



Audio Engineering Society Convention Paper

Presented at the 125th Convention
2008 October 2–5 San Francisco, CA, USA

The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

QESTRAL (Part 3): system and metrics for spatial quality prediction

P.J.B. Jackson¹, M. Dewhurst^{1,2}, R. Conetta², S. Zielinski², F. Rumsey², D. Meares³, S. Bech⁴ and S. George²

¹Centre for Vision, Speech & Signal Processing, University of Surrey, UK

²Institute of Sound Recording, University of Surrey, UK

³DJM Consultancy, Sussex, UK, on behalf of BBC Research

⁴Bang & Olufsen a/s, Peter Bangs Vej 15, 7600 Stuer, Denmark

Correspondence should be addressed to Philip Jackson (p.jackson@surrey.ac.uk)

ABSTRACT

The QESTRAL project aims to develop an artificial listener for comparing the perceived quality of a spatial audio reproduction against a reference reproduction. This paper presents implementation details for simulating the acoustics of the listening environment and the listener's auditory processing. Acoustical modelling is used to calculate binaural signals and simulated microphone signals at the listening position, from which a number of metrics corresponding to different perceived spatial aspects of the reproduced sound field are calculated. These metrics are designed to describe attributes associated with location, width and envelopment attributes of a spatial sound scene. Each provides a measure of the perceived spatial quality of the impaired reproduction compared to the reference reproduction. As validation, individual metrics from listening test signals are shown to match closely subjective results obtained, and can be used to predict spatial quality for arbitrary signals.

1. INTRODUCTION

Motivation

Models that predict the sound quality impair-

ments of speech and audio coding systems based on the timbral and temporal aspects of reproduced sound have already been developed and established (the PESQ [19] and PEAQ [24] models have been

adopted by the International Telecommunications Union). Compression algorithms based on perceptual models, such as MP3 and AAC, have demonstrated how audio signals can be cut down with minimal effect on perceived attributes of the reproduced sound, and underline the importance of the listener's perception in designing audio reproduction systems. However, the increased use of multi-channel reproduction systems has raised demand for methods to assess the spatial aspects of reproduced sound. The ability to predict spatial attributes of a sound scene is useful because it is costly in time and resources to perform exhaustive subjective listening tests for all the conditions one would wish to investigate. The model described here enables detail of reproduced sound fields to be examined, and the results used to assist in the design of audio compression, transmission and reproduction systems. Furthermore, the availability of simulation results makes possible comparison of theoretical predictions with physical measurements that could lead to advances in our explanation of human sound perception.

Applications

Our approach, although centred on particular reproduction systems, programme material and listening environments, is designed using general principles of acoustic propagation and spatial sound perception for a wide range of potential applications. Audio processing devices, such as downmixers, codecs and spatial effects, occur in speech and music broadcasting, movies, games, auditory displays, and have implications for the acquisition, editing, encoding and rendering of media content with audio. Quality measures do not need to provide a complete model of human perception, and certainly PEAQ and PESQ do not claim to do so. We are aiming at predicting a global measure of spatial quality for reproduced sound systems, nevertheless incorporating metrics that relate to low level spatial attributes. In particular, we have developed metrics that can be used for sound localisation, as well as for evaluating source width and the sense of listener envelopment. This paper describes those metrics, which are used in the prediction of overall spatial quality, and in measuring spatial distortions arising from audio processing.

Attributes

Localisation can be considered the primary spatial attribute. It is innate, relevant for survival, and

an essential sensory input giving us context in the world, including the suggestion that “the ears are for pointing the eyes”. Of the other, secondary, spatial attributes, sound source width and the sense of envelopment are typically judged to be the most important for the overall spatial perception, and these attributes have been the subject of the largest amount of research, both in the field of concert hall acoustics and in the field of reproduced audio [3, 14, 16]. The need to assess perceived spatial attributes at multiple locations across the listening area comes from the fact that media consumers do not comply with ITU standards in positioning of loudspeakers, they move around, and often multiple people are listening together. There are many open questions about how stable a reproduced sound scene may be and how significant degradations are under a variety of listening conditions and environments, such as in a home cinema or lounge. While there has been previous research into predicting spatial attributes away from the sweet spot at the centre of the listening area, these have been confined to varying the listener location in just a single dimension and have also only been concerned with directional localisation [16, 20]. One of the key objectives of this research is to investigate perceived spatial attributes, including some secondary spatial attributes, at multiple locations across the listening area. Hence, our system entails developing an artificial listener able to predict several perceived spatial attributes at different locations in the listening area of an audio reproduction system.

