Audio Engineering Society

Convention Paper

Presented at the 120th Convention
2006 May 20–23 Paris, France

Relating auditory attributes of multichannel sound to preference and to physical parameters

Sylvain Choisel[1,2], and Florian Wickelmaier[1]

[1] *Sound Quality Research Unit, Department of Acoustics, Aalborg University, 9220 Aalborg, Denmark*

[2] *Bang & Olufsen A/S, 7600 Struer, Denmark*

Correspondence should be addressed to Sylvain Choisel (`sc@acoustics.aau.dk`)

ABSTRACT

Sound reproduced by multichannel systems is affected by many factors giving rise to various sensations, or auditory attributes. Relating specific attributes to overall preference and to physical measures of the sound field provides valuable information for a better understanding of the parameters playing a role in sound quality evaluation. Eight selected attributes are quantified by a panel of 39 listeners using paired-comparison judgments and probabilistic choice models, and related to overall preference. A multiple-regression model predicts preference well, and some similarities are observed within and between musical program materials, allowing for a careful generalization regarding the perception of spatial audio reproduction. Finally, a set of objective measures is derived from analysis of the sound field at the listening position in an attempt to predict the auditory attributes.

1. INTRODUCTION

Sound quality research endeavors to gain insight into the mechanisms which underly listener preference. Complex stimuli are typically involved in sound quality assessments, giving rise to various sensations, or *auditory attributes*, which potentially contribute to perceived overall quality. The investigation of the relationship between (specific) auditory attributes and global preference requires quantifying (or scaling) these perceived magnitudes, which is usually time consuming and costly. Therefore, objective parameters of the stimuli are sought after, which predict the auditory attributes and are easy to obtain and quick to measure. While objective parameters have been extensively studied in the field of concert-hall acoustics (see [1] for a review), and have been standardized to a large extent [2], the development of objective measures of spatial audio quality is still at an early stage.

Apart from pioneering studies on multichannel recording and playback [3], most work on quality of reproduced sound has focused on timbral aspects of monophonic reproduction (e. g., [4]). As multichannel audio formats are growing in popularity, the question arises how the various reproduction modes influence the listener's perception. Of particular interest is how spatial auditory sensations are affected by the introduction of center and surround loudspeakers in a multichannel setup, or by various processing algorithms. More recent studies have addressed the problem of identifying and quantifying auditory attributes which are relevant to sound quality in the context of multichannel reproduced sound [5, 6, 7, 8].

The present study aimed at investigating more specifically the perceptual differences between reproduction modes typically encountered in home audio systems: Selected musical excerpts—originally produced for five-channel reproduction—were reproduced in various formats (mono, stereo and several multichannel formats). In a recent study [9], Zieliński *et al.* have focused on the overall perceptual evaluation (the so-called *basic audio quality*) of reproduction modes similar to the ones used in the present work. Rumsey *et al.* [10] investigated the influence of timbral, frontal and surround fidelity changes on basic audio quality. The present study, however, intended to seek explanations for such global differences in terms of more specific auditory attributes and in terms of objective parameters.

Several studies have attempted to apply objective parameters from concert-hall acoustics to the field of reproduced sound [11, 12, 13, 14]. These rely on the characterization of the recording or reproduction chain by a set of impulse responses, which makes it possible to account for variations in the recording environment, the recording technique, the loudspeaker/room interaction during playback, or some artificial reverberation algorithms. Unfortunately, the complexity of the stimuli encountered in popular multichannel music, and of the processes involved in their production (such as down- or upmixing), makes it difficult to characterize them by impulse responses. In this study, analysis was performed on the musical signals recorded at the listening position, and the resulting parameters were related to auditory attributes obtained from listening tests.

In summary, the goals of this study were to (1) verify that listeners can consistently judge upon auditory attributes which are relevant in the context of multichannel music reproduction, (2) quantify these attributes on meaningful scales, (3) determine their relation to overall preference, and (4) relate them to objective parameters of the sound field at the listening position.

2. METHOD

2.1. Subjects

Thirty-nine listeners (27 males, 12 females) took part in this study. They were all native Danish speakers, selected among 78 candidates according to their auditory and verbal abilities. The selection procedure (detailed in [15]) consisted of pure-tone audiometry, a stereo-width discrimination task and a verbal fluency test. This was done in order to ensure that the listeners selected could (1) appreciate spatial differences in sound and (2) readily produce a description of their sensations.

2.2. Setup and stimuli

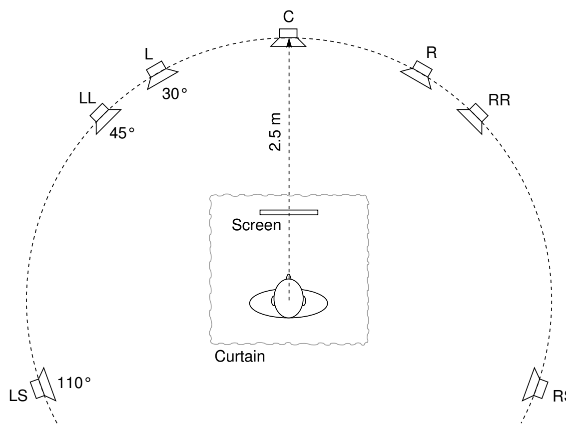

Figure 1: Playback setup consisting of seven loudspeakers: left (L), right (R), center (C), left-of-left (LL), right-of-right (RR), left surround (LS) and right surround (RS). This setup was symmetrically placed with respect to the width of the room and was hidden from the subject by an acoustically transparent curtain. A computer flat screen was used as a response interface.

