A ROADMAP FOR ASSESSING THE QUALITY OF EXPERIENCE OF 3D AUDIO BINAURAL RENDERING

Rozenn Nicol, Laetitia Gros, Cathy Colomes
Orange Labs
Lannion, France
rozenn.nicol@orange.com

Olivier Warusfel, Markus Noisternig, Hélène Bahu
UMR STMS IRCAM-CNRS-UPMC
Paris, France
olivier.warusfel@ircam.fr

Brian FG Katz, Laurent S. R. Simon
LIMSI-CNRS
Orsay, France
brian.katz@limsi.fr

ABSTRACT

Today there are 2 major evolutions in spatial audio. First, an enhanced 3D audio experience, where virtual sound sources can be accurately synthesized in any direction, is possible with technologies such as binaural, Wave Field Synthesis, Higher Order Ambisonics or Vector Base Amplitude Panning. Second, 3D audio is on the way to being democratized through binaural adaptation for headphone listening. These evolutions call for revisiting the methods and tools used to assess the perception of spatial sound reproduction. The first objective of this paper is to delineate the problem, by exploring the potential dimensions and the related attributes underlying the perception of spatial sound, mainly within the context of binaural reproduction. Secondly, assessment methods, including both standard and less conventional ones, are listed, and their relevance for the measure of the attributes previously identified is discussed.

1. INTRODUCTION

Sound spatialization is undergoing major evolutions with promises of enhanced 3D audio experience and the recent inclusion of height information. Beyond discrete channel audio, sound field representation formats such as Wave Field Synthesis (WFS) and Higher-Order Ambisonics (HOA) are now appearing in the professional community. These technologies typically require large sophisticated installations. However, binaural audio playback using Head Related Transfer Functions (HRTFs) makes 3D audio immersion possible for any listener using nothing more than standard headphones.

The use of 3D audio technologies allows for synthesizing virtual sound sources at any position in space with an accuracy that is very close to natural listening [1]. Binaural audio with headphones could be the first technology which makes immersive 3D audio mass consumable. Spatial information becomes then an inherent feature of the experience. This raises the question of to what extent we can quantify and optimize the quality of experience (QoE) in immersive audio systems.

The consequence is that methods and tools for audio assessment should be revisited. At least, a deeper investigation of spatial attributes is required. Perceptual attributes, such as "spatial impression", "depth", "envelopment", "width" [2, 3], which were identified by past studies on room acoustics and linked to spatial perception, are no longer sufficient. More generally, our main concern here is to measure how a 3D sound scene is perceived by a listener, whatever technologies or listening setups are used. The objective is to assess the perceived "quality", i.e. to identify the perceptual dimensions used by the listeners to make their judgments. Two main categories of dimensions are already identified: 1- the spectral content (i.e. timbre), 2- the spatial location of the sound [4]. More global audio quality ratings rely on other perceptual dimensions which have to be investigated, and for which new tools of assessment are yet to be identified. Naturalness, or plausibility, of the virtual sound sources is one example of such additional dimensions.

It is beyond the scope of this article to answer these questions on a global scale. In the following we will mainly focus on the context of binaural sound reproduction, in the case of a direct binaural synthesis of the different audio objects of a sound scene as well as in the case of the binaural decoding of any channel based audio format using the virtual speaker paradigm. It is well known from literature, for instance, that the use of non-individualized HRTFs degrades both the spectral and spatial "quality" of the reproduced sound sources. Besides the use of individual or generic HRTFs, the decoding of a stereo or surround format through the virtual speaker paradigm is also less convincing than a direct binaural rendering of the sound sources. Thus our concern may be the assessment of the "quality" of any given set of HRTFs. By "quality" it is meant to measure how a source is perceived by a listener when processed by a given HRTF or how a sound scene is perceived when rendered or decoded through binaural synthesis. Usually the quality of HRTFs is assessed by localization tests, which mainly focus on the localization accuracy and often do not measure other perceptual dimensions (such as timbre).

