PERCEPTUAL EVALUATION OF SPATIAL AUDIO: WHERE NEXT?

Jon Francombe, Tim Brookes, and Russell Mason
Institue of Sound Recording, University of Surrey, Guildford, GU2 7XH, UK
email: j.francombe@surrey.ac.uk

From the early days of reproduced sound, engineers have sought to reproduce the spatial properties of sound fields, leading to the development of a range of technologies. Two-channel stereo has been prevalent for many years; however, systems with a higher number of discrete channels (including rear and height loudspeakers) are becoming more common and, recently, there has been a move towards loudspeaker-agnostic methods using audio objects. Perceptual evaluation, and perceptually-informed objective measurement, of alternative reproduction systems can inform further development and steer future innovations. It is important, therefore, that any gaps in the field of perceptual evaluation and measurement are identified and that future work aims to fill those gaps. A standard research paradigm in the field is identification of the perceptual attributes of a stimulus set, facilitating controlled listening tests and leading to the development of predictive models. There have been numerous studies that aim to discover the perceptual attributes of reproduced spatial sound, leading to more than fifty descriptive terms. However, a literature review revealed the following key problems: (i) there is little agreement on exact definitions, nor on the relative importance of each attribute; (ii) there may be important attributes that have not yet been identified (e.g. attributes arising from differences between real and reproduced audio, or pertaining to new 3D or object-based methods); and (iii) there is no model of overall spatial quality based directly on the important attributes. Consequently, the authors contend that future research should focus on: (i) ascertaining which attributes of reproduced spatial audio are most important to listeners; (ii) identifying any important attributes currently missing; (iii) determining the relationships between the important attributes and listener preference; (iv) modelling overall spatial quality in terms of the important perceptual attributes; and (v) modelling these perceptual attributes in terms of their physical correlates.

1. Introduction

From the early days of reproduced sound, audio engineers have been aware of the requirement to capture and reproduce the spatial dimensions of audio; the first record of a stereo transmission is Clement Ader's experiment performed in 1881, in which two spaced microphones were used to drive two 'telephone receivers' at a remote location (the Paris Electrical Exhibition) [1]. Since then, a wide range of recording, encoding, and reproduction systems have been developed that aim to more or less accurately represent the spatial properties of the captured sounds. For many years, two-channel stereo—capable of producing phantom images of sound sources between two frontal loudspeakers—was the most widely accepted method. More recently, systems with a higher number of discrete channels (such as 5.1 surround sound) have been implemented in an attempt to provide localisation of sound sources across a greater range of positions and to enable a feeling of envelopment in an environment.

As spatial audio becomes more prevalent in the home, it is ever more important to perform detailed and relevant evaluation of systems to inform further development and drive future innovation. Physical evaluation is undoubtedly important, but ultimately a system is only as good as a listener thinks it is; therefore, perceptual evaluation is always of paramount importance. Perceptual evaluation is generally performed through carefully controlled listening tests. When a panel of trained listeners is used, the results of such tests are considered accurate and objective [2, 3]. However, listening tests are expensive and time consuming to perform; therefore, the development of predictive perceptual models that can provide similar results to a panel of listeners is often the goal of psychoacoustic research.

It is important to identify the facet of experience that should be evaluated and modelled. Consequently, the method of descriptive analysis (DA) is widely used in the sensory sciences (perceptual audio evaluation has adopted many methodological practices from food science). In DA, the relevant perceptual dimensions are identified using qualitative elicitation methods, and then ratings are made on the elicited scales to provide quantitative analysis. An introduction to elicitation methods used in the sensory sciences is provided by Bech and Zacharov [2].

The relevant attributes may depend on the reproduction content and context; therefore, it is also important to consider the aim of spatial audio reproduction. This is discussed in Section 1.1, before a review of methods used and attributes determined in Section 2. Based on the findings of the literature review presented in these sections, a discussion of the current state of perceptual audio evaluation and the requisite future directions is presented in Section 3.