The listening tests took place in a 60-m^2 sound-insulated listening room complying with the ITU-BS1116 requirements [16]. Seven loudspeakers (Genelec 1031A) were placed as shown in Figure 1, at a distance of 2.5 m to the listening position. Five of them were arranged in accordance with the ITU-R recommendation BS.775-1 [17]; two additional speakers were placed at ±45° for the reproduction of stereo over a wider base angle (defined as the bearing angle between the loudspeaker pair, as seen from the listening position). The setup was hidden from the subject by an acoustically transparent curtain.

Four musical excerpts (two pop, two classical) of about 5 s duration were selected from commercially available multichannel material (Table 1), and played back in mono, stereo, and various multichannel formats summarized in Table 2. All eight reproduction modes were derived from the original 5-channel format by downmixing (to stereo, mono and phantom mono) and upmixing from stereo (to matrix, Dolby Pro Logic II and DTS Neo:6). More details on the stimulus generation can be found in [18].

Table 2: Reproduction modes: full name, abbreviation and loudspeakers used for playback (see Figure 1).

Name	Abbr.	Speakers
mono	*mo*	C
phantom mono	*ph*	L,R
stereo	*st*	L,R
wide stereo	*ws*	LL,RR
matrix upmixing	*ma*	L,R,LS,RS
Dolby Pro Logic II	–*	L,R,C,LS,RS
DTS Neo:6	–*	L,R,C,LS,RS
original 5.0	*or*	L,R,C,LS,RS

*referred to as *u1* and *u2* (in no specific order) in the rest of this paper.

2.3. Scaling auditory attributes and preference

An attempt was made to quantify eight selected auditory attributes (see [18] for details on the attribute elicitation) as well as overall preference using a paired-comparison procedure. For each pair of reproduction modes, the subjects were asked (in Danish) "Which of the two sounds is more…" followed by one of the following adjectives: *wide (bred)*, *elevated (høj oppe)*, *spacious (rummelig)*, *enveloping (omsluttende)*, *far ahead (langt foran)*, *bright (lys)*, *clear (tydelig)* and *natural (naturlig)*. Definitions of these attributes were generated by the experimenters so as to represent as much as possible the subjects' descriptors elicited.

Two buttons on a computer screen, labeled A and B, were visually emphasized in turn (by changing their size) during playback to indicate which sound was playing. The response was made by clicking the button corresponding to the chosen sound. Each pair was judged only once. The within-pair order was balanced across subjects and the between-pair order was random. Each attribute was evaluated for all four program materials in a single block lasting for about 25 minutes. Each subject evaluated two attributes in a session lasting for one hour, including a break in the middle. Thus, four sessions were required for all eight attributes. The order of the attributes and program materials was balanced across subjects using five different 8×8 Graeco-Latin squares. Each subject gave 28 judgments per program material and auditory attribute.

Preference was quantified in a similar manner. For each pair of reproduction modes the subjects were instructed to indicate which one they preferred.

Table 1: List of musical program material.

Disc	Title	Medium	Track	Time
Beethoven: Piano Sonatas Nos. 21, 23 & 26 – Kodama	Sonata 21, op. 53 (Rondo)	SACD	03	1'51 – 1'56
Rachmaninov: Vespers – St. Petersburg Chamber Choir conducted by Korniev	Blazen Muzh	SACD	03	2'04 – 2'09
Steely Dan: Everything Must Go	Everything Must Go	DVD-A	09	0'52 – 0'57
Sting: Sacred Love	Stolen Car	SACD	06	1'55 – 2'00

Each pair was presented in both within-pair orders (AB and BA). Thus, each subject gave 56 preference judgments per program material in a fifth session of approximately one hour duration.

The paired-comparison data were aggregated across subjects in order to estimate pairwise probabilities, P_{xy}, of choosing sound x over sound y. These probabilities were analyzed using so-called *probabilistic choice models*. The first model used was the Bradley-Terry-Luce (BTL) model [19, 20], which predicts P_{xy} as a function of parameters associated with each sound

$$P_{xy} = \frac{u(x)}{u(x) + u(y)}, \tag{1}$$

where $u(\cdot)$ is a ratio scale of the attribute under study. The BTL model imposes strong restrictions on the choice frequencies and, therefore, requires highly consistent judgments.

The second, less restrictive, model was the so-called *elimination-by-aspects* (EBA) model [21, 22], which is a generalization of the BTL model. According to EBA, one sound is chosen over a second one because of a certain *aspect* which belongs to the first but not to the second sound. EBA predicts P_{xy} by

$$P_{xy} = \frac{\sum\limits_{\alpha \in x' \setminus y'} u(\alpha)}{\sum\limits_{\alpha \in x' \setminus y'} u(\alpha) + \sum\limits_{\beta \in y' \setminus x'} u(\beta)}, \tag{2}$$

where $\alpha, \beta, \ldots$ are the aspects (or features) of the sounds, and $x' \setminus y'$ denotes the set of aspects belonging to sound x but not to sound y. As for the BTL model, $u(\cdot)$ is a ratio scale of the attribute. EBA can to some extent cope with attributes which are composed of multiple aspects. Parameter estimation and model testing was performed using special-purpose software [23].

2.4. Calculation of objective parameters

2.4.1. Recordings at the listening position

In search of physical parameters reflecting relevant properties of the sound field, the stimuli (as used in the listening tests) were recorded at the listening position, using several recording techniques. First, binaural recordings were made, using an artificial head (Brüel & Kjær 4100). Second, recordings were made using an AKG C34 stereo microphone, with one capsule having an omnidirectional pattern, and the other one a bipolar pattern (Figure 2) with the null pointing towards the center loudspeaker.