Two attributes of spatial impression have been recognised in concert hall acoustics. Apparent source width (ASW) occurs when the early lateral reflections fuse with the direct sound image, causing the image of the sound source to become wider. Listener envelopment (LEV) is more associated with the late lateral reflections (>80 ms) for this kind of sound scene. ASW and LEV are based on the ratio of the lateral energy (due to the reflections) to the total energy [2, 3]. The inter-aural cross-correlation (IACC) has also been used to predict spatial impression, since lateral reflections cause the signals at the ears to become decorrelated [1]. The measures developed for concert hall acoustics rely on impulse responses, yet researchers have shown that source signals can change the sound's spatial impression.

Griesinger's diffuse-field transfer function and Mason's interaural cross-correlation fluctuation function extract measures for other source signal types [14, 16].

Metrics

In order to obtain measures from a real or simulated sound field, signals are recorded using real or virtual microphones, which can be placed, for example, at the ears of an artificial listener. Two kinds of metric are calculated from the captured signals, to represent distortions in the foreground and background audio streams, respectively [4, 14]. The foreground stream would likely consist of a dominant sound source that was the focus of attention, whereas any other sources (e.g., independent, less-prominent sounds) would comprise the background stream. While not all of the metrics employed in the model directly correspond to a single perceived spatial attribute, the rationale for including each of the metrics is defined in terms of its ability to capture information relevant to the spatial impression. For a given source, the location and width metrics do correspond directly to perceived spatial location and width, which are of primary and secondary importance in evaluating foreground distortions. Metrics were also developed to describe the influence of the background stream, particularly the effects of direct and indirect sound in creating the impression of envelopment, validated using formal listening tests [5, 12]. Maintaining relevance to human perception of spatial sound, a number of binaural metrics (i.e., metrics that use the sound pressure signals at the two ears of the listener) have been incorporated to predict the perceived spatial attributes of reproduced sound [8].

System overview

The system architecture we use is outlined in [21], and involves several stages of processing, up to the prediction of a measure of spatial quality. Descriptions of the reproduction systems used to generate the two sound fields (reference and impaired) are input to the model, which includes any process that transforms or degrades the signals with respect to the reference reproduction system. For example, a five-channel loudspeaker layout (ITU-R BS.775, [18]) and a two-channel loudspeaker layout could be used for the reference and impaired reproduction respectively, and the mapping between the signals for

the two reproduction systems could be achieved using a standard down-mix algorithm. The model then generates two renderings of the sound field that allow it to identify distortions in the foreground and background audio streams with respect to the reference reproduction system. The source signals can be arbitrary, and the artificial listener positioned and oriented as required, to yield simulated microphone and binaural signals ready for further processing. The use of explicit acoustic modelling enables the model to predict the response at different positions within the listening area and also allows us to model the results in different listening environments.

Components of this paper

Here, we will focus on the stages leading to the production of foreground and background metrics: generation of audio signals, reproduction of the sound field, capture of microphone and binaural signals, and extraction of metrics. Generating the audio signals encompasses all the activities associated with capturing, panning, mixing and encoding them into a given format. The reproduction consists of performing an acoustic simulation for the specified reproduction system within the specified listening environment. An artificial head and virtual microphones, of defined directivity, are placed within the simulation of the reproduced sound field to record simulated sound pressure signals. This paper will concentrate on describing the conversion of those microphone and binaural artificial listener signals into metrics that can be used in the prediction of spatial quality. Subsequent inclusion of these metrics within the model to prediction spatial sound quality is covered in the fourth of this group of papers [8].

2. COMPONENTS OF THE MODEL

This section describes the modelling framework that was used, including the calculation of binaural signals in the reproduced sound field. The coordinate system is described, followed by a discussion of the standard reproduction systems that were employed.

