Acta Acustica united with Acustica
Vol. 100 (2014) 945 – 955

DOI 10.3813/AAA.918774

Vector-Base and Ambisonic Amplitude Panning: A Comparison Using Pop, Classical, and Contemporary Spatial Music

G. Marentakis, F. Zotter, M. Frank

Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Inffeldgasse 10/3, 8010 Graz, Austria. [marentakis, zotter, frank]@iem.at

Summary

Vector-Base Amplitude Panning (VBAP) and Ambisonics are commonly used in 3D audio reproduction via loudspeakers. While research has been investigating their properties using psychoacoustic test signals, there is only a small number of investigations employing musical material. Considering the musical application of these spatialization methods, we present an experimental study characterizing quality aspects using excerpts that belong to three different musical genres (popular, classical, and contemporary spatial music). The study compares seven configurations of vector-base and Ambisonic amplitude panning in a hemispherical listening environment that is permanently installed in the IEM CUBE. Four configurations thereof used 24 loudspeakers, and the others used a subset of 12 loudspeakers. In pairwise comparisons, participants rated each configuration pair on a quasi-continuous scale in terms of preference, envelopment, spatial clarity, sound quality, and stability. Perceptual scales were constructed which revealed how configurations ranked in terms of each attribute. The ranking of the tested configurations on the perceptual scales was dependent on the musical material. In the case of the popular and the classical music piece, results were relatively consistent and participants tended to prefer the configurations that used 12 loudspeakers. Results indicate that preference judgements are correlated to envelopment, sound quality, and spatial clarity.

PACS no. 43.38.Md,43.38.Vk,43.75.Wx

1. Introduction

The majority of spatial audio rendering evaluation studies employ simple stimuli, such as noise, and they mainly focus on 2D or 3D localization and the spatial extent of the auditory events [1, 2, 3, 4, 5, 6], and rarely sound coloration [5, 7]. Evaluation using more ecologically valid stimuli, such as music or soundscapes, has received less attention due to the multidimensional nature of the stimuli. In such cases, evaluation is mostly performed by employing direct scaling of perceptual attributes that have either been found to be relevant in the literature or elicited for the purpose of the study [8, 9, 10, 11]. Both attribute elicitation as well as direct scaling studies have been mostly performed using relatively simple spatialization algorithms such as stereo and 5.1 and primarily in the horizontal plane. Choisel *et al.* [11] has criticized the use of direct scaling procedures in experiments involving multidimensional stimuli, such as music, because 1. subjects might not be able to combine the different dimensions into a single one, 2. subjects tend to use scale extent in an idiosyncratic way, e.g. concentrating on the top or bottom of the scales, 3. the relative inability of such procedures to reliable encode the perceptual distance among the stimuli, and 4. the poor way with which judgment intransitivity is encoded. He proposed using indirect scaling methods for this purpose to help overcome aforementioned problems.

In this study, we seek to evaluate 3D spatial audio renderers using ecologically valid musical stimuli. Following Choisel *et al.* [11], we employ indirect scaling of Preference and the following perceptual attributes: Envelopment, Spatial Clarity, Sound Quality, and Stability. These attributes were selected from attributes that have been proposed in the literature because they form an integral part of what listeners consider to be important for a spatial audio listening [8, 9, 10, 11, 12]. In the literature such attributes were generated using verbal elicitation procedures in response to popular and classical music that has been reproduced primarily using stereo and surround sound rendering in the horizontal plane. To what extent they are applicable in the case of more sophisticated 3D rendering techniques and in the case of contemporary music has not been examined yet.

In the experiment we present here, three three-dimensional spatial audio renderer configurations (VBAP, energy-preserving and all-round Ambisonics) realized on a hemispherical fixed loudspeaker setup were compared us-

Received 12 May 2014,
accepted 5 July 2014.

ing musical material of substantial variability, ranging from pop, to classical, and contemporary music. Based on the results we make inferences on the feasibility of the undertaking, the extent to which judgments were affected by the material itself, the way preference can be explained on the basis of the other four attributes, and on how the different renderers were rated by the listeners. We present first the 3D audio renderers we have used, among them state-of-the-art hemispherical Ambisonics decoding [13, 14], and then the experiment and its results.

2. Amplitude panning

Amplitude panning aims at steering the perceived direction of an individual auditory object by distributing its sound signal $s(t)$ to loudspeakers using frequency-independent, real-valued, and largely positive weights g_l. To obtain stable loudness in common listening environments, weights should be normalized $\sum_l g_l^2 = 1$.

The signal of the lth loudspeaker is simply obtained by

$$x_l(t) = g_l\, s(t). \tag{1}$$

The amplitude-panned sound at its desired direction is called a virtual source. The amplitude panning algorithm calculates the weights $g_l = g_l(\boldsymbol{\theta}_\mathrm{s})$ for the virtual source direction $\boldsymbol{\theta}_\mathrm{s}$.

We denote any direction, be it of virtual source or loudspeakers, by its $\mathbb{R}^3$ unit direction vector $\boldsymbol{\theta} = [\cos\varphi\cos\vartheta, \sin\varphi\cos\vartheta, \sin\vartheta]^\mathrm{T}$ which depends on the azimuth angle φ and the elevation angle ϑ in the spherical coordinate system.

2.1. Vector-Base Amplitude Panning (VBAP)

VBAP [15] defines weights $\tilde{g}_l$ fitting a weighted superposition of the loudspeaker direction vectors $\sum_l \tilde{g}_l \boldsymbol{\theta}_l$ to the direction of the virtual source $\boldsymbol{\theta}_\mathrm{s}$. Amplitude panning weights are a normalized version thereof, $g_l = \tilde{g}_l / \sqrt{\sum_l \tilde{g}_l^2}$. Weights should be positive $\tilde{g}_l \geq 0$ and activate only the fewest loudspeakers around the virtual source. Finding the weights can be formalized as ℓ_1-minimization under equality and non-negativity constraints,

$$\min \sum_l |\tilde{g}_l| \tag{2}$$
$$\text{subject to} \quad \sum_l \tilde{g}_l\, \boldsymbol{\theta}_l = \boldsymbol{\theta}_\mathrm{s}, \text{ and } \tilde{g}_l \geq 0, \quad \forall l.$$

Solutions are either obtained by numerical optimization, e.g. cvx [16, 17], or by constructing the convex hull spanned by the loudspeaker direction vectors. In the common second case, the hull is searched for the facet, line, or vertex, which yields an all-positive solution. The all-positive solution exists as long as the corresponding loudspeaker vertices enclose an angle $\leq 90°$. Figures 1a–c specify the vertices of either the 24 or 12 loudspeakers employed in the experiment below. The figures present the corresponding convex hulls in azimuthal equidistant projection (similar to a view from the z-axis, but with elevation steps mapped differently). The Table in Figure 1 provides the exact loudspeaker angles.