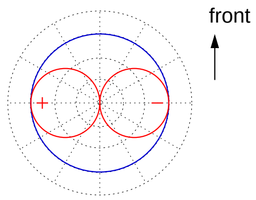

Figure 2: Polar response of a the stereo microphone composed of coincident omnidirectional and bidirectional (or *figure-of-eight*) microphones.

2.4.2. Interaural cross-correlation coefficient (IACC)

The IACC is frequently used in concert-hall acoustics, as a measure of spaciousness and apparent source width (ASW, cf. [24]). It is calculated as the maximum of the interaural cross-correlation function Φ_{lr}:

$$\Phi_{lr}(\tau) = \frac{\int_{t_1}^{t_2} p_l(t)\, p_r(t+\tau)\, dt}{\sqrt{\int_{t_1}^{t_2} p_l^2(t)\, dt \int_{t_1}^{t_2} p_r^2(t)\, dt}}, \tag{3}$$

where $p_l(t)$ and $p_r(t)$ are either time signals recorded at the ears of a dummy head, or, in room acoustics, binaural room impulse responses. The latter offer the possibility of separating the early and late reflections, which have been found to contribute to different qualities of concert halls (see, e. g., [25] for a review). In the case of multichannel audio, however, impulse responses (e. g., from each loudspeaker) do not represent adequately the contribution of each channel, which is highly dependent on the program material and the way it has been mixed. Therefore, binaural recordings of the stimuli will be used, which do not directly allow for a separation of early and late energy.

Typically, the maximum cross-correlation is calculated on the absolute value of $\Phi_{lr}(\tau)$, for $-1\,\mathrm{ms} <$

$\tau < +1\,\text{ms}$. In order to account for possible negative correlations[1], the extremum of the signed cross-correlation function is used instead. Thus, in this paper, IACC is calculated as

$$IACC = \operatorname*{extr}_{-1\,\text{ms}<\tau<+1\,\text{ms}} \Phi_{lr}(\tau),$$

where extr is either min or max, whichever has a greater absolute value.

The calculation of IACC in binaural models often includes a half-wave rectification followed by a low-pass filter, to simulate the envelope extraction taking place in the auditory system at high frequencies (see, e. g., [26]). Such a simple model was shown to be more adequate than the classical IACC in predicting auditory source width of narrow-band sounds [27]. This modified IACC, denoted by $IACC_f$, was implemented using a third-order Butterworth low-pass filter with a 1-kHz cutoff frequency. After rectification and low-pass filtering of the at-ear signals, the regular cross-correlation was used (Equation 3).

Both $IACC$ and $IACC_f$ were calculated over the duration of the stimuli (about 5 s) in a sliding window of length 50 ms, with 50% overlap. The time-varying IACC was then averaged to obtain a single value per stimulus.

2.4.3. Lateral energy fraction

Lateral energy fraction was introduced by Barron and Marshall [28] as a measure of spatial impression. It is defined as the ratio of *early* sound energy arriving laterally over sound energy arriving from all directions, and is in practice calculated as [25]

$$LF_E = \frac{\int_{5\,\text{ms}}^{80\,\text{ms}} h_8^2(t)\,dt}{\int_{0\,\text{ms}}^{80\,\text{ms}} h_0^2(t)\,dt}\,, \tag{4}$$

where $h_0(t)$ and $h_8(t)$ are impulse responses measured in the room with an omnidirectional and bipolar microphone, respectively. From the impulse response measurements, only the early reflections (hence the subscript E), are included. When dealing with recordings, the early and late energy cannot be easily separated; therefore, the *total* lateral fraction will be calculated as

$$LF_T = \frac{\int_{t_1}^{t_2} p_8^2(t)\,dt}{\int_{t_1}^{t_2} p_0^2(t)\,dt}\,, \tag{5}$$

where $p_0(t)$ and $p_8(t)$ are recorded with an omnidirectional and a bipolar microphone, respectively, as described in Section 2.4.1.

2.4.4. Simulated sound field

In some situations it can be useful to estimate physical parameters directly from the loudspeaker signals, rather than conducting recordings at the listening position. For simplicity, let us assume free field conditions (alternatively, room impulse responses could be used to account for the effects of the room). The signal recorded by a virtual omnidirectional microphone can be expressed as the sum of the loudspeaker signals $x_c(t)$ of each active channel c,

$$p_0(t) = \sum_c x_c(t).$$

The signal recorded by an ideal figure-of-eight microphone includes the loudspeaker direction θ_c, and is obtained as

$$p_8(t) = \sum_c x_c(t)\,\cos(\theta_c).$$

The parameter LF_{sim} is then identical to LF_T (Equation 5), but using the simulated sound field instead of the real recorded signals.

Finally, $IACC_{sim}$ is defined as $IACC_f$, but with left and right signals derived from the virtual stereo microphone (as in Mid/Side stereophony): $p_l(t) = p_0(t) + p_8(t)$ and $p_r(t) = p_0(t) - p_8(t)$.

2.4.5. Spectral centroid

The spectral centroid, f_c, is the center of gravity of the frequency spectrum, and has been used as a correlate of brightness of musical instruments [29, 30]. It is calculated as

$$f_c = \frac{\sum_{i=1}^{N} f_i A(f_i)}{\sum_{i=1}^{N} A(f_i)}\,, \tag{6}$$

where $A(f_i)$ is the amplitude of the spectrum in frequency band i. In this paper, the 1/3-octave band spectra of the binaural recordings are used.

