# Training of Listeners for Evaluation of Spatial Attributes of Sound

Juha Merimaa[1*] and Wolfgang Hess[1†]

[1]*Institute of Communication Acoustics, Ruhr-Universität Bochum, 44780 Bochum, Germany*

Correspondence should be addressed to Juha Merimaa (`juha.merimaa@hut.fi`)

**ABSTRACT**
A group of listeners were engaged in training to learn to evaluate auditory source width (ASW) and listener envelopment (LEV). The training consisted of discussions on perception of spatial sound and visualization of both attributes with drawings. After each session the subjects evaluated the ASW and LEV of a set of stimuli consisting of different source signals simulated in a few chosen acoustical environments. Most subjects developed consistent criteria for their judgements and maintained them throughout the training and a subsequent control two months later. However, considerable individual differences were found. Analysis of the data revealed that large part of the differences was due to different judgements between the chosen source signals. The training also suggested that some differences could have been caused by the translation from multi-dimensional perception to the unidimensional judgements. A further graphical evaluation of the stimuli showed that this was not the case.

## 1. INTRODUCTION

Spatial impression in concert halls is traditionally divided into auditory (or apparent) source width (ASW) and listener envelopment (LEV). Early work on spatial impression concentrated mainly on the ASW dimension [1, 2, 3] produced by early lateral reflections, although Marshall [1] described the investigated "spatial responsiveness" as "a sense of en-

velopment in the sound". Blauert and Lindemann [4] first showed that "auditory spaciousness" is a multidimensional perceptual attribute affected differently by early and late reflections. Existence of the ASW and LEV dimensions was later verified in listening experiments by Morimoto and Maekawa [5] and Bradley and Soulodre [6]. Similar attributes have also been found in studies aiming at identifying the spatial dimensions of reproduced sound. Berg and Rumsey [7, 8] clearly describe source width and envelopment and Koivuniemi and Zacharov [9] found the attributes "broadness" and "sense of space". For

---

*Also with: Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Finland

†Also with: Harman/Becker Automotive Systems, 76307 Karlsbad-Ittersbach, Germany

a scene-based paradigm for description and assessment of spatial sound, see Rumsey [10].

General evaluation of ASW and LEV appears to be very difficult and to require trained listeners. The main problem is that the concepts of ASW and LEV are not clear to naive listeners. Training of listeners with samples including changes that are as unidimensional as possible in the desired attributes has been described by Koivuniemi and Zacharov [9] and Neher et al. [11]. However, in this study a different approach was chosen. The training utilized a method somewhat similar to verbal elicitation techniques reported in the literature [7, 8, 9] but in a different context. More specifically, terms for the dimensions were given, and the task of the listeners was to find these attributes from the training samples and to learn to identify them. Possible learning was monitored in a listening experiment conducted after each training session.

In addition to collecting data of the progress of the training, the experiment also addressed source signal dependence of spatial impression in rooms. The results showed considerable individual differences between the listeners, part of which were hypothesized to be due to the utilized direct scaling paradigm. This hypothesis was further investigated using graphical assessment of the stimuli in Experiment II.

The paper is organized as follows. Sec. 2 describes the training process. Sec. 3 presents the results of the control experiment denoted Experiment I. The graphical assessment of the stimuli is described in in Sec. 4, followed by summary and conclusions in Sec. 5.

## 2. TRAINING

As discussed in the introduction, the objective of the training was to explain the spatial attributes of interest (ASW, LEV) to naive subjects, as well as to help them to find consistent criteria for evaluating these attributes. Altogether 16 subjects, aged between 16–27 years, participated in the training. Half of the subjects were male and half female, and they were paid an hourly rate for both the training and the subsequent experiments. None of the subjects had earlier experience in listening tests involving spatial sound, although some subjects had

participated earlier in other experiments at the Institute of Communication Acoustics. Tested pure tone audiometry levels of all the subjects were within 20 dB.

The training was conducted in groups of four subjects in four sessions on four consecutive days. The groups were reorganized on each day so that each subject would get to interact with as many others as possible. Each training session consisted of approximately 35–40 minutes of group work, after which a control listening experiment was conducted individually to each subject at a time. During the group work, the subjects listened to binaural stimuli reproduced with Stax Lambda Pro headphones. The stimuli included diotic signals, different anechoic source signals convolved with measured binaural room impulse responses (BRIRs), as well as noises with a varying degree of interaural crosscorrelation.

In the first session the ASW and LEV of the stimuli were discussed. After the session some subjects indicated that the meanings of ASW and LEV were not yet clear, so an additional task of visualizing the stimuli with drawings was introduced to the last three sessions. The subjects were given a piece of paper with a head (as seen from above) depicted in the middle, and they were asked to illustrate the sound source and the envelopment. The drawings were then discussed. In the third and fourth session the drawings were limited to a method naturally chosen by most subjects, consisting of visualization of the direction and width of the source with a line or an arc and the envelopment with an ellipse drawn around the head. Note that Mason et al. [12] have also proposed this form of training in their discussion of verbal and nonverbal elicitation methods.

Since all the subjects were German speaking, the training was conducted in German. The terms "(wahrgenommene) Breite der Schallquelle" ((perceived) width of the sound source) and "(wahrgenommene) Umhüllung" ((perceived) envelopment) were used. The authors moderated the discussions. In order not to teach the subjects to answer according to the expectations of the authors but to learn to identify their own perceptions instead, the training was conducted in a double blind manner: the stimuli of each session were played back

in a randomized order with headphones to the listeners only, so that the authors did not know which stimulus was currently being discussed. The authors also restricted their role to only asking questions about the stimuli and the perception of the listeners, thus avoiding any direct commentation based on their own impressions.

In the group discussions, special emphasis was placed on comparison of ASW and LEV. The reasoning behind the chosen method was that if the ASW and LEV are clear perceptual dimensions which can be found in verbal elicitation experiments [7, 8, 9], a group of subjects should be able to recognize and learn them using a set of diverse training samples. Furthermore, the method avoided the difficult task of generating and validating a set of training samples with unidimensional changes in ASW and LEV.

The starting point for the discussions was that no perception is wrong and individual differences can exist. Individual differences did indeed exist, although based on informal observations, the conversations and the drawings appeared to converge somewhat towards the end of the training. The possible convergence was formally assessed in the listening experiment described in the next section.

## 3. EXPERIMENT I

The purpose of Experiment I was twofold. An important goal was, of course, to characterize the learning during the training. However, evaluating the ASW and LEV of different source signals in a few selected acoustical environments was also considered interesting as such. Altogether 12 stimuli were assessed. The same 16 paid subjects as in the training participated in the experiment. However, one of the male subjects was not available during the last run of the experiment, so his data was left out of the analysis.

The experiment was conducted in a sound proof booth. The stimuli were played back from a PC with a RME Digi96/8 Pad sound card at 48 kHz sampling frequency using a Stax SRM Monitor headphone amplifier with diffuse field equalization and Stax Lambda Pro headphones. No head-tracking was used.

### 3.1. Method

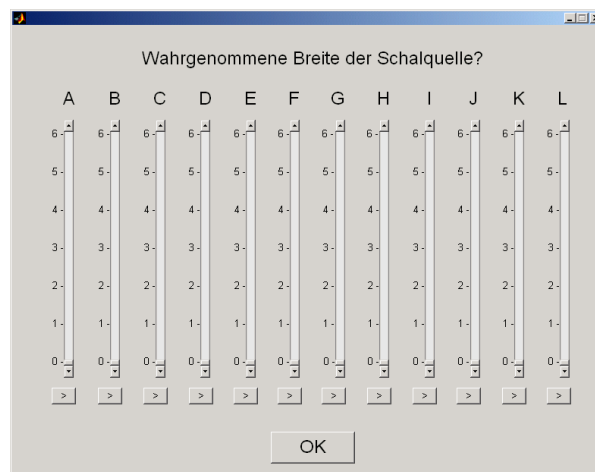The stimuli were assessed using a graphical user interface (GUI) and a direct scaling paradigm. A



Fig. 1: GUI used in Experiment I.

screenshot of the GUI is shown in Fig. 1. Each slider corresponded to a single stimulus, and the order of the stimuli was randomized. One run consisted of first evaluating the ASW and then the LEV of the 12 stimuli using similar GUIs. Between the evaluations of ASW and LEV, the order of the stimuli was changed and the sliders were reset to zero. The subjects were able to play the stimuli in any order and as many times as necessary with the buttons below the sliders. Scale values between 0–6 were used, where 0 denoted a point source or no envelopment.

The whole experiment consisted of five runs, each lasting approximately 5–10 minutes. The first four runs were conducted directly after the training sessions on consecutive days. In order to study possible long term effects, the fifth run was performed two months later. Before the fifth run, the subjects were not given a chance to listen to the stimuli or any other training samples. During the experiment, the subjects had written definitions of ASW and LEV available. ASW was defined as the perceived horizontal extension of the sound source and LEV as the perceived degree of envelopment by the sound field. None of the subjects reported having perceived multiple sound sources in any of the stimuli.

