Chapter 7

User Testing

User studies can evaluate and validate the performance of a sound spatialization system, including a sound spatialization backend and a procedure for sound spatialization resource management. Since the developed resource management is based on human perception, an evaluation using objective tests would not measure its effectiveness. Testing system performance using subjective tests is not easy, and difficult to reproduce because the test conditions are difficult to control. Auditory experiments and results are comprehensively described in [Blauert, 1996]. Subjects' abilities to localize sound vary across subjects, experimental conditions and tasks. Localization ability depends on the stimulus surrounding sounds and room features. Performance tests can only be done for specific tasks of certain applications.

For the perceptual space discussed in Chapter 3, which is used for the decision process of the resource assignment, user capabilities must be estimated, confirmed and then generalized. More important than system evaluation is system calibration. This is the process of tuning all system parts to give best performance to the user. Elegant resource management also takes into account the performance capabilities of the spatialization backends.

7.1 Task-based performance tests for sound spatialization backends

User performance is measured for the same applications using different spatialization backends. Besides comparing the performance data for backends, it can be put into relation to user performance in a "natural" environment. Simple tasks can be:

- estimating sound direction

- comparing object distance

Monitoring of user performance is typically done using forced-choice selection or applying a tracking device.

7.2 Evaluation criteria used by subjects

Absolute or physical judgments are not appropriate for experimental responses. The questions must be relative and oriented to subjective dimensions of the sound space [Kendall and Martens, 1984, p. 121]:

Relative direction (azimuth and elevation)

Relative distance (range)

Definition clarity and impression of the size of a sound source

Spaciousness room characteristics, i.e., liveness, size, shape, etc.

Spatial texture changes of the sound perception itself through the environment (i.e., room, other objects like occluders)

7.3 Taxonomy of psychoacoustic validations

7.3.1 Comparing sound spatialization backends to reference recordings

- Binaural recording of a defined stimulus through the ears of the subject or dummy head
- Simultaneous recording (i.e., pick up) of a stimulus as a monaural anechoic reference signal
- Binaural spatialization of a stimulus (using a spatialization backend)
- Alternating presentation to the subject of both binaural versions, using an equalized headphone
- Questioning the subject regarding the difference between the stimuli

7.3.2 Comparing sound spatialization backends to a reference impulse response

For psychoacoustic verification of computer models for binaural room simulation [Pompetzki, 1993], the following procedure was employed:

- Measurement of a reference impulse response using a dummy head
- Convolution of impulse response with a monaural anechoic reference signal
- Binaural spatialization of stimulus (using a spatialization backend)
- Alternating presentation to the subject of both binaural versions, using an equalized headphone (ABA order)
- Questioning the subject regarding the difference between the stimuli

7.3.3 Direct comparison between sound spatialization backends

Objective direct comparison is not an easy task because many parameters are involved which can distort the results of such a test. The output devices of the backends might be different and needs to be harmonized. In case of a loudspeaker system the room influence the result through its reverberation.

- Binaural recording of a defined stimulus through the ears of the subject or dummy head
- Simultaneous recording of the stimulus as monaural anechoic reference signal
- Binaural spatialization of the stimulus using the spatialization backends
- Alternating presentation to the subject of both spatialization backends using equalized headphone to the subject
- Questioning the subject regarding the difference between the stimuli

Figure 7.1 shows a schematic of a test for comparing two different spatialization backends. In practice, such a test can be simplified by recording the produced spatialization.