1.1 The aims of spatial audio reproduction

The evaluation criteria for a spatial audio system will depend on the aim of the reproduction; it may be that different physical metrics or perceptual facets will be relevant in different scenarios. It is therefore worth investigating the different purposes of spatial audio reproduction.

Theile [4] suggests that a system for stereophonic reproduction should *"... satisfy aesthetically and it should match the tonal and spatial properties of the original sound at the same time"*. In some situations, it is likely that there will be a trade-off between these two aspects; in other situations, both may be possible. Theile notes that as psychoacoustic understanding develops and recording and reproduction techniques become more flexible, we are more likely to achieve a successful balance.

In some areas of audio reproduction—such as auralisation for listening tests, room design, or archival purposes—it is certainly desirable to simulate a real listening situation. However, Rumsey [5] suggests that reproduced sound can only approximate real sound. The degree to which this is true is unclear; the difficulties of making controlled comparisons between a real sound scene and a reproduction mean that the two have rarely been compared in the literature, although a reasonably direct comparison experiment was recently performed by Francombe *et al.* [6].

It is often held that reproducing an exact replica of real life audio is not desirable or possible; there may be benefits in reproducing audio in a way that does not try to simulate the corresponding real experience, or in a way that enhances particular audio cues to compensate for a lack of other modalities. It is also possible to recreate sound scenes that do not, or could not, exist in the real world.

The aim of a spatial audio system—and therefore the relevant evaluation parameters—will also be determined by the programme material, listening context, and type of listener; for example, Guastavino and Katz [7] suggest that expert listeners may prefer good localisation whilst consumers prefer immersion.

There is clearly a division in the aim of a spatial audio reproduction system. In some scenarios, it is desirable to mimic a real life sound field, and an ideal system would be able to produce a listening experience that is perceptually indistinguishable from real life. In other situations, it is not necessary or desirable for the listening experience to mimic real life. The goal may be to produce an impossible auditory experience that sounds believable; to create a completely novel and unbelievable experience; or simply to produce a more enjoyable listening experience than the comparative real life version.

2. Perceptual evaluation of spatial audio

Spatial quality has long been considered an important aspect of sound quality; Letowski [8] divides sound quality into 'timbral' and 'spatial' quality, and Rumsey *et al.* [9] suggested that spatial quality accounts for approximately 30% of basic audio quality. However, the two domains do overlap [10]; this is accounted for in Letowski's heirarchical multilevel auditory assessment language (MURAL) for description of an 'auditory image' [8], in which a number of the outer categories cross the boundary between 'timbre' and 'spaciousness'. Many of the numerous studies of spatial attributes begin by stating that spatial aspects have been overlooked in favour of timbral aspects. Whilst this may have been the case in the past, the number of studies considering spatial attributes suggests that the balance has to some extent been redressed. A range of such studies are discussed in Section 2.1, followed by research that has considered the relationships between different attributes in Section 2.2.

2.1 Attribute elicitation for spatial quality evaluation

Nakayama *et al.* [11] studied the effects of reproducing audio through an increasing number of channels (one to eight). They used eight loudspeakers in the horizontal plane surrounding the listener in an anechoic chamber, and fed the loudspeakers from recordings made using eight unidirectional microphones located at the same relative positions as the loudspeakers. Participants were asked to make preference ratings in a single-stimulus presentation and similarity judgements in a paired comparison test. Multidimensional scaling was performed on the similarity judgements and three factors were found to contribute to preference (explaining 77% of the variance): the factors were labelled *depth of image source*, *sensation of fullness*, and *sensation of clearness*. The eight-channel reproduction was found to be 'near' and 'full', whilst the two-channel reproductions had low 'fullness'.

Zacharov and Koivuniemi [12] performed a DA experiment to discover ways of assessing spatial sound perception from systems including monophonic, different stereo recording techniques, five-channel, and eight-channel periphonic reproduction (employing first-order ambisonics). They collected preference ratings and conducted a language development task using individual and group vocabulary methods. Twelve attributes were produced: *sense of direction*, *sense of depth*, *sense of space*, *sense of movement*, *penetration*, *distance to events*, *broadness*, *naturalness*, *richness*, *hardness*, *emphasis*, and *tone colour*. Partial least squares regression (PLS-R) was used to map the preference ratings to the attributes, resulting in extraction of two dimensions: the first was positively loaded by *movement*, *depth*, and *space * naturalness*, and negatively loaded by *emphasis*; the second was positively loaded by *broadness * tone colour*, *broadness*, and *penetration*, and negatively loaded by *penetration * distance*, *direction*, and *distance*.