2.1. Reproduction systems

The coordinate system was centred within the listening area with the origin at the sweet spot, and by default the listener faced forwards at 0° . The reproduction systems investigated include Mono, Two Channel Stereo (TCS), Five Channel Stereo (FCS,

equivalent to 5.0) and Wave Field Synthesis (WFS, 32-channel). Loudspeakers were implemented here as monopole point sources, although more accurate directivity patterns are planned for future experiments.

2.2. Rendering of reproduced sound field

The reproduction of sound in the simulated acoustical environment can be modelled as a linear invariant system, where the sound pressure at any point is the superposition of pressures due to each sound source. For both microphones and our artificial listener, the transfer functions from each source to each sensor were modelled directly: in Matlab for the case where the recording environment was anechoic, and an acoustical simulation package (either CATT-Acoustics or ODEON) for the case where the recording environment was reflective. In all cases, these allowed for modelling of the directivity of the sensors. For the artificial listener, directivity was encapsulated within HRTFs [10], measured with a KEMAR dummy head and torso and compensated for the source distance. The system was designed to work with arbitrary source signals and reproduction systems. The Matlab implementation of the model framework was validated by accurately reproducing the WFS pressure plots in [6].

2.3. Auditory processing

The processing of the artificial listener's binaural signals followed a conventional model of human auditory processing, as in [23]. It includes the division into critical bands, envelope smoothing, calculation of IID, calculation of IACC and derived ITD, duplex and loudness weighting, frequency-wise fusion, and combination of ITD and IID cues for localisation.

The binaural signals are first separated into critical bands and these signals are then half-wave rectified and low pass filtered. The IID cues for each critical band are calculated from the ratio of energy in the left and right signals for a given frame. The ITD cues are derived from the cross-correlation of the rectified and filtered left- and right-ear signals in each critical band, according to the time at which the peak in inter-aural cross-correlation is attained. The IID and ITD cues are then converted to angle scores using a database of IID and ITD values for known

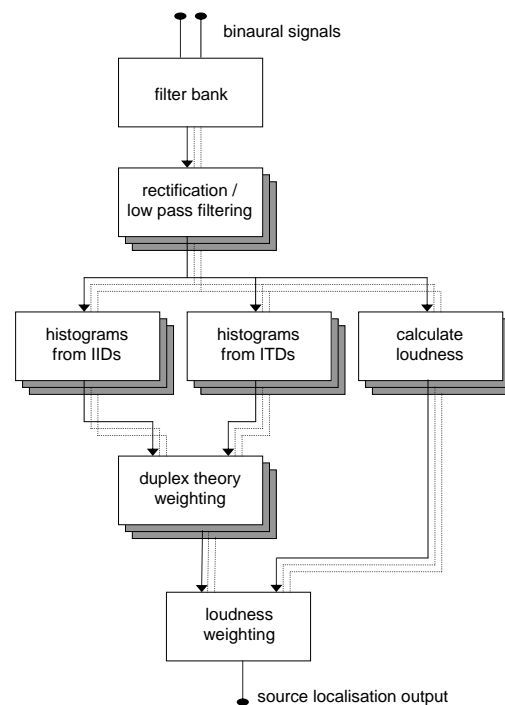


Fig. 1: Process of calculating source localisation scores across azimuth from binaural signals [23].

angles from a database of head-related transfer function (HRTF) measurements. These scores are combined firstly across the critical bands, and then the IID and ITD summary scores fused to give a single angle of localisation at the peak. An overview of the auditory processing is shown in Figure 1. Details of how metrics related to localisation were obtained will be described in the next section.

3. METRICS

Two main categories of metrics were considered in order to produce a variety of metrics, including ones that have proven to be useful in previous experiments and ones designed to capture perceptually-important changes in the spatial impression of reproduced sound. The first category involves signals that can be from either real or virtual microphones located in the reproduced sound field. The second category incorporates signals from an artificial head, such as a KEMAR, either directly recorded or indirectly calculated using its HRTFs within the acoustical simulation. In all cases, the real or virtual signal

capture can be placed and oriented arbitrarily, giving the system the capability to extract metrics at multiple locations throughout the listening area.