Another conventional method of assessment is the measure of Basic Audio Quality (BAQ) [5], where "Quality" here refers to the fidelity with which a signal is transmitted or rendered by a system. In other words, degradations are measured in comparison to a given reference. It has been widely used in perceptual audio experiments as a quick means to compare different alterations of a signal, and was initially developed to classify audio codecs. Applying them to the assessment of the quality of HRTFs is questionable. Assessing binaural reproduction and in particular the suitability of a set of HRTFs can be considered as a measure of degradation. In most cases, the reference is unknown to the listener, unless the subject can directly compare the real audio scene (e.g., sound played from a real loudspeaker at the same position as the virtual sound source, played from a virtual loudspeaker) with its binaural representation. In addition, evaluating BAQ gives little, if any, information about the artifacts of a given binaural technique. The exception to this rule is the assessment of modeled HRTFs, in which the reference is the measured individual HRTF. However, this choice of reference is not necessarily the ideal case, insofar as real world HRTF measurements are not free from errors and approximations. Besides, the claim that measured individual HRTFs provide the best overall audio experience (i.e. not only in terms of localization, but for all the other perceptual dimensions) remains to be verified.

We propose using the term Quality of Experience (QoE) as an alternative to BAQ. QoE measures how a subject experiences a given system. Since the range of situations that can be covered by binaural reproduction is very large, we believe this QoE should remain multidimensional and multicontextual.

The remainder of this article is an attempt to draw a roadmap of the different dimensions and methods that should be investigated to delineate the numerous dimensions of the Quality of Experience. In the following, two principal questions will be addressed. Section 1 will first explore what are the potential dimensions underlying the perception of binaural sound. Then, Section 2 will present an overview of available methods of perceptual assessment. Both conventional and new tools will be considered. Three different groups of methods to assess the perceptual audio quality will be used for classification: a) direct assessment of a perceptual attribute without any reference; b) direct assessment with a reference; and c) indirect assessment, in which case the quality is inferred from the subject's behavior (e.g., measure of task performance). For each perceptual dimension the relative merits of the different methods will be discussed.

2. PERCEPTUAL DIMENSIONS OF BINAURAL SOUND

Binaural audio consists of a left and right ear signal that can be directly played back over headphones. These signals can be directly recorded (e.g., using a dummy head or in-ear microphones), or synthesized using individual or non-individual HRTF filters. QoE addresses the questions of a) how can a listener describe his/her perception, and b) what are the objective features (especially acoustical) and how do they correspond to the perceptual dimensions. Some dimensions are already known. For instance, the physical properties of the sound scene that the binaural sound intends to reproduce have clearly an influence on perception, namely: the frequency content of the sound signal, the location of the sound source, the acoustic environment (room effect), etc. The perceptual attributes related to these physical parameters are called "physical-related attributes" in the following discussions. Another category of attributes concerns the effect on the psychic or affective state of the listener: are the virtual sound sources plausible, to what extent does the listener feel immersed in the virtual sound scene, what are his/her emotion(s), etc.? More generally, perceptual studies need to be reconsidered in order to take into account the specific context of listening, involving perception-action feedback according to a given task or cognitive situation [6].

2.1. How to investigate perceptual dimensions?

Different methods to identify the perceptual dimensions and associated attributes have been proposed in the literature. The two most prevalent approaches are: 1- Multi-Dimensional Scaling (MDS) and 2- verbal elicitation techniques, e.g. Descriptive Analysis (DA), Repertory Grid Technique (RGT), or Free-Choice Profiling (FCP). MDS measures the perceptual dissimilarities between a large set of stimuli and derives the most relevant perceptual parameters from these distances. However, MDS methods do not always guarantee that the dimensions revealed by the analysis correspond actually to perceptual attributes. The INDSCAL method has been introduced to overcome this limitation [7] and has been used for musical timbre analysis [8] as well as for perceptual studies in room acoustics [9]. Alternatively, DA, RGT and FCP are direct verbal methods for eliciting and evaluating people's subjective experiences and perceived differences between the stimuli of a similar large set of stimuli [10, 11]. They originate from the food industry, where they've been used for creating dictionaries of words for flavour qualification. These methods help in constructing a space corresponding to the perceptual attributes elicited in the given set of stimuli, and can also infer the words that best describe each end of a perceptual dimension. However, DA, RGT, and FCP are time-consuming methods, both for the experimentor as well as for the subjects of the experiment.

