Russell Mason and Francis Rumsey , *University of Surrey, Guildford, England*

Presented at
the 106th Convention
1999 May 8–11
Munich, Germany

AN AUDIO ENGINEERING SOCIETY PREPRINT

An investigation of microphone techniques for ambient sound in surround sound systems

Russell Mason and Francis Rumsey, Institute of Sound Recording,
School of Performing Arts, University of Surrey,
Guildford, Surrey, GU2 5XH, UK
`r.mason@surrey.ac.uk`
`f.rumsey@surrey.ac.uk`

Abstract

A controlled subjective test was carried out to assess selected qualities of three ambient microphone techniques for surround sound. The effects of signal delay and microphone distance were explored. The tests indicate that the perceived results are programme dependent, but that a compromise can be found using delayed close microphones, giving similar quality for the range of programme items used.

0   Introduction

This paper is part of the work carried out under the Eureka 1653 Medusa (Multichannel Enhancement of Domestic User Stereo Applications) project. The Medusa projects involves collaborative research between the following partners: the British Broadcasting Corporation, the Music Department of the University of Surrey, Nokia Research Centre, Genelec Oy, and Bang & Olufsen A/S.

The purpose of the project is to examine the variables of the domestic multichannel sound system, with and without picture, to carry out the essential optimisation leading to consumer end products. These products will combine the requirements of multichannel reproduction together with the less complex modes of reproduction, such as stereo and mono. This involves linked studies of programme production and perceptual elements, leading to a single optimised approach to domestic reproduction.

With the advent of new discrete multichannel delivery formats, there is a need for the creation of discrete multichannel programme material. Since the invention of 2-channel stereophony, much experimentation has taken place into optimum microphone techniques, especially in the area of recording classical music. These techniques now need to be reassessed for application in multi-channel surround sound.

There has already been a large amount of experimentation by recording engineers into possible solutions for multichannel recording. However, this has been mostly informal listening: there has been little in the way of structured tests which examine fundamental aspects of recording techniques.

At the 103rd AES Convention in New York, Messrs A. Fukada, K. Tsujimoto and S. Akita of the NHK Broadcast Centre, Japan, presented a paper on 'Microphone Techniques for Ambient Sound on a Music Recording' [1]. They found that if omnidirectional microphones were used at a distance to capture the ambient sound for the rear channels, the sound was inferior to using rear-pointing cardioid microphones quite close to the main microphone array.

The main difference between these two techniques, distance, causes two effects: added delay and a change in the direct to reverberant ratio of the sound. The effect of these two parameters on the subjective evaluation of the sound has not yet been fully explored. Because of this, one can only hypothesise about why one may be preferred over the other.

The present authors carefully considered the reasons for preferring the closer microphone technique. This led to a theory that this technique may contain more direct sound, which will then be reproduced from the rear speakers, resulting in a 'fuller' sound. However, this may lead to a problem of confusing front/back localisation due to the direct sound being perceived both front and back.

If the optimum ambient microphone technique should use rear pointing cardioid microphones close to the main array as indicated by Fukada et al [1], then one must consider the problems that may arise which are unique to this approach. If one solely concentrates on a conventional 'concert' layout in which the musical sources are located on a stage in front of the listener, then the main issue is that the direct sound from the front may leak into the microphones dedicated to the rear channels. This is partly due to the imperfect cardioid response of microphones.

The simplest method of reducing the annoyance of this acoustic crosstalk is to attenuate the channels feeding the rear speakers. However, this is a less than adequate solution, as the level of the desired ambient information will also be reduced. A more favourable technique would be to make use of the perceived suppression caused by the precedence effect.

Haas quantified the extent of the suppression effect by asking subjects to raise the gain of a delayed source until it appeared to be at the same level as a primary source [2]. He used both speech and band-limited white noise as the signal source. From this, he found that the suppression rose rapidly from 0 dB at a delay time of 0 ms to a maximum of between 8 and 12 dB at approximately 7ms. This was then constant up to a delay period of approximately 25ms where the suppression appeared to gradually roll off.