2.2. Ambisonic panning/decoding

Ambisonic panning [18, 19, 20] considers a continuous excitation of surrounding sources in terms of a finite-order expansion in spherical harmonic functions $Y_n^m(\boldsymbol{\theta})$. Spherical harmonics depend on the direction vector $\boldsymbol{\theta}$, and they have two integer indices $0 \leq n \leq \infty$ and $|m| \leq n$. An expansion into spherical harmonics of limited order, $n \leq \mathrm{N}$, can represent any directional function $g(\boldsymbol{\theta})$ whose directional resolution is uniformly limited. Such an N-th order function with optional weights a_n expands the continuous pattern representing a virtual source at $\boldsymbol{\theta}_\mathrm{s}$

$$g(\boldsymbol{\theta}) = \sum_{n=0}^{\mathrm{N}} \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta})\, a_n\, Y_n^m(\boldsymbol{\theta}_\mathrm{s}). \tag{3}$$

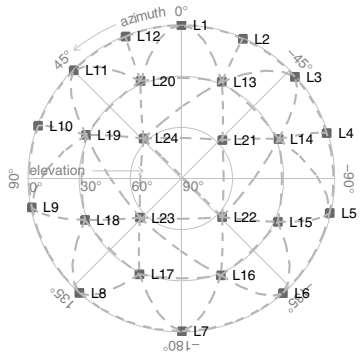
Typically, so-called max-$\boldsymbol{r}_\mathrm{E}$ weights [21] are used, which can be approximated as $a_n \approx P_n(137.9°/(\mathrm{N}+1.51))$, using the Legendre polynomials $P_n(\mu)$, cf. [14]. The continuous expansion is represented by the loudspeakers, which are, however, located at discrete directions. Customizing the discretization to the given facility, the final amplitude panning weights are obtained through a decoder [18] d_{nm}^l whose coefficients are only known after a suitable decoding rule is defined for the given playback facility

$$g_l = \sum_{n=0}^{\mathrm{N}} \sum_{m=-n}^{n} \underbrace{d_{nm}^l}_{\text{decoder}} \underbrace{a_n\, Y_n^m(\boldsymbol{\theta}_\mathrm{s})}_{\text{encoder}}. \tag{4}$$

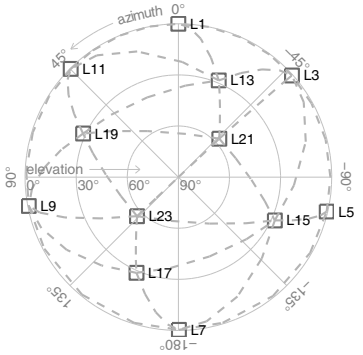
In many cases, decoding to irregular or incomplete spherical loudspeaker layouts is necessary, as addressed in [22, 23]. Ambisonic panning weights obtained by any of the above decoding rules are not strictly but largely positive, and they roughly fulfill both linear and square vector proportionalities $\sum_l g_l \boldsymbol{\theta}_l \propto \boldsymbol{\theta}_s$, $\sum_l g_l^2 \boldsymbol{\theta}_l \propto \boldsymbol{\theta}_s$ and a rough normalization constraint $\sum_l g_l^2 \approx$ const. The difficulty of decoding to the hemisphere and of fulfilling different constraints is handled differently by the decoding rules stated below. Most importantly: only a little is known about their perceptual aspects.

2.2.1. Energy-preserving decoding for the hemisphere (EP-Ambisonics)

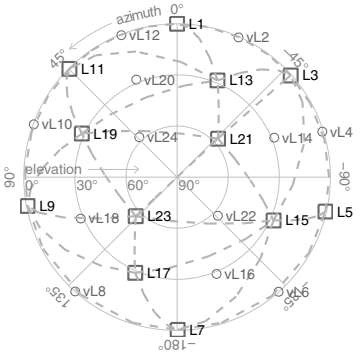
As one decoding rule, we employ energy-preserving Ambisonic decoding as described in [13]. The investigated configurations decode to all 24 lousdpeakers with $\mathrm{N} = 5$ and to the subset of 12 loudspeakers with $\mathrm{N} = 3$. This decoding rule exactly normalizes the sum of the squared loudspeaker weights $\sum_l g_l^2 =$ const. To achieve this for the hemisphere, the limited-order spherical harmonics are transformed to a smaller set of basis functions called *Slepian functions*, cf. [24]. These functions are obtained by numerical integration to form the matrix $\boldsymbol{G} =$



(a) 24 channel setup



(b) 12 channel subset



(c) 12 channel subset plus 12 virtual

	1	2	3	4	5	6	7	8	9	10	11	12
φ_l (°)	0	-23.7	-48.2	-72.6	-103.1	-138.5	-179.8	138.3	100.9	69.8	44.8	21.4
ϑ_l (°)	0.0	0.5	0.6	0.6	0.6	0.2	0.4	0.1	0.6	0.4	0.5	0.5

	13	14	15	16	17	18	19	20	21	22	23	24
φ_l (°)	-22.7	-67.9	-114.2	-157.8	156.4	113.3	65.4	22.7	-46.8	-133.4	133.4	43.7
ϑ_l (°)	28.5	28.5	27.9	28.7	28.7	28.4	28.5	28.0	57.0	57.0	56.6	57.7

(d) Azimuth angles φ_l and elevation angles ϑ_l of the 12+8+4 loudspeaker arrangement in the IEM CUBE

Figure 1. Azimuthal equidistant projection and exact polar coordinates of the loudspeaker setups employed for VBAP and Ambisonics. Dashed lines connecting loudspeakers indicate the convex hull used in the implementation of the Hybrid and the VBAP, AllRAD renderers.

$[\iint_{\vartheta\geq 0}^{\pi} Y_n^m(\boldsymbol{\theta}) Y_{n'}^{m'}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}]_{nm}^{n'm'}$, and by truncating its eigendecomposition $\boldsymbol{G} = \boldsymbol{Q}\Sigma\boldsymbol{Q}^{\mathrm{T}}$ to only those eigenvectors $\boldsymbol{Q}_>$ of sufficiently large eigenvalue. The energy-preserving decoding rule for the sampled Slepian functions uses their singular value decomposition $\boldsymbol{Q}_>^{\mathrm{T}}[Y_{nm}(\Theta_l)]_{nm}^{l} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}$ and defines a decoder $[d_{nm}^{l}]_{l}^{nm} = \boldsymbol{V}[\boldsymbol{I},\,\boldsymbol{0}]^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{Q}_>^{\mathrm{T}}$ that provides $\sum_l g_l^2 = 1$ on the upper hemisphere, cf. [13].

