A new method for B-format to binaural transcoding

Svein Berge[*], Natasha Barrett

AES 40th international conference, Tokyo, Japan, October 8–10, 2010

Abstract

A frequency-domain parametric method for transcoding first-order B-format signals to a binaural format is introduced. The method provides better spatial sharpness than linear methods allow. A high angular resolution planewave decomposition of the B-format establishes two independent direction estimates per time/frequency bin. This alleviates the requirement that the sound sources in a mix are W-disjoint orthogonal, implicit in previous nonlinear methods. The characteristics and causes of audible artifacts are discussed. Methods are introduced that suppress the different types of artifacts. A listening test is presented that ranks the sound quality of the method between third-order and fifth-order linear ambisonics systems.

1 Introduction

The B-format for spatial audio was developed as part of the ambisonic system of sound recording and reproduction [8]. The format consists of four channels, W, X, Y and Z. The W channel represents the acoustic pressure at a point in space, while the other channels represent the components of the pressure gradient at the same point.

Much work has gone into the development of methods for decoding B-format signals for playback over various loudspeaker layouts [9, 6]. The problem of decoding, or *transcoding* to a binaural format intended for playback over headphones has received less attention. This problem is often considered a corollary to loudspeaker decoding, since loudspeaker feeds can be converted into a binaural signal using virtual loudspeakers, i.e. convolution with head-related impulse responses [17]. Several authors [15, 16] have noted that gains in computational efficiency can be made by combining the linear decoding and convolution operations. In that case, the number of virtual loudspeakers has no influence on computational cost.

The minimum number of virtual loudspeakers is four, which renders sharply localized auditory events when sound sources roughly coincide with one of the virtual loudspeakers, but very blurry or dual auditory events when sound sources fall between virtual loudspeakers. The maximum number of virtual loudspeakers is equal to the size of the HRTF dataset. This gives an isotropic, but somewhat blurry rendering of the auditory scene.

Nonlinear frequency-domain methods have the potential to produce sharp, isotropic renderings [14, 18]. These methods use a direction estimate and diffuseness measure for each time/frequency bin to steer each bin to a small subset of a large number of virtual loudspeakers. This approach works well when the different sounds in the mixture are W-disjoint orthogonal, i.e. occupy distinct time-frequency bins. When this is not the case, as in most musical mixtures, it can be argued that the human auditory system has a limited capability to localize simultaneous sounds that overlap in time and frequency. Although limited, this capability is not entirely absent [10].

The current paper proposes a parametric decomposition of the B-format signal which is capable of localizing two sound sources per time-frequency bin. A method is presented that uses this decomposition to create a sharp, isotropic, stable and artifact-free binaural transcoding of B-format signals.

2 Parametric decomposition

The method operates in the time/frequency domain. In this section, we will only consider a single frequency band, where the B-format signal is represented by eight numbers: The real and imaginary part of each channel. Considering nothing but the number of degrees of freedom, it seems plausible that the signal should be

[*] Berges Allmenndigitale Rådgivningstjeneste, Oslo, Norway

possible to decompose into two planewaves, since each is represented by four independent numbers: The real and imaginary part of the amplitude and a three-element unit vector representing its direction of travel. If w, x, y and z are the complex-valued signals, then the decomposition can be written as

$$\underbrace{\begin{bmatrix} \sqrt{2}w \\ x \\ y \\ z \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} 1 & 1 \\ x_1 & x_2 \\ y_1 & y_2 \\ z_1 & z_2 \end{bmatrix}}_{\mathbf{V}} \underbrace{\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{\mathbf{A}}, \tag{1}$$

where the bottom three rows of **V** contain real-valued unit vectors pointing in the directions of arrival and **A** contains the complex amplitudes of those waves. To find this decomposition, the real and imaginary parts of **X** must first be split into separate columns. The resulting 4-by-2 matrix is decomposed with a QR decomposition [12]:

$$\begin{bmatrix} \Re(\mathbf{X}) & \Im(\mathbf{X}) \end{bmatrix} = \mathbf{QR} \tag{2}$$

The QR decomposition finds two matrices: The 4-by-2 matrix **Q**, whose columns are orthonormal, and the 2-by-2 matrix **R**, which is upper triangular. Through some basic matrix operations, we will transform **Q** into the matrix **V**, which satisfies the following conditions:

1. For each column, the sum of the square of the bottom three elements is equal to the square of the top element
2. The top element is equal to 1