### 3.2. Stimuli

The 12 evaluated stimuli consisted of three anechoic source signals played back diotically (denoted the anechoic condition), as well as convolved with three

|       | Anechoic | Room | Small hall | Large hall |
|-------|----------|------|------------|------------|
| Cello | 1        | 2    | 3          | 4          |
| Noise | 5        | 6    | 7          | 8          |
| Snare | 9        | 10   | 11         | 12         |

Table 1: Combinations of source signals and environments used as stimuli in the experiments.

measured binaural room impulse responses (BRIRs). The source signals and the acoustical environments are listed in Table 1. Some similar although longer source signals in the same acoustical environments were used as part of the training samples. The first source signal was a single 1 s long note played on a cello (the first note on track 22 of the Archimedes CD [13]). The snare drum sample consisted of a single snare drum hit recorded in the anechoic chamber of Laboratory of Acoustics and Audio Signal Processing at Helsinki University of Technology. Both the cello and the snare drum sample were upsampled to 48 kHz. For the noise, 1 s long pink noise gated on and off with 10 ms raised cosine ramps was utilized. The same "frozen" noise sample was used throughout the experiments.

The binaural impulse responses of the three different acoustical environments were measured with an L-Acoustics MTD108a loudspeaker and a diffuse-field equalized dummy head based on the Neumann KU 80 head. The custom pinnae of the dummy head were asymmetrical and chosen as a result of an exploration study by Hudde and Schroeter [14]. The dummy head was placed on a manikin, and it was in all cases facing the sound source. The responses were measured using a sweep excitation and 48 kHz 24 bit A/D conversion. The signal-to-noise ratio of all acquired impulse responses was above 80 dB, rendering thus the background noise inaudible with the utilized stimulus levels. The free-field response of the measurement loudspeaker was equalized in the frequency range of 70–20000 Hz using an inverse minimum phase filter smoothed with 1/3 octave resolution.

The measured rooms and halls are described in more detail in Table 2. In the large hall, there are only few low level early reflections from the sides, and the sound is not considered very spacious. The listed values of early interaural cross-correlation ($IACC_E$)

are also very high. The small hall, on the other hand, is fairly diffuse (considerably lower IACCs), and it should thus yield relatively high ASW and LEV. The $IACC_E$ values of the room are between those of the two halls but the reverberation time (RT) is considerably smaller, which might change the criteria for the judgements. The anechoic case (not listed in Tables 2) was expected to provide an achor as the least spacious stimuli having no reverberation and IACC = 1 at all frequencies.

The IACCs calculated over the whole duration of the stimuli (source signals convolved with the BRIRs) including the reverberation tail are listed in Table 3. There is an obvious interaction between the sound source and the acoustical environment especially in the room and the small hall. The noise in the small hall has the lowest IACCs and would be expected to appear as the most spacious stimulus. The anechoic stimuli again have IACC = 1 at all frequencies.

For level calibration, the level of the anechoic noise was set to 72 dB SPL, and the subjective loudness of the other stimuli was equalized to that of the anechoic noise. A normalization gain for each stimulus was determined in a preliminary experiment using an adaptive up-down procedure [16]. The experiment started with 3 dB steps, which were reduced to 1 dB after 4 reversals. The assessment of each stimulus was terminated after 10 reversals. All stimuli were randomly interleaved in order to prevent the subjects anticipating the sequential changes. Three experienced listeners participated in the experiment, and the averages of their gain factors were used to equalize the loudness of the stimuli for the main experiment.

### 3.3. Results
Since there was a defined zero point on the scale, the experiment was expected to yield ratio scale data. However, this did not seem to be the case. Also the anechoic samples were often judged as having a considerable width and/or envelopment, and it appears that different subjects were using the zero point differently. Hence, the data were treated as interval scale judgements. In order to transform the data from all of the listeners to same scale, the means and standard deviations of the judgements of each subject during each run were scaled to the mean and standard deviation calculated over all subjects and all runs, as recommended by ITU [17].

**Large Hall**
Audimax, Ruhr-Universität Bochum, Germany

Fan shaped surround multi-purpose hall. 1872 seats. Source was located on the stage and receiver on the main floor at a distance of 9 m from the source.

|  | Frequency / Hz | | | | | |
|---|---|---|---|---|---|---|
|  | 125 | 250 | 500 | 1k | 2k | 4k |
| $RT_{30}$ | 2.4 | 2.5 | 2.7 | 2.6 | 2.1 | 1.7 |
| $IACC_E$ | 0.99 | 0.98 | 0.98 | 0.99 | 0.97 | 0.96 |
| $IACC_L$ | 1.00 | 0.96 | 0.72 | 0.56 | 0.28 | 0.29 |

**Small Hall**
Folkwang-Hochschule für Musik, Duisburg, Germany

T-shaped chamber music hall with seating areas in front of and on both sides of the stage. Approximately 150 movable seats. Source was located on the stage and receiver on the main floor at a distance of 8 m from the source.

|  | Frequency / Hz | | | | | |
|---|---|---|---|---|---|---|
|  | 125 | 250 | 500 | 1k | 2k | 4k |
| $RT_{30}$ | 1.8 | 1.7 | 1.5 | 1.6 | 1.4 | 1.3 |
| $IACC_E$ | 0.96 | 0.86 | 0.56 | 0.40 | 0.34 | 0.45 |
| $IACC_L$ | 0.92 | 0.71 | 0.16 | 0.11 | 0.07 | 0.10 |

**Room**
Studio, Institute of Communication Acoustics
Ruhr-Universität Bochum, Germany

Rectangular acoustically treated listening room. Source was located at the position of the center speaker of the 5.1 system and receiver in the "sweet spot" at a distance of 1.7 m from the source.

|  | Frequency / Hz | | | | | |
|---|---|---|---|---|---|---|
|  | 125 | 250 | 500 | 1k | 2k | 4k |
| $RT_{30}$ | 0.5 | 0.4 | 0.5 | 0.5 | 0.5 | 0.5 |
| $IACC_E$ | 0.91 | 0.83 | 0.43 | 0.59 | 0.66 | 0.79 |
| $IACC_L$ | 0.59 | 0.58 | 0.20 | 0.16 | 0.22 | 0.11 |

Table 2: Description of the acoustical environments and measurement positions for the BRIRs. Reverberation times ($RT_{30}$) and interaural cross-correlations (IACC) were calculated at octave bands as specified in [15], with the exception that binaural impulse responses were used also for the RT estimation. $IACC_E$ was evaluated over the first 80 ms from the beginning of the direct sound in the impulse response, and $IACC_L$ between 80 and 1000 ms.

|  |  | Frequency / Hz | | | | | |
|---|---|---|---|---|---|---|---|
|  |  | 125 | 250 | 500 | 1k | 2k | 4k |
| room | cello | 0.96 | 0.94 | 0.78 | 0.81 | 0.79 | 0.76 |
|  | noise | 0.93 | 0.85 | 0.48 | 0.61 | 0.66 | 0.77 |
|  | snare | 0.89 | 0.82 | 0.57 | 0.61 | 0.66 | 0.73 |
| small hall | cello | 0.91 | 0.90 | 0.57 | 0.50 | 0.42 | 0.15 |
|  | noise | 0.93 | 0.79 | 0.36 | 0.27 | 0.23 | 0.31 |
|  | snare | 0.91 | 0.81 | 0.46 | 0.35 | 0.29 | 0.26 |
| large hall | cello | 0.99 | 0.99 | 0.97 | 0.91 | 0.87 | 0.92 |
|  | noise | 1.00 | 0.97 | 0.93 | 0.92 | 0.92 | 0.94 |
|  | snare | 0.99 | 0.98 | 0.94 | 0.93 | 0.92 | 0.94 |

Table 3: Interaural cross-correlations calculated at octave bands over the whole duration of the stimuli.

The transformed ASW and LEV data of the individual subjects are plotted in Figs. 2 and 3, respectively. There are clear differences between the subjects. Based on visual inspection of the data, most subjects were performing fairly consistently. However, subjects 1, 4, and 13 did not appear to develop very consistent criteria for evaluating the ASW, and during several runs they judged the anechoic stimuli as widest. Nevertheless, the LEV judgements of subjects 1 and 4 are consistent, whereas subject 12 had problems. The results in the following sections are presented excluding subjects 1, 4, and 13 from the ASW data and subjects 12 and 13 from the LEV data. To test the validity of this exclusion, the whole analysis was also repeated including all subjects. The only notable difference was an increase in the error variances in the analysis of variance (ANOVA).

### 3.3.1. Learning Effects
The data of the individual listeners in Figs. 2 and 3 do not show very clear learning effects apart from the LEV judgements of the excluded subject 12. She has judged the first two runs differently compared to the last three runs. She was also one of the subjects who indicated that the meanings of ASW and LEV were not clear after the first training session. Fig. 4 shows the data averaged over the included subjects. The means and standard deviations of different runs are, indeed, very similar, although the ASWs of the cello samples from the last run appear slightly lower.