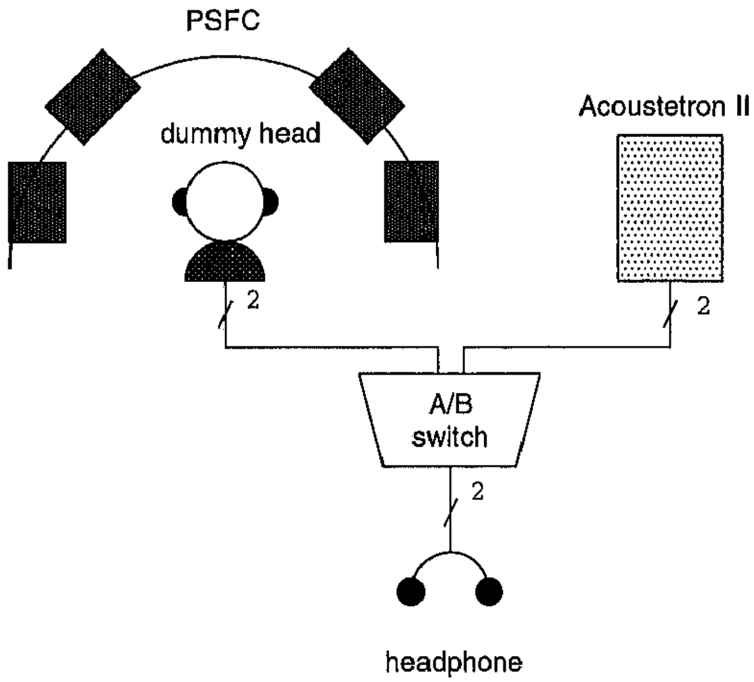

Figure 7.1: A/B test for spatialization backends via dummy head

7.4 Psychoacoustic evaluation of the clustering algorithm

The validity of the presented clustering algorithm in Chapter 4 can be demonstrated for specific configurations. The test falls into the category of the previous section, which directly compares two spatialization backends. Here the same spatialization backend was used, but spatialization was done with and without clustering enabled.

7.4.1 Method

A scene with three sound sources was prepared. A timed script activated the sound sources using MIDI commands. In this evaluation study, the spatialization backend Acoustetron II (see Section 6.1.2) was used. As sound generator, 4 MIDI synthesizer Roland SoundCanvas SC-55mkIIs were used. The sounds were a bird (instrument 124), a telephone (instrument 125), and a gun shot (instrument 128). The sources were triggered with 500 ms delay in between so that the onsets did not overlap. Two of the MIDI synthesizer produced the identical signal for reverberation, which was passed via a mixer to a reverberator Yamaha REV 500 as monaural signal. The reverberant signal mixed with directionalized sound from the spatialization backend. The configuration of the reverberator was setup to simulate a medium sized room. (Parameters were Effect only, 24 ms predelay, 1 s reverb time high-ratio 0.4, and ER level 100.) The reverberator improved externalization [Begault, 1994, p. 97] (see also p. 33 in this thesis) and was used to produce ambient sound for sound sources which could not be spatialized.

stimuli	spatialization channels	number of clusters	ambient sound sources
no restrictions	3	0	0
clustering	2	1	0
ambient	2	0	1

Table 7.1: Stimuli use of spatialization resources

Listening was done using headphones in an anechoic chamber. Five listeners participated. One trial consisted out of a ABA sequence. A stimuli was either three sound source processed with three spatialization channels (N), processed using the developed clustering algorithm (C), or processed using two spatialization channels and one sound source was presented ambient (A).

stimuli	label	x	y	sound
no restrictions	source 1	7.87	-7.87	gun shot
	source 2	7.87	7.87	phone ringing
	source 3	7.87	11.81	bird call
clustering	source 1	7.87	-7.87	gun shot
	cluster 1	7.87	9.84	bird call and phone ringing
ambient	source 1	7.87	-7.87	gun shot
	source 2	7.87	7.87	phone ringing
	ambient	-	-	bird call

Table 7.2: Stimuli source description (using the coordinate system of the CRE API)

This is summarized in Table 7.2 (coordinates are represented using the CRE API [CRE, 1994]). The total number of trials was 72; each stimulus combination was presented 8 times. The nine trial combinations are listed in Table 7.3. The listeners were asked to rate the dissimilarity of the spatial imagery. They marked "1" when the spatial images were judged equal and "5" for the largest difference. The stimuli combinations with itself were included to check if users response is randomly. Also the ratings for a stimuli pair in different order should give similar ratings.

stimulus	abbreviation	first	second
1	NCN	non-restricted	clustered
2	CNC	clustered	non-restricted
3	NAN	non-restricted	ambient
4	ANA	ambient	non-restricted
5	CAC	clustered	ambient
6	ACA	ambient	clustered
7	NNN	non-restricted	non-restricted
8	CCC	clustered	clustered
9	AAA	ambient	ambient

Table 7.3: Trial combinations

7.4.2 Results and discussion

The averaged rating for all lister for each stimulus is shown in Table 7.4. The mean ratings on the diagonal show that not always the listener detected that the same stimulus was presented three times.

		second		
		N	C	A
first	N	1.075	2.225	4.500
	C	2.175	1.000	4.375
	A	4.375	4.525	1.500

Table 7.4: Dissimilarity between intervals: non-restricted (N), clustered (C), and ambient (A)

Average dissimilarity rating regardless the order of listening is shown in Figure 7.2. The average dissimilarity for clustered and ambient processing (CA) was 4.45, for non-restricted to ambient processing (NA) was 4.4375, and for non-restricted to clustered processing was 2.2. The averaged dissimilarly for comparison of all three stimuli with itself was 1.1917.

Sound spatialization with no restrictions in the number of spatialization channels or using clustering for (based on the specific configuration) were rated very dissimilar to the processing with one ambient sound source. The dissimilarity judgments between processing with no restrictions and clustering were ranked half compared to the ratings for ambient processing with the others.

Assuming spatialization resource limitations and specific configuration, processing using clustering improves the spatial imagery. Clustering, as implemented, did not give the same spatial image to processing without limitations.

Bibliography

[Begault, 1994] Durand R. Begault. *3-D Sound for Virtual Reality and Multimedia.* Academic Press, 1994. ISBN 0-12-084735-3.

[Blauert, 1996] Jens Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press, revised edition, 1996. ISBN 0-262-02413-6.

[CRE, 1994] Crystal River Engineering, Inc. *CRE_TRON Library Reference Manual,* August 1994. Revision B.

[Kendall and Martens, 1984] Gary S. Kendall and William L. Martens. Simulating the cues of spatial hearing in natural environments. In *ICMC: Proc.*

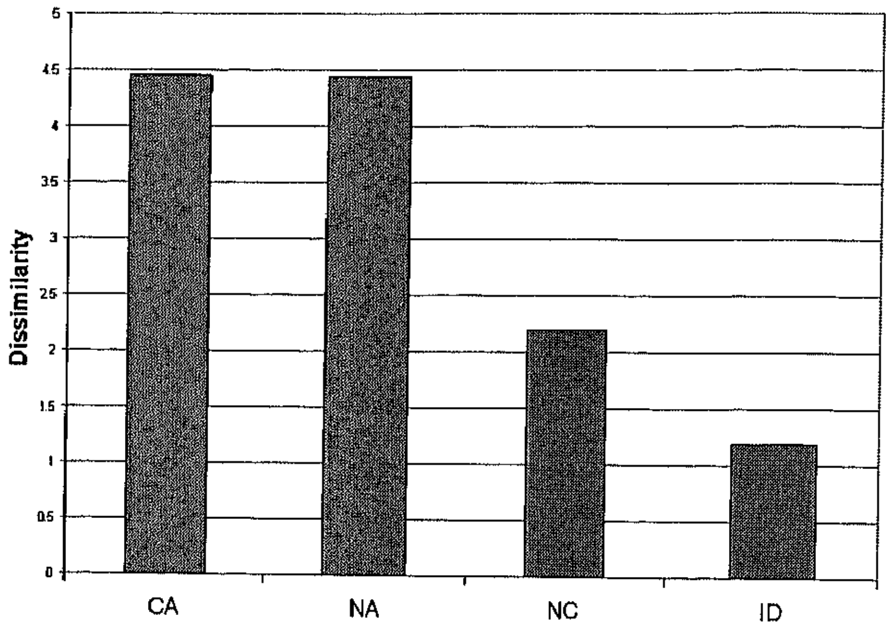

Figure 7.2: Dissimilarity for non-restricted (N), clustered (C), and ambient (A) processing

Intnl. Comp. Music Conf., pages 111–126, Paris, 1984. Computer Music Association.

[Pompetzki, 1993] Wulf Pompetzki. *Psychoakustische Verifikation von Computermodellen zur binauralen Raumsimulation*. PhD thesis, Ruhr-Universität Bochum, 1993. (In German).