Guastavino and Katz [7] investigated the differences between spatial audio systems with ambisonic recordings of soundscapes and live musical performances replayed over various systems. A free verbalisation task was used to describe the experience of listening to 2D and 3D reproductions with and without subwoofer. This process produced seven categories: *presence/immersion*, *readability of the scene/sense of space*, *distance to the scene*, *timbre*, *stability*, *localisation*, and *hedonic judgements*. The 2D systems were generally preferred—they were found to be enveloping and spatially well defined, and they provided a good sense of immersion (equating to high presence). The 3D systems were not enveloping and sounded further away. The subwoofer was good for realism in traffic noise but otherwise the methods including the subwoofer were found to be too rich in bass frequencies. In a subsequent experiment, 2.1, 6.1, and 12.1 reproductions were compared using six musical and environmental stimuli. Alongside free descriptions, ratings were made on six scales from the first experiment. *Readability* and *localisation* were found to be correlated, as were *presence* and *distance*, and *distance* and *coloration*. *Presence* and *readability* were found to play an important role in overall quality.

Berg and Rumsey [13] performed an experiment to elicit verbal descriptors and reduce them to attribute scales for representation of perceived quality of a spatial audio stimulus set with music,

speech, and environmental material, reproduced over variants of monophonic to five-channel systems. A repertory grid technique (RGT) elicitation was performed, and a total of 342 constructs were elicited. The most common utterance was related to the fact that the sounds had been produced by loudspeakers, not natural sources. Other common descriptions were *width* and *source location*, and the sense of being surrounded. Two cluster analyses were performed; it is difficult to determine exact attribute labels following this type of analysis, but the following terms were used to describe one or more clusters: *externalisation*, *distance/depth*, *phase*, *localisation (front–back)*, *envelopment*, *source depth*, *room perception*, *(source) width*, *detection of background sounds*, and *frequency spectrum*.

Choisel and Wickelmaier [14] also elicited attributes for evaluation of multichannel reproduced sound, utilising paired and triadic comparison versions of the RGT method alongside perceptual structure analysis (PSA). They used commercially available classical and pop recordings and tested a variety of reproduction systems, from mono to five-channel. Following the RGT elicitation, ratings were made in a multiple stimulus presentation, and cluster analysis used to reduce the attributes. Twelve spatial categories were determined: *width*, *envelopment*, *spaciousness*, *elevation*, *vertical spread*, *distance*, *depth*, *homogeneity*, *focus/blur*, *skew*, *stability*, and *presence*. Additionally, four timbral categories (*brightness*, *spectral balance*, *sharpness*, and *bass*) and four other categories (*naturalness*, *clarity*, *loudness*, and *miscellaneous*) were established. Results from the PSA method were found to be similar, suggesting that there is a consistent underlying perceptual structure and the exact elicitation method is of secondary importance. Eight attributes were found to be common between the methods: *width*, *envelopment*, *elevation*, *spaciousness*, *brightness*, *distance*, *clarity*, and *naturalness*.

Guastavino *et al.* [15] performed an experiment to compare the spatial quality of various soundscapes (including a live concert) reproduced by three methods: binaural recordings over stereo loudspeakers (with crosstalk cancellation), ambisonic recordings reproduced over six regularly-spaced loudspeakers, and stereo recordings (ORTF pair) replayed over standard two-channel stereo. For one sample (an outdoor recording of traffic noise), subjects were asked to listen to the three reproduction methods and freely describe the experience, choose the item that sounded most similar to day-to-day listening, and justify their choice. For the remaining samples, ratings were made on six scales (from previous research): *envelopment*, *immersion*, *representation*, *readability*, *realism*, and *overall quality*. Responses from the open questionnaire were classified into the following categories: *immersion/envelopment*, *distance*, *rear sound*, *low frequencies*, *readability*, *phasing effect*, and *timbre*.