2.2.2. All-Round Ambisonic Decoding (AllRAD)

As described in [14], All-Round Ambisonic Decoding (AllRAD) is a hybrid alternative combining VBAP and Ambisonics (cf. [23]). The investigated configurations use an internal layer of either 70 or 180 virtual sources $\{\mathring{\Theta}_j\}$ in an optimal spherical t-design [25] arrangement, for direct sampling of equation (3) to obtain the virtual Ambisonic part of the decoder $\mathring{d}_{nm}^{j} = Y_n^m(\mathring{\Theta}_j)$. The j^{th} internal virtual source is represented on the given loudspeakers by a set of static VBAP gains $\hat{g}_l^j$ according to equation (2). As described in [14], the VBAP part for the upper hemispherical playback uses an extra internal virtual source direction at nadir $\Theta_{\mathrm{J}+1} = [0, 0, -1]^{\mathrm{T}}$ to enlarge the convex hull. This is done to preserve some of the signal amplitude on the lower hemisphere by mapping to the closest horizontal loudspeaker pairs. The AllRAD decoder consisting of the VBAP part and virtual Ambisonic part finally is $d_{nm}^{l} = \sum_{j=1}^{\mathrm{J}} \hat{g}_l^j \mathring{d}_{nm}^{j}$.

2.2.3. Hybrid VBAP-EP-Ambisonics

As a third alternative, the energy-preserving Ambisonic decoder $d_{\mathrm{EP}nm}^{l}$ of the order $\mathrm{N} = 5$ designed for 24 loudspeakers was used, but only to feed the 12 odd-numbered loudspeakers. For this purpose, odd-numbered loudspeakers were fed by the decoder directly, while the additional signals of even-numbered loudspeakers were represented as virtual VBAP sources. Accordingly, the decoder was defined as $d_{nm}^{2j+1} = d_{\mathrm{EP}nm}^{2j+1} + \sum_i \hat{g}_{2j+1}^{2i}\, d_{\mathrm{EP}nm}^{2i}$, see Figure 1c.

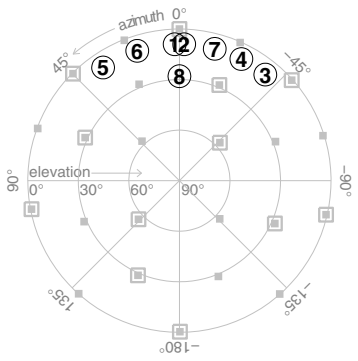
3. Experiment

3.1. Participants

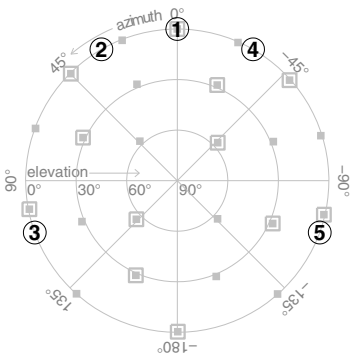
Thirty participants took part in the experiment (mean age 27 years, standard deviation 3.5 years, 11 females); all participants were members of a trained listening panel [26]. Participants were randomly allocated in three groups, each of which rated one of the three musical excerpts.

3.2. Apparatus and materials

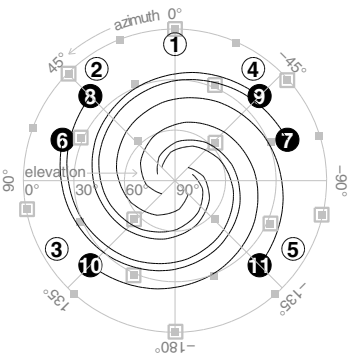
The experiment took place in IEM CUBE, where 24 loudspeakers are permanently installed. Twelve of them are located in the horizontal plane, eight at approximately 30° elevation and four at approximately 60° elevation, cf. Figure 1. The configurations listed in Table I were evaluated. Participants rated the configurations in pairs using the Graphical User Interface (Figure 3). The GUI contained five quasi-continuous scales, one for each attribute, that could obtain values between 0 and 128 with a step of 1, two buttons (A and B) that played back each of the two musical excerpts, a red arrow to proceed to the next test pair, the current trial number, and total number of trials. The GUI was displayed on a screen with a resolution of



(a) Radiohead



(b) Mahler



(c) Kühr

	Radiohead								Mahler				
Ch	1	2	3	4	5	6	7	8	1	2	3	4	5
$\varphi(°)$	2	-2	-39	-27	34	18	-15	0	0	30	110	-30	-110
$\vartheta(°)$	9	9	9	9	9	9	9	28	0	0	0	0	0

	Kühr										
Ch	1	2	3	4	5	6	7	8	9	10	11
$\varphi(°)$	0	35	120	-35	-120	70	-70	45	-45	135	-135
$\vartheta(°)$	9	9	9	9	9	19	19	19	19	19	19

(d) Azimuth φ and elevation angles ϑ of static/starting positions for the three pieces

Figure 2. Azimuthal equidistant projection of the sound sources in the three pieces used in the experiment. Black numbers in white disks correspond to the static virtual source positions, while white numbers in black disks correspond to the starting positions of the virtual source trajectories in the Kühr piece. For reference, gray circles indicate elevations of 0°, 30°, and 60°, in which the 24 (small filled squares) or 12 loudspeakers (large squares) are arranged.

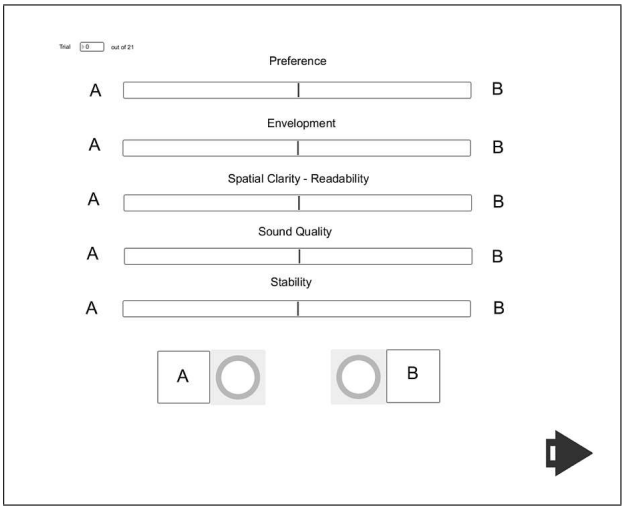


Figure 3. The Graphical User Interface of the experiment.

1280x1024 placed on a desk in front of the listener that was sitting in the sweet spot in the center of the room.

3.3. Stimuli

Stimuli were: 1. a 27 s long excerpt of the song Reckoner from the band Radiohead, 2. a 27 s long excerpt of Mahler's 3rd Symphony 5th Movement, and 3. a 24 s long excerpt of Gerd Kühr's, Revue Instrumentale et Électronique, a composition for an instrument ensemble and electroacoustic music (2004/05).

Table I. The renderers/configurations in the experiment.

Renderer	Ord.	Loudspk.	Abbreviation
VBAP	-	24	VBAP/24
VBAP	-	12	VBAP/12
EP-Ambisonics	5	24	EP/24/5
EP-Ambisonics	3	12	EP/12/3
AllRAD(180)	5	24	AllRAD/180
AllRAD(70)	5	24	AllRAD/70
Hybrid VBAP-EP	5	12	EP/12/5

Figure 2 presents the spatial mix used in each piece. Piece 1 (Radiohead) was an 8-channel close-mic multitrack recording, including only static sources that were rendered with an elevation of either 9° or 28°. Piece 2 (Mahler) was spaced-mic surround recording from the Salzburger Festspiele (Tonmeister Edwin Pfanzagl-Cardone) that was statically rendered on the horizontal plane. Piece 3 (Kühr) contained two parts: i. a 5.1 recording of the instrument ensemble that was statically rendered with

Table II. Range of obtained scales for the five attributes and three pieces used in the experiment. Kendall's coefficient of concordance for each attribute across the three pieces in the last column.