Blauert also researched the effect of delay on the localisation of sounds in the median plane [3]. He found that with a single speaker in front of the listener and a single speaker behind both emitting an identical signal, the sound source was perceived to be located at the rear. This he attributed to the spectral effects of comb filtering. When a gradually increasing delay was added, the changes initially resulted in confusion for the listener. Soon after, the suppression effect of precedence stabilised the location to be either front or back, depending on which was the leading loudspeaker.

From this it is clear that delay can be used as a psychoacoustic tool to lessen the perceived effect of any acoustic crosstalk that may arise. However, the questions that remain are whether this artificial delay will have a detrimental effect on the overall quality of the sound reproduction, and how this compares with distant ambient microphones that have an inherent delay. Subjective tests were carried out in order to clarify this.

1   Listening Test Material

To test the microphone techniques, it was necessary to record musical extracts in an organised and well-documented way, using two separate microphone techniques for the ambient sound. These were:
- a pair of rear-pointing cardioid microphones quite close to the main microphone array
- a pair of omnidirectional microphones at a distance

The close microphones could then have a delay added in order to make use of the suppression of the precedence effect as described above.

1.1   Recording of material

The authors proposed to keep an identical technique for all the excerpts to be used for listening tests, as this would keep extraneous variables to a minimum. This meant that the technique had to be versatile enough to cope with the wide range of situations.

There were a number of coincident or near-coincident techniques available that could be used. The microphone array which was used consisted of two coincident ORTF pairs (cardioids 17cm apart, 110° angle), one facing forwards and the other backwards. This choice was based on the research of Ceoen [4], whose results showed that the ORTF technique was a good compromise between the factors of stereo image, timbre, liveness, intimacy and dynamic range. This was supplemented by a centre cardioid microphone facing forwards. It is recognised that the addition of the centre channel may have caused a narrowing of the stereo image. However, as this was the same for all the stimuli, it should have minimal effect on the experimental results. The distant ambient microphone technique was chosen to be a pair of near coincident (20 centimetres apart) omnidirectional microphones as used by Fukada et al [1].

The microphones used for the main array were AKG C-414 B-ULS. Each channel was recorded discretely onto a Tascam DA-88 via DDA microphone pre-amps. The ambience microphones were Schoeps CMC-5U omnis. All of the microphone channels in the main array were deliberately set to give the same output level from a reference acoustical tone generator held at the grill of each microphone. This may not be optimum in terms of not being able to 'tweak' the levels for a better subjective result. However, it gives a firm repeatable standard which is based on keeping the signal path as natural and unaltered as possible. The distant ambient microphones were set to give the same output level as the close rear pointing microphones for the intended sound source at the performance location.

For all of the recordings the main array was placed at approximately the critical distance of the acoustical environment. The ambient microphones were then placed 10 metres behind the main array. This is shown in figure 1.

1.2   Choice of material

The recording opportunities available included a wide range of styles. Ideally, the musical extracts would have included at least five different programme sections including a full symphony orchestra playing fortissimo, other combinations of small numbers of instruments, and a male voice talking in an anechoic chamber [5]. For this experiment the stimulus of a voice in an anechoic environment was unsuitable, therefore this was altered to be a speaking voice in a reverberant acoustic. Due to limited time and opportunities, the chance to record some of the types of musical extracts was not possible.

In the end, a good compromise was reached with the following:
1. A church organ played fortissimo. This has the full range of frequencies and sound intensities in a very reverberant environment that gives the listener a better chance of judging the ambient field.
2. A string quartet concert in a classical recording studio. This is a less reverberant environment, but with ambient noise created by an audience quite close to the performers.
3. A medium-sized choir with organ accompaniment. This is again a very reverberant environment with a lot of ambience that can be judged.
4. A male speaking voice recorded in a reverberant environment. This extract is particularly important due to the sensitivity of human perception to reflection and delays on speech. The speaker was deliberately chosen to be a voice that all the subjects could easily recognise so they had a mental reference of the sound in real life.

2 Experimental design

It was important to design the listening tests as carefully as possible to obtain data which was 'repeatable, a measure of the prespecified aspect of the perceived sound, and reported on an appropriate scale' [6]. This is essential in the quest for an accurate result without bias. Every attempt was made to abide by the relevant standards of IEC 268-13 [5] and ITU-R BS 1116 [7].