[1] Out-of-phase signals may be produced by the surround loudspeakers in some of the upmixing algorithms.

2.4.6. Zwicker sharpness

Taking into account the specific loudness in critical bands, a more psychoacoustically based measure can be devised, such as the sharpness model proposed by Zwicker and Fastl [31]:

$$S = 0.11 \, \frac{\int_0^{24\,\mathrm{Bark}} N'(z) g(z) z \, dz}{\int_0^{24\,\mathrm{Bark}} N'(z) \, dz} \text{ acum}, \qquad (7)$$

where $N'(z)$ is the specific loudness and $g(z)$ a weighting factor as a function of critical-band rate, increasing exponentially above 16 Bark. Sharpness was calculated from the binaural recordings using Brüel & Kjær's PULSE Sound Quality software.

3. RESULTS

3.1. Relation between specific auditory attributes and overall preference

Figure 3 shows the derived ratio scales for each auditory attribute and the four types of program material. The results for *naturalness* are not shown, because this attribute was disregarded in the further analysis (see below). According to goodness-of-fit tests, the simple, but restrictive BTL model was generally found to fit the choice data well, which indicates that judgments were highly consistent. For the attributes *width* and *envelopment* in the Steely Dan excerpt, however, the BTL model had to be rejected, but two EBA models accounted for the data. This suggests that several aspects might have played a role when judging upon these attributes (more details on the analysis of the choice data are reported in [32]). Within each attribute, considerable similarity of the scales was observed across program materials, which was even more pronounced within musical genre (classical and pop music). For example, *ws* was perceived to be strongly elevated in comparison with the other reproduction modes in the pop material (Steely Dan and Sting); the effect was less distinct, but still visible, for the classical material. The stimuli showed the smallest perceptual differences with respect to *distance*; the mono sounds (*mo* and *ph*) were perceived to be nearest to the listener only for the pop music, for the classical music they were further away than most of the other reproduction modes. Except for distance and brightness, *mo* and *ph* were located at the lower end of the sensation scales, which induces correlation also *across* the attributes. Especially the correspondence between *spaciousness* and *envelopment* is striking. Clearly, these attributes did not vary independently in the stimuli under study.

Table 3: Attribute loadings on the factors ($F1$ and $F2$) obtained from principal component analysis, and variance explained by these factors after varimax rotation. Loadings higher than 0.6 are indicated in boldface.

	Classical		Pop	
Attribute	F1	F2	F1	F2
width	.50	**.75**	**.94**	.17
spaciousness	**.68**	**.68**	**.93**	.26
envelopment	.56	**.77**	**.94**	.17
distance	−.16	**−.88**	**.84**	.13
brightness	**.91**	.24	.24	**.92**
elevation	**.83**	.41	.15	**.93**
clarity	**.90**	.35	**.78**	.47
Var. explained (%)	48	39	58	30

To circumvent problems of collinearity of the attribute scales, principal component analysis (PCA) with varimax rotation was used to reduce the attribute space to fewer independent factors (or *components*). In order to increase the generalizability of the model, the data were aggregated within musical genre, i. e., classical music (Beethoven and Rachmaninov) and pop music (Steely Dan and Sting), thereby doubling the number of data points to be predicted. This was justified given the similarities observed in the attribute scales across program materials (see Figure 3). *Naturalness* was excluded from the analysis because it was considered more global than the other specific attributes and not sufficiently separate from preference, the correlation coefficients between *naturalness* and preference ranging from 0.94 (Steely Dan) to 0.98 (Rachmaninov).

The PCA was performed on the remaining seven attributes. In the case of the classical music, 87% of the variance in the scale values was explained by the first two factors which, after rotation, accounted for 48 and 39% of the variance, respectively. For the pop music, the first two components accounted for 58 and 30% (88% cumulated) after rotation. The loadings of the attribute scales on the first two factors, calculated as correlation coefficients, are reported in Table 3. Although the relationship between the attributes and the two factors is more clearcut for the