The possible learning effects related to each stimulus were investigated further with repeated measures ANOVA. Significant differences ($p < 0.05$) were found for the ASW of the noise source in the
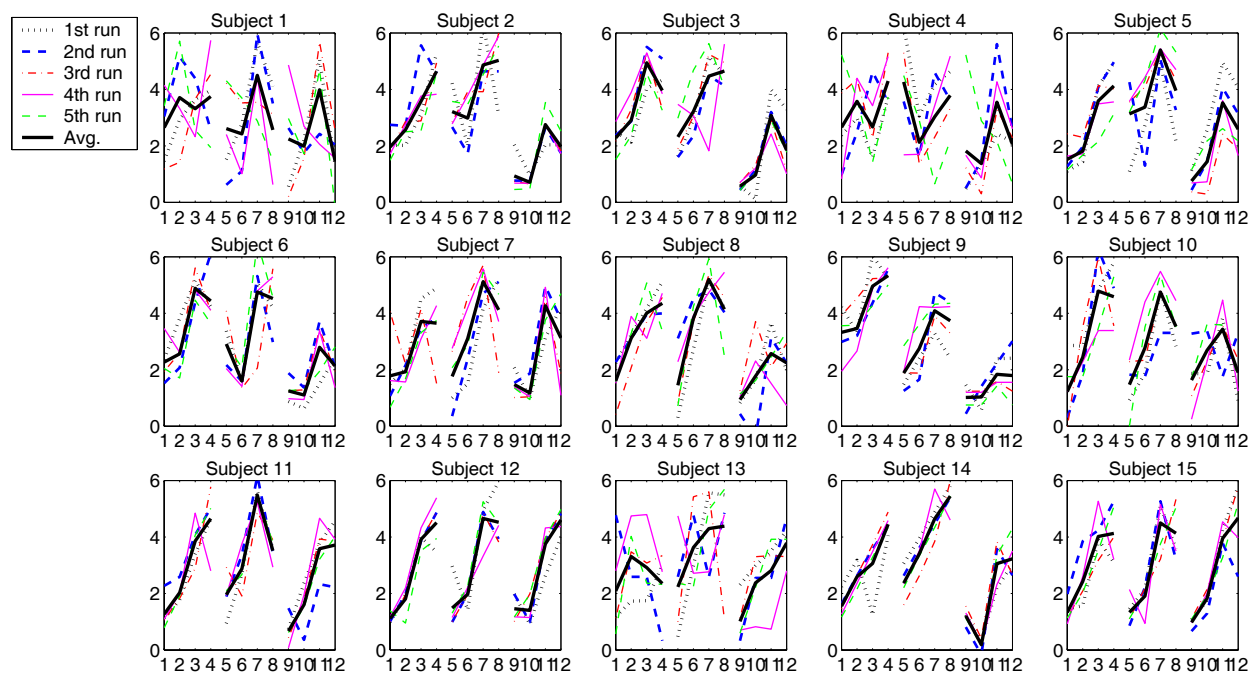
Fig. 2: Experiment I: Transformed ASW data. The stimuli on the x-axes are as listed in Table 1.
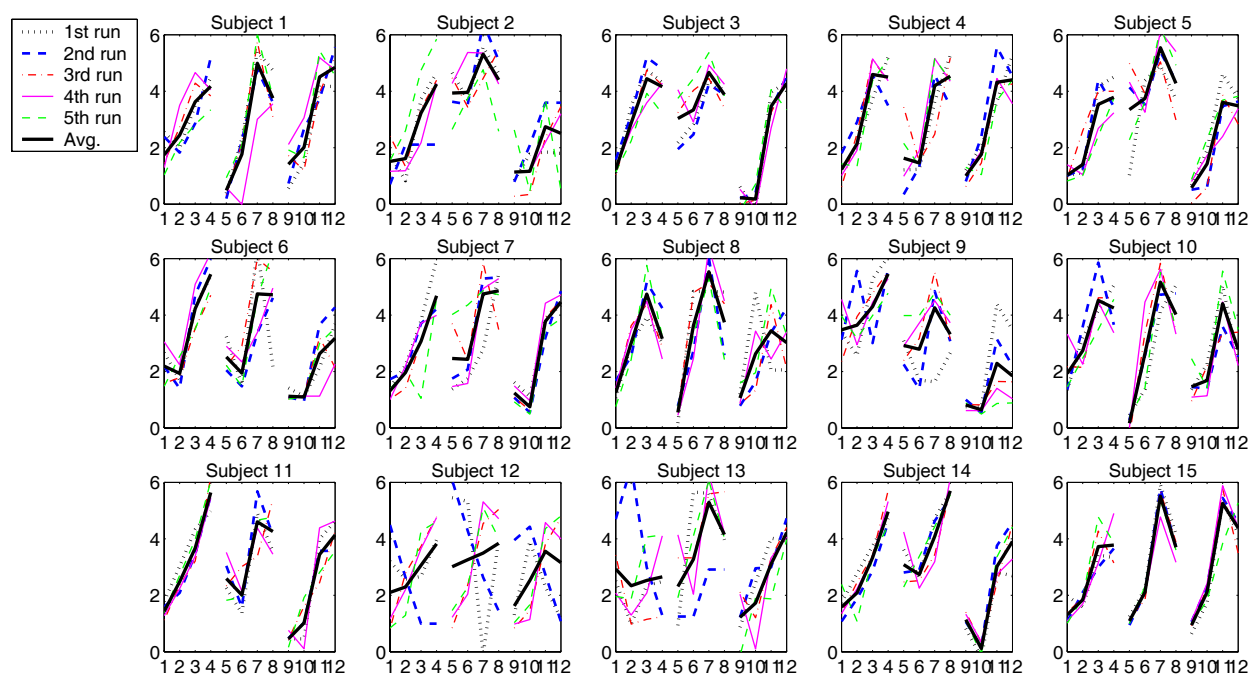


Fig. 3: Experiment I: Transformed LEV data. The stimuli on the x-axes are as listed in Table 1.

room and the snare drum source in the large hall. None of the LEV judgements were significantly different over the runs. It should be noted that with the chosen significance level it is expected that on the average one out of twenty tests will yield significant results by chance. The noise in the room shows a linear increasing trend, which could be a real learning effect. However, since this was the only considerable change over different runs, it can be concluded that the overall learning was negligible during the extended training. Furthermore, the subjects were able to evaluate the stimuli similarly after a break of two months. Nevertheless, the first training session was, of course, necessary in order to explain the concepts of ASW and LEV to the subjects.

### 3.3.2. ASW and LEV of the Stimuli

As no considerable differences between the five runs were found, the data from all runs were taken to further analysis. The averages over subjects and repetitions are shown in a reorganized order in Fig. 5. It is apparent, that the curves for the different sources have fairly similar shapes but different offsets. The snare drum samples have overall lowest ASWs and LEVs. Furthermore, the ASW and LEV are judged almost the same.

In order to gain more insight into the data, a three-way ANOVA of the ASW and LEV data was performed. The stimulus factor was split into room and source effects, and the subject factor was included as a fixed effect in the analysis. Due to the normalization of the scales of the individual listeners, the subject factor cannot have a main effect. However, individual differences should result in interactions with the subject factor. The results are shown in Tables 4 and 5. All factors and interactions are highly significant. The room factor accounts in both cases for the largest proportion of variance, and there is a small interaction between the room and the source. The interactions with the subject factor are fairly large and will be investigated further in the next section.

### 3.3.3. Individual Differences

As a first method to address the individual differences, factor analysis of the subject space was tried. However, the covariance matrices of the subject means are not positive definite, which makes the analysis difficult. Principal component analysis (PCA) revealed that fewer components than the number of subjects are sufficient to explain 100%

of the variance, which suggests multivariate dependences between the data of the subjects. Nevertheless, PCA also gives useful information about the underlying dimensions, and it was thus chosen as the exploration method.

PCA of both ASW and LEV yielded two components with eigenvalues above 1. Figs. 6 and 7 show the factor loadings for these two components. The first component is in both cases largely common to all subjects, accounting for 80% (ASW) and 76% (LEV) of the total variance. Notable individual differences appear in the second components, accounting for 8% (ASW) and 11% (LEV) of the total variance. Note that means over repetitions were used in this analysis, and thus the proportions of variance accounted for are not directly comparable with those reported in the ANOVA results in Tables 4 and 5.

Figs. 6 and 7 also illustrate the effects of the second components on the judgements. The component 2 in the LEV diagram is affected be different stimulus features than component 2 in in the ASW subject space, although both are related to differences between the sources signals. Due to the similarity of the ASW and LEV judgements both components probably exist for both attributes. However, PCA searches for components accountin for maximum amount of variance, and the weights of these two components may have been different in the ASW and LEV assessments.

The effect of the second ASW component is best understood from a plot of the data for different acoustical environments against the source signals (three rightmost panels in Fig. 6). The component describes an interaction between the subjects and the source signals. Subjects 12 and 15 have judged the stimuli mainly based on the acoustical environments. Subjects 5, 10, and 14 have perceived the noise source as widest, whereas subject 9 has perceived the cello as widest. Although less clear from the individual data, the second component of the LEV judgements appears to describe the relative weights of the source and room factors. The average data of subjects 1, 8, 10, and 15 is little influenced by the source signal, and the weight increases towards higher values of component 2.