Lorho [16] investigated the perceptual characteristics of stereo loudspeaker systems found in mobile multimedia devices. A range of popular music stimuli were used, replayed on mono and stereo devices. Individual attribute sets were developed by a group of listeners, leading to a total of 111 terms. The experimenter made a qualitative grouping of the terms and identified five groups: *loudness aspects*, *spatial aspects* (*width*, *stereo effect*, *spatial focussing*, *spaciousness*, *distance*, *presence*, *echo*, *effect output*, *spaciousness in distance*, *symmetry of stereo picture*), *timbral aspects*, *sound disturbance aspects*, and *sound articulation aspects*.

Recently, Lindau *et al.* [17, 18] proposed a forty-eight item semantic differential scale for assessment of *"... real, imagined, and simulated acoustic scenes."* Stimuli are considered to constitute foreground sources, background sources, the simulated room acoustical environment, the reproduction system, and the laboratory environment. The forty-eight scales fall into eight categories: *timbre*, *tonalness*, *geometry*, *room*, *time behaviour*, *dynamics*, *artefacts*, and *general*. The attributes are more relevant to the description of a scene than to a comparative evaluation of different reproduction methods.

2.2 Attribute grouping

The research presented above shows that whilst the dimensions of spatial perception have been investigated for a variety of stimuli and reproduction methods, the resulting picture is complex and there is no consensus on the correct attributes to use for a detailed dimensional evaluation of spatial

audio. In fact, where spatial audio is evaluated, a general rating of *overall spatial quality* or similar is often used (e.g. [19]). Conetta *et al.* [20] considered a set of low-level spatial attributes but ultimately collected ratings of overall spatial quality (and went on to model overall quality in terms of metrics that did not directly measure specific perceptual attributes [21]). The ITU BS.1116 standard [22] recommends that *basic audio quality* is evaluated in every case, but allows further assessment of *stereophonic image quality* for stereo systems, and *front image quality* and *impression of surround quality* for multichannel systems.

Perhaps in response to this, Rumsey [5] advocates a *"horses for courses"* approach to selecting appropriate attributes. The downside to this approach is a large degree of redundancy in research: it seems unnecessary for researchers to develop a new set of attributes for evaluation of each new spatial audio system, especially given the apparent considerable overlap between terminology.

Berg and Rumsey [13] attempted to relate their findings to those of other elicitation experiments. Whilst they found some apparent overlap, this rather informal method of comparing similar semantic themes is perhaps symptomatic of the challenging nature of the task of identifying the relevant perceptual attributes; various researchers follow their own rigorous and scientific method of determining attributes, but when it comes to comparing between studies there is often little more consensus than "we feel that our definition of x corresponds with their definition of y".

Regardless of this complex picture, it seems evident from the similarity in language used and the results of studies employing different methodologies and with different reproduction systems that there is overlap between the produced attribute sets, and that it might be possible to approach an underlying perceptual representation of the experience of spatial audio replay. Consequently, a number of studies have focussed on trying to determine attribute categories.

Rumsey [23] presents a 'scene-based paradigm' for evaluation of spatial audio scenes. The first major division of spatial quality separates *imaging quality* and *spatial impression* [5]. The former category refers to the accuracy with which reproduced sources can be localised as well as the impression of details of the physical space such as perceived *width* and *depth* (and by extension, *height*). The latter category can be further divided into two: *envelopment* and *naturalness*. Both are difficult to define, but envelopment seems to refer to perceived immersion in the sound field, whilst naturalness is linked to *ecological validity* [15] or *presence* [23, 24, 25], defined by Rumsey as *"... the sense of being inside an (enclosed) space.* [...] *In other words, subjects feel present within the space rather than absent from the space."*.