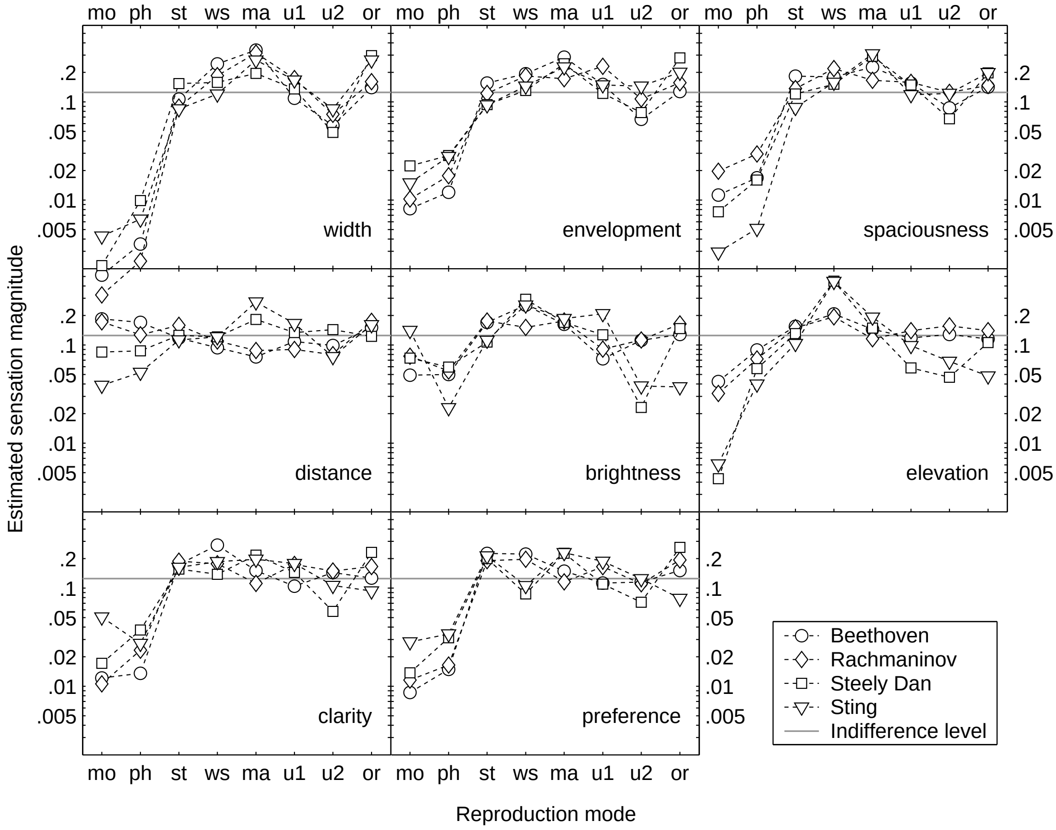

Figure 3: Ratio scales of seven auditory attributes and overall preference, estimated using BTL and EBA models for four types of program material. See Table 2 for abbreviations of the reproduction modes. The indifferent line shows the theoretical location of the scale values when all pairwise probabilities are 50%.

pop music (because the intercorrelation between the attributes is not as strong as for the classical music), similarities can be observed between the two genres: *brightness* and *elevation* load on the same factor, while the other factor is closely related to *width*, *spaciousness*, *envelopment* and *distance* (note that *distance* loads negatively for the classical music; see also Figure 3). Thus, an analogy can be made between Factor 1 in the PCA for classical music and Factor 2 for the pop music, and vice-versa, with the following exceptions: *clarity* which loads on Factor 1 in both cases, and *spaciousness* which loads equally on both factors for the classical material. Figures 4 and 5 show a graphical representation of the attribute loadings and stimulus scores in the two-dimensional factor spaces. The coordinates of the arrow endpoints are calculated as two times the factor loadings.

A multiple regression was performed on the two factors (F_1 and F_2) obtained from PCA in order to predict the preference scale values (P). The resulting regression equations are

$$\hat{P} = .138 + .075F_1 + .017F_2 - .014F_1^2 \quad \text{(Classical)} \qquad (8)$$

$$\hat{P} = .155 + .057F_1 + .058F_2 - .032F_2^2 \quad \text{(Pop)} \qquad (9)$$

all three terms in each equation being significant. In both genres, the quadratic term refers to the factor correlating with *brightness* and *elevation*, and

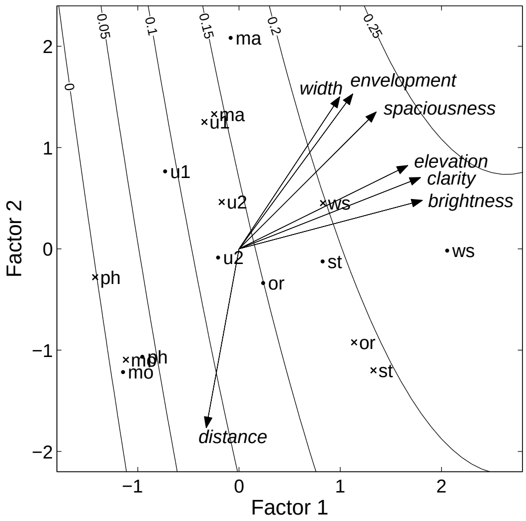

Figure 4: Graphical representation of the factor space obtained from principal component analysis of the attribute scales, and predicted preference (Equation 8) for the classical music material. Factor loadings of the attributes are shown as arrows, and the scores of the reproduction modes along the two factors are represented as dots (Beethoven) or crosses (Rachmaninov). The preference estimated from the two factors is represented by contour lines.

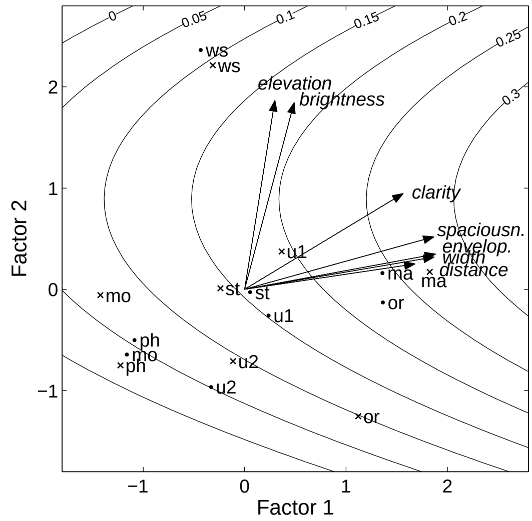

Figure 5: Graphical representation of the factor space obtained from principal component analysis, and predicted preference (Equation 9) for the pop music material: Steely Dan (dots) and Sting (crosses). See Figure 4.

is mainly due to *ws* which was both bright and elevated, but only moderately preferred. This gives rise to an inverse u-shaped relation between this factor and preference which was modeled by the quadratic term. The predicted preferences are illustrated by contour lines in Figures 4 and 5 for classical and pop music, respectively; the values written along the equal-preference contours follow from Equation 8 and 9. In Figure 4, for example, predicted preference increases when moving from the left to the upper right part of the panel. Generally, the two models were found to predict preference quite well with a total explained variance of 94% (classical) and 84% (pop). The largest prediction errors were obtained for *u1* in the classical music, and *st* in the pop music, both being underestimated.