3.1. Microphone-based metrics

The first category of metrics is derived using signals from one or more microphones. Here we describe two microphone configurations, a single omni-directional microphone placed at the location of interest and a coincident array consisting of an omni plus two figure-of-eight microphones. By convention, we use discrete signals at a standard audio sampling rate of 44.1 kHz.

3.1.1. Signal intensity

The mono signal captured by a single omnidirectional microphone, $m_W(n)$, is used to give a measure related to the total energy arriving at the listening position, which is calculated as the root-mean-square amplitude:

$$\text{TotEnergy} = \sqrt{\frac{\sum_{n=1}^N m_W^2(n)}{N}}, \quad (1)$$

where N is the size of the signal frame in samples.

3.1.2. Directional coherence

The virtual microphone array was based on supplementing the omni-directional microphone with two figure-of-eight (velocity) microphones at right angles to one another in the horizontal plane. These x and y directions in plan view correspond to a line pointing directly ahead for a listener facing forward (i.e., orientation of 0°) and the axis through the listener's ears, respectively. The correlation between the omni-directional signal and each of the directional signals, $m_X(n)$ and $m_Y(n)$, indicates how directional the sound field is. These B-format signals are combined to give cardioid microphone signals [13]. The metric is computed by combining the x and y components through a principal components analysis (a.k.a. Karhunen-Loève Transform) and examining the size of the largest eigenvalue λ_1^2 , in proportion to the total energy in the signal:

$$\text{CardKLT} = 100 \left(1 - \frac{\lambda_1^2}{\text{TotEnergy}} \right). \quad (2)$$

Figure 2 gives a block diagram.

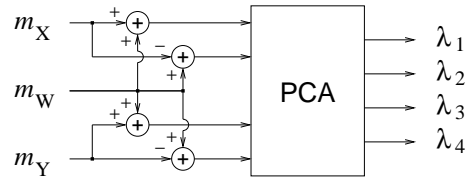


Fig. 2: Block diagram of CardKLT metric.

3.2. Ear-based metrics

While microphone-based metrics go some of the way to describe the spatial characteristics of a sound field, human perception is based on signals arriving at the ears, which are attenuated and coloured by the effects of the torso, head and pinnae. Hence, we have included in our set of metrics a number of measures derived from ear signals recorded or simulated by an artificial listener.

3.2.1. Monaural entropy

Although one might assume that the spatial impression of a sound field depends exclusively on the spatial characteristics of the sound field, other factors can heavily influence one's interpretation of a reproduced sound scene. For instance, a piano may be perceived to be wider than a flute despite being played back through a single loudspeaker, and many voices more enveloping than one. Equally, the division of signal components into foreground and background streams, mediated to some extent by higher cognitive processes, can affect the way that those components are perceived. Therefore, the signal entropy was introduced as a measure of the amount of information in the signal, which is expected to correlate with these factors. The entropy measure used was calculated for the signal at the left ear, $a_L(n)$:

$$\text{EntropyL} = - \sum_{n=1}^N P(a_L(n)) \ln P(a_L(n)), \quad (3)$$

where the probability of a sample value $P(a_L(n))$ is estimated from the histogram of the sample distribution [17].

3.2.2. Binaural cues

The most important spatial cues listeners receive are obtained from differences between the signals at the two ears, the binaural signals. These inter-aural differences are quantified in terms of time, intensity

and the strength of the cross-correlation, yielding a range of binaural metrics. Biologically-inspired preprocessing of the binaural signals was introduced in Section 2.3 and will now be expanded to explain the range of metrics extracted that relate to localisation.

As in Figure 1, the binaural signals are processed in frequency bands corresponding to a bank of gammatone filters with approximately $1/4^{\text{rd}}$ -octave bandwidth ($F = 24$ bands).¹ The left and right signal envelopes, $b_L(n)$ and $b_R(n)$, are generated by rectifying and smoothing the band-limited signals with a 1.1 kHz low-pass filter (to mimic hair cell behaviour). Hence, a set of IACCs is obtained for each frequency band f and any time t :

$$\text{IACC}(t, f) = \max_{\tau} \left(\frac{\sum_{n=1}^N b_L(t+n)b_R(t+n+\tau)}{\sqrt{\sum_{n=1}^N b_L^2(t+n) \sum_{n=1}^N b_R^2(t+n)}} \right), \quad (4)$$

where τ is the lag between the two signals in samples, and its value at the maximum is the corresponding ITD for that frame and band, $\text{ITD}(t, f)$. The lag is normally limited to lie within the range ± 1 ms.