### 3.3.4. Relation of the Results to Stimulus Properties

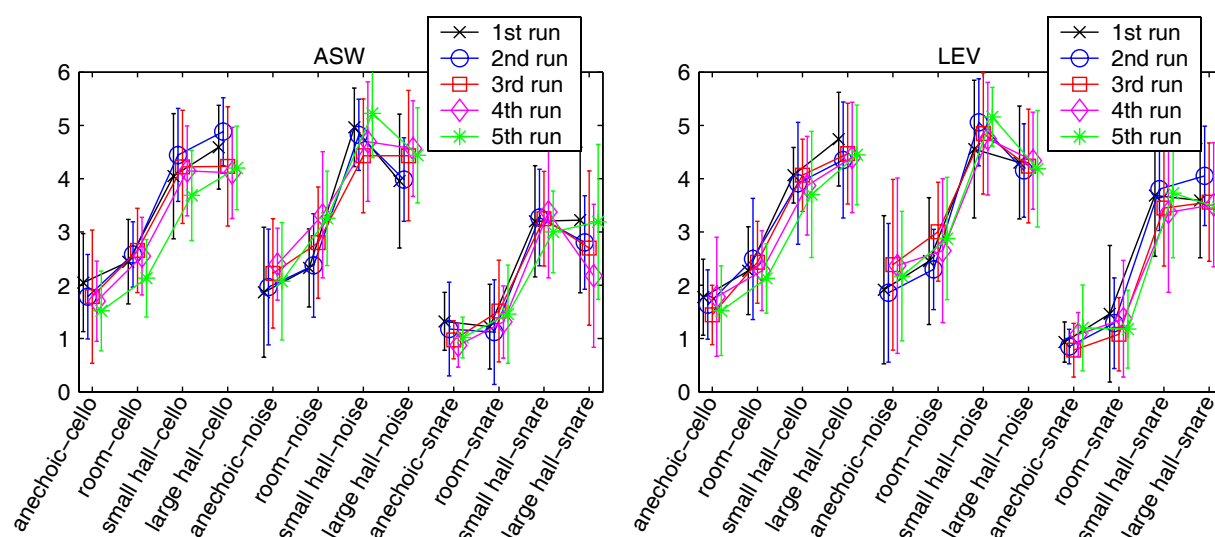Comparison of the results of the different acousti-

Fig. 4: Experiment I: Means and standard deviations of ASW (left panel) and LEV (right panel) averaged over 12 and 13 subjects, respectively. The lines connect judgements of the same stimuli in different acoustical environments.
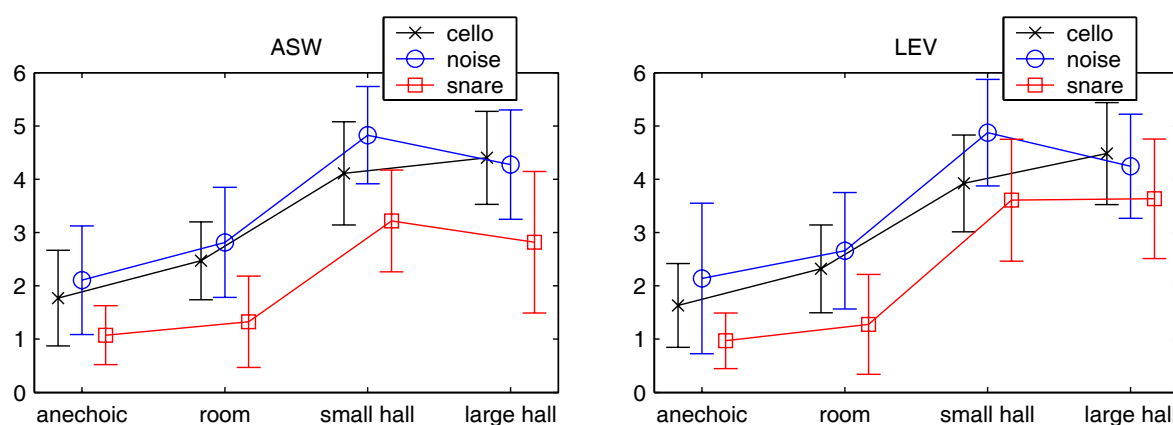


Fig. 5: Experiment I: Means and standard deviations of ASW (left panel) and LEV (right panel) averaged over subjects and repetitions.

cal environments in Fig. 5 to the interaural cross-correlations (IACCs) listed in Tables 2 and 3 reveals that the IACCs do not correspond well to the experimental data. If the IACC were the only descriptor for spatial impression, the small hall should have yielded by far larger ASWs and LEVs than the large hall. Also the ASW in the room should have been larger than in the large hall, although this or-

der would not have been expected for LEV due to the low energy of the late part of the response of the room. The data are insufficient to draw general conclusions. However, it seems that the size of the acoustical environment (or possibly the reverberation time) has affected the judgements such that smaller spaces have been perceived less spacious.

| Source | Sum Sq. | d.f. | Mean Sq. | $F$ | $p$ | % of Var. |
|---|---|---|---|---|---|---|
| Room | 762.55 | 3 | 254.183 | 391.56 | 0.000 | 45.6 |
| Source | 257.22 | 2 | 128.612 | 198.12 | 0.000 | 15.4 |
| Room*Source | 19.82 | 6 | 3.304 | 5.09 | 0.000 | 1.2 |
| Room*Subject | 73.01 | 33 | 2.212 | 3.41 | 0.000 | 4.4 |
| Source*Subject | 121.17 | 22 | 5.508 | 8.48 | 0.000 | 7.2 |
| Room*Source*Subject | 65.72 | 66 | 0.996 | 1.53 | 0.006 | 3.9 |
| Error | 373.91 | 576 | 0.649 | | | 22.3 |
| Total | 1673.40 | 719 | | | | |

Table 4: Experiment I: Analysis of variance of the ASW data.

| Source | Sum Sq. | d.f. | Mean Sq. | $F$ | $p$ | % of Var. |
|---|---|---|---|---|---|---|
| Room | 1051.18 | 3 | 350.393 | 609.84 | 0.000 | 52.4 |
| Source | 164.02 | 2 | 82.012 | 142.74 | 0.000 | 8.2 |
| Room*Source | 29.53 | 6 | 4.921 | 8.57 | 0.000 | 1.5 |
| Room*Subject | 137.98 | 36 | 3.833 | 6.67 | 0.000 | 6.9 |
| Source*Subject | 171.78 | 24 | 7.158 | 12.46 | 0.000 | 8.6 |
| Room*Source*Subject | 92.36 | 72 | 1.283 | 2.23 | 0.000 | 4.6 |
| Error | 358.53 | 624 | 0.575 | | | 17.9 |
| Total | 2005.39 | 779 | | | | |

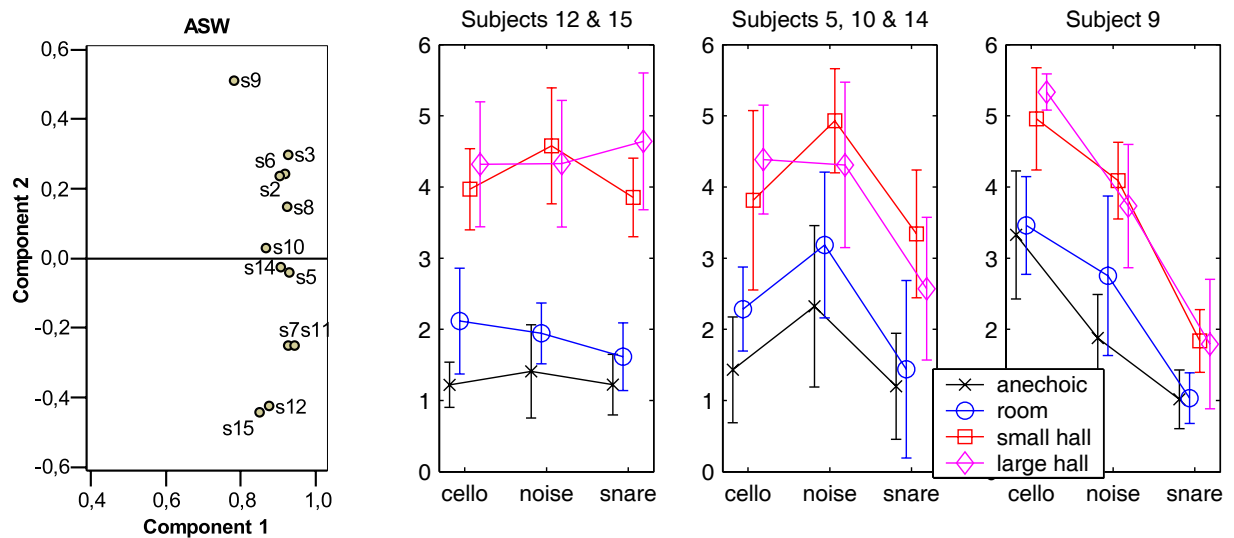Table 5: Experiment I: Analysis of variance of the LEV data.