3.2. Relation between auditory attributes and objective parameters

The relation between each of the auditory attributes and each of the objective parameters was estimated using linear regression. In order to increase the generalizability of the results, the data were combined within genres, as in the previous section, resulting in 16 data points. Figure 6 illustrates the negative correlation between the interaural cross-correlation coefficient $IACC_f$ and perceived *spaciousness*. As a result, $IACC_f$ explained 64% (respectively 84%) of the variance in the *spaciousness* scale values, for classical (resp. pop) music. Figure 7 shows that there is a positive correlation between lateral fraction and *envelopment*; the explained variance amounts to 68 (classical) and 71% (pop music). Generally it can be seen that $IACC_f$ and LF_T values cover a larger range in the classical recordings than in the pop music, although this does not yield a higher correlation.

The squared correlations between all objective and subjective data are reported in Table 4. As expected, interaural cross-correlation coefficient and lateral energy fraction correlate mostly with the spatial attributes (*width*, *envelopment* and *spaciousness*). In addition, some correlation with *clarity* can be observed for pop music, which is much less for classical music. Note that the PCA results in Table 3 and Figures 4 and 5 also suggest a closer relation between *clarity* and the spatial attributes for pop than for classical music.

The spectral centroid f_c and sharpness S, however,

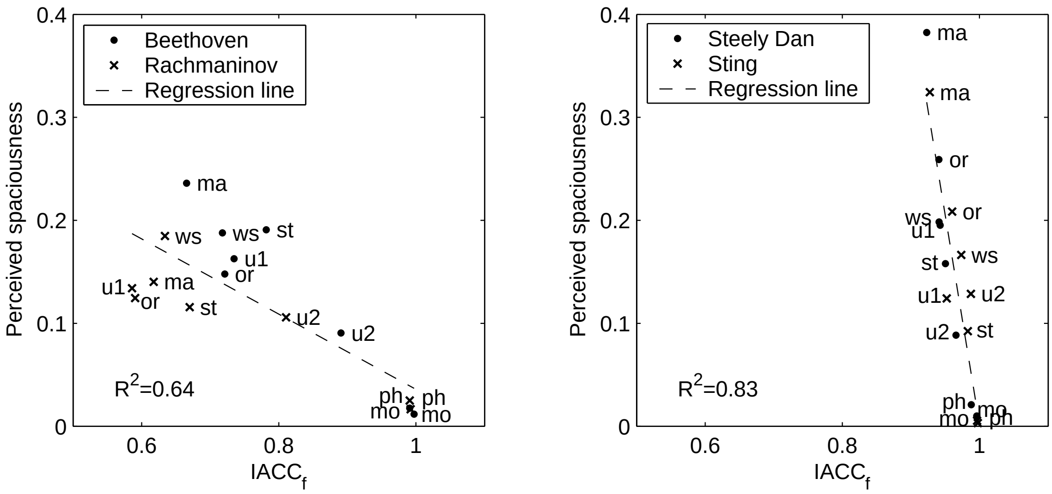

Figure 6: Relation between *spaciousness* and $IACC_f$. Regression on the aggregated data for classical (left) and pop music (right). R^2 indicates the squared correlation between predicted and observed values.

Table 4: Squared correlations (R^2) between objective parameters and auditory attributes *width, envelopment, spaciousness, distance, brightness, elevation* and *clarity*, for data combined within musical genre: classical (C) and pop (P). See Section 2.4 for a definition of the objective parameters. Note: R^2 higher than 0.6 is indicated in boldface.

	wid		env		spa		dis		bri		ele		cla	
	C	P	C	P	C	P	C	P	C	P	C	P	C	P
$IACC$	**.75**	.14	**.67**	.39	.56	.51	.23	.22	.42	.05	.34	.26	.27	.38
$IACC_f$	**.60**	.44	**.71**	**.68**	**.64**	**.83**	.17	.57	.46	.22	.51	.22	.43	**.62**
LF_T	**.88**	.38	**.68**	**.71**	.57	**.90**	.34	.48	.39	.25	.23	.18	.23	**.66**
$IACC_{sim}$	**.75**	.53	**.78**	.59	**.67**	**.74**	.24	**.81**	.30	.27	.32	.01	.31	.57
LF_{sim}	**.90**	.53	**.74**	**.77**	**.65**	**.93**	.34	**.65**	.40	.23	.28	.07	.28	**.71**
f_c	.01	.20	.02	.05	.01	.01	.00	.09	.10	.03	.01	.29	.00	.00
S	.02	.09	.01	.00	.01	.01	.03	.03	.04	.01	.02	.40	.00	.01

did not relate substantially to any of the auditory attributes, with a maximum correlation ($R^2 = 0.40$) being observed between sharpness and *elevation* for pop music. For *brightness* in particular, where some correspondence was expected, the correlation was low: a maximum of only 10% of the variance was accounted for by the spectral centroid (in classical music).

4. DISCUSSION

4.1. Predicting listener preference

Predicting listener preference from specific subjective attributes and, ultimately, from objective measures, is one of the ongoing challenges in research on sound quality. It was not the ambition of this exploratory study to develop a general sound quality model; however, the relation between specific auditory attributes and overall preference established in this paper provides some insight in which sensations could play a role when assessing the overall quality of reproduced sound.