	Radiohead	Mahler	Kühr	W
Preference	0.54	0.51	0.25	0.25
Envelopment	0.49	0.44	0.33	0.29
Spatial Clarity	0.30	0.24	0.12	0.30
Sound Quality	0.43	0.32	0.13	0.30
Stability	0.21	0.20	0.17	0.13

an elevation of 9° and ii. 6 mono tracks in which mono sources moved along a spiral trajectory originating from 19° and terminating at 79° elevation after rotating 360 degrees in azimuth. The Mahler piece was the only piece that was rendered on the horizontal plane. Consequently, in the case of local-panning VBAP renderers the horizontal plane loudspeakers were exclusively activated. However, in the case of the global Ambisonic panning algorithms elevated loudspeakers were also active.

3.4. Procedure

Pieces were tested as a *between subjects* variable by each of the three participant groups, while renderers as *within subjects* in each group. Prior to starting the experiment, an explanation to all four attributes was given. Participants were instructed to move the slider towards A or B according to the extent to which A or B provided more of the particular attribute (Figure 3). The sliders for all attributes were presented in the same screen for each pair in the experiment.

Participants were asked to rate Preference in proportion to the degree each listening experience was favorable. Participants were asked to rate Spatial Clarity in proportion to the degree each listening experience provided clarity and precision with respect to localization. Participants were asked to rate Envelopment in proportion to the degree they were feeling immersed in the sound scene. Participants were asked to rate Sound Quality in proportion to the degree sounds were rendered without audible coloration or distortions. Participants were asked to rate Stability in proportion to the degree the scene was stable with regard to head movements and movement in the room.

Stability and Spatial Clarity were inspired from [10]. Stability was used in a similar sense as in [10], but was also extended in this study to not only include head movements but also movement within the listening area. Spatial Clarity related to the readability/localization attribute used by [10] and the Spatial Clarity attribute in [27]. The use of Envelopment is the typical one in the case of the Kühr piece, which contained sounds all around the listener, as well as the Mahler piece, which contained room information rendered by the rear loudspeakers. In the case of the Radiohead piece, in which close-mic frontal sounds were used, with this attribute we aimed to capture possible differences in immersion due to sound being emitted by peripheral loudspeakers in Ambisonics rendering. Within

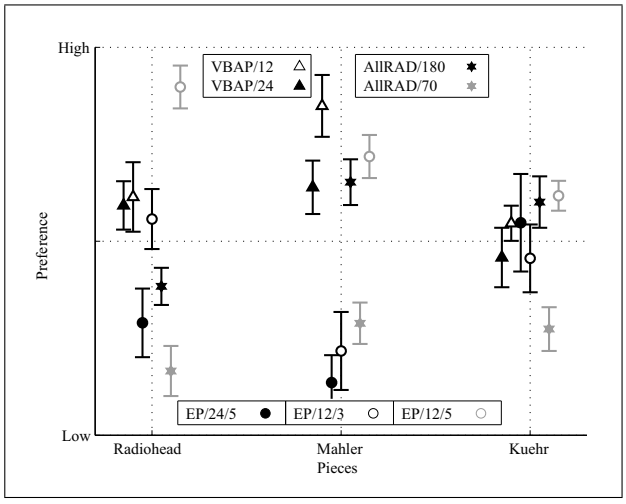


Figure 4. Preference scales and standard errors for the seven different renderers and the three different pieces.

each pairwise comparison trial, participants could listen to each renderer realization for as many times as they wanted. However, each pair was only rated once per participant to keep the experiment within a reasonable time frame; there were no trial repetitions. Typically a participant would listen to each excerpt in the sweet spot again and again in order to rate each of the attributes, and then walk around in the space to judge the stability of the percept. After providing ratings for all attributes, participants pressed a button to proceed to the next trial.

3.5. Results

Scales were constructed for each participant individually, based on the pairwise comparison data (Figures 4 and 5). The normalized (between 0 and 1) rating in each scale was used to represent the frequency with which each participant would choose the corresponding algorithm as this would emerge in a typical scaling experiment using A/B comparisons. This allowed the calculation of the scales according to the standard Thurstone Case V procedure [28]. Mean scales in Figures 4 and 5 are plotted unnormalized as their range depicts the magnitude of the perceived difference among the renderers; standard error of the means are plotted to provide an overview of the variability in the dataset. The average scale range for each attribute is provided in Table II. It is evident that the variation in the judgments of the participants was reduced in the third piece and in addition for certain attributes, most notably Stability and Spatial Clarity. On the other hand, the variation in the participants' judgments was highest for preference, envelopment, and sound quality judgments. The relative variability across the three pieces is also reflected in Kendall's coefficient of concordance, which remained low when considering the rank order of the renderers in the three different pieces (Table II).

For a more detailed analysis, a one-way repeated-measures Analysis of Variance (ANOVA) was performed independently for each piece and attribute in the experiment, using the renderer type as independent variable and