Fig. 6: Experiment I: Principal component analysis of the ASW subject space (left panel) and illustrations of average judgements at three different levels of component 2.
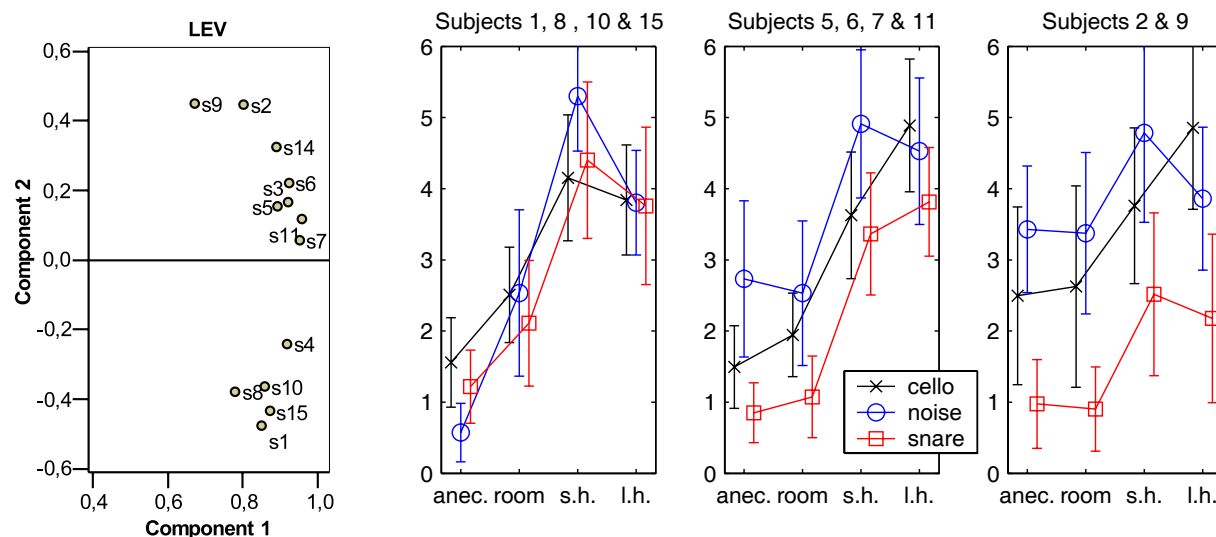
Fig. 7: Experiment I: Principal component analysis of the LEV subject space (left panel) and illustrations of average judgements at three different levels of component 2.

As can be seen from Table 3, the noise stimuli have considerably lower IACCs than the cello in the room and in the small hall, wheras the differences are small in the large hall. This is in line with the overall experimental results, although the effect is relatively small in the averages. However, the data for subjects 5, 10, and 14 in Fig. 6 show clear trends corresponding to the IACCs of the cello and noise stimuli. Also in the other panels of Fig. 6 the order of the cello and the noise is reversed between the small and the large hall. Nevertheless, the large offsets between the stimuli as seen, for instance, for subject 9, as well as the ordering of the anechoic stimuli (which all had IACC = 1 and should thus score the same) are difficult to explain with anything but cognitive processes possibly related to the meaning of the sounds.

The snare drum samples have been obviously judged using different criteria compared to the continuous source signals. It is not surprising that the ASWs of such impulsive stimuli are smaller. However, also the LEV judgements seem to have been affected, although up to smaller extent in the two halls.

### 3.4. Discussion

The changes in ASW due to different sound sources have been earlier investigated by Mason and Rumsey [18] and Usher and Woszczyk [19], although Mason and Rumsey actually studied microphone techniques and Usher and Woszczyk different delays between a pair of loudspeakers. The existence of differences between the source signals contradicts the results of Mason and Rumsey [18] but is in line with the findings of Usher and Woszczyk [19].

The experiment was not easy. The choice of different source signals for a single experiment implies the assumption that ASW and LEV can be unambiguously evaluated despite other differences in the stimuli. As discussed in the previous section, all listeners were not able to do this. However, the experimental data do show similar trends for almost all listeners within a source signal, suggesting that in a more typical experiment involving one source in several modified acoustical environments, the listeners would agree better.

What is more surprising than the individual differences, is the high correlation between the ASW and LEV judgements. The snare drum samples in the reverberant environments seem to have slightly smaller ASWs than LEVs. Furthermore, the source factor accounts for more variance in the judgement of ASW than LEV, which could suggest that (at least part of) the subjects have interpreted ASW

and LEV as traditionally discussed in the literature. The high correlation could be due to one or more of several reasons: (1) ASW and LEV were highly correlated in the chosen (natural) acoustical environments, (2) ASW and LEV are generally highly correlated, and/or (3) the inexperienced subjects were not able to properly differentiate between ASW and LEV, or they have — despite the training — judged, for instance, envelopment by the sound source [10] which would be expected to correlate more with ASW. Since Morimoto and Maekawa [5] and Bradley and Soulodre [6] have been able to better separate the two dimensions in listening tests, reason (2) cannot be the only explanation. Further investigations into this topic are part of our future work.

## 4. EXPERIMENT II

In Experiment I considerable individual differences were observed. A large part of the differences was attributed to different judgements of the source signals. However, informal observations during the training suggested another possible reason for the inter subject disagreement. As mentioned in Sec. 2, the drawings and discussions appeared to converge during the training. Nevertheless, interpretations of the drawings seemed to differ. For instance, the LEV of the anechoic samples was often visualized as a circle inside of or very close to the head, but some subjects still considered them to be highly enveloping due to an even distribution of sound around the head contrary to some other samples. Therefore, it was hypothesized that the translation from the "spatial images" of the stimuli to the scale values could be a source for individual differences. This hypothesis was tested with a graphical assessment of the same stimuli. The same 15 paid subjects that concluded Experiment I participated in the experiment. The same sound proof booth and hardware were used.

### 4.1. Method

The task of the listeners was to visualize the sound source with an arc of a circle and the envelopment with an ellipse, as in the training sessions but this time with a GUI. A screenshot of the GUI is shown in Fig. 8. A fixed head, as seen from the top facing the zero azimuth, was scetched in the center of the drawing area. The arc and the ellipse were adjusted by dragging the shown control points with a mouse. In order to make the task as similar as possible to the training, several degrees of freedom
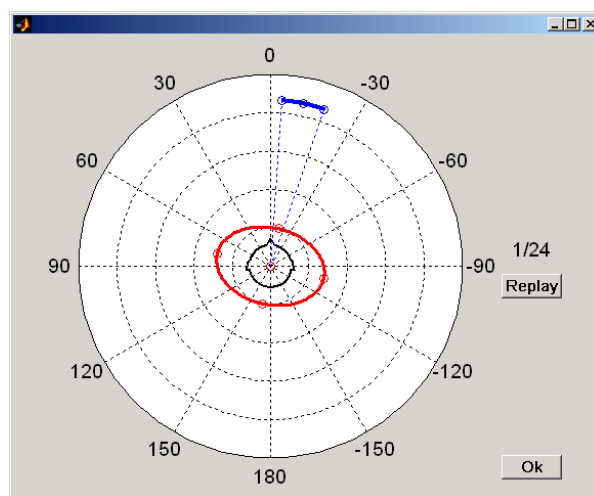


Fig. 8: GUI used in Experiment II.

were allowed. The subjects were able to adjust the width, direction, and distance of the arc, as well as the length of the main axes, rotation, and the center point of the ellipse.

The experiment was conducted directly after the fifth run of Experiment I. The experiment consisted of four consecutive runs, each lasting approximately 5–10 minutes. During each run, the subjects evaluated each stimulus once. A short break was held between Experiment I and the first run, as well as between the second and third runs. The stimuli were presented in a randomized order, and the subjects were able to replay each stimulus as many times as necessary. The GUI was reset to default values after each judgement.

Apart from the method of assessing ASW and LEV, the experimental design differed from Experiment I in two significant ways: The ASW and LEV of each stimulus were assessed simultaneously using slightly different methods, which might have helped in emphasizing the differences between the two attributes. However, as a drawback the subjects were not able to move back and forth between the stimuli and fine tune their judgements based on pairwise comparisons, which might lead to larger error variances.

### 4.2. Stimuli

The same stimuli as in Experiment I were used.

### 4.3. Results

The first run was considered practice and excluded from the analysis. The analysis procedure follows the same guidelines as in Experiment I. However, the data extraction and transformation are a bit more complicated.

Two different measures for ASW were studied: the angular width and the absolute length of the arc. LEV was evaluated as the area of the ellipse regardless of its shape and orientation. The standard deviations of the ASW and LEV measures for each stimulus were (in accordance with Weber's law) approximately proportional to the stimulus means, which suggests a logarithmic transformation. The distance scores, on the other hand, were approximately normally distributed on a linear scale.

The data range used by different subjects for the different measures differed considerably. Since the focus of this study lies in differences between the stimuli, the means and standard deviations of the data of each subject during each run were again scaled to a chosen mean and standard deviation. For overall graphical evaluation and numerical treatment of the distance, the reference means and standard deviations were calculated over all subjects and runs, and the ASW and LEV were scaled back to linear scale. In order to facilitate comparison between the two experiments, the numerical (logarithmic) ASW and LEV data were also scaled to the same the means and standard deviations as in Experiment I.