Each musical extract was played as each of three versions. Each version contained one type of stimulus delivered through the rear loudspeakers. This was either:
- the unprocessed rear facing microphones of the array (R);
- the delayed rear facing microphones of the array (RD);
- the ambient microphones that were placed at a distance (A).

These were combined into pairs for comparison, with each variation of the musical excerpt being compared with every other version of the same musical extract. The subject could freely switch between the two versions presented to the rear loudspeakers. In each case the 3 channels feeding the front speakers were fixed.

There was limited time available to carry out the listening tests, so each subject was given a 30-minute slot. This was enough to present the subject with 90 seconds of programme at a time, with a 30-second gap for thinking, writing and preparation for the next excerpt. Therefore it was possible to fit 14 pairs of auditory events in the time with a two-minute gap for finishing and changing to the next listener.

The auditory events occurred in a pseudo-random order. They were chosen carefully not to repeat, and to space out the stimuli. Within each pair, the versions were randomly allocated to either source A or B to create a balance. This is to counteract the influence of fatigue, practice, or other variations that may occur over time. Unfortunately the equipment needed to provide each listener with his or her own random order was not available. Table 1 shows the order of the extracts.

It is essential to test the listener's ability to judge the material reliably and repeatedly. Therefore it was necessary for some of the pairs of auditory events to be heard another time, adding further opportunity for checking the reliability of judgements. Table 1 shows the order of the extracts.

3 Experimental details

3.1 Equipment

The material was compiled using a Sony OXF-R3 digital console. This enabled the audio to be kept entirely in the digital domain, with all levels set precisely at unity gain, and the delays set to sample accuracy. The period of the delay was set at 29.1ms. This is equivalent to the 10 metres distance of the ambient microphones based on the speed of sound in dry air at 20°C of 343 ms^{-1}.

The tests were undertaken in Studio 2 of the PATS building at the University of Surrey. The dimensions of the room are 6.7 metres by 8.5 metres with a height of 4.5 metres. This is within the specification of ITU-R BS 1116 and close to that of IEC 268-13. The reverberation time of this room is specified as being approximately 0.4 seconds at the low frequencies (63-250 Hz), falling to 0.32 seconds at 4000 Hz. The noise floor is very low, with a noise criterion (NC) of 20. The reverberation time and background noise levels are well within the criteria of IEC 268-13 though slightly outside ITU-R BS 1116. The air conditioning in the room keeps the ambient temperature, humidity and air pressure within the specified limits.

The tests were played back from a Tascam DA-88 DTRS multitrack recorder into a Neve V-series console so that the levels could be set as precisely as needed. The outputs from the console fed Quad 405 amplifiers that were powering five matched Rogers LS 3/5 loudspeakers. There was a switching box in the circuit of the rear speakers for the auditioning of the different versions.

The loudspeakers were placed approximately 1.5 metres from the listener, pointing towards the listener, and in the configuration stated in ITU-R BS 1116. They were at ear height for the average seated listener and positioned so that the listener was at the centre of the room (within the limits of IEC 268-13). One subject at a time undertook the test.

The sound intensity of the loudspeakers was set from tone on the tape, using a B&K SPL meter. In retrospect, noise would have been a more suitable signal due to its reduced position and frequency dependency. However, within acceptable limits this ensured, along with the level testing at earlier stages, that all the channels maintained a unity gain with respect to each other through all the stages of the recording, editing and replay process. The aim was for all of the sources to be played back at a natural level, and that the voice was at a level where there was no detectable difference between this and the person talking in the room, as was suggested by Bech [8]. This equated to a level from each loudspeaker at the central listening position of 88dBA from a signal at –4dBFS.

3.2   Test procedure

The listening tests were of a blind A - B paired comparison type. The listeners were asked to rate the auditory events according to two criteria: stereo image and spaciousness. The criteria were carefully chosen to reflect the research carried out earlier and to provide the listeners with a point of reference. They were defined as below.

Stereo image - How easy is it to localise the sources presented? Does the image fill the frontal soundstage? How stable is the image? How natural is the image? How close is it to being at the concert?
Spaciousness - How enveloping is the sound? How natural is the ambience? How deep is the image? How close is it to being in the venue?