We begin by satisfying condition 1, which implies that the square of the ℓ^2-norm of each column is twice the square of its top element. Since the columns of **Q** are orthonormal, the square of the ℓ^2-norm of a linear combination of these columns is equal to the sum of the square of the mixing coefficients. It is therefore relatively easily to verify that condition 1 is satisfied by matrix **D**, where **Q** has been multiplied with the “mixing matrix” **C**:

$$b = \sqrt{2(Q_{11}^2 + Q_{12}^2) - 1} \tag{3}$$

$$\mathbf{C} = Q_{11} \begin{bmatrix} 1 & 1 \\ -b & b \end{bmatrix} + Q_{12} \begin{bmatrix} b & -b \\ 1 & 1 \end{bmatrix} \tag{4}$$

$$\mathbf{D} = \mathbf{QC} \tag{5}$$

Once condition 1 is satisfied, each column can be multiplied or divided by a scalar without violating condition 1. Satisfying condition 2 is now simply a matter of dividing each column by its top element:

$$\mathbf{V} = \mathbf{D} \begin{bmatrix} D_{11}^{-1} & 0 \\ 0 & D_{12}^{-1} \end{bmatrix} \tag{6}$$

Although the amplitudes of the planewaves are not actually used in the proposed method, we show here how they can be calculated to complete the decomposition. Since **V** was obtained by right-multiplying **Q** with two matrices, **A** can be obtained by left-multiplying **R** with the inverse of these matrices. To collect the real and imaginary parts into a single column, we finally multiply by the vector $[1, i]^T$

$$\mathbf{A} = \begin{bmatrix} D_{11} & 0 \\ 0 & D_{12} \end{bmatrix} \mathbf{C}^{-1} \mathbf{R} \begin{bmatrix} 1 \\ i \end{bmatrix} \tag{7}$$

There are cases ($Q_{11}^2 + Q_{12}^2 < \frac{1}{2}$) where the decomposition does not exist. In an isotropic noise field this concerns 1/4 of all samples. In real sound recordings, however, this percentage is lower, and low-energy frequency bands are over-represented. The energy contained in these frequency bands usually sums up to around 2 to 3 percent of the total energy. These cases must be handled with an alternative method. Since they concern a small fraction of the signal, the choice of alternative method makes no perceptible difference to the sound, and these methods will not be treated further in this paper.

3 Transcoder implementation

In order to apply the parametric planewave decomposition to a signal, the signal needs to be transformed into the time/frequency domain. The signal is first split into overlapping blocks of samples. Each block is then multiplied with a window function. Before transforming the block into the frequency domain with an FFT, zero padding is added to reduce the effects of wrap-around.

A straight-forward transcoder design would then decompose each time/frequency bin into planewaves and multiply the complex amplitude of each planewave with the HRTF coefficients associated with the relevant direction and frequency. Finally one would convert the blocks back into the time domain with an IFFT and add them to the output of the overlapping, neighbouring blocks.

The problem with this design is that it produces severe artifacts. The output is effectively related to the

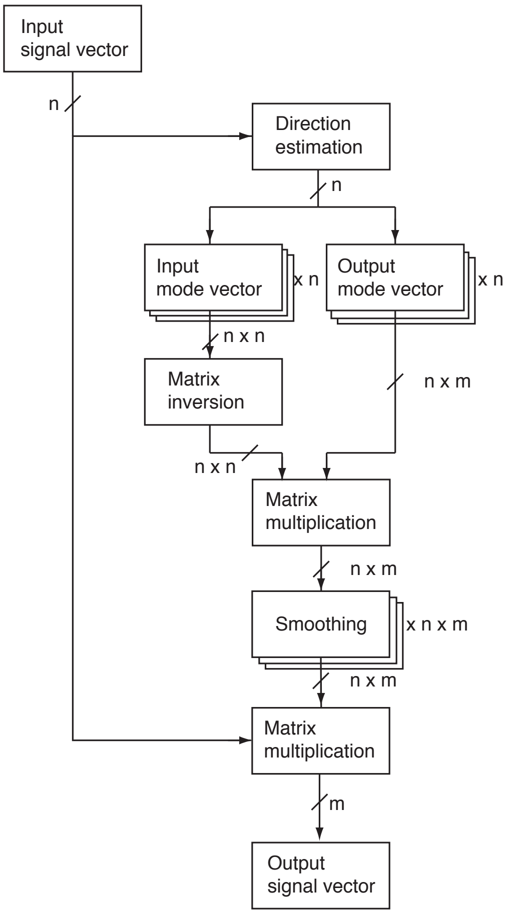


Figure 1: Complete transcoder, reproduced from [3]. In this case, the number of input channels $n = 4$ and the number of output channels $m = 2$.

input through transfer functions that may have sharp edges in the frequency domain and change rapidly in the time domain. These edges and changes need to be smoothed, but in order to do that we need access to the explicit transfer functions that connect the input to the output.