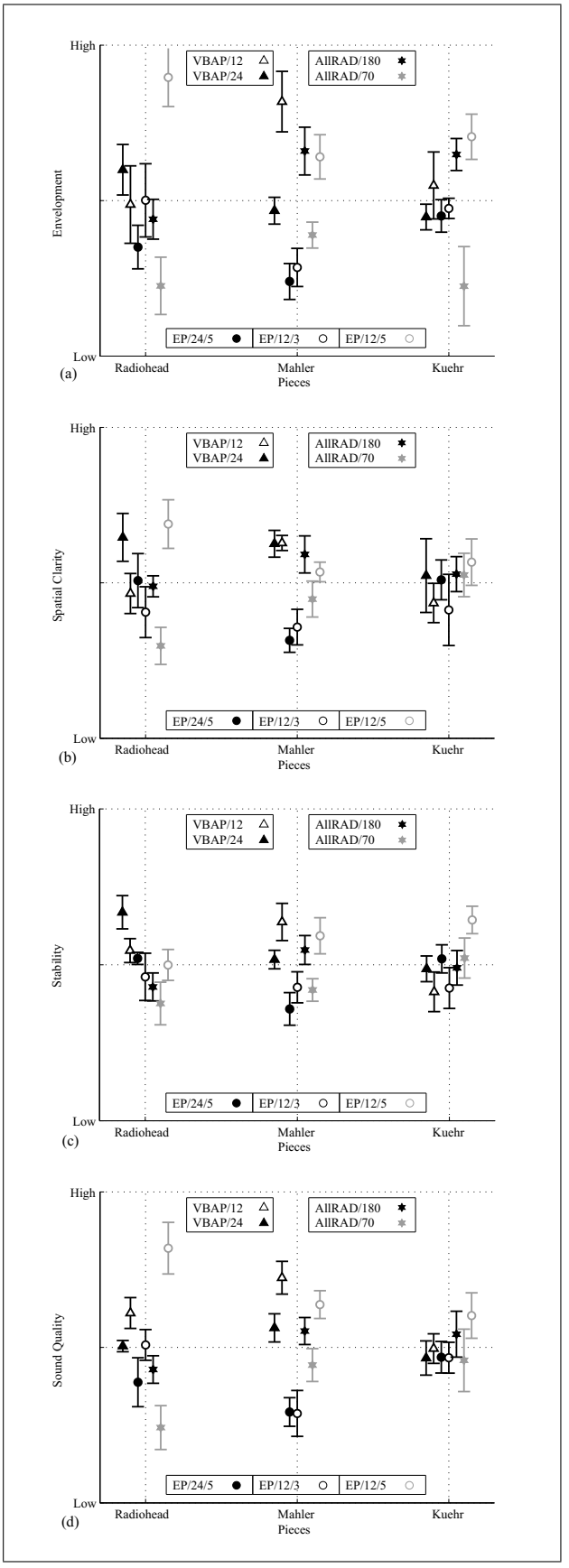


Figure 5. Scales and corresponding standard errors for the seven different renderers and the three pieces. (a) Envelopment, (b) Spatial Clarity, (c) Stability, (d) Sound Quality.

Table III. a: Summary of the statistical analysis for the three pieces in the experiment: One-way ANOVA for each attribute and piece.

Attribute	Piece	F-ratio, p-value
Preference	Radiohead	$F_{6,54} = 10.0,\ p = 0.001$
	Mahler	$F_{6,54} = 13.2,\ p = 0.001$
	Kühr	$F_{6,54} = 2.06,\ p = 0.073$
Envelopment	Radiohead	$F_{6,54} = 4.32,\ p = 0.001$
	Mahler	$F_{6,54} = 8.79,\ p < 0.001$
	Kühr	$F_{6,54} = 3.59,\ p = 0.005$
Spatial Clarity	Radiohead	$F_{6,54} = 3.09,\ p = 0.011$
	Mahler	$F_{6,54} = 6.56,\ p = 0.001$
	Kühr	$F_{6,54} = 0.37,\ p = 0.892$
Sound Quality	Radiohead	$F_{6,54} = 7.85,\ p = 0.001$
	Mahler	$F_{6,54} = 8.79,\ p < 0.001$
	Kühr	$F_{6,54} = 0.56,\ p = 0.763$
Stability	Radiohead	$F_{6,54} = 2.70,\ p = 0.023$
	Mahler	$F_{6,54} = 3.76,\ p = 0.003$
	Kühr	$F_{6,54} = 1.65,\ p = 0.153$

Table III. b: Linear regression coefficients of Preference regressed on i. the other four attributes ($C_{SC,SQ,E,ST}$) and ii. a two-dimensional representation of the space spanned by the same four attributes, obtained using PCA ($PC_{1,2}, C_{PC_{1,2}}$). (E) = Envelopment, (SC) = Spatial Clarity, (SQ) = Sound Quality, (ST) = Stability.

Radiohead, $R^2_{SC,SQ,E,ST} = 0.77$, $R^2_{PC_{1,2}} = 0.77$					
	SC	SQ	E	ST	
$C_{SC,SQ,E,ST}$	0.18	0.44	0.43	0.15	$C_{PC_{1,2}}$
PC_1	**0.72**	0.32	0.08	**0.61**	0.40
PC_2	0.00	0.34	**0.89**	-0.30	0.48
Mahler, $R^2_{SC,SQ,E,ST} = 0.80$, $R^2_{PC_{1,2}} = 0.80$					
	SC	SQ	E	ST	
$C_{SC,SQ,E,ST}$	0.55	0.55	0.28	0.38	$C_{PC_{1,2}}$
PC_1	**0.65**	**0.75**	-0.11	0.10	0.77
PC_2	0.00	0.09	**0.95**	0.31	0.43
Kühr, $R^2_{SC,SQ,E,ST} = 0.36$, $R^2_{PC_{1,2}} = 0.33$					
	SC	SQ	E	ST	
$C_{SC,SQ,E,ST}$	0.43	0.30	0.36	0.03	$C_{PC_{1,2}}$
PC_1	**0.94**	**0.34**	0.03	0.01	0.52
PC_2	0.00	-0.07	**0.95**	-0.31	0.31

the scale values for each participant and renderer as the dependent variable (Table III.a). Significant main effects emerged for all attributes for the Radiohead and Mahler pieces, but only for Envelopment in the case of the Kühr piece. Following a significant main effect of renderer, pairwise t-tests between the different renderer pairs were performed within each piece and attribute to spot which differences among the compared rendering techniques accounted for the globally significant differences observed

in ANOVAs. Significant differences (t-tests, $p < 0.05$) are reported next; no pairwise comparisons are reported for the Kühr piece due to the lack of significant main effects.

Preference:

Radiohead: EP/12/5 was the most preferred renderer, significantly more than all others. There was no difference among VBAP/24, VBAP/12, EP/12/3 and AllRAD/180 renderers, with the exception of AllRAD/180 which was significantly less preferred than VBAP/24. AllRAD/70 and EP/24/5 were the least preferred renderers, significantly less than all others. EP/24/5 was significantly less preferred than AllRAD/70 but not from AllRAD/180.

Mahler: VBAP/12, EP/12/5 and AllRAD/180 were the most preferred renderers, their difference not being significant. VBAP/24 follows, significantly less preferred than VBAP/12 and EP12/3 but not than AllRAD/180. AllRAD/70, EP/12/3 and EP/24/5 follow being significantly less preferred than all the aforementioned renderers but with no significant difference among them.

Envelopment:

Radiohead: EP/12/5 yielded significantly higher Envelopment compared to all other renderers except VBAP/24. VBAP/24 was at par with VBAP/12 and EP/12/3 and significantly better than EP/24/5, AllRAD/180, AllRAD/70. There was no significant difference among VBAP/12, EP/24/5, AllRAD/180 and AllRAD/70. EP/12/3 was not significantly different than VBAP/12 and AllRAD/180 provided significantly more Envelopment than AllRAD/70. Similar results were obtained in the case of the Kühr piece.

Mahler: VBAP/12 obtained the highest ranking, which was however not significantly different than neither AllRAD/180 nor EP12/5. Following, EP/12/5 and AllRAD/180 were in addition not different to VBAP/24 and better than EP/24/5, EP12/3 and AllRAD/70. On the lower tail, EP/12/3 was not significantly different than AllRAD/70, EP/24/5 and VBAP/24, while AllRAD/70, EP/24/5 provided less envelopment than VBAP/24 but not than EP/12/3.

Spatial Clarity:

Radiohead: VBAP/24 and EP/12/5 renderers yielded significantly more Spatial Clarity compared to both AllRAD/70 and EP/12/3, EP/12/5 performing better than VBAP/12. AllRAD/70 ranked significantly lower than EP/24/5, AllRAD/180.