Rumsey points out that it is hard to ensure that the descriptors are completely orthogonal. For example, it is difficult to imagine some sources being simultaneously localizable and enveloping: *"... at what point does the attribute we call source width become another one called envelopment? (The correct answer is probably, 'when subjects say that it does.')"* [23]. This lack of orthogonality—and to some extent the lack of agreement in language between studies—is not considered to be a major problem; Rumsey [5] states that *"Independence and one-dimensionality may have to be sacrificed to usefulness and meaning of the results in many practical situations, particularly when one is more interested in assessing products or systems than one is in learning more about the minutiae of human perception."*

Berg and Rumsey [26] collected ratings of simple stimuli on twelve attributes from previous elicitation studies: *naturalness*, *presence*, *preference*, *envelopment*, *source width*, *localisation*, *source distance*, *room width*, *room size*, *room spectral bandwidth*, *room sound level*, and *background noise level*. They found that every attribute was able to be used to distinguish between stimuli, suggesting that there is a consensus of understanding between participants. Cluster analysis on the results revealed three groups: *naturalness*, *presence*, *preference*, *envelopment*, *room spectral bandwidth*, and *source width*; *background noise level*, *room width*, *room size*, *room level*, and *source distance*; and *localisation*. Three dimensions were also extracted from a principal component analysis (PCA): the first component concerned *general aspects*, *naturalness*, and *preference*. The second was related to *source distance*, *room level*, and *presence*. The third concerned *source image focus*. These categories were

simplified to *general*, *source*, and *room* factors; however, it was noted that the participants may have been biased by having prior knowledge of the attributes and approximately how they fell into these groups. It was also found that *naturalness* and *presence* were highly correlated, as were *preference* and *envelopment*.

Choisel and Wickelmaier [27] used a paired comparison method to collect preference ratings and attribute ratings using previously elicited descriptors [14]. They used probabilistic choice models to develop ratio scales, and found that preference could be predicted using two principal components: a timbral component and a spatial component. However, one attribute loading heavily onto the timbral dimension was *elevated*, described as follows: *"Some sounds might appear to be positioned at the same level as your ears. Some others might be lower (closer to the floor) or higher (toward the ceiling). Indicate which sound you perceive as being higher in space."* This description seems very much like a spatial rather than timbral attribute.

Le Bagousse *et al.* [28] attempted to reduce a list of attributes from the literature [29] into attribute families. A simple free categorisation task was performed by subjects, who were asked to determine between two and five groups from the original twenty-nine attributes, and to label their groups. Seventy-three groups were produced by eighteen participants. The resultant data matrix was analysed using agglomerative hierarchical clustering, and three families were identified (with one split into two subgroups). The families were named *defaults* (interfering elements or nuisances—*noise*, *distortion*, *background noise*, *hum*, *hiss*, *disruption*); *space* (spatial impression related characteristics—*depth*, *width*, *localisation*, *spatial distribution*, *reverberation*, *spatialisation*, *distance*, *envelopment*, *immersion*); and *timbre* (sound colour—*brightness*, *tone colour*, *coloration*, *clarity*, *hardness*, *equalisation*, *richness*, and timbre/other—*homogeneity*, *stability*, *sharpness*, *realism*, *fidelity*, and *dynamics*).

Whilst such analysis is useful for categorising attribute types, it still does not answer the questions posed above as to which of the individual descriptors are orthogonal or of particular importance to the overall listening experience.

3. Discussion

The review presented above highlighted three problems in current spatial audio evaluation.

There is little agreement on the exact definitions of the attributes, nor on the relative importance of each attribute. Whilst the various studies performed have often produced similar terms, and authors have attempted to synthesise full attribute sets (most notably Rumsey's 'scene based paradigm' [23]), there has been no conclusive agreement reached about the definitions that should be used, and consequently studies that purport to analyse similar concepts may introduce important differences. Similarly, whilst large attribute sets have been produced, there is no clear picture of which of these attributes are important, or on the number of perceptual dimensions that listeners can differentiate on. This has led to either the approach of asking for ratings on one overall term (such as *basic audio quality* or *spatial audio quality*), or the development of large and cumbersome attribute sets (such as Lindau's spatial audio quality inventory [17]) which are time-consuming to use.