According to [12], the first step of dimensional analysis methods is to generate a large set of stimuli that is representative of the differences encountered in the area of interest, so that many or all perceptual differences can be expected to be found when comparing all the stimuli.

In the context of binaural sound reproduction, building a set of representative stimuli is a non-trivial task. Binaural stimuli are highly individual: HRTFs vary widely from one subject to another and using non-individual HRTFs leads to large perceptual differences. Sets of stimuli that include non-individual binaural signals will therefore be perceived differently from one subject to another, and may lead to different constructs from one subject to another, making the statistical analysis of the elicitation process more complex. A solution would be to generate a set of sufficiently different stimuli so that similar constructs are found. It would therefore need to include individual HRTFs, deteriorated individual HRTFs and non-individual HRTFs. The deteriorations would need to be perceptually identical from one subject to another, which is not possible as long as the perceptual attributes and their underlying models remain unknown. Finding the perceptual dimensions in the context of binaural recordings should therefore be considered as an iterative process, where several consecutive experiments of perceptual dimension identification should be conducted. In these conditions, an exhaustive identification of all the dimensions of the perception of binaural sound is a very difficult task, if possible at all.

2.2. Physical-related attributes

Physical-related attributes describe perceptual attributes that can be directly linked to a physical or mathematical property of either the sound source, the acoustic space, or the sound reproduction system.

2.2.1. *Timbral attributes*

According to the British Oxford dictionary, timbre is *the character or quality of a musical sound or voice as distinct from its pitch and intensity*. According to this definition, spatial characteristics should be part of a sound's timbre. Indeed the acoustic response of the room contributes to the timbre of sound at the listener's place. At different positions in a room listeners will not perceive the same timbre of a sound. What's more, in binaural reproduction, timbre has an ambiguous role: spectral features are partly interpreted as localization cues. However, a number of studies separate timbre properties of a sound from its spatial properties [13, 4]. In this paper, we will consider timbral attributes separately from spatial attributes, though some attributes may overlap.

Several lists of timbral attributes have been designed. [14] proposed a list of timbral attributes that can be used in machine learning. However, attributes such as the zero crossing rate or centroid temporal peakedness can hardly be used by subjects to describe timbral qualities of sound. For this reason, lists of specific perceptual attributes of timbre have been developed for music instruments [15], speech [16], and loudspeakers [10]. Timbral attributes are generally related to the spectro-temporal properties of the sound.

2.2.2. *Source location*

The location of a sound source is, to some extent, directly perceptible by listeners. It is generally expressed in terms of azimuth, elevation and distance. Spherical coordinate systems are preferred to cartesian coordinate systems as the sound source position is perceived relatively to the subject itself. Localizability, or spatial definition, refers to the ease of localizing the sound and is an additional aspect worth considering.

Perception of distance is rarely assessed in listening tests (stereophonic or multichannel reproduction). However, it should be highlighted that, in the specific case of binaural sound, distance perception is of greater importance. This is related on one hand to the ability of binaural reproduction to render varying distances and on the other hand to the problem of inside-the-head localization (IHL) and externalisation as a common artifact of binaural reproduction is that the virtual sound source is localized inside or close to the head. Therefore, when judging auditory distance for binaural sound, this phenomena must be carefully examined.