The intensity, or level, difference is also calculated from the binaural envelope signals and typically expressed in decibels:

$$\text{IID}(t, f) = 10 \log_{10} \left(\frac{\sum_{n=1}^N b_R^2(t+n)}{\sum_{n=1}^N b_L^2(t+n)} \right). \quad (5)$$

3.2.3. Derived binaural metrics

The binaural cues at a given listening position provide a wealth of perceptually-relevant information about the sound field at that location. In particular, the ITD and IID cues are usually combined for estimating the perceived location of a sound source. However, the degree of correlation of the signals at the two ears has been shown to contain information about the width of the source and the sense of envelopment [15, 16, 11]. So, by taking an average over F frequency bands, one metric represents a summary of the IACC values for an orientation of 0° :

$$\text{IACC0} = \left(1 - \frac{1}{F} \sum_{f=1}^F \left(\max_t \text{IACC}(t, f) \right) \right). \quad (6)$$

¹The gammatone filter bank was based on Slaney's efficient implementation [22]. Low and high cutoff frequencies for each filter were taken from Gaik's cross-correlation model [9].

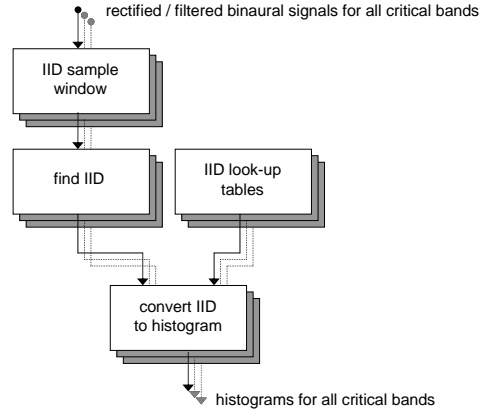


Fig. 3: Use of inter-aural intensity difference (IID) cues and look-up tables to give confidence scores for source localisation angles.

Similar metrics can be obtained for other orientations of the artificial listener's head, e.g., IACC90 when facing 90° to the right.

When evaluating how well a sound scene has been reproduced for a given programme item, it is useful to consider the spatial distribution of the dominant phantom sources. Thus, our final set of metrics are based on estimated azimuth characteristics. For each critical band, the inter-aural difference is converted to an array of confidence scores for each angle θ , using look-up tables trained on HRTF data. Figure 3 shows the architecture for IID cues; a similar architecture exists for ITD cues. A peak in the confidence score indicates a likely angle for a sound source. The confidence scores are weighted by Duplex theory and by loudness within each band and added, to yield a summary score at that time for each cue, $c_{\text{ITD}}(t, \theta)$ and $c_{\text{IID}}(t, \theta)$. The ITD and IID cues are normally then combined to give an overall score across θ . While the cues are not entirely independent, our experiments have indicated more accurate azimuth predictions forming the product, $c_{\text{Both}}(t, \theta) = c_{\text{ITD}}(t, \theta)c_{\text{IID}}(t, \theta)$. However, a pair of metrics is computed from the ITD and IID confidences that describes the spread of sources by averaging over time and then taking the standard deviation, treating the scores as a histogram:

$$\text{std_itd} = \text{std} \left(\frac{1}{T} \sum_{t=1}^T (c_{\text{ITD}}(t, \theta)) \right)$$