Fig. 9 shows an overview of the data from all listeners in the form of graphical averages. The averages were formed by filling the ellipses and overlaying all judgements of a stimulus on top of each other. The darker colors describe a higher number of objects at the same position. In addition to equalization of the means and standard deviations of the length of the arc, the distance, and the area of the ellipse, front-back confusions were resolved by mirroring the images relative to the frontal plane whenever the source had been localized behind the listener. The reason for the choice of the length of the arc instead of angular width normalization will become clear in the next subsections.

### 4.3.1. ASW and LEV of the Stimuli

Fig. 10 shows a comparison of the average ASW and LEV results with Experiment I. ANOVA results of the ASW and LEV data from the current experiment are presented in Tables 6, 7, and 8. As explained in the previous section, the data are logarithms of the graphical measures with means and standard deviations equalized to the data from Experiment I. For easier comparison with Experiment I in the presence of individual differences, subjects 1, 4, and 13 were again excluded from the ASW and subjects 12 and 13 from the LEV data, as in Experiment I. The judgements of the excluded listeners are discussed further in Sec. 4.3.4.

Overall, the LEV results are almost identical to Experiment I, suggesting that the area of an ellipse is a valid measure for LEV. Out of the two ASW measures, the length of the arc corresponds better to the earlier ASW judgements. However, the ANOVA reveals that the room factor accounts for less variance and the interaction between subjects and rooms is larger than in Experiment I. This effect is even more pronounced in the ANOVA of the ASW derived from the angular width of the arc, in which case the source signal has larger main effect than the room. Note that the Room*Source interactions are this time not significant for the length of the arc and LEV.

### 4.3.2. Direction and Distance

Before getting into discussion of the differences between the two ASW measures it is helpful to investigate the distance data. Although evaluation of direction or distance was not the purpose of this study, the data are an interesting side product worth analyzing. All listeners apart from subject 13 had used the possibility to adjust the direction and distance of the arc to describe their spatial perception. Subject 13 was thus the only listener excluded from the analysis.

The azimuthal angle of the center of the arc was chosen to represent the source direction. Since the direction judgements are in absolute angles, no normalization was used. However, front-back confusions were resolved by mirroring sources localized behind the listeners to the front. The analysis results are shown in the left panel of Fig. 11. On the average, the azimuthal angles are close to 0. The source signal appears to have affected also the localization, and there were again large individual differences. Part of the source signal dependent displacements might be due to a larger broadening of some sources on one
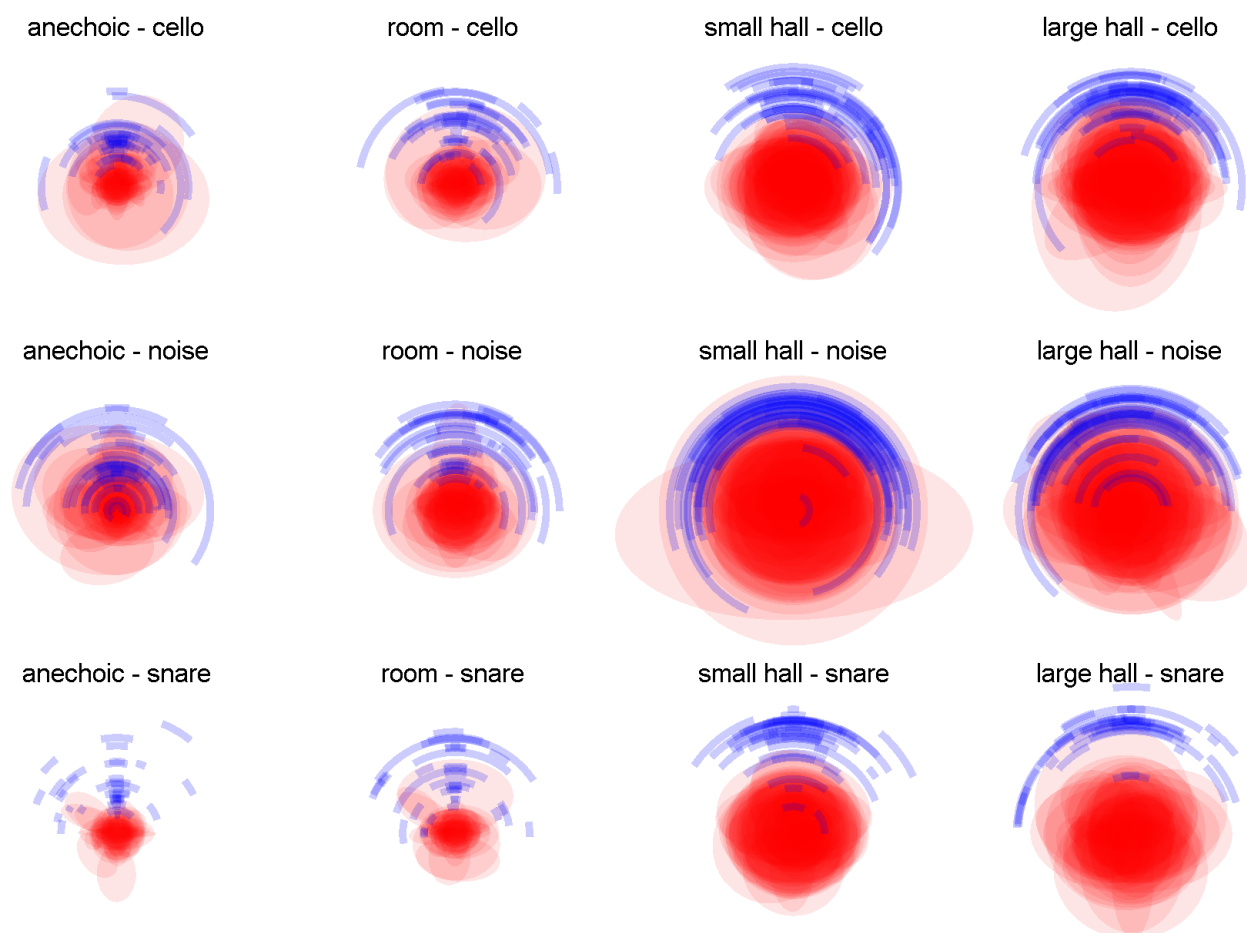
anechoic - cello          room - cello          small hall - cello          large hall - cello

anechoic - noise          room - noise          small hall - noise          large hall - noise

anechoic - snare          room - snare          small hall - snare          large hall - snare

Fig. 9: Graphical averages of the stimuli formed by overlaying all transformed graphical images of a stimulus on top of each other.

| Source | Sum Sq. | d.f. | Mean Sq. | $F$ | $p$ | % of Var. |
|--------|---------|------|----------|-----|-----|-----------|
| Room | 268.41 | 3 | 89.471 | 146.38 | 0.000 | 29.1 |
| Source | 208.62 | 2 | 104.308 | 170.65 | 0.000 | 22.6 |
| Room*Source | 7.17 | 6 | 1.194 | 1.95 | 0.072 | 0.8 |
| Room*Subject | 99.84 | 33 | 3.026 | 4.95 | 0.000 | 10.8 |
| Source*Subject | 88.72 | 22 | 4.033 | 6.60 | 0.000 | 9.6 |
| Room*Source*Subject | 72.60 | 66 | 1.100 | 1.80 | 0.001 | 7.9 |
| Error | 176.04 | 288 | 0.611 | | | 19.1 |
| Total | 921.40 | 431 | | | | |

Table 6: Experiment II: Analysis of variance of the ASW data extracted from the length of the arc.
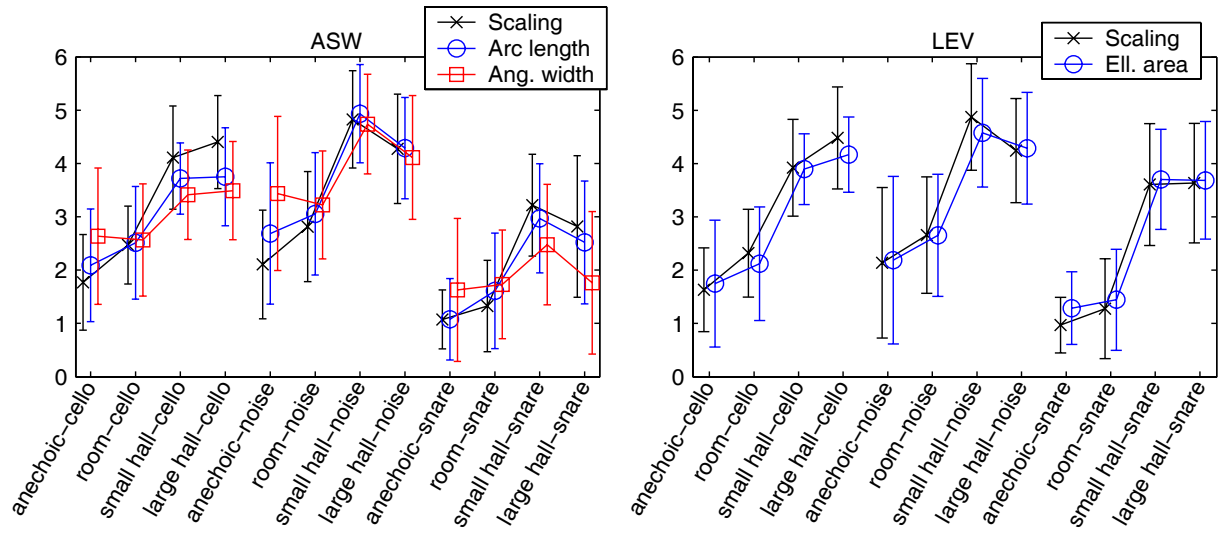
Fig. 10: Comparison of means and standard deviations of ASW (left panel) and LEV (right panel) judgements from Experiment I and Experiment II.