In order to deal with the collinearity of the elicited, and subsequently scaled attributes, this relation was obtained by regression of the preference scale values on two (orthogonal) principal components extracted from the attribute scales. It is not possible from this study to determine whether this collinearity results from a common underlying sensation, or whether distinct sensations are involved but co-vary in the context of the selected stimuli. Therefore, the rela-

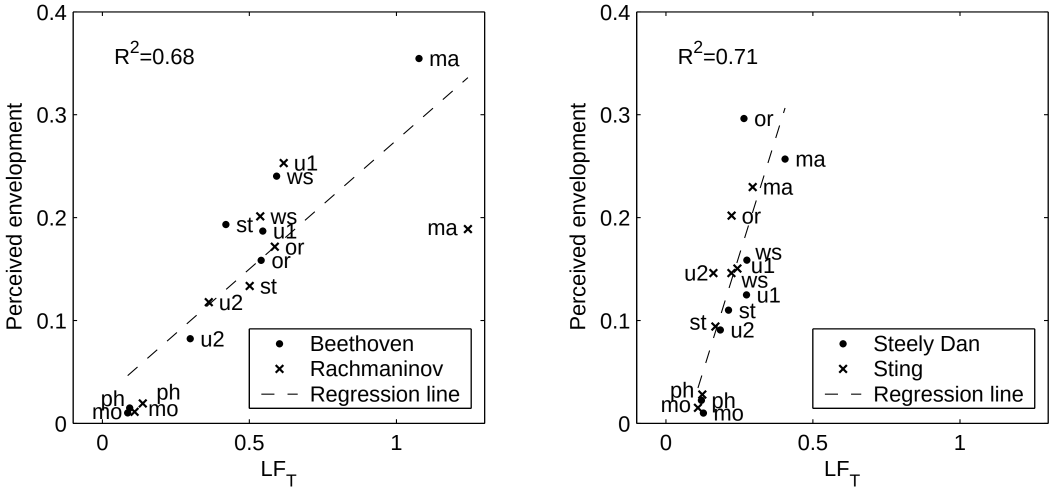

Figure 7: Relation between *envelopment* and lateral fraction for classical (left) and pop music (right).

tion between single attributes and overall preference must be interpreted with care. It should be seen as an indicator of the possible contribution of each specific attribute, which should be confirmed in further studies.

The four recordings were grouped into two musical genres, resulting in two models, one for classical music (Equation 8) and the other one for pop music (Equation 9), which accounted for 94 and 84% (respectively) of the variance in the preference scale values. The similarities between the classical and pop genres in Table 3 and in Equations 8 and 9 are encouraging, as they suggest that similar sensations might have played a similar role in the preference judgments across program material. From the quadratic term in the regression equations, the tentative conclusion might be drawn that, for the two attributes *elevation* and *brightness*, there exists an optimal value above which preference starts to decrease. However, considering the exploratory nature of this study, and the limited number of stimuli, it will be incumbent upon future research to gain a clearer picture of the functional relations between preference in multichannel sound and the underlying auditory attributes.

4.2. Deriving objective correlates

The derivation of objective correlates of auditory attributes is of interest at least for two reasons: First, objective measures might provide further insight into auditory perception of complex stimuli. Second, such measures might be easier and cheaper to obtain than the auditory attributes themselves. In this study it was possible to apply parameters originally devised for room acoustics to reproduced sound. For spatial attributes (*width*, *envelopment*, *spaciousness*) the correlations observed were substantial (Table 4). The linear relationship, however, was not always perfect. In particular, none of the objective parameters correlated highly with *width* in pop music, partly due to the original 5-channel format (*or*) being underestimated; for *spaciousness* and *envelopment*, however, such outliers were not observed. This underestimation might be explained by discrete events on the side of the listener, contributing to the impression of width but not strongly affecting $IACC$ and LF. It is likely that a better prediction would result if more involved methods of deriving objective parameters were employed, such as including variations in interaural time difference [33], or using a more complex binaural model (e. g., [34]). Future work will include such analyses.

For the trimbral attributes—especially *brightness*—frequency centroid and sharpness showed a poor performance. The reason might be that these parameters are based on the spectra of the reproduction modes. Overall, however, the differences in the spectra were only subtle and do not reflect the changes in *brightness*. It is conceivable that listeners when judging *brightness* or *elevation* focus on those aspects of the stimuli which maximize their differences, e. g., on the timbre of the voice of the lead singer, or on the vertical position of the saxophone. It will be a challenge for more advanced objective measures of timbral attributes to take such strategies into account.

4.3. Concluding remarks

The following conclusions can be drawn based on the results reported in this paper: (1) Listeners' judgments upon selected auditory attributes of spatially reproduced sound and overall preference were found to be highly consistent. (2) Consequently, it was possible to derive scales of sensation strength from the collected binary paired-comparison data. (3) The relationship between specific auditory attributes and preference was expressed in multiple regression models which predict the data well. (4) Objective parameters of the sound field based on interaural cross-correlation and lateral fraction were able to predict spatial auditory attributes. Predictors of the timbral attributes are yet to be found.

5. ACKNOWLEDGMENTS

This research was carried out as part of the "Centerkontrakt on Sound Quality" which establishes participation in and funding of the "Sound Quality Research Unit" (SQRU) at Aalborg University. The participating companies are Bang & Olufsen, Brüel & Kjær, and Delta Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP). The authors would like to thank Wookeun Song for helping with calculating instrumental measures of sharpness.

6. REFERENCES

[1] L. L. Beranek. Concert hall acoustics. *J. Acoust. Soc. Am.* 92(1):1–39, 1992.

[2] ISO 3382. Acoustics – measurement of the reverberation time of rooms with reference to other acoustical parameters. ISO, Geneva, Switzerland, 1997.