Figure 1 outlines the proposed transcoder design, omitting the transformations into and out of the time/frequency domain. For each time/frequency bin, a traditional decoding matrix is computed that decodes the input signal to four virtual loudspeakers. This matrix is obtained by inverting a matrix whose columns are the input mode vectors,

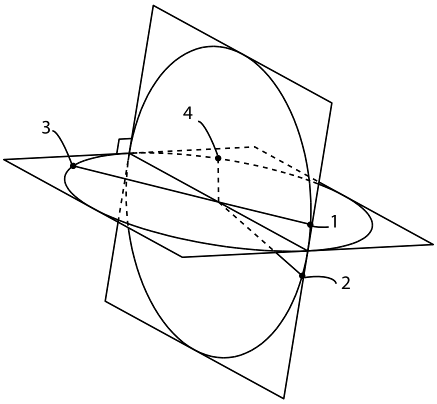


Figure 2: Two direction estimates (1) and (2) are complemented with two "opposite" directions (3) and (4)

$$m_i = \begin{bmatrix} 2^{-1/2} \\ x_i \\ y_i \\ z_i \end{bmatrix} \qquad (8)$$

where $< x_i, y_{i,}, z_i >$ are unit vectors pointing towards the virtual loudspeakers. Two of these are placed in the precise directions determined by the planewave decomposition. The direction of the other two are not estimates of sound source locations, and are only chosen to ensure good conditioning of the matrix inversion, as illustrated in Figure 2.

Output mode vectors are computed that convert each virtual loudspeaker signal into a pair of headphone signals. These are head-related transfer functions that have undergone some processing which will be treated in the next section.

The decoding matrix and the matrix of output mode vectors are multiplied before the resulting matrix is multiplied with the input signal vector. This matrix decodes and re-encodes the signal in one operation. It contains the explicit transfer functions that we need to smooth. The required smoothing can be ensured with simple one-pole filters, both along the frequency axis and between consecutive blocks.

4 Head-related transfer functions

In principle, any panning function can be used as output mode vector, depending on the loudspeaker config-

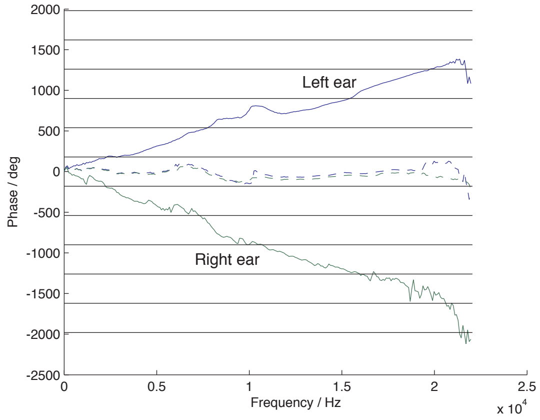


Figure 3: Measured, unwrapped phase response (from [21]) for sound sources straight ahead (– –) and at 45° left azimuth (–) in left and right ear canal. Horizontal lines are 360° apart. The roll-off above 20 kHz can more easily be explained by filters in the signal chain than physics related to the HRTFs.

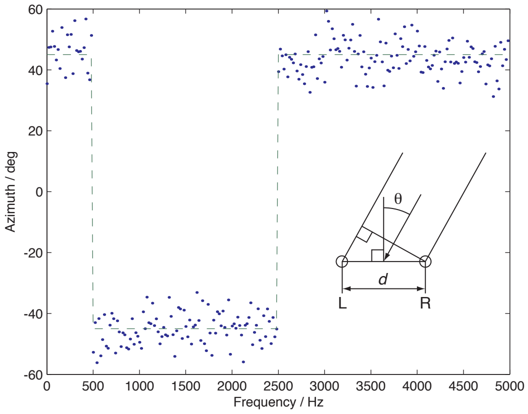


Figure 4: Ideal (– –) and noisy (·) azimuth

uration. Using HRTFs as mode vectors presents challenges not encountered with most other panning functions. Firstly, since HRTF sets are measured in discrete directions, some form of interpolation may be necessary to create a continuous panning function. This problem is not trivial and much work has gone into its solution [11]. When using large HRTF sets, one may simply choose to pick the closest measurement in each instance.