Mahler: The two VBAP configurations yielded highest Spatial Clarity, not significantly different than AllRAD/180, but significantly higher than AllRAD/70, EP12/3 and EP/24/5 (not VBAP/12) and EP/12/5. AllRAD/180 comes third, yielding higher clarity than EP/24/5, EP/12/3, and marginally AllRAD/70. EP/12/5 comes fourth, better than EP/24/5 and EP/12/3, but not significantly different than neither AllRAD/180 nor AllRAD/70.

Sound Quality:

Radiohead: EP/12/5 yields significantly better Sound Quality than all other renderers. VBAP/24, VBAP/12, AllRAD/180 and EP/12/3 follow with no significant differences among them, with the exception that Sound Quality for EP/12/3 was worse than VBAP/12. At the lowest end of the scale, AllRAD/70 and EP/24/5 not were not significantly different but worse than all the rest of the renderers.

Mahler: VBAP/12 leads the scale yielding significantly better Sound Quality than all renderers save VBAP/24. VBAP/24, AllRAD/180 and EP12/5 follow, not significantly different to each other. EP/12/3, AllRAD/70 and EP 24/5 follow not significantly among each other but significantly worse than the rest.

Stability:

Radiohead: VBAP/24, VBAP/12 and EP/24/5 were the most stable renderers, with no significant difference among them. VBAP/24 was significantly more stable than all the remaining renderers. VBAP/12, EP/24/5 were significantly more stable than the AllRad renderers. EP12/3, AllRAD/70, AllRAD/180 were not significantly different to each other.

Mahler: Few differences were significant, VBAP/12, VBAP/24, AllRAD/180 and EP/12/5 were significantly more stable than AllRAD/180 and EP/24/5, and VBAP/12 was more stable than AllRAD/70.

3.6. Preference prediction and dimensionality reduction

Multiple linear regression was applied to the dataset of each piece, with Preference as the dependent and Envelopment, Spatial Clarity, Sound Quality, and Stability as the predictor variables (Table III.b). The obtained regression coefficients indicate a high weighting of Envelopment and Sound Quality for the Radiohead piece, Spatial Clarity and Sound Quality for the Mahler piece and a more uniform weighting of the attributes for the Kühr piece. Results are similar when applying stepwise regression. The order with which terms come into the model is: Envelopment ($R^2 = 0.65$), Sound Quality ($R^2 = 0.74$), and Spatial Clarity($R^2 = 0.77$) for the Radiohead piece; Sound Quality ($R^2 = 0.68$), Envelopment ($R^2 = 0.75$), Spatial Clarity ($R^2 = 0.78$) and Stability ($R^2 = 0.80$) for Mahler; and Spatial Clarity ($R^2 = 0.16$), Envelopment ($R^2 = 0.29$), Sound Quality ($R^2 = 0.36$) for the Kühr piece. For the Radiohead and Kühr piece, adding Stability as a predictor variable does not improve the model significantly.

It is relatively common to observe correlations between the perceptual attributes [10, 11, 29], which can obscure the interpretation of the linear regression results. In our case, the correlation coefficient among the aggregated individual scales used for the prediction was low, below 0.65 for all pieces, with the exception of the correlation coefficient between Sound Quality and Spatial Clarity in the case of the Mahler piece, which was 0.8. Although sliders for all attributes appeared simultaneously, this did not seem to increase intra-attribute correlation significantly, high correlations were also observed in [11], where only two attributes were presented at a time, as well as in [10, 29], where only one attribute was presented at a time.

To reduce the impact of the collinearity problem, it has been suggested [11] to construct the Preference model us-

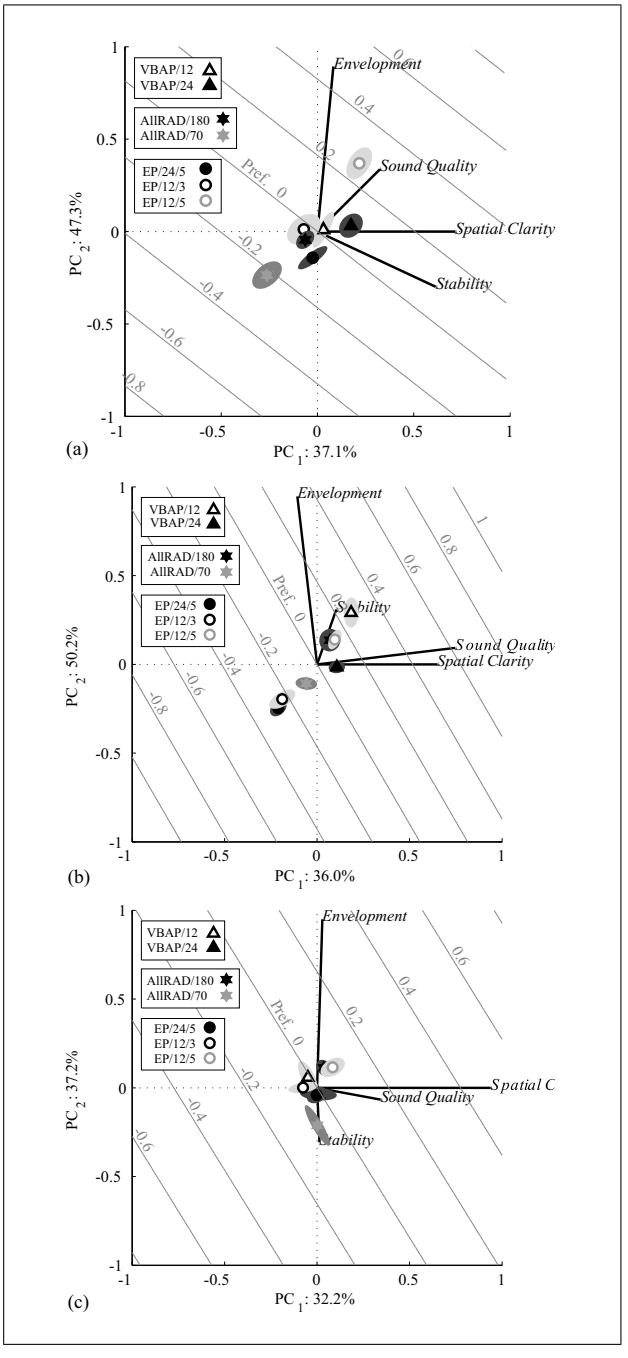


Figure 6. Results of a principle component analysis for each piece; rotated so that Spatial Clarity is aligned with the horizontal axis (PC_1). Whenever necessary, the resulting PC_2 was mirrored so that Envelopment points upwards. Depicted Preference contours were calculated using the regression coefficients in Table III.b. Percentages along the axes indicate the amount of variance that can be explained by the first two components, ellipses indicate standard errors. (a) Radiohead 84.4%, (b) Mahler 86.2%, (c) Kühr 69.4%.

ing a projection of the dataset upon an orthonormal subspace of the attribute space, obtained through Principal Component Analysis (PCA). When performing PCA on Envelopment, Spatial Clarity, Sound Quality, and Stability, it was found that two components explain about 85% of the variance in the case of the Radiohead and Mahler pieces, while three components would be necessary to explain 89% of the variance in the case of the Kühr piece. The subsequent analysis is performed on the subspace spanned by the first two components. To facilitate interpretation, this subspace was rotated so that Spatial Clarity aligns with PC_1. Whenever necessary, the resulting second subspace component was mirrored so that a positive mapping of Envelopment onto PC_2 was obtained, cf. Figure 6. In the figure, by projecting each attribute vector onto the two axes their weighting in each principal component is obtained, while the cosine of the angle between each two of them indicates the extent to which they are correlated. The ellipse centre indicates each average renderer score and the ellipse itself the standard error of the average. By projecting the symbols orthogonally onto an attribute vector, the ranking of each renderer with regard to each attribute can be observed.