[3] T. Nakayama, T. Miura, O. Kosaka, M. Okamoto, and T. Shiga. Subjective assessment of multichannel reproduction. *J. Audio Eng. Soc.* 19(9):744–751, October 1971.

[4] A. Gabrielsson and H. Sjögren. Perceived sound quality of sound-reproducing systems. *J. Acoust. Soc. Am.* 65(4):1019–1033, 1979.

[5] F. Rumsey. Subjective assessment of the spatial attributes of reproduced sound. *Proceedings of the AES 15th International Conference: Audio, Acoustics & Small Spaces*, pp. 122–135, 1998.

[6] J. Berg and F. Rumsey. Spatial attribute identification and scaling by repertory grid technique and other methods. *Proceedings of the AES 16th International Conference: Spatial Sound Reproduction*, pp. 51–66, 1999.

[7] N. Zacharov and K. Koivuniemi. Audio descriptive analysis & mapping of spatial sound displays. *Proceedings of the 2001 International Conference on Auditory Displays, Espoo, Finland*, pp. 95–104, 2001.

[8] C. Guastavino and B. F. G. Katz. Perceptual evaluation of multi-dimensional spatial audio reproduction. *J. Acoust. Soc. Am.* 116(2):1105–1115, 2004.

[9] S. K. Zieliński, F. Rumsey, and S. Bech. Effects of down-mix algorithms on quality of surround sound. *J. Audio Eng. Soc.* 51(9):780–798, September 2003.

[10] F. Rumsey, S. K. Zieliński, R. Kassier, and S. Bech. On the relative importance of spatial and timbral fidelities in judgements of degraded multichannel audio quality. *J. Acoust. Soc. Am.* 118(2):968–976, 2005.

[11] D. Griesinger. Objective measures of spaciousness and envelopment. *Proceedings of the AES 16th International Conference: Spatial Sound Reproduction*, April 1999.

[12] R. Mason and F. J. Rumsey. A comparison of objective measurements for predicting selected subjective spatial attributes. *112th Convention of the Audio Engineering Society, Munich, Germany, May 10–13*, 2002. Preprint 5591.

[13] G. A. Soulodre, M. C. Lavoie, and S. G. Norcross. Objective measures of listener envelopment in multichannel surround systems. *J. Audio Eng. Soc.* 51(9):826–840, September 2003.

[14] M. Deschamps, O. Warusfel, and A. Baskind. Investigation of interactions between recording/mixing parameters and spatial subjective attributes in the frame of 5.1 multichannel. *115th Convention of the Audio Engineering Society, New York, USA, October 10–13*, 2003. Preprint 5930.

[15] F. Wickelmaier and S. Choisel. Selecting participants for listening tests of multichannel reproduced

sound. *118th Convention of the Audio Engineering Society, Barcelona, Spain, May 28–31*, 2005. Preprint 6483.

[16] ITU-R BS.1116. Methods for the subjective assessment of small impairment in audio systems including multichannel sound systems. International Telecommunications Union, Geneva, Switzerland, 1997.

[17] ITU-R BS.775-1. Multichannel stereophonic sound system with and without accompanying picture. International Telecommunication Union, Geneva, Switzerland, 1994.

[18] S. Choisel and F. Wickelmaier. Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound. *118th Convention of the Audio Engineering Society, Barcelona, Spain, May 28–31*, 2005. Preprint 6369.

[19] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* 39:324–345, 1952.

[20] R. D. Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

[21] A. Tversky. Elimination by aspects: a theory of choice. *Psychol. Rev.* 79:281–299, 1972.

[22] A. Tversky and S. Sattath. Preference trees. *Psychol. Rev.* 86:542–573, 1979.

[23] F. Wickelmaier and C. Schmid. A Matlab function to estimate choice model parameters from paired-comparison data. *Behav. Res. Meth. Instr. Comp.* 36:29–40, 2004.

[24] T. Okano, L. L. Beranek, and T. Hidaka. Relations among interaural cross-correlation coefficient (IACC_E), lateral fraction (LF_E), and apparent source width (ASW) in concert halls. *J. Acoust. Soc. Am.* 104(1):255–265, July 1998.

[25] T. Hidaka, L. L. Beranek, and T. Okano. Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls. *J. Acoust. Soc. Am.* 98(2):988–1007, 1995.

[26] J. Blauert and W. Cobben. Some consideration of binaural cross correlation analysis. *Acustica* 39(2):96–104, 1978.

[27] R. Mason, T. Brookes, and F. Rumsey. Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli. *J. Acoust. Soc. Am.* 117(3):1337–1350, March 2005.

[28] M. Barron and A. H. Marshall. Spatial impression due to early reflections in concert halls: The derivation of a physical measure. *J. Sound Vib.* 77(2):211–232, 1981.

[29] J. M. Grey and J. W. Gordon. Perceptual effects of spectral modifications on musical timbres. *J. Acoust. Soc. Am.* 63(5):1493–1500, 1978.

[30] J. Beauchamp. Synthesis by spectral envelope and brightness matching of analyzed musical instrument tones. *J. Audio Eng. Soc.* 30(6):396–406, 1982.

[31] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models*. Springer, second updated edition, 1999.

[32] S. Choisel and F. Wickelmaier. Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference. *Submitted*, 2006.

[33] R. Mason, T. Brookes, and F. Rumsey. Integration of measurements of interaural cross-correlation coefficient and interaural time difference within a single model of perceived source width. *117th Convention of the Audio Engineering Society, San Francisco, USA, October 28–31*, 2004. Preprint 5591.

[34] W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *J. Acoust. Soc. Am.* 80(6):1608–1622, 1986.