Secondly, HRTFs are complex-valued functions, where the phase part encodes inter-aural time delay. The phase shift is roughly proportional to frequency (see Figure 3), so any uncertainty in the direction estimates translates into a phase uncertainty that is also proportional to frequency. At low frequencies, this is not problematic, and the resulting phase noise is sufficiently reduced in the already established smoothing process. At high frequencies, the required level of smoothing would adversely affect the discrimination between sources. This problem has been encountered before, with DirAC-based binaural transcoding [13]. A new solution is presented in the following, based on the well established fact that our auditory system is insensitive to inter-aural phase delay at high frequencies. The ability to notice interaural phase delay deteriorates gradually from 600 Hz and vanishes above approximately 1.4–1.6 kHz, while inter-aural group delay contributes to localization over the full audible range [5].

4.1 Suppression of phase noise

For the sake of clarity, let us study the process by way of example, using only a single direction estimate and simplify the HRTFs to an idealized form. The frequency-dependent source direction in our example is

$$\theta(2\pi f) = \begin{cases} \pi/4 & f < 500\,\text{Hz} \\ -\pi/4 & 500\,\text{Hz} \leq f < 2.5\,\text{kHz} \\ \pi/4 & 2.5\,\text{kHz} \leq f \end{cases} \tag{9}$$

Let us sample this function at discrete frequencies and add noise to simulate errors caused by interfering sources, noise and reverberation:

$$\tilde{\theta}_i = \theta(\omega_i) + X_i, \tag{10}$$

where i is the index of the frequency band and X is a Gaussian process with $\sigma = 0.1$ and zero mean. This azimuth and its estimate are illustrated in Figure 4.

Without losing features relevant to the current explanation, we can approximate the HRTFs with transfer functions corresponding to omni-directional microphones spaced $d = 17\,\text{cm}$ apart. We will only study the left function, the right function is simply its complex conjugate.

$$H(\omega, \theta) = \exp\left(\frac{id\omega}{2c}\sin\theta\right) \tag{11}$$

$$\Phi(\omega,\theta) = \arg\{H(\omega,\theta)\} = \frac{d\omega}{2c}\sin\theta \qquad (12)$$

$$\tilde{\phi}_i = \Phi(\omega_i, \tilde{\theta}_i) \qquad (13)$$

It follows from Equation 12 that the recovered phase estimate, $\tilde{\phi}_i$, has a noise level proportional to frequency as shown in Figure 5a, which becomes problematic at high frequencies. Another phase estimate can be made by using the group delay of the HRTFs instead of their phase shift. Group delay is defined as the negative frequency-derivative of the phase, in this case:

$$\tau_g(\omega,\theta) = -\frac{\partial}{\partial\omega}\Phi(\omega,\theta) = -\frac{d}{2c}\sin\theta \qquad (14)$$

A second phase estimate can be recovered by summing the group delays over the frequency axis:

$$\tilde{\tilde{\phi}}_i = -\sum_{j=0}^{i} \tau_g(\omega_j, \tilde{\theta}_j)\Delta\omega, \qquad (15)$$

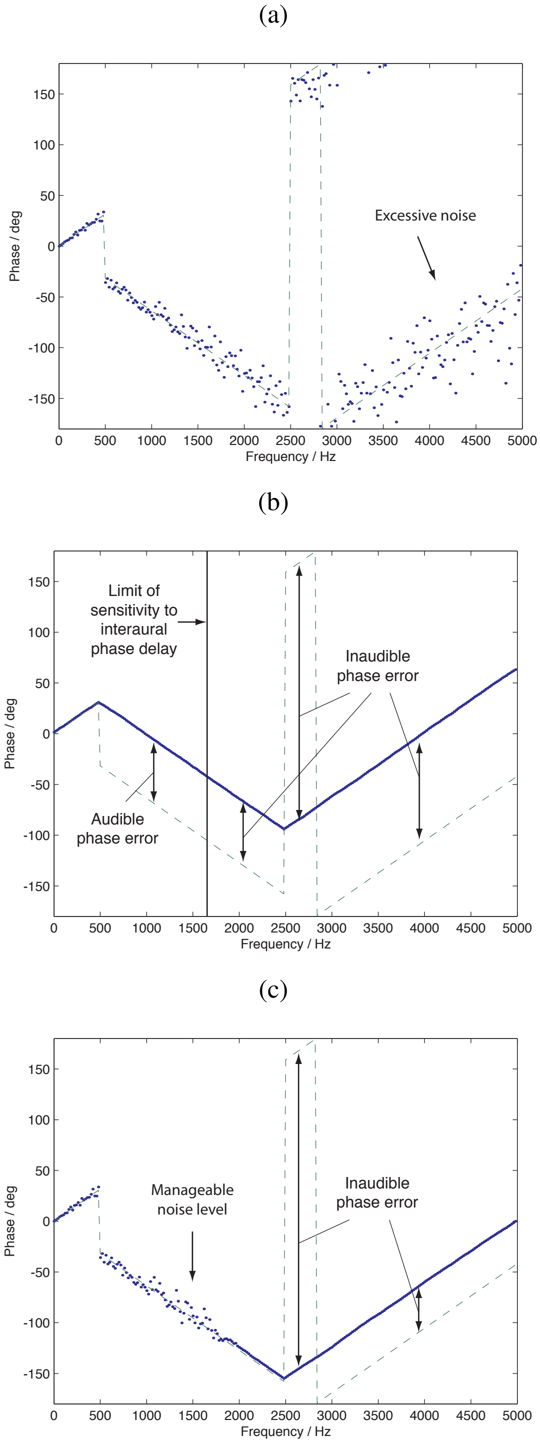
where $\Delta\omega$ is the center-to-center separation between frequency bands. This phase estimate has much less noise, as shown in Figure 5b, and although it approximates the correct group delay, it can deviate far from the correct interaural phase delay. Since we do not perceive interaural phase delay at frequencies above 1.6 kHz, the idea is to use $\tilde{\phi}_i$ at low frequencies and $\tilde{\tilde{\phi}}_i$ at high frequencies. To obtain a smooth transition between the two, it is possible to combine them in the following way:

$$\tilde{\tilde{\tilde{\phi}}}_i = \begin{cases} \tilde{\phi}_i & i = 0 \\ \tilde{\phi}_i(1 - s(\omega_i)) + \ldots & i \geq 1 \\ \left(\tilde{\tilde{\tilde{\phi}}}_{i-1} - \tau_g(\omega_i, \tilde{\theta}_i)\Delta\omega\right) s(\omega_i) & \end{cases} \qquad (16)$$

where $s(\omega)$ is a function that transitions smoothly from 0 to 1 over a range of about 1.6–1.8 kHz. For example,

$$s(2\pi f) = \frac{1}{1 + \exp\left((f_c - f)/f_0\right)}, \qquad (17)$$

where $f_c = 1700\,\text{Hz}$ and $f_0 = 100\,\text{Hz}$. The resulting phase estimate is plotted in Figure 5c.


(a)

(b)

(c)

Figure 5: Ideal (– –) and recovered (·) phase shift

4.2 Matching direction estimates

The decomposition into planewaves gives no hint about the identity of the planewaves, which can be swapped freely within each frequency band. When summing group delays across frequency bands as in Equations 15 and 16, the resulting phase would not be correct if delays relating to different sound sources were interleaved. Therefore it is necessary to match direction estimates in neighboring frequency bands. This matching can be done based on amplitude and/or direction.

5 Listening test

A formal listening test was conducted to assess the sound quality produced by the method. The test largely followed the procedures set out in [4], which in turn were based on the MUSHRA recommendations [1] with the exception that no hidden anchor was used.

5.1 Experimental setup

The same stimuli were used as in [3], which consisted of six horizontal sound scenes, shown in Figure 6. A seventh scene was also created and only used for training (not shown). Two of the scenes (enfant and cuisine) were identical to scenes used in [4]. Since a possible weakness of any nonlinear method is the reproduction of scenes with multiple overlapping sounds, all scenes were chosen such that there were always at least three, usually more, overlapping sound sources. Each scene lasted between 10 and 17 seconds.

A reference signal was created by applying the measured HRTFs of subject 1033 in the Listen database [21] to the monophonic sounds. Five systems were tested and compared to the reference:

1-4: 1st order, decoded to four virtual loudspeakers

3-8: 3rd order, decoded to eight virtual loudspeakers

5-12: 5th order, decoded to twelve virtual loudspeakers

H: 1st order, decoded with the proposed method

REF: Hidden reference

The virtual loudspeaker feeds were filtered using the same HRTF set as the reference signals. In preparing the test stimuli, different decoder flavors were tested. It was clear that both max-r_E and in-phase decoding led to a significant loss of treble, which would most likely