Multiple regression of Preference on the two most significant principal components in the different conditions of the experiment led to very similar results as in the case of using four predictor variables (Table IIIb). The contours in Figure 6 show the predicted Preference based on the aforementioned linear regression model. It can be further verified that Sound Quality is an important Preference predictor (cf. Table III.b) and the attribute that aligns best with Preference in the case of Mahler and Radiohead pieces (cf. Figure 6). Envelopment and Spatial Clarity are also important but their contribution to the models as single attributes and their degree of alignment to Preference depends on the musical material used. It is however important to note that they are in all cases nearly orthogonal and the vector emerging from their combination is pointing consistently in the Preference direction.

4. Discussion

While the use of non-musical test stimuli in spatial audio renderer evaluation has its merit, the use of musical material is also important to provide for an ecologically valid task. We have performed a renderer evaluation using musical material in order to obtain insight on the quality of the listening experience with respect to the examined renderers and also to investigate the process of spatial audio evaluation using musical material. We found that such an approach is promising as long as the musical material is familiar and its spatial complexity is low. This is evidenced by the fact that in the case of the Radiohead and Mahler pieces our investigation yielded significant findings and a clear influence of the musical material itself on the listeners' judgments. In such cases, listeners can therefore provide judgments consistent enough to enable selection, discrimination, and characterization of renderers.

Although variation of the listeners' judgments across musical material is to be expected [10, 11], the one observed between the electro-acoustic piece (Kühr) and the other two is pronounced. Apparently the spatial complexity and the lack of familiarity with the sound material obscured the results of the evaluation by blurring the differ-

ences inherent in spatial audio renderers. Furthermore, the relative inability to obtain consistent ratings in the case of the Kühr piece might also relate to the fact that the attributes used here were derived from the literature on two-dimensional reproduction of static musical sound material which did not include contemporary electroacoustic music. Consequently, they might not be appropriate for the purpose of evaluating more complex three-dimensional musical scenes involving auditory movement and contemporary electroacoustic music material. This hypothesis needs however to be investigated further and cannot be answered based on the results of this study.

Concerning the Radiohead and Mahler excerpts, higher-order Ambisonics and VBAP renderers using twelve loudspeakers were preferred to the 24 loudspeaker ones, arguably because Sound Quality was a critical determinant of the Preference judgments. By inspecting Figure 5(d), it is clear that listeners judged the sparser loudspeaker configurations to be superior in terms of Sound Quality. Such a result is not surprising considering that fewer loudspeakers yield a cleaner response that might help preserve transients and reproduce the high frequency content better. An interesting observation emerges when considering its prevalence in Preference of VBAP/12 over VBAP/24 in the Mahler piece. This is mainly due to significantly more Envelopment and a marginally better Sound Quality ($p = 0.06$). The importance of Envelopment can be attributed to the fact that in the case of VBAP/12 the energy from the dominating frontal signal channels 2 and 4 was distributed to wider loudspeaker angles ($\sim 90°$ between loudspeaker 11 and 3) in comparison to VBAP/24 ($\sim 40°$ between loudspeaker 12 and 2), which, at least for monophonic sounds, is known to result in a broadening of the sound image that could account for the increased envelopment ratings [5]. In addition, in VBAP/12 a smaller total number of loudspeakers carries loud-enough signals. Thus, Sound Quality slightly improves.

The most preferred renderer in the case of the Mahler piece was VBAP/12, while in the case of the Radiohead piece it was EP/12/5. Interestingly, in the first case, the most critical Preference predictors are Sound Quality and Spatial Clarity, and in the second Sound Quality and Envelopment. One may therefore deduce that if Spatial Envelopment is already encoded in the recording, as in the case of the Mahler excerpt, Spatial Clarity is what becomes important. Additional Envelopment gained through the use of global panning algorithms is less crucial. By contrast, Spatial Clarity is less important in the reproduction of a spatially clear, close-mic recording (Radiohead excerpt), in which case Envelopment makes rendering successful. This explains the sucess of global panning algorithms, which tend to increase Envelopment by disturbing the signal over a larger number of loudspeakers. Accordingly in Figure 6. Envelopment is almost orthogonal to Preference in the case of the Mahler piece, but not in the case of the Radiohead piece.

Another contributing factor to the prevalence of VBAP/12 in the case of the Mahler piece could be that EP decoders are not performing well for non-coincident microphone recordings. This hypothesis is also supported by the drop in the ranking of the EP decoders in the Mahler piece in comparison to the Radiohead piece, for all attributes. EP decoders tend to broaden and mix signals panned close to the horizon. This property could account for the aforementioned deterioration.

Interestingly, although Sound Quality is typically the predominant factor for Preference, see e.g. [29], it appears that the importance of the attributes can vary depending on the musical material. For example, the importance of Spatial Clarity was low in the case of the Radiohead piece, significant in the case of the Mahler piece, and predominant in the case of the Kühr piece. In the second case, this is arguably due to the difficulties in reproducing a surround recording with Spatial Clarity, while in the third it may relate to the difficulty in reproduction of sound movement. A similar explanation can be made for Envelopment based on the arguments and observations in the previous paragraphs. Further insight in the relationship between Preference and the remaining attributes is provided by the results of the principal component analysis, presented in Figure 6, according to which Spatial Clarity remains orthogonal to Envelopment for all three pieces. As a general rule for succesful rendering one would need to provide both to maximize Preference. This may prove difficult, as based on our results it appears that global and local panning algorithms trade off Spatial Clarity for Envelopment, at least for low-order global panning algorithms.

Stability had a weak influence on the Preference judgements and Stability scales yielded few significant differences. Notably, the VBAP/24 algorithm performs best in the case of the Radiohead piece, in the same way as VBAP/12 and EP/24/5. Concerning the VBAP algorithms this can be attributed to the smaller number of active loudspeakers, which results in less distortion for listening positions away from the sweet spot. In addition, VBAP/24 has a smaller inter-loudspeaker distance compared to VBAP/12, which further supports the Stability of this renderer. The ranking of EP/24/5 is also much better in terms of Stability in comparison to its ranking for other attributes, implying that a higher order and a larger number of loudspeakers helps stabilize the image in global panning algorithms. The results are more difficult to explain in the case of the Mahler and Kühr pieces. This can be attributed on the non-coincident recording technique used for the Mahler recording and in addition on the spatial complexity of the Kühr piece and cannot be explained based on the results of this study. Finally, it is worth mentioning that the AllRAD/70 configuration was consistently judged to be worse than the AllRAD/180 renderer, an assumption that has been made in [14] further verifying the validity of the evaluation method.

