, and R.A. Kennedy, “Modal expansion of HRTFs: continuous representation in frequency-range-angle,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Japan, 2009.

[5] D. Menzies and M. Al-Akaidi, “Nearfield binaural synthesis and Ambisonics,” *Journal of the Acoustic Society of America*, vol. 121, no. 3, pp. 1559–1563, March 2007.

[6] M. Noisternig, A. Sontacchi, T. Musil, and R. Höldrich, “A 3D ambisonic based binaural sound reproduction system,” in *AES 24th International Conference on Multichannel Audio*, Banff, Canada, June 2003, Audio Engineering Society (AES).

[7] A.J. Berkhout, “A holographic approach to acoustic control,” *Journal of the Audio Engineering Society*, vol. 36, pp. 977–995, December 1988.

[8] S. Spors, R. Rabenstein, and J. Ahrens, “The theory of wave field synthesis revisited,” in *124th AES Convention*. Audio Engineering Society (AES), May 2008.

[9] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens, “Physical and perceptual properties of focused sources in wave field synthesis,” in *127th AES Convention*. Audio Engineering Society (AES), October 2009.

[10] S. Spors and R. Rabenstein, “Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis,” in *120th AES Convention*, Paris, France, May 2006, Audio Engineering Society (AES).

[11] G. Dickins, M. Williams, and L.W. Hanlen, “On the dimensionality of spatial fields with restricted angle of arrival,” in *International Symposium on Infomation Theory*, Adelaide, SA, September 2005, pp. 1033 – 1037.

[12] S. Spors and J. Ahrens, “Analysis and improvement of pre-equalization in 2.5-dimensional wave field synthesis,” in *128th AES Convention*, London, UK, May 2010, Audio Engineering Society (AES).

[13] A. Lindau and S. Weinzierl, “FABIAN - An instrument for software-based measurement of binaural room impulse responses in multiple degrees of freedom,” in *Proc. of the Tonmeistertagung*, Hannover, Germany, Nov. 2008.

[1] audio samples can be downloaded at http://audio.qu.tu-berlin.de/?p=456