The scales used are limited in the fact that they are multidimensional attributes, and as such it is possible for listeners to interpret them differently. Whilst this may be a problem, they are broad enough to cover the scope of the investigation in a form which is simple enough to limit errors in judgement. For a further discussion of this, the reader is referred to [9].

The ratings were judged as to how closely they were 'true-to-nature'. The scale used followed the recommendation of IEC 268-13 (shown in Appendix A). A rating of 10 means that the auditory event could not possibly be any better. At the other end of the scale, a rating of 0 means that the auditory event could not possibly be any worse. The listener was also asked to express a simple preference between the two auditory events.

3.3   Subjects

The listeners were all unpaid volunteers from the Tonmeister Music and Sound Recording course at the University of Surrey. There was a mixture of first, second and final years. This means that there was a mixture of relatively inexperienced and experienced listeners. The listeners were all given identical instructions (see Appendix A). These were based on the research of Bech [8].

The instructions were deliberately set out to prepare the listener about what aspects of the sound to listen for and grade. There was a careful avoidance of any indication of the type of processing which had taken place on the audio.

Unfortunately there was no time to have any practice runs in the test. Also, the instructions were written and not communicated verbally with the listener. These both contravene the guidelines in IEC 268-13, but were due to time limits. However, the repeats of auditory event pairs built into the test worked to evaluate the reliability of the listener.

4   Analysis

4.1   Post-screening of subjects

The results from each subject were entered into the statistical software package SPSS for analysis. The first stage involved selecting the most reliable subjects. Ideally, the subjects used in the experiment would have been trained and pre-selected. Unfortunately this was not possible in the time scale. Therefore the listeners were untested in terms of their reliability and ability to hear subtle differences.

The post-selection was attempted based on a number of different criteria. These were:
- Individual ANOVA error variance of less than 1.5
- Individual ANOVA error variance of less than 3
- Individual $F_{process}$ scores (termed F_L in [8])
- Ranking subjects based on individual error variance and ANOVA F statistics for the microphone technique variable [10]

Each of these lowered the overall error variance of all the subjects by varying amounts. The most effective of these was the ranking procedure (as described in [10]) in terms of giving the lowest error value and the most confident results that the different microphone techniques were perceivably different.

However, when examining the results from all the above methods, it was clear that the process of filtering did not reveal any additional information that was not present when including all the listeners. Therefore, because of this and in order not to bias the results in any way, the authors felt it most appropriate to include data from all the subjects for use in the analysis.

The analysis was also carried out considering the subjects as either a fixed factor or a random factor. Again, this made no difference to the conclusions so for convenience the fixed factor version is included here.

4.2   Scale judgements

A multivariate ANOVA was carried out on the judgement data to examine the main effects and any interactions between the variables in the test. This is shown in table 2.

The error variance is a little high, above the level noted previously by Rumsey and others [10]. However, this is most likely to be a result of not filtering the less reliable subjects in this case.

The most significant single factor for each attribute is 'PROCESNO', the variable of microphone technique. However nearly all of the factors are significant ($p<0.01$), with 'SUBJECNO' and 'EXCERPNO' both having good significance.

Figure 2 shows the mean value and the associated 95% confidence intervals for the stereo image grade for each microphone technique for all musical excerpts and all listeners. As can be seen, the undelayed rear pointing microphone extracts were rated significantly lower than the other two types. The main difference between these groups is the factor of delay; the two types rated better have delay present, either deliberately or inherently. This appears to prove the hypothesis as being correct.

Figure 3 shows the mean value and the associated 95% confidence intervals for the spaciousness grade for each microphone technique for all musical excerpts and all listeners. In this case, the ambient microphone technique is rated worse than either of the rear pointing close microphone techniques. It seems that in this case, rear pointing microphones fairly close to the main pair results in an increased sense of spaciousness. This confirms the findings of Fukada et al as discussed above [1].

There is also a strong interaction (EXCERPNO*PROCESNO) shown in the ANOVA. The high significance in this section indicates that the effect of the different microphone techniques on the front image is dependent on the programme material. It could also indicate that various types of musical extract and various recording venues have a differing ability to reveal the strengths and weaknesses of systems or processes.