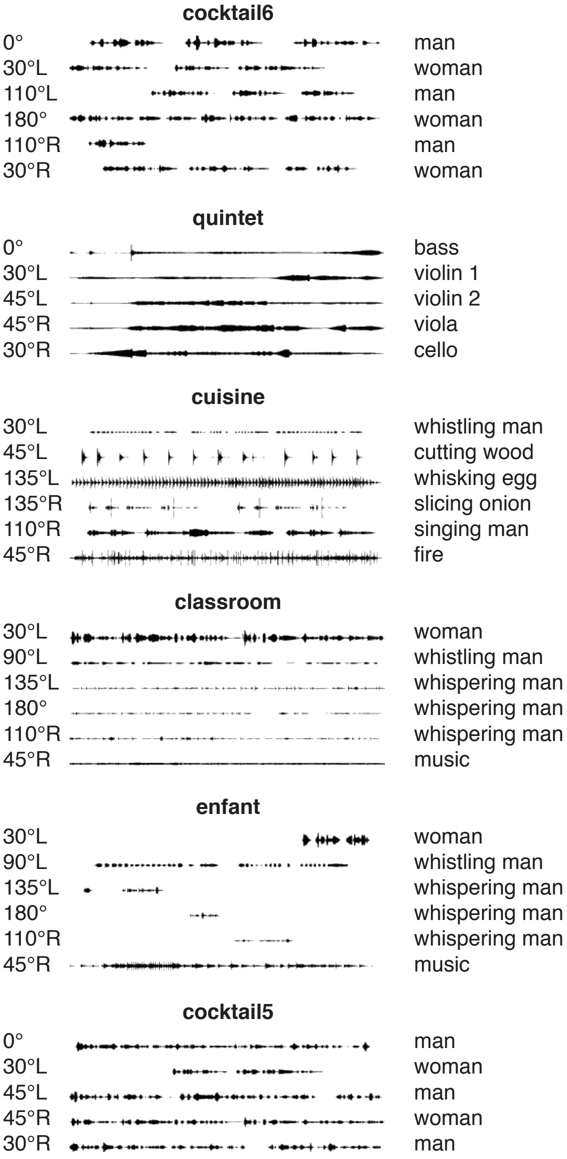


Figure 6: Program material

overshadow the participants' judgment of other aspects of sound quality. We therefore chose to use basic decoding. Using more than $2M+2$ virtual loudspeakers, the minimum recommended number [4], results in a similar loss of treble. Graham Wakefield's Max/MSP externals[20] were used for encoding and decoding the 1-4 and 3-8 systems, while the 5-12 system was processed with a custom-made Max/MSP patch.

The reference version of each sound scene was presented to each participant, along with the five test systems in randomized order, labeled "A" through "E". Participants were asked to rate each of the five signals in each of the six scenes on a scale from 0 to 100, associated with the following guidelines. The adjectives were given in English and the explanation in Norwegian.

80-100: "Excellent," no degradation

60-80: "Good," little change in position

40-60: "Fair," deviation from original position, sources widening

20-40: "Poor," substantial deviation from original position, sources widening, difficult to localize sound sources

0-20: "Bad," sources are completely out of their original position, very hard to localize.

Participants were told about the hidden reference and asked to identify it if possible by giving it 100 points. No hidden anchor was used.

The test was implemented as an Adobe Flash applet, which required the sound to be encoded in MP3 files at 224 kbps. The applet was posted on a social internet forum [19]. There was no pre-screening of participants, who were only required to have headphones, silent surroundings and normal hearing, none of which was possible to verify. Participants were promised a free movie ticket. A total of 28 participants registered and received recorded instructions. Of these, 18 went on to complete the test. Only one participant was able to correctly identify all six hidden references, while 9 were able to correctly identify at least half of them. Only their scores were used in the subsequent analysis, under the assumption that these listeners provided more consistent data.

5.2 Results

The results are summarized in Figures 8 and 9. The difference in score between the 1-4 and 3-8 systems is not statistically significant. The differences between other systems are, however, statistically significant using a 5% significance level in a one-way ANOVA test with post-hoc Tukey's HSD. The mean scores were

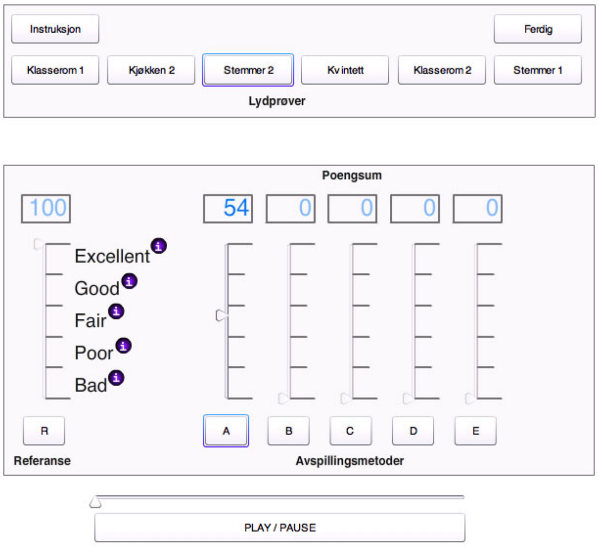


Figure 7: User interface for the listening test

REF: 90 points (excellent)

5-12: 75 points (good)

H: 61 points (good)

3-8: 39 points (poor)

1-4: 37 points (poor)