2.2.3. *Perceived width and Apparent Source Width (ASW)*

The perceived width of a sound source is the measure of an auditory event's spatial extent, which can be expressed in terms of an angular span (i.e. spherical sector) and depth. An auditory event does not usually have clear limits, which makes it difficult to define without a reference. In room acoustics, Apparent Source Width (ASW) is affected by from where and at what times the early reflections arrive [17, 18, 19]. ASW is also affected by the physical extent of the source (when not a point source) and to the sound reproduction system [20, 21]. Objective measures have found ASW to be correlated to the interchannel or interaural cross-correlation [22, 23]. Several other attributes are perceptually close to ASW yet different enough to require distinction: locatedness [24] and diffuseness of the source, as well as spatial unity

2.2.4. *Room-related attributes*

The room in which sounds are played and/or recorded has a strong influence on their perceptual attributes. Rooms add reflections to the original sound (the direct sound), which may result in a change in perceived timbre or spatial properties. The source directivity also influences the perceived sound [25]. Some spatial attributes and their characterization can be found in, e.g., [17, 26, 10]. However, the number of perceptual dimensions remains unclear. Early work of Sabine identified 3 attributes ("loudness", "distorsion of complex sounds: interference and resonance", and "confusion: reverberation, echo and extraneous sounds"). Present literature rely on at least 7-9 necessary attributes [27, 28]. The study of their correlation with acoustic criteria showed that the perception of room quality is mainly influenced by the energy of the direct sound (including early reflections), the overall energy of reverberated sound, the decay time of reverberation (i.e. reverberation time), the time and spatial distribution of early reflections, and the frequency balance of each criteria, all of which have well-defined metrics.

2.2.5. *Discussion*

Several physical parameters which potentially represent a sound scene are listed in the previous section. For some, it is unknown if and how they affect auditory perception. The intensity of the perceptual effect is probably not the same for each parameter: their relative weights need also to be determined. In addition, this physical description should be revisited in light of studies concerning auditory scene analysis [29]. The brain distorts the "physical reality" for building the associated percept (concept of auditory streams based on grouping or segregation of auditory events). Information is ordered by mental and cognitive processing, sometimes independently from the physical properties. Some studies suggest for instance that frequency features are of primary importance, above spatial properties [29].

2.3. Psychic and affective attributes

Other attributes may be related more to the listening experience than to physical properties of the sound sources or the properties of the room. Psychic and affective attributes refer to the results of further processing and analysis of the sound scene by the brain. That which is of interest is no longer the "pure acoustic information", but the way in which the psychic state of the listener is modified by the sound. This question is rather new, and we are far from having a clear understanding of all the dimensions involved. It should be noted however that these effects are obviously highly dependent on the audio content and personal experience of the subject. As a first contribution, we propose here to consider 3 potential attributes: naturalness (and its correlates), readability, and emotion.

2.3.1. *Naturalness*

Binaural technology offers the ability to reproduce at the ears of a listener the exact sound that would have arrived at his/her ears if he/she had been located in the original environment. Alternatively, binaural synthesis can be used to create an auditory environment, in which case the only reference a listener may have is an expectation of the auditory scene which might be induced from memory or derived from visual information. Since binaural technology aims at mimicking natural listening, the realism, the naturalness, and the fidelity to the original sound scene or to the expected one, often seems to be a more important question in this context than for other spatial sound technologies.

[30] showed that naturalness is a desirable characteristic of an environment. In addition, one can expect to find naturalness ratings being strongly correlated to preference ratings, as [31] showed. Thus the rating of naturalness might be influenced by preference, even if a preference judgment is not explicitly asked from the listeners. Nevertheless, in some situations, naturalness might not be as desirable as listeners may think. For instance, it happens that, when attending a concert, our natural experience is poor (e.g., the sound scene is perceived as ill-defined and narrow), maybe due to the listening's position. A recording of such a sound scene could then be poorly rated in terms of preference and possibly in terms of realism, in contrast to a more processed recording, where each individual source is clearly definable, and which could be perceived as more natural even though it would be physically impossible.

Naturalness of a sound scene is regularly used in perceptual evaluations [10] but is should be treated carefully. Naturalness is essentially a comparison between an unknown reference (the original sound scene, where the listener often was not present or has an old and potentially erroneous memory of it) and a known signal (the binaural reproduction). [32] questions the desirability of fidelity and naturalness. Nuances of the concept of naturalness, such as plausibility and presence, have been developed to address additional perceptual attributes.