When figure 4 is examined carefully, a pattern emerges. Even though three of the excerpts (Organ, String Quartet and Choir) have no significant difference in terms of stereo image, there is a large difference in the scores for the speaking voice. This can be explained by a description from the author's own informal listening. The example with the very poor rating had very poor front/back imaging. As such, the listeners found it difficult to determine whether the source was located front left or rear right and the result was quite uncomfortable. The reason for this is the direct sound from the front leaking into the rear right channel; a problem that was solved by delaying the same feed.

The spaciousness judgements as shown in figure 5 also contain some interesting trends. It is apparent that the perceived spaciousness of the delayed rear pointing microphones is fairly similar across all programme items. The trend also shows that the mean scores for this microphone technique are higher than most of the ratings for the other microphone techniques. However, this is not by a significant amount. The ambient microphone technique rates consistently worse and the results with the undelayed rear pointing microphones are very varied. Again, this latter effect may be explained by a description from the author's informal listening. The low register of the organ with the undelayed rear pointing microphones was very enveloping due to the leakage of the direct sound into the rear channels. Unlike the previous example, the much more diffuse sound of the organ did not cause the problem of confusing localisation. The other three musical excerpts for this technique do not show results that are as significant.

4.3   Preference

To analyse the preferences, the raw data was converted to a more convenient form. This consisted of entering a '1' if the excerpt was preferred out of the pair presented, and a '-1' if it was not preferred. This produced better results than using a '1' and '0' as a negative number gives a better indication of a dislike. This is similar to the well-established statistical technique of the sign test, and is also used in the early stages of calculating a Wilcoxon matched-pairs signed ranks test [11].

The sum of the preference scores is shown in figure 6. From this, it is clear that the most preferred microphone technique uses delayed rear pointing close microphones. This is as would be expected from the results of the judgement scales. The ambient microphone technique comes second, receiving approximately an equal amount of preferences and non-preferences. The undelayed rear pointing microphone technique is significantly worse.

If this is subdivided into musical extract types, as was done with the judgement scores, it shows that the delayed rear pointing microphone technique is the best due to its consistency over all the musical extracts (see figure 7). The other two techniques vary greatly. The authors assume this is due to the same reasons as given for the judgement ratings. It is also interesting to note that the extract with the least reverberant information (String Quartet) has the least clear-cut preferences. This is something to bear in mind in future tests.

5   Conclusion

By using controlled subjective listening tests, the results suggest the following for this particular case:
- a surround sound microphone technique which includes an amount of delay between the front and rear channels is less likely to suffer from problems with the stereo image.
- a surround sound microphone technique which uses close rear pointing cardioid microphones for the rear channels is likely to have a better spatial impression than a technique which uses distant omnidirectional microphones.
- stimuli with a lower level of reverberance (both in terms of level and reverberation time) are more difficult to judge in terms of spaciousness.
- stimuli which are more diffuse in their nature are less likely to suffer from problems of localisation.

However, the current results may only have limited external validity. It is accepted that this investigation covers one set of microphone techniques and one set of listeners. How well this will translate to multi-microphone or other techniques is still unknown.

Whilst these results show trends which are significant and others that are not, they relate to this set of subjects who were a mixture of experienced and inexperienced listeners. If the listeners were pre-selected on the basis of their reliability and ability to hear and accurately rate small differences [8], then some of the trends shown in the data may have achieved significance. This would have made the sample more representative of the population of audio engineers who are trained and experienced in listening to minor effects on the audio.

However, the results do give a starting point for those who are interested in experimenting.

6   Further research

There are further tests that could be carried out.

A definite improvement in the quality of the tests would be the use of a randomised double blind test. The results may have been influenced by the order in which the stimuli occurred. Whilst every effort was made to randomise this, there may have been aspects of acclimatisation over time which affected the results. In an ideal system, the excerpts would have been presented to each listener in a different order so as to limit the influence of any bias that may have occurred.

Another relevant factor to define is the optimum length of delay. A listening test in which the subjects can set preferred delay time on a blind scale would be interesting. The results of this could be compared to the findings of Haas for the different lengths of delay. There may also be a correlation between the type of signal and the preferred delay time. This may be combined with

finding an optimum position for the rear pointing microphones which is a trade-off between the spaciousness of close microphones and the improved stereo image of delayed microphones.