5.3 Discussion

The use of unscreened and untrained (9 out of 18, self-reported) participants in non-controlled environments with 18 different headphone models are factors that are all expected to increase the variance in the data. Several participants reported difficulty in discriminating between the different systems. Despite this, when aggregating the data, all systems apart from 1-4 and 3-8 were discriminated with statistical significance and in the expected order. It seems clear that the H system produces sound closer to the reference than the 1-4 system, which uses the same input signal and even the 3-8 system, which has access to more than twice the bandwidth (7 channels vs. 3). The latter system does not score significantly higher than the 1-4 system. This is

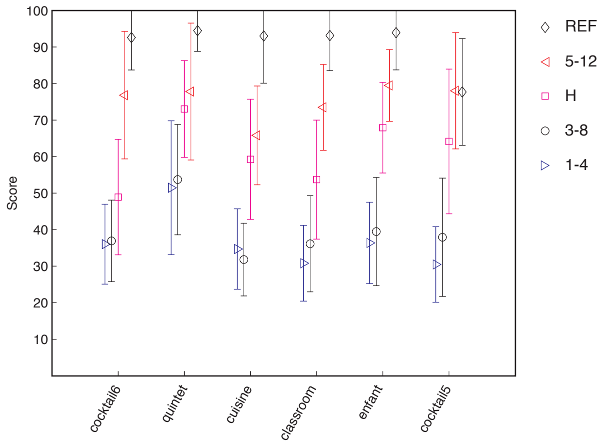


Figure 8: Mean scores across the nine participants for each system in each scene. Error bars indicate 95% confidence interval.

surprising, considering that first order and third order systems were clearly distinguished in previous tests using loudspeaker arrays [4, 3]. One possible explanation is that third order loudspeaker playback is more resilient to head movement than first order. With headphone presentation, this advantage disappears. Further, it is likely that participants compressed the lower part of their evaluation scale, compared to the previous tests, in order to accommodate the higher quality of the fifth order system.

It also seems clear that the 5-12 system produces sound closer to the reference than the H system. However, the difference in their median scores is only barely significant at the 5% level, and without post-screening this difference would have decreased below the level of statistical significance.

As with any listening test, one may question the generality of the results, given that only a few test sounds can be presented to the listeners. The most likely weakness of non-linear methods is their reproduction of multiple overlapping sound sources, and the scenes that were tested were considerably "busier" than the typical sound scene encountered in broadcasting or telecommunication. The result for the H system should therefore be a conservative estimate.

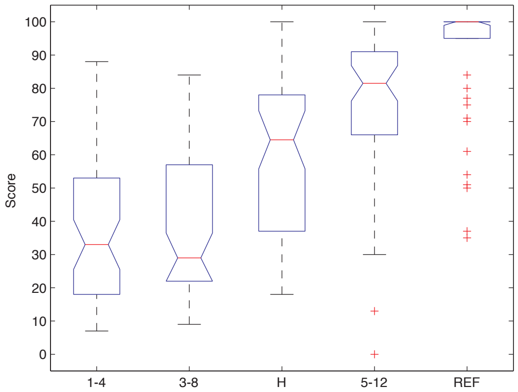


Figure 9: Box plot showing results across all six scenes and nine participants. Red lines indicate median scores. The top and bottom of each box are the 25th and 75th percentiles of samples. The width of the notches is calculated so that boxes whose notches do not overlap have different medians at the 5% significance level. Whiskers show minimum and maximum scores, and red crosses indicate outliers.

6 Other observations

The following observations are based on the authors' own listening. The material used were the 208 sound files currently available at the Ambisonia web site [2], Angelo Farina's simultaneous binaural and B-format recordings[7] and our own field recordings and synthesized sound scenes. In informal blind testing, using the few available instances of simultaneous B-format and binaural recording techniques, the proposed method seems to provide binaural audio of similar quality to dummy head recordings.

Without smoothing or suppression of phase noise, two classes of artifacts become apparent. Most prominent are the artifacts related to filter dispersion. Sharp transients are transformed into noise bursts and intermittent ticking occurs at the frame period, as the contents of some frames wrap around the frame boundaries. Somewhat less noticeable are artifacts related to the time variation in filter coefficients. In combination with overlap-add processing, this gives rise to flutter.

6.1 Frequency smoothing

Smoothing the phase and amplitude response of each transfer function as described in section 3 improves the reproduction of transient attacks. Excessive smoothing leads to a loss of discrimination between sources which in turn causes the sensation of auditory events that move in response to each other. Excessive smoothing of this kind also leads to a loss of sensation of space and distance.