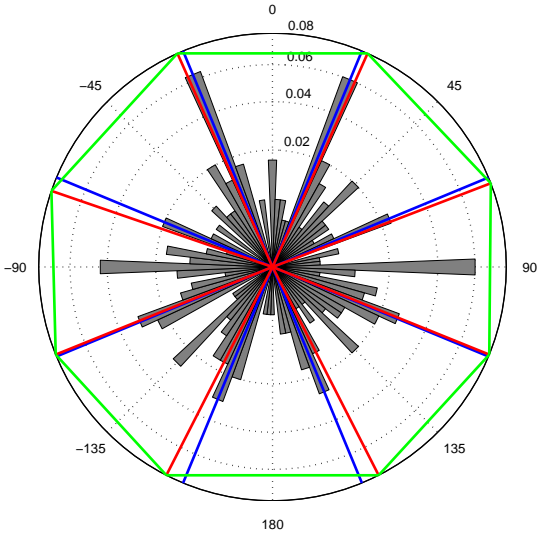


Fig. 4: Example of area calculation for the hull metric for a set of $\hat{\theta}$ values (red lines), estimating the directions of eight sources, based on peaks in the localisation scores from individual source components. Blue lines show the panned locations of the sources. Green lines outline the hull which sits within the unit circle [7].

$$\text{std_iid} = \text{std} \left(\frac{1}{T} \sum_{t=1}^T (c_{\text{IID}}(t, \theta)) \right) \quad (7)$$

Another binaural metric used in the model is a measure that evaluates the ability of the reproduction system to render a complete scene around the listener.

$$\text{hull} = \text{area}_{t=1}^T \exp(j\hat{\theta}(t)). \quad (8)$$

where $\hat{\theta}(t) = \arg \max_{\theta} c_{\text{Both}}(t, \theta)$ is the estimated angle of localisation at any time t in radians. The function $\text{area}(\cdot)$ returns the area of the polygon (convex hull) connecting all the angles projected onto the unit circle, which ranges from zero to 2π . An example is shown in Figure 4.

It is known from concert hall acoustics, and our own listening experiments, that lateral sound energy tends to have a significant effect on the sense of immersion and envelopment. Thus, we define a metric that records the angle of the dominant source closest to the sides at $\pm 90^\circ$:

$$c90 = \min_t \left| \frac{\pi}{2} - |\hat{\theta}(t)| \right|. \quad (9)$$

When evaluating an impaired reproduction (DUT) to a reference reproduction (Ref), direct comparison can be made of the azimuths of localisable source from frame to frame. From this, we extract two metrics that capture the average and the maximum localisation error between the reproductions respectively:

$$\begin{aligned} \text{MeanAngDiff} &= \frac{1}{T} \sum_{t=1}^T \left| \hat{\theta}_{\text{DUT}}(t) - \hat{\theta}_{\text{Ref}}(t) \right| \\ \text{MaxAngDiff} &= \max_{t=1}^T \left| \hat{\theta}_{\text{DUT}}(t) - \hat{\theta}_{\text{Ref}}(t) \right|. \quad (10) \end{aligned}$$

4. DISCUSSION

The rationale for selecting these foreground and background metrics was informed by the observed changes in perceived spatial feature values when altering multi-channel audio material with typical audio processes. Metric selection was a mixture of informed guesswork, inspiration from previous work on spatial metrics, knowledge of audio processes, attempts to account for specific low level attributes and pragmatic evaluation of what worked. This paper does not claim to present the definitive final set of metrics, yet it provides a holistic approach to the development of spatial metrics which we hope will yield additional improvements in the future.

5. SUMMARY

Within the context of predicting an overall measure of spatial sound quality, we motivate an approach that considers important attributes of foreground and background streams in the perception of a reproduced sound field. Herein are described a range of metrics: TotEnergy, CardKLT, EntropyL, IACC0, IACC90, std_itd, std_iid, hull, c90, MeanAngDiff and MaxAngDiff. Some of these metrics can be related to individual spatial attributes, such as localisation angle, sound source width or listener envelopment. Further work evaluates the ability of these metrics to predict subjective mean opinion scores of the spatial quality of sound reproduction [8].

6. REFERENCES

- [1] M. Barron. Objective measures of spatial impression in concert halls. In *Proceedings of the 11th International Congress on Acoustics*, Paris, 1983.