As mentioned above, the tests could be repeated using a larger pre-selected panel of listeners to give the tests greater statistical power. This may uncover significant results that currently are only presented as trends in the mean values.

The tests could also be repeated to include listeners seated in alternative positions. This would show how relevant the system is to listening in a real situation where the listeners are not always seated at the optimum location.

A large amount of other variables exist which could have been tested within the scope of the experiment but which were excluded to keep complexity to a minimum. The scope of the experiment serves as a reference point for further work and as a guide for those starting out in this relatively new field.

References

[1] Fukada, Akira et al., 1997: 'Microphone Techniques for Ambient Sound on a Music Recording', **Audio Engineering Society Preprint**, 103rd Convention - New York (26-29 September), No. 4540.

[2] Haas, Helmut, 1972: 'The Influence of a Single Echo on the Audibility of Speech', **Journal of the Audio Engineering Society**, Vol. 20, No. 2, (March), pp. 145-159. (Translated from 'Über den Einfluss des Einfachechos auf die Horsamkeit von Sprache', 1951).

[3] Blauert, Jens, 1971: 'Localization and the Law of the First Wavefront in the Median Plane', **Journal of the Acoustical Society of America**, Vol. 50, No. 2 (Part 2), pp. 466-470.

[4] Ceoen, Carl, 1972: ' Comparative Stereophonic Listening Tests', **Journal of the Audio Engineering Society**, Vol. 20, No. 1, (January/February), pp. 19-26.

[5] IEC 268-13, 1987: 'Sound system equipment - Part 13. Guide for listening tests on loudspeakers'.

[6] Bech, Søren, 1989: 'The Influence of Room Acoustics on Reproduced Sound Part 1- Selection and Training of Subjects for Listening Tests', **Audio Engineering Society Preprint**, 87th Convention - New York, (18-21 October), No. 2850.

[7] ITU-R BS 1116, 1994: 'Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems', **International Telecommunications Union**, Recommendations ITU-R BS 1116, pp. 276-297.

[8] Bech, Søren, 1992: 'Selection and Training of Subjects for Listening Tests on Sound-Reproducing Equipment', **Journal of the Audio Engineering Society**, Vol. 40, No. 7/8, (July/August), pp.590-609.

[9] Rumsey, Francis, 1998: 'Subjective assessment of the spatial attributes of reproduced sound', **Proceedings of the AES 15th International Conference**, Audio Engineering Society, Copenhagen, Denmark, Oct 1998.

[10] Rumsey, Francis, 1998: 'Controlled subjective assessments of 2-to-5 channel surround sound processing algorithms', **Audio Engineering Society Preprint**, 104th Convention - Amsterdam, (16-19 May), No. 4654.

[11] Clegg, Frances, 1995: **Simple Statistics: A course book for the social sciences**, (Cambridge: University Press).

Appendix A   Instructions for listeners

The purpose of these listening tests is to judge the factors of stereo image and spaciousness for various surround sound processing. The test will be conducted in five channel surround sound.

The factors which you are grading can be defined as follows.
Stereo image - how easy is it to localise the sources presented? Does the image fill the frontal soundstage? How stable is the image? How natural is the image? How close is it to being at the concert?
Spaciousness - How enveloping is the sound? How natural is the ambience? How deep is the image? How close is it to being in the venue?

Each factor should be regarded separately and there should be no overlap in the criteria. The stereo image factor should not include image depth and other such factors, and the spaciousness factor should not include stereo width.

The test is of an A - B form. This means that there will be two concurrent sources which you can switch between at any time. To do this, there is a box on the table in front of you with switches labelled A and B. You are asked to grade both sources A and B. The grading however is not to be a comparison between the two sources, but grading on an absolute scale as shown below. You are also asked to express a preference between A and B.

Each extract will be 90 seconds long. During this time, you are free to switch between examples A and B. After each extract will be a 30 second gap for thought and writing before the start of the next extract. There are 14 extracts in all.

The grading will be on the scale as shown on the next page. Number 10 represents that the source is perfect and cannot be any better. Number 0 denotes that the source could not possibly be any worse. You are asked to draw a line representing your opinion of the quality of each factor as described above. Your preferred source should be circled at the bottom.

If you make a mistake, simply clearly cross out the error and draw a new choice.