2.3.2. Readability

When listening to a complexe sound scene, i.e. composed of many sources conveying various information, the question of its readability is an important parameter. Here this term refers to the ability to discriminate the different concurrent sound sources, in order to focus on one specific component [33]. Speech intelligibility is one particular example (i.e. cocktail-party effect [34]), but readability it pertinent for any audio content. In classical music, it describes the ability of the listener to dynamically focus on one instrument (or group of instruments) [35]. More generally, readability is one aspect of auditory scene analysis, which allows one to separate the overall scene into several streams with various levels of processing. Readability is affected for instance by frequency or spatial separation.

2.3.3. Emotion

The emotional dimension corresponds to any emotion that is felt by the listener, whether positive or negative. Our primary concern is to acknowledge that the listener is "touched" by the sound, which is an indication of a certain degree of immersion. Nevertheless, the nature of the emotion is also relevant information. After Wundt [36], emotion is described by 3 dimensions: valence (pleasure vs unpleasure), vigilance or dominance (no control vs maximal control), and arousal (excitation vs relaxation). However, valence and arousal have been shown as the more reliable dimensions [37]. Measuring and analyzing emotions is a difficult task, all the more that it is highly dependent on the content and the individual, though some bio-metric data has been shown to provide correlations to certain aspects.

3. WHICH EXPERIMENTAL PROCEDURE FOR WHICH PERCEPTUAL PARADIGM?

Evaluating the above-mentioned perceptual attributes requires suitable experimental protocols. This section and the following one review the most common approaches to perceptual evaluation in the context of binaural experience and discuss for each one the perceptual paradigms they may help to evaluate. From experimental psychology, two main categories of evaluation are distinguished: direct evaluation, where the subject is directly asked to rate the attributes under study, and indirect evaluation, where the subject's perceptual rating is inferred [12].

3.1. Direct measure of attributes

This method intends to assess each perceptual attribute separately. For example, if we consider the perception of timbre and focus on the descriptor "clarity" [38], the subject is asked to rate the clarity of audio stimuli. For each stimulus, he / she gives a score within an appropriate scale ranging between 0%, meaning a muffled sound, and 100%, meaning a clear sound. For each attribute, the choice of the scale (grading, labeling) must be carefully designed as a function of the considered attribute. A single attribute or a set of attributes can be measured at the same time. Another example is localization accuracy testing, where the subject is asked to report the azimuth, elevation, and/or distance of a sound source. In that case, the scales are in degrees or meters. A third example is the assessment of room quality by a questionnaire, in which the subject is asked to rate a set of attributes ("presence", "room effect", "reverberance", etc.).

3.2. Direct measure of attributes with a reference

A reference is defined here as any audio or visual stimulus which is provided to the subject and to which he/she is asked to compare to the signal under assessment. It is not necessarily a reference representative of a high quality standard. The task of the subject is to rate an attribute of an audio sample in comparison to the given reference. In addition, specific anchors, corresponding to various grades along the judgment scale, can be included. This inclusion allows for a check of the reliability of the subject (median anchor) or to insure that the listeners are using the scale consistently through the whole experiment (low anchor). Anchors should be chosen so that the stimuli are distributed along the whole scale that listeners have to use. If not, one may encounter various kinds of biases as described in [39]. [40] describes two types of anchors: direct anchors, that are explicitly given to the listener, similar to references, and indirect or "hidden" anchors, which listeners are not aware of. The main advantage of direct anchoring is that it can help stabilize the range of estimations given by the subjects. Indirect anchors can also be used for that, though less efficiently. Indirect anchors are useful to estimate the reliability of the subject or biases related to the non-uniformity of the results.

Examples of direct measure with a reference are AB or ABX tests, recommended when the differences between stimuli are small, or MUSHRA tests [41] if the differences are moderate, though they were designed to evaluate degradations caused by audio codec compression. A third example is a relative localization test through alignment, where the listener needs to match the perceived direction or distance of a test stimulus with that of a reference stimulus.