- [2] M. Barron and A.H. Marshall. Spatial Impression Due to Early Lateral Reflections in Concert Halls: the Derivation of a Physical Measure. *J. Sound and Vibration*, 77(2):211–232, July 1981.
- [3] J.S. Bradley and G.A. Soulodre. The Influence of Late Arriving Energy in Spatial Impression. *J. Acoust. Soc. Am.*, 97(4):2263–2271, April 1995.
- [4] A.S. Bregman. *Auditory Scene Analysis: The Perceptual Organisation of Sound*. MIT, Cambridge, MA, 1990.
- [5] R. Conetta, P. J. B. Jackson, S. Zielinski, and F. Rumsey. Envelopment: What is it? a definition for multichannel audio. In *1st SpACE-Net Workshop, York, UK*. 2007.
- [6] J. Daniel, R. Nicol, and S. Moreau. Further investigations of High Order Ambisonics and Wavefield Synthesis of holophonic sound imaging. page 5788, Amsterdam, The Netherlands, March 2003, 114th Conv. Audio Eng. Soc., Preprint 5788.
- [7] M. Dewhirst. *Modelling perceived spatial attributes of reproduced sound*. PhD thesis, CVSSP/IoSR, University of Surrey, 2008.
- [8] M. Dewhirst *et al.* QESTRAL (Part 4): Test signals, combining metrics and the prediction of overall spatial quality. Presented at the *125th AES Convention*, San Francisco, October 2008. Audio Engineering Society.
- [9] W Gaik. Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modelling. *J. Acoust. Soc. Am.*, 94(1):98–110, July 1993.
- [10] W.G. Gardner and K.D. Martin. HRTF measurements of a KEMAR. *J. Acoust. Soc. Am.*, 97(6):3907–3908, June 1995.
- [11] S. George, S. Zielinski, and F. Rumsey. Feature extraction for the prediction of multichannel spatial audio fidelity. *IEEE Transactions on Audio, Speech and Language Processing*, 13(6):1994–2005, November 2006.
- [12] S. George, S. Zielinski, F. Rumsey, and S. Bech. Evaluating the sensation of envelopment arising from 5-channel surround sound recordings. In *Presented at the AES 124th Convention, Amsterdam*, 2008.
- [13] Michael A. Gerzon. Periphony: With-height sound reproduction. *J. Audio Eng. Soc.*, 21(1):210, 1973.
- [14] D. Griesinger. Objective Measures of Spaciousness and Envelopment. In *Proceedings of the AES 16th International Conference, Rovaniemi, Finland*, April 1999.
- [15] R. Mason, T. Brookes, and F. Rumsey. Integration of measurements of interaural cross-correlation coefficient and interaural time difference within a single model of perceived source width. San Francisco, California, October 2004, 117th Conv. Audio Eng. Soc., Preprint 6317.
- [16] R. Mason, T. Brookes, and F. Rumsey. Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli. *J. Acoust. Soc. Am.*, 117(3):1337–1350, March 2005.
- [17] R. Moddemeijer. On estimation of entropy and mutual information of continuous distributions. *Signal Processing*, 16(3):233–246, 1989.
- [18] Rec. ITU-R BS.775-1. Multichannel stereophonic sound system with and without accompanying picture, 1994.
- [19] A.W. Rix, J.G. Beerends, M.P. Hollier, and A.P. Hekstra. PESQ - The New ITU Standard for End-to-End Speech Quality Assessment. 109th Conv. Audio Eng. Soc., Preprint 5260, Los Angeles, California, September 2000.
- [20] J. Rose, P. Nelson, and T. Takeuchi. Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations. *J. Acoust. Soc. Am.*, 112(5):1992–2002, November 2002.
- [21] F. Rumsey *et al.* QESTRAL (Part 1): Quality Evaluation of Spatial Transmission and Reproduction using an Artificial Listener. Presented at the *125th AES Convention*, San Francisco, October 2008. Audio Engineering Society.

- [22] M. Slaney. An efficient implementation of the petterson-holdsworth auditory filter bank. Apple computer technical report #35, 1993.
- [23] B. Supper. *An onset-guided spatial analyser for binaural audio*. PhD thesis, Institute of Sound Recording, University of Surrey, 2005.
- [24] T. Thiede, W.C. Treurniet, R. Bitto, C Schmidmer, T. Sporer, J.G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten. PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality. *J. Audio Eng. Soc.*, 48(1/2):3–29, January/February 2000.