5. Conclusions

Overall, we observed that musical material and the recording technique influence Preference, Envelopment, Spatial

Clarity, Sound Quality, and Stability ratings for different rendering techniques. It was found that Preference could be explained on the basis of the other four attributes. The way in which each of these attributes contributes depends on the musical material and the recording technique used. There is a consistently strong correlation between Preference and Sound Quality, while the importance of Spatial Clarity and Envelopment varies according to the material used. The space spanned by our attributes could be reduced to a two-dimensional subspace using PCA, which explained nearly 90% of the judgments' variance for two of our pieces and more than two-thirds for the third piece.

Of the rendering techniques explored, EP/12/5 and VBAP were found to be consistently yielding high Preference ratings. The first prevailed for the reproduction of close-mic sound or individual tracks and the second when considering spaced-mic surround recorded material. EP/24/5, the AllRAD/70 and EP/12/3 were overall at the lower ends of the scales in terms of Preference, while AllRAD/180 was consistently in the middle of the Preference scales.

Acknowledgements

We acknowledge the help of Thomas Musil in setting up the Kühr piece and the experiment in the IEM CUBE. This study was supported by the AAP project, funded by Austrian ministries BMVIT, BMWFJ, the Styrian Business Promotion Agency (SFG), and the departments 3 and 14 of the Styrian Government. The Austrian Research Promotion Agency (FFG) conducts the funding under the Competence Centers for Excellent Technologies (COMET, K-Project), a program of the above-mentioned institutions. Support was also provided by project Klangräume: Situated Usability Evaluation of Interactive 3D Audio Environments funded by Zukunftsfonds Steiermark, Land Steiermark, Austria.

References

[1] E. Benjamin, A. Heller, R. Lee: Localization in Horizontal-Only Ambisonic Systems. Audio Engineering Society Convention 121, San Francisco, CA, USA, October 2006.

[2] P. Stitt, S. Bertet, M. Van Walstijn: Perceptual investigation of image placement with ambisonics for non-centred listeners. Proc. of the 16th Int. Conference on Digital Audio Effects (DAFx-13), Maynooth, Ireland, September 2013.

[3] G. Kearney, E. Bates, F. Boland, D. Furlong: A comparative study of the performance of spatialization techniques for a distributed audience in a concert hall environment. Audio Engineering Society Conference: 31st International Conference: New Directions in High Resolution Audio, London, UK, June 2007.

[4] V. Pulkki: Localization of amplitude-panned virtual sources II: Two- and three-dimensional panning. J. Audio Eng. Soc **49** (2001) 753–767.

[5] M. Frank: Phantom sources using multiple loudspeakers in the horizontal plane. Dissertation. University of Music and Performing Arts Graz, Austria, 2013.

[6] V. Pulkki: Uniform spreading of amplitude panned virtual sources. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 1999, 187–190.

[7] K. Ono, V. Pulkki, M. Karjalainen: Binaural modeling of multiple sound source perception: Coloration of wideband sound. Audio Engineering Society Convention 112, Munich, Germany, April 2002.

[8] F. Rumsey: Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm. J. Audio Eng. Soc **50** (2002) 651–666.

[9] H.-K. Lee, F. Rumsey: Elicitation and grading of subjective attributes of 2-channel phantom images. Audio Engineering Society Convention 116, Berlin, Germany, May 2004.

[10] C. Guastavino, B. F. G. Katz: Perceptual evaluation of multi-dimensional spatial audio reproduction. Journal of the Acoustical Society of America **116** (2004) 1105–1115.

[11] S. Choisel, F. Wickelmaier: Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference. Journal of the Acoustical Society of America **121** (2007) 388–400.

[12] F. Rumsey, J. Berg: Verification and correlation of attributes used for decribing the spatial quality of reproduced sound. 19th AES International Conference: Surround Sound - Techniques, Technology and Percpeption, 2001.

[13] F. Zotter, H. Pomberger, M. Noisternig: Energy-Preserving Ambisonic Decoding. Acta Acustica united with Acustica **98** (2012) 37–47.

[14] F. Zotter, M. Frank: All-round ambisonic panning and decoding. J. Audio Eng. Soc **60** (2012) 807–820.

[15] V. Pulkki: Virtual sound source positioning using vector base amplitude panning. J. Audio Eng. Soc **45** (1997) 456–466.

[16] M. Grant, S. Boyd: CVX: Matlab software for disciplined convex programming, version 2.0 beta. `http://cvxr.com/cvx`, September 2013.

[17] C. B. Barber, D. P. Dobkin, H. T. Huhdanpaa: The quickhull algorithm for convex hulls. ACM Transactions on Mathematical Software **22** (1996) 469–483.

[18] D. H. Cooper, T. Shiga: Discrete-matrix multichannel stereo. J. Audio Eng. Soc. **20** (1972) 346–360.

[19] J. Daniel: Représentation de champs acoustiques, application à la transmission et à la reproduction de scénes sonores complexes dans un contexte multimédia. Dissertation. Université de Paris 6, France, 2001.

[20] Y. Wu, T. Abhayapala: Theory and design of soundfield reproduction using continuous loudspeaker concept. IEEE Transactions on Audio, Speech, and Language Processing **17** (2009) 107–116.

[21] J. Daniel, J.-B. Rault, J.-D. Polack: Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions. Audio Engineering Society Convention 105, San Francisco, CA, USA, September 1998.

[22] M. Poletti: Robust two-dimensional surround sound reproduction for nonuniform loudspeaker layouts. J. Audio Eng. Soc. **55** (2007) 568–610.

[23] J.-M. Batke, F. Keiler: Using VBAP-derived panning functions for 3D Ambisonics decoding. 2nd Ambisonics Symposium, Paris, France, May 2010.

[24] R. Pail, G. Plank, W. Schuh: Spatially restricted data distributions on the sphere: the method of orthonormalized functions and applications. Journal of Geodesy **75** (2001) 44–56.

[25] R. H. Hardin, N. J. A. Sloane: t-designs. http://www2.research.att.com/~njas/sphdesigns/dim3/, June 2012.

[26] M. Frank, A. Sontacchi: Performance review of an expert listening panel. Fortschritte der Akustik, DAGA, Darmstadt, Germany, March 2012.

[27] G. Paine, R. Sazdov, K. Stevens: Perceptual investigation into envelopement, spatial clarity, and engulfment in reproduced multi-channel audio. Audio Engineering Society

Conference: 31st International Conference: New Directions in High Resolution Audio, Jun 2007.

[28] L. Thurstone: A law of comparative judgment. Psychological Review **34** (1927) 273–286.

[29] F. Rumsey, S. Zieliński, R. Kassier, S. Bech: On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. The Journal of the Acoustical Society of America **118** (2005) 968–976.