| Source | Sum Sq. | d.f. | Mean Sq. | $F$ | $p$ | % of Var. |
|---|---|---|---|---|---|---|
| Room | 78.26 | 3 | 26.086 | 33.79 | 0.000 | 8.5 |
| Source | 283.66 | 2 | 141.830 | 183.72 | 0.000 | 30.8 |
| Room*Source | 15.56 | 6 | 2.593 | 3.36 | 0.003 | 1.7 |
| Room*Subject | 171.80 | 33 | 5.206 | 6.74 | 0.000 | 18.6 |
| Source*Subject | 78.86 | 22 | 3.584 | 4.64 | 0.000 | 8.6 |
| Room*Source*Subject | 70.92 | 66 | 1.075 | 1.39 | 0.035 | 7.7 |
| Error | 222.34 | 288 | 0.772 | | | 24.1 |
| Total | 921.40 | 431 | | | | |

Table 7: Experiment II: Analysis of variance of the ASW data extracted from the angular width of the arc.

| Source | Sum Sq. | d.f. | Mean Sq. | $F$ | $p$ | % of Var. |
|---|---|---|---|---|---|---|
| Room | 545.65 | 3 | 181.885 | 280.53 | 0.000 | 49.4 |
| Source | 62.81 | 2 | 31.405 | 48.44 | 0.000 | 5.7 |
| Room*Source | 6.18 | 6 | 1.031 | 1.59 | 0.150 | 0.6 |
| Room*Subject | 84.39 | 36 | 2.344 | 3.62 | 0.000 | 7.6 |
| Source*Subject | 120.36 | 24 | 5.015 | 7.73 | 0.000 | 10.9 |
| Room*Source*Subject | 82.51 | 72 | 1.146 | 1.77 | 0.000 | 7.5 |
| Error | 202.29 | 312 | 0.648 | | | 18.3 |
| Total | 1104.19 | 467 | | | | |

Table 8: Experiment II: Analysis of variance of the LEV data extracted from the area of the ellipse.
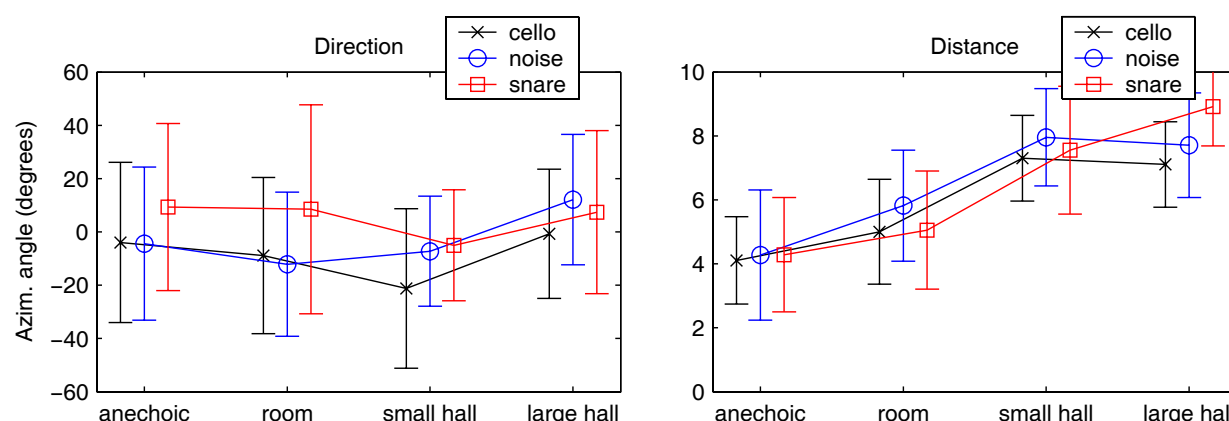
Fig. 11: Experiment II: Means and standard deviations of perceived directions (left panel) and distances (right panel) of the sound sources averaged over 14 subjects.

side. However, in Fig. 9 it can be seen that some subjects have also localized certain sources directly on one side, and the large scattering of the directions of the anechoic stimuli is quite surprising.

For distance analysis, the radius of the arc (the distance of the center point of the arc from the center of the head) was used. As described earlier, the means and standard deviations of the (linear) distance scores of each subject during each run were scaled to the means and standard deviations calculated over all subjects and runs. The results are shown in the right panel of Fig. 11, where a value of 10 corresponds to the largest possible distance. The radius of the head was 1.1. The distance appears to be mainly determined by the room, and the distances are in correct order compared to the real distances (Table 2), although they are considerably offset from 0. Some of the individual differences are again visible in Fig. 9.

### 4.3.3. Comparison of the ASW Measures
So far in the description of Experiment II, the two ASW measures have been described in parallel. It was originally expected that the angular width would be a better measure. However, the ANOVA results in Sec. 4.3.1 already showed that the length of the arc has smaller error variances and interactions with the subject factor. For a more rigorous argument, we have to once again have a look at the data of the individual listeners.

The transformed ASW data of the individual listeners derived from the length and the angular width of the arc are shown in Figs. 12 and 13, respectively. The length of the arc before the normalization equals the angular width (in radians) multiplied by the distance. Thus, the data interact with the perceived distance. The interaction becomes especially important for sources very close to or inside the head. Indeed, it can be seen from Fig. 13 that using the angular width exaggerates the ASWs of the anechoic sources. In some cases, such as for subject 11, it even reverses the order of several stimuli. On the other hand, the difference in the measures for the small and large hall is small, since the ratio of their perceived distances (as shown in Fig. 11) is small.

It is obvious that, for instance, a small circle in the middle of the head should not correspond to the widest source. Therefore, in the presence of the anechoic stimuli, the length of the arc is a better overall descriptor of ASW. For the other stimuli, the measures characterize slightly different features and it is difficult to conclude which one should be preferred (see also discussion on egocentric perspective in [12]). However, for all listeners the lengths of the arcs correlate better with the scaling judgements of Experiment I. The difference is smaller excluding the anechoic samples, but the average correlation coefficient is still 0.82 vs. 0.73 for the angular width. Thus, it can be concluded that when asked for a single scale value, the listeners did more often answer
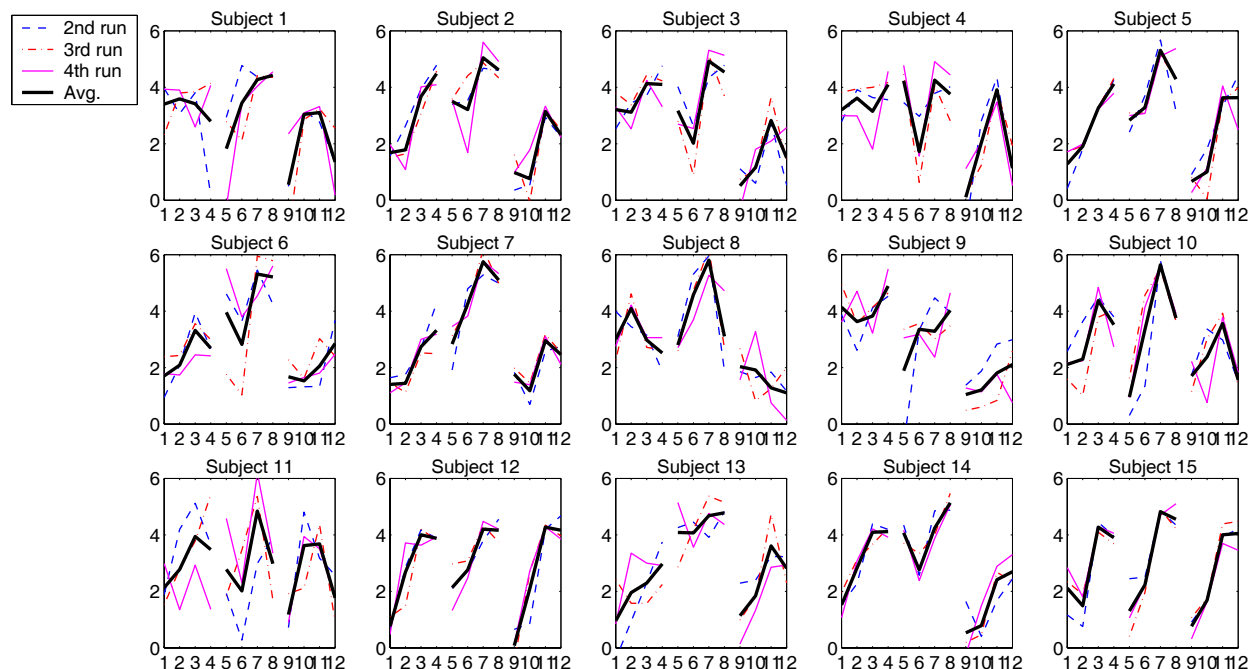
Fig. 12: Experiment II: Transformed ASW data derived from the length of the arc. The stimuli on the x-axes are as listed in Table 1.
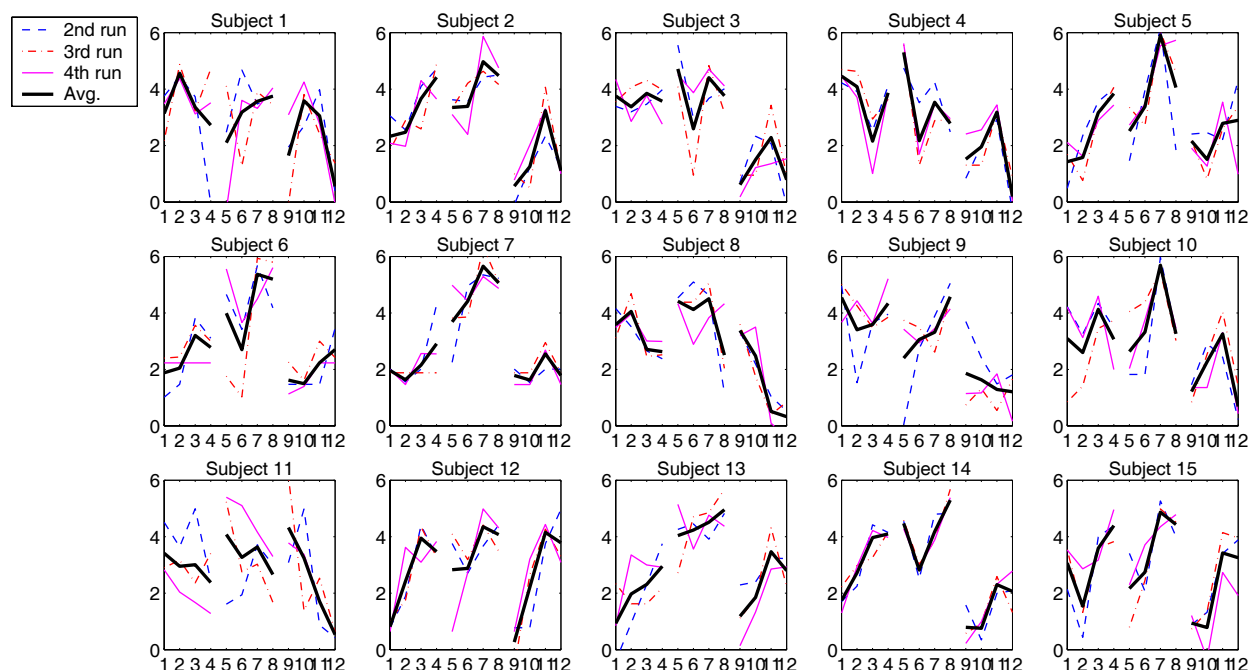


Fig. 13: Experiment II: Transformed ASW data derived from the angular width of the arc. The stimuli on the x-axes are as listed in Table 1.

according to the visualized length of the arc, thus taking the perceived distance into account.

### 4.3.4. Individual Differences in ASW and LEV judgements

Also on the level of individual subjects, the LEV data of the Experiment II are very similar to Experiment I, and the same conclusions apply apart from the inconsistency of subject 12, which was earlier due to learning. Using the graphical method, the subjects excluded from the ASW analysis in Experiment I seemed to produce somewhat more consistent and meaningful results, although they still had large variances. Principal components analysis also revealed similar structures as in Experiment I for both ASW (length of the arc) and LEV, although the positions of some subjects were slightly shifted.

### 4.4. Discussion

Comparison of the results of the two experiments shows that on the average the listeners have translated the visual sound images in the same way to the unidimensional ASW and LEV values. Since the results of Experiments I and II were so similar and both took approximately same time to conduct, it is impossible to make statements about the superiority of either method. For an extensive discussion on the differences in the methodologies, see [12]. Nevertheless, it should be noted that the graphical evaluation is more intuitive and in uncertain cases it may help in assuring that the listeners are concentrating on the correct attributes. The feedback from the listeners was also positive. Several subjects considered it easier to draw the perception than to assign scale values to ASW and LEV.

### 5. SUMMARY AND CONCLUSIONS

In this paper, training of listeners for assessment of two common attributes of spatial impression, auditory source width (ASW) and listener envelopment (LEV), was described. The training consisted of group discussions and visualization of the stimuli by drawings. In Experiment I it was found that the first training session was sufficient to explain the concepts to the listeners. The analysis of the data revealed that most subjects had formed consistent criteria for evaluating ASW and LEV. However, both attributes were highly correlated.

The experiment included all combinations of three anechoic source signals and four acoustical environ-

ments as stimuli. The environments were found to have most effect on the judgements, although they did not appear in the order of interaural cross-correlations (IACCs) calculated from binaural room impulse responses (BRIRs). Instead, the results were also affected by the size of the acoustical environment. Furthermore, considerable individual differences did exist. A large part of the individual differences were attributed to different judgements between the source signals.

It was also hypothesized that the direct scaling method could have been the reason some individual differences, but this hypothesis was shown wrong in Experiment II. Both direct scaling and graphical assessment of the stimuli produced comparable results. The obtained ASW scale values corresponded to the length of an arc describing the sound source, and the LEV scale values to the area of an ellipse describing the envelopment.

### 6. ACKNOWLEDGEMENTS

### 7. REFERENCES

[1] A. H. Marshall. A note on the importance of room cross-section in concert halls. *J. Sound and Vibration*, 5(1):100–112, 1967.

[2] M. Barron. The subjective effects of first reflections in concert halls — the need for lateral reflections. *J. Sound and Vibration*, 15(4):475–494, 1971.

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

[4] J. Blauert and W. Lindemann. Auditory spaciousness: Some further psychoacoustic analyses. *J. Acoust. Soc. Am.*, 80(2):533–542, 1986.

[5] M. Morimoto and Z. Maekawa. Auditory spaciousness and envelopment. In *Proc. 13th International Congress on Acoustics*, volume 2, pages 215–218, Belgrade, Yugoslavia, 1989.

[6] J. S. Bradley and G. A. Soulodre. The influence of late arriving energy on spatial impression. *J. Acoust. Soc. Am.*, 97(4):2263–2271, 1995.

[7] J. Berg and F. Rumsey. Identification of perceived spatial attributes of recordings by repertory grid technique and other methods. In *AES 106th Convention*, Munich, Germany, 1999. Preprint 4924.

[8] J. Berg and F. Rumsey. In search of the spatial dimensions of reproduced sound: Verbal protocol analysis and cluster analysis of scaled verbal descriptors. In *AES 108th Convention*, Paris, France, 2000. Preprint 5139.

[9] K. Koivuniemi and N. Zacharov. Unravelling the perception of spatial sound reproduction: Language development, verbal protocol analysis and listener training. In *AES 111th Convention*, New York, NY, USA, 2001. Preprint 5424.

[10] F. Rumsey. Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm. *J. Audio Eng. Soc.*, 50(9):651–666, 2002.

[11] T. Neher, F. Rumsey, and T. Brookes. Training of listeners for the evaluation of spatial sound reproduction. In *AES 112th Convention*, Munich, Germany, 2002. Preprint 5584.

[12] R. Mason, N. Ford, F. Rumsey, and B. de Bruyn. Verbal and nonverbal elicitation techniques in the subjective assessment of spatial sound reproduction. *J. Audio Eng. Soc.*, 49(5):366–384, 2001.

[13] V. Hansen and G. Munch. Making recordings for simulation tests in the Archimedes project. *J. Audio Eng. Soc.*, 39(10):768–774, 1991.

[14] H. Hudde and J. Schröter. Verbesserungen am Neumann-Kunstkopfsystem. *Rundfunktechnische Mitteilungen*, 1:1–6, 1981.

[15] ISO 3382. Acoustics — measurement of the reverberation time of rooms with reference to other acoustical parameters. International Organization for Standardization, 1997.

[16] H. Levitt. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.*, 49(2):467–477, 1970.

[17] ITU-R. Recommendation BS.1284-1, General methods for the subjective assessment of sound quality. International Telecommunication Union Radiocomminication Assembly, 2003.

[18] R. Mason and F. Rumsey. A comparison of objective measurements for predicting selected subjective spatial attributes. In *AES 112th Convention*, Munich, Germany, 2002. Preprint 5591.

[19] J. Usher and W. Woszczyk. Design and testing of a graphical mapping tool for analyzing spatial audio scenes. In *AES 24th Conference*, Banff, Canada, 2003.