There may be important attributes that have not yet been determined. As loudspeaker layouts develop to include more channels, new signal processing methods are developed, and the object-based audio production chain becomes more prominent, new attributes that affect the listening experience may be determined. Many studies have focussed on the difference between alternative reproduction methods, but there has been little focus on the difference between real and reproduced audio. It is important that attribute elicitation in the spatial audio literature keeps up to date with the current state of the art in production, transmission, and reproduction of spatial audio, and does not neglect live (non-reproduced) sound.

There is no model of overall spatial quality based on the most important attributes. Listening tests are expensive and time-consuming to run, and this problem is exacerbated where there are large numbers of individual attributes to rate. Where perceptual models exist, these can save a great deal of

time; therefore, it would be beneficial if the perceptual correlates of the important attributes could be ascertained in order that perceptual models could be produced that predicted ratings for each of these attributes as well as combining them to predict overall quality.

3.1 Proposed future work

As a consequence of the convoluted picture presented in the literature, lower-level attributes are generally neglected in favour of collecting ratings of overall spatial quality or basic audio quality. In order to clarify the literature and provide useful guidelines for researchers investigating spatial audio, the authors contend that future research should focus on the following goals.

Ascertaining which attributes of reproduced spatial audio are most important to listeners. It would be beneficial to reduce redundancy and highlight only those attributes that contribute significantly to the quality of listening experience, as these are the aspects that could improve the listening experience and drive innovation in spatial audio systems. It is also necessary to determine contextual effects on the important attributes, or to at least make clear the intended application areas of any attribute sets.

Identifying any important attributes that are currently missing. As noted above, audio systems are continually advancing, and the important attributes may therefore change. It is important that any models can accommodate currently relevant technologies. Evaluation of radically new systems should be sensitive to the fact that new attributes may be important, but elicitation experiments should also be wide-ranging and consider all relevant degradations in order to provide robust results.

Determining the relationships between the important attributes and listener preference. As well as determining the attributes that are important to listeners, quantitative relationships should also be determined. This will help to identify advances that have the potential to make a significant impact on quality of listening experience, and avoid the situation in which an improvement in one attribute gives a degradation in another (and therefore negatively affects the overall experience).

Modelling overall spatial quality in terms of the important perceptual attributes, and modelling these perceptual attributes in terms of their physical correlates. It would save considerable time, money, and effort in running listening tests if there existed models of the important perceptual attributes of spatial quality that could be used to make perceptually-informed predictions. The first stage of such models involves determining the physical correlates and their relationship to the attributes.

4. Acknowledgements

This work was supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1) and the BBC as part of the BBC Audio Research Partnership

REFERENCES

1. B.F. Hertz. 100 years with stereo-the beginning. *J. AES*, 29(5):368–372, May 1981.
2. S. Bech and N. Zacharov. *Perceptual audio evaluation: theory, method and application*. Wiley, Chichester, 2006.
3. S. Zielinski, F. Rumsey, and S. Bech. On some biases encountered in modern audio quality listening tests - a review. *J. AES*, 56(6):427–451, June 2008.
4. G. Theile. On the naturalness of two-channel stereo sound. In *AES 9th Int. Conf.: Television Sound Today and Tomorrow (Paper No. 9-024)*, February 1991.
5. F. Rumsey. Subjective assessment of the spatial attributes of reproduced sound. In *AES 15th Int. Conf.: Audio, Acoustics & Small Spaces (Paper No. 15-012)*, October 1998.