6.2 Time smoothing

A small amount of averaging between transfer functions in consecutive frames is sufficient to remove all flutter. Excessive amounts of smoothing in the time domain causes difficulties in tracking moving sources and with localization of onsets, which is particularly important because of the precedence effect [5].

6.3 Phase noise suppression

The effect of phase noise suppression as described in section 4.1 is similar to frequency smoothing: It improves reproduction of transients, and leads to a subtle improvement in the perception of depth and clarity of simultaneous sources. Excessive suppression at low frequencies causes audible errors in inter-aural phase delay, which gives rise to the sensation of auditory events that move randomly from side to side. However, when used in moderation at high frequencies and in combination with moderate frequency smoothing, it is possible to obtain perfectly sharp transient reproduction without the adverse side-effects of either method.

6.4 Hardware requirement

The method has been implemented in C and compiled for a 2.66 GHz Intel Core 2 Duo. Real-time transcoding at 48 kHz requires about 25% of one CPU core's time.

7 Conclusions

It is known that nonlinear transcoding from B-format to a binaural format provides better sharpness than first order linear transcoding. We have shown how the artifacts that arise from such processing can be suppressed and shown that the resulting sound quality ranks between third order and fifth order linear transcoding. When used in combination with well-known B-format recording techniques, this method provides an alternative to dummy-head recording, with the added advantage of B-format transformations and tailor-made HRTFs.

8 Acknowledgments

The authors would like to thank Orange Labs for providing the program material from their experiments.

References

[1] Method for the subjective assessment of intermediate quality level of coding systems. *ITU-R BS*, pages 1534–1, 2003.

[2] Ambisonia. `http://www.ambisonia.com`.

[3] S. Berge and N. Barrett. High Angular Resolution Planewave Expansion. In *Proceedings of the 2010 Ambisonics Symposium*, 2010.

[4] S. Bertet, J. Daniel, E. Parizet, and O. Warusfel. Influence of Microphone and Loudspeaker Setup on Perceived Higher Order Reproduced Sound Field. In *1st Ambisonics Symposium*, 2009.

[5] J. Blauert. *Spatial hearing: the psychophysics of human sound localization*. The MIT Press, 1997.

[6] J. Daniel. Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. *Université Pierre et Marie Curie (Paris VI): Paris*, 2000.

[7] A. Farina. String quartet recorded with dummy head and soundfield microphone. `http://pcfarina.eng.unipr.it/Public/B-format/Prodi/`.

[8] M.A. Gerzon. Periphony: With-height sound reproduction. *J. Audio Eng. Soc*, 21(1):2–10, 1973.

[9] M.A. Gerzon and G.J. Barton. Surround sound apparatus, May 26 1998. US Patent 5,757,927.

[10] M.D. Good and R.H. Gilkey. Sound localization in noise: The effect of signal-to-noise ratio. *The Journal of the Acoustical Society of America*, 99:1108, 1996.

[11] K. Hartung, J. Braasch, and S.J. Sterbing. Comparison of different methods for the interpolation of head-related transfer functions. In *AES 16th Int. Conf. on Spatial Sound Reproduction*, pages 319–329, 1999.

[12] A. S. Householder. *The Numerical Treatment of a Single Nonlinear Equation*. McGraw-Hill, 1970.

[13] M.V. Laitinen. Binaural reproduction for directional audio coding. Master's thesis, Helsinki University of Technology, 2008.

[14] D.S. McGrath and A.R. McKeag. Wavelet conversion of 3-D audio signals, September 30 2003. U.S. Patent 6,628,787.

[15] A. McKeag and D. McGrath. Sound field format to binaural decoder with head tracking. *AES Convention 6r*, 1996.

[16] D. Menzies. W-panning and O-format, tools for object spatialization. In *Proceedings of the 22nd International Conference of the AES on Virtual Synthetic and Entertainment Audio, Espoo, Finland*, 2002.

[17] M. Noisternig, T. Musil, A. Sontacchi, and R. Höldrich. A 3D real time Rendering Engine for binaural Sound Reproduction. In *ICAD*, volume 9, pages 107–110, 2003.

[18] V. Pulkki, MV Laitinen, J. Vilkamo, J. Ahonen, T. Lokki, and T. Pihlajamäki. Directional audio coding - perception-based reproduction of spatial sound. In *International Workshop on the Principles and Applications of Spatial Hearing*, 2009.

[19] Underskog. `http://www.underskog.no`.

[20] G. Wakefield. Ambisonics for Max/MSP. `http://www.grahamwakefield.net/soft/ambi~/index.htm`, 2006.

[21] Warusfel, O. Listen hrtf database. `http://recherche.ircam.fr/equipes/salles/listen/index.html`, 2003.