3.3. Indirect measure

3.3.1. *Task performance*

In this situation, the subject is asked to perform a task in the context of binaural sound. His/her QoE is inferred from his/her success. For instance, the listener's task can consist in describing the sound scene, that is to report the number, the nature, and the location of the sound sources [42]. Another example is given by [43] who inferred QoE by asking the listener to explore the sound scene and find targets. In [44], Guillon described a localization test where the listener had to localize the virtual sound source as fast as possible, recording the response time for various sets of HRTFs corresponding to different levels of HRTF modeling. Results showed a correlation between the response time and the modeling quality. The general intent of these types of experiments is to derive information about the naturalness and readability of the sound scene from observations of the listener's behavior.

3.3.2. *Physiological measures*

Psychophysiology studies the relationships between physiological responses and psychological changes. The principle is to observe cognitive, emotional, or behavioral phenomena by analyzing the physiological responses of the subject. Electrodermal activity, heart pulse, skin temperature, or eye activity are examples of physiological observables that can be recorded and linked to the psychic state of the listener [45, 46]. Particularly, electrodermal activity and heart pulse are considered as relevant measure of emotions, the former being associated to the "arousal" dimension, the latter to the "valence" dimension.

3.3.3. *Brain imagery*

Magnetic Resonance Imagery (MRI), electroencephalogram or magnetoencephalogram are useful tools for observing brain activity. Particularly, technological progress has made electroencephalograms easier to measure with a simple headset. Understanding the link between neuronal activity and the underlying mental process is improving daily. Currently, we are not able to translate brain activity maps directly into what subjects think or feel. However, some information about their emotions can be inferred from knowledge of neuronal activity and connections. For instance, in [47], brain activity of ferrets listening to virtual sounds was compared between individual and non-individual HRTFs, showing that the spatial selectivity of neurons is strongly altered in the case of non-individual synthesis. This is a potential measure of naturalness or plausibility. Brain activity can also give information on the mental effort required by the listening task, which could be linked to readability. This measure could be of interest when subjects are sound engineers being observed during a post-production session. Brain imagery appears thus as a promising tool to investigate perception of spatial sound in general, and binaural sound in particular.

4. CONCLUSIONS

Assessment of the QoE of spatial sound reproduction is the question raised by this article. Rather than providing answers, this paper attempts to clarify what is known and what requires further investigations, for which several promising tools are recommended. Complementary assessments are needed: on the one hand, overall ratings where all dimensions are taken into account, and on the other hand, unidimensional measures focused on one specific attribute. In order to accomplish this however, the first step is to clarify the number and semantic interpretation of the numerous perceptual dimensions. An open question is whether these dimensions depend on the sound reproduction system, namely the audio spatialization technology used. In addition, it is important to identify or develop objective criteria correlated with the perceptual dimensions.

5. ACKNOWLEDGMENTS

This work was carried out within the context the FUI funded project "BiLi" (www.bili-project.org) with support from "CAP DIGITAL - PARIS REGION".

6. REFERENCES

[1] McAnally K. Martin R. and Senova M., "Free-field equivalent localization of virtual audio," *J. Audio Eng. Soc.*, vol. 49, pp. 14–22, 2001.

[2] "ITU-R BS.1284-1: general methods for the subjective assessment of sound quality," Tech. Rep., 2003.

[3] "EBU Tech 3286 assessment methods for the subjective evaluation of the quality of sound programme material - music," Tech. Rep., 1997.

[4] Sarah Le Bagousse, Mathieu Paquier, Catherine Colomes, and Samuel Moulin, "Sound quality evaluation based on attributes - application to binaural contents," Oct. 2011, Audio Engineering Society.

[5] "ITU-R BS.1387-1: method for objective measurements of perceived audio quality," Tech. Rep., 2001.

[6] J. Blauert, D. Kolossa, K. Obermayer, and K. Adiloğlu, "Further challenges and the road ahead," in *The Technology of Binaural Listening*, Jens Blauert, Ed., Modern Acoustics and Signal Processing, pp. 477–501. Springer Berlin Heidelberg, Jan. 2013.

[7] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition," *Psychometrika*, vol. 35, pp. 283–319, 1970.

[8] J.M. Grey, "Multidimensional perceptual scaling of musical timbres," *J. Acoust. Soc. Am.*, vol. 61, pp. 1270, 1977.