6. J. Francombe, T. Brookes, and R. Mason. Elicitation of the differences between real and reproduced audio. In *AES 138th Conv., Warsaw, Poland*, May 2015.
7. C. Guastavino and B.F.G. Katz. Perceptual evaluation of multi-dimensional spatial audio reproduction. *J. Acoust. Soc. Amer.*, 116(2):1105–1115, August 2004.
8. T. Letowski. Sound quality assessment: concepts and criteria. *AES 87th Conv., New York, USA (Paper No. 2825)*, October 1989.
9. F. Rumsey, S. Zieliński, R. Kassier, and S. Bech. On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *J. Acoust. Soc. Amer.*, 118(2):968–976, August 2005.
10. F.E. Toole. Subjective measurements of loudspeaker sound quality and listener performance. *J. AES*, 33(1/2):2–32, February 1985.
11. T. Nakayama, T. Miura, O. Kosaka, M. Okamoto, and T. Shiga. Subjective assessment of multichannel reproduction. *J. AES*, 19(9):744–751, October 1971.
12. N. Zacharov and K. Koivuniemi. Audio descriptive analysis & mapping of spatial sound displays. In *Proceedings of the 2001 Int. Conf. on Auditory Display, Espoo, Finland*, June 2001.
13. J. Berg and F. Rumsey. Identification of quality attributes of spatial audio by repertory grid technique. *J. AES*, 54(5):365–379, May 2006.
14. S. Choisel and F. Wickelmaier. Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound. *J. AES*, 54(9):815–826, September 2006.
15. C. Guastavino, V. Larcher, G. Catusseau, and P. Boussard. Spatial audio quality evaluation: comparing transaural, ambisonics and stereo. In *13th Int. Conf. on Auditory Display, Montreal, Canada*, June 2007.
16. G. Lorho. Perceptual evaluation of mobile multimedia loudspeakers. *AES 122nd Conv., Vienna, Austria (Paper No. 7050)*, 2007.
17. A. Lindau. Spatial audio quality inventory (SAQI). Test manual. *Audio Communication Group, TU Berlin*, February 2014.
18. A. Lindau, V. Erbes, S. Lepa, H-J. Maempel, F. Brinkman, and S. Weinzierl. A spatial audio quality inventory (SAQI). *Acta Acustica united with Acustica*, 100(5):984–994, September 2014.
19. A. Härmä, M. Park, and A. Kohlrausch. Data-driven modeling of the spatial sound experience. In *AES 136th Conv., Berlin, Germany (Paper No. 9025)*, April 2014.
20. R. Conetta, T. Brookes, F. Rumsey, S. Zielinski, M. Dewhirst, P. Jackson, S. Bech, D. Meares, and S. George. Spatial audio quality perception (part 1): impact of commonly encountered processes. *J. AES*, 62(12):831–846, January 2014.
21. R. Conetta, T. Brookes, F. Rumsey, S. Zielinski, M. Dewhirst, P. Jackson, S. Bech, D. Meares, and S. George. Spatial audio quality perception (part 2): a linear regression model. *J. AES*, 62(12):847–860, January 2014.
22. ITU-R. Recommendation BS.1116-1: methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. 1997.
23. F. Rumsey. Spatial quality evaluation for reproduced sound: terminology, meaning, and a scene-based paradigm. *J. AES*, 50(9):651–666, September 2002.
24. K. Ozawa, Y. Chujo, Y. Suzuki, and T. Sone. Contents which yield high auditory-presence in sound reproduction. *Kansei Engineering International*, 3(4):25–30, September 2002.
25. K. Ozawa, Y. Chujo, Y. Suzuki, and T. Sone. Psychological factors involved in auditory presence. *Acoustical Science and Technology*, 24(1):42–44, January 2003.
26. J. Berg and F. Rumsey. Verification and correlation of attributes used for describing the spatial quality of reproduced sound. In *AES 19th Int. Conf.: Surround Sound – Techniques, Technology, and Perception, Schloss Elmau, Germany*, June 2001.
27. S. Choisel and F. Wickelmaier. Evaluation of multichannel reproduced sound: scaling auditory attributes underlying listener preference. *J. Acoust. Soc. Amer.*, 121(1):388–400, January 2007.
28. S. Le Bagousse, M. Paquier, and C. Colomes. Families of sound attributes for assessment of spatial audio. *AES 129th Conv., San Francisco, CA, USA (Paper No. 8306)*, November 2010.
29. S. Le Bagousse, Catherine C., and M. Paquier. State of the art on subjective assessment of spatial sound quality. In *AES 38th Int. Conf.: Sound Quality Evaluation (Paper No. 5-3)*, June 2010.