Grades can be made at any point on the line, not just at the integer points.

It is important to remember that there is no right or wrong answer. Your judgement is the correct answer and you will not be marked on your choices. You will not be told anything about the sources involved or the processing carried out, if any. If you are interested, you can find out after all the listening tests have been carried out.

To repeat: I am looking for your evaluation of the sources heard based on the criteria of stereo image and spaciousness. The sources should be regarded separately and graded as such. The grading scale is from 0 to 10, with the extremes being the worst and best possible. The preference indication is the choice which you find most aesthetically pleasing.

Please feel free to ask any questions, preferably before the test begins.

Enjoy the listening, and thank you for taking part.

Table 1: Order of extracts as played to the listeners.

| Extract No.	Excerpt	Source A	Source B
1	Organ	RD	R
2	Chamber	RD	A
3	Quartet	R	A
4	Forbes	RD	R
5	Organ	A	R
6	Forbes	A	R
7	Quartet	RD	A
8	Organ	RD	R
9	Chamber	R	RD
10	Forbes	RD	R
11	Quartet	A	R
12	Forbes	RD	A
13	Organ	A	RD
14	Chamber	A	R

Table 2: Multivariate ANOVA results table for all listeners with Stereo Image Grade and Spaciousness Grade as dependant variables, and SUBJECNO, EXCERPNO and PROCESNO as fixed factors.

| Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
SUBJECNO	Stereo image grade	318.077	19	16.741	10.889	.000
	Spaciousness grade	182.959	19	9.629	5.597	.000
EXCERPNO	Stereo image grade	35.115	3	11.705	7.614	.000
	Spaciousness grade	21.778	3	7.259	4.220	.006
PROCESNO	Stereo image grade	43.960	2	21.980	14.297	.000
	Spaciousness grade	22.915	2	11.457	6.660	.001
SUBJECNO * EXCERPNO	Stereo image grade	175.292	57	3.075	2.000	.000
	Spaciousness grade	186.042	57	3.264	1.897	.000
SUBJECNO * PROCESNO	Stereo image grade	84.306	38	2.219	1.443	.050
	Spaciousness grade	82.043	38	2.159	1.255	.152
EXCERPNO * PROCESNO	Stereo image grade	119.224	6	19.871	12.925	.000
	Spaciousness grade	41.314	6	6.886	4.003	.001
SUBJECNO * EXCERPNO * PROCESNO	Stereo image grade	201.431	114	1.767	1.149	.175
	Spaciousness grade	180.882	114	1.587	.922	.690
Error	Stereo image grade	491.953	320	1.537		
	Spaciousness grade	550.502	320	1.720		
Total	Stereo image grade	21557.500	560			
	Spaciousness grade	22454.210	560			
Corrected Total	Stereo image grade	1560.370	559			
	Spaciousness grade	1319.096	559			

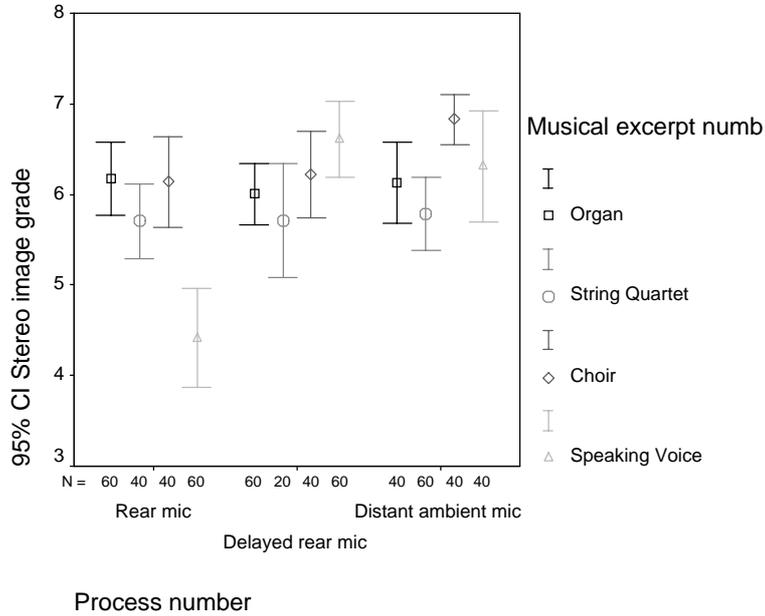**Figure 1: Diagram of microphone layout.**
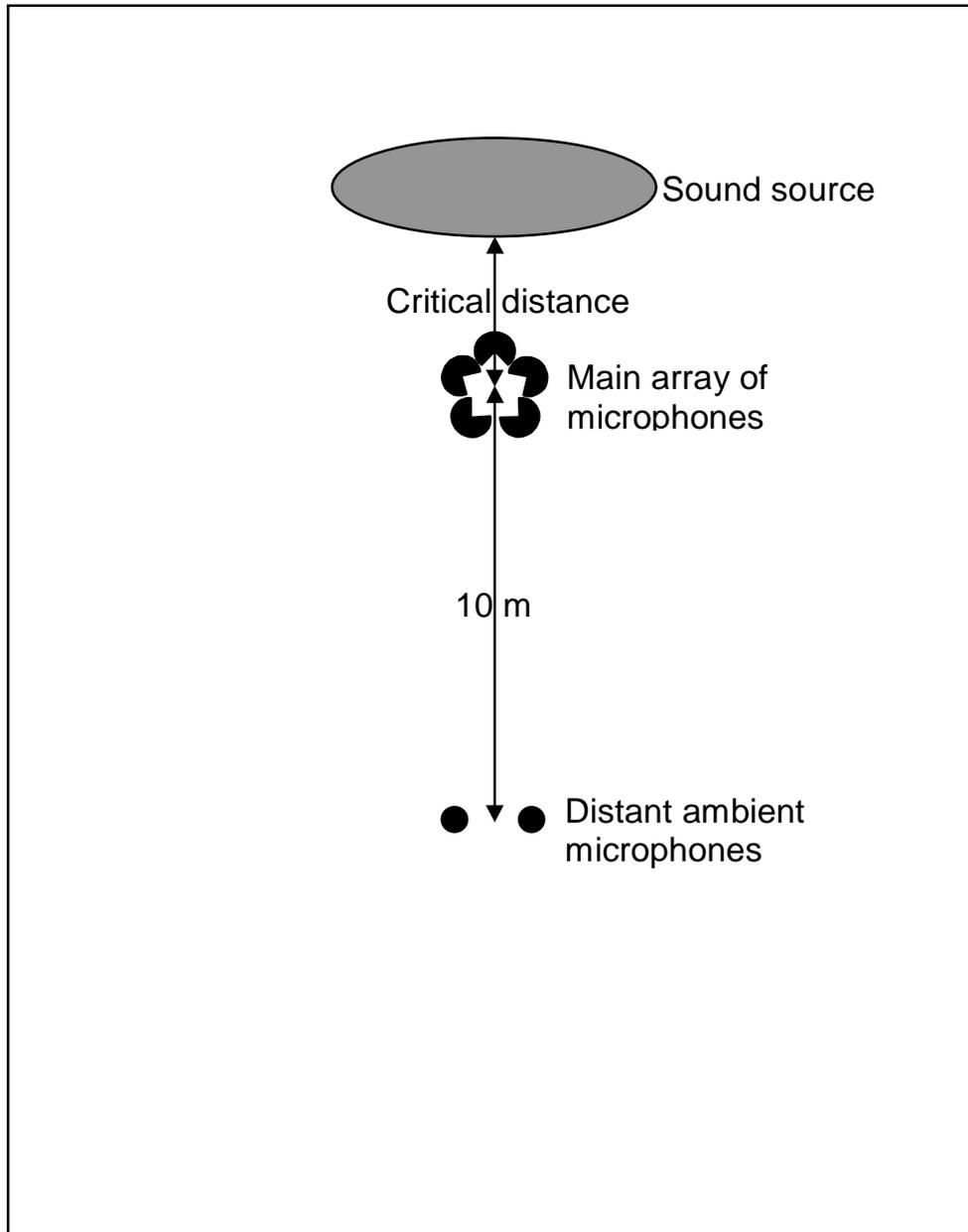
Figure 2: Mean value and the associated 95% confidence intervals of Stereo Image Grade for all listeners separated by microphone technique.