[9] C. Lavandier, *Validation perceptive d'un modèle objectif de caracérisation de la qualité acoustique des salles*, Ph.D. thesis, Université du Maine, Le Mans, 1989.

[10] Nick Zacharov and Kalle Koivuniemi, "Unravelling the perception of spatial sound reproduction: Language development, verbal protocol analysis and listener training," Nov. 2001, Audio Engineering Society.

[11] Gaetan Lorho, *Perceived quality evaluation: an application to sound reproduction over headphones*, Ph.D. thesis, Aalto University, Aalto, 2010.

[12] Nick Zacharov and Soren Bech, *The Perceptual Audio Evaluation: Theory, Method and Application*, John Wiley & Sons, 2006.

[13] Francis Rumsey, Slawomir Zielinski, Rafael Kassier, and Soren Bech, "On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality," *Journal of the Acoustical Society of America*, vol. 118, no. 2, pp. 968–976, Aug. 2005.

[14] James McDermott, Niall J. L. Griffith, and Michael O'Neill, "Timbral, perceptual, and statistical attributes for synthesized sound," *IN: PROC. OF THE INTERNATIONAL COMPUTER MUSIC CONFERENCE. (2006*, 2006.

[15] Taffeta M. Elliott, Liberty S. Hamilton, and Frédéric E. Theunissen, "Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones," *The Journal of the Acoustical Society of America*, vol. 133, no. 1, pp. 389–404, Jan. 2013.

[16] Ville-Veikko Mattila, "Descriptive analysis of speech quality in mobile communications: Descriptive language development and external preference mapping," Nov. 2001, Audio Engineering Society.

[17] David Griesinger, "General overview of spatial impression, envelopment, localization, and externalization," in *15th Audio Engineering Society International Conference: Audio, Acoustics & Small Spaces*, Copenhage, Denmark, Oct. 1998, Preprint 15-013.

[18] David Griesinger, "Objective measures of spaciousness and envelopment," in *Audio Engineering Society 16th International Conference: Spatial Sound Reproduction*, Rovaniemi, Finland, Mar. 1999, Preprint 16-003.

[19] Ingo B. Witew and Johannes A. Buechler, "The perception of apparent source width and its dependence on frequency and loudness," *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 3224–3224, Nov. 2006.

[20] B F G Katz and C. d'Alessandro, "Apparent source width and the church organ," in *7ème congrès de la Société Française d'Acoustique et 30ème congrès de la Société Allemande d'Acoustique (CFA/DAGA 2004)*, Strasbourg, 2004, pp. 1235–1236.

[21] Kimio Hamasaki, Toshiyuki Nishiguchi, Hiroyuki Okubo, Yasushige Nakayama, Reiko Okumura, and Masakazu Iwaki, "Natural reproduction of symphony orchestra music by an advanced multichannel live sound system," in *121st Audio Engineering Society Convention*, Oct. 2006, Preprint 6966.

[22] Russell Mason, Tim Brookes, and Francis Rumsey, "Development of the interaural cross-correlation coefficient into a more complete auditory width prediction model," in *Proceedings of the 18th International Congress on Acoustics*, Kyoto, Japan, 2004, vol. IV, pp. 2453–2456.

[23] Toshiyuki Okano, Leo L. Beranek, and Takayuki Hidaka, "Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls," *The Journal of the Acoustical Society of America*, vol. 104, no. 1, pp. 255–265, July 1998.

[24] Thomas Lund, "Enhanced localization in 5.1 production," in *Audio Engineering Society 109th convention*, Los Angeles, United States, 2000, Preprint 5243.

[25] Markus Noisternig, Franz Zotter, and Brian FG Katz, "Reconstructing sound source directivity in virtual acoustic environments," in *Principles and Applications of Spatial Hearing*, Yôiti Suzuki, Douglas S Brungart, and Hiroaki Kato, Eds., pp. 357–373. World Scientific Publishing Co. Pte. Ltd., 2011.

[26] Jan Berg and Francis Rumsey, "Identification of perceived spatial attributes of recordings by repertory grid technique and other methods," in *Audio Engineering Society 106th convention*, Münich, May 1999, Preprint 4924.

[27] L. L. Beranek, "Concert hall acoustics," *J. Acoust. Soc. Am.*, vol. 92, pp. 1–39, 1992.

[28] E. Kahle, *Validation d'un modèle objectif de la perception de la qualité acoustique sur un ensemble de salles de concert et d'opéra*, Ph.D. thesis, Université du Maine, Le Mans, 1995.

[29] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*, The MIT Press, 1990.

[30] Paul Rozin, "The meaning of "Natural". process more important than content," *Psychological Science*, vol. 16, no. 8, pp. 652–658, 2005.

[31] Jan Berg and Francis Rumsey, "Correlation between emotive, descriptive and naturalness attributes in subjective data relating to spatial sound reproduction," in *109th Audio Engineering Society Convention*, Sept. 2000, Preprint 5206.

[32] Francis Rumsey, "Faithful to his master's voice? questions of fidelity and infidelity in music recording," in *Recorded Music: Philosophical and Critical Reflections*. 2008.

[33] Catherine Guastavino and Brian F G Katz, "Perceptual evaluation of multi-dimensional spatial audio reproduction," *The Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 1105–1115, Aug. 2004, PMID: 15376676.

[34] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions.," *ACUSTICA - acta acustica*, vol. 86, pp. 117–128, 200.

[35] J. Sanson, *Contrôle musical et perceptif de la spatialisation sonore en zone étendue*, Ph.D. thesis, UPMC, Paris, 2011.

[36] W. Wundt, *Gundriss der Psychologie [Outlines of Psychology]*, Leipzig, Germany: Entgelman, 1896.

[37] Bradley M. M. Lang P. J., Greenwald M. K. and Hamm A. O., "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261– 273, 1993.

[38] Gaetan Lorho, "Evaluation of spatial enhancement systems for stereo headphone reproduction by preference and attribute rating," May 2005, Audio Engineering Society.

[39] Slawomir Zielinski, Francis Rumsey, and Søren Bech, "On some biases encountered in modern audio quality listening tests-a review," *Journal of the Audio Engineering Society*, vol. 56, no. 6, pp. 427–451, June 2008.

[40] "ITU-R BT.500-10: methodology for the subjective assessment of the quality of television pictures," Tech. Rep., 2000.

[41] "ITU-R BS.1534-1: method for the subjective assessment of intermediate quality levels of coding systems," Tech. Rep., 2003.

[42] J. Faure, "Evaluation de la synthese binaurale dynamique," Tech. Rep., France Telecom, 2005.

[43] A. Gonot, *Conception et évaluation d'interfaces de navigation dans les environnements sonores 3D*, Ph.D. thesis, CNAM, Paris, 2008.

[44] P. Guillon, *Individualisation des indices spectraux pour la synthèse binaurale: recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF*, Ph.D. thesis, Université du Maine, Le Mans, France, 2009.

[45] J. Lassalle, *Etude de l'influence de la qualité audiovisuelle sur la qualite d'expérience du spectateur: Combinaison d'indicateurs subjectifs, physiologiques et oculaires*, Ph.D. thesis, Telecom Bretagne, 2013.

[46] Cédric R. André, Jean-Jacques Embrechts, Jacques G. Verly, Marc Rebillat, and Brian F. G. Katz, "Sound for 3D cinema and the sense of presence," in *Proceedings of the 18th International Conference on Auditory Display*, Atlanta, USA, June 2012.

[47] Jenison R. Mrsic-Flogel T., King A. and Schnupp J., "Listening through different ears alters spatial response fields in ferret primary auditory cortex," *J. Neurophysiol.*, vol. 86, pp. 1043–1046, 2001.

[48] Gunther Theile, "On the naturalness of two-channel stereo sound," *The Journal of the Audio Engineering Society*, vol. 39, no. 10, pp. 761–767, Oct. 1991.

[49] L. L. Beranek, *Music, Acoustics & Architecture*, John Wiley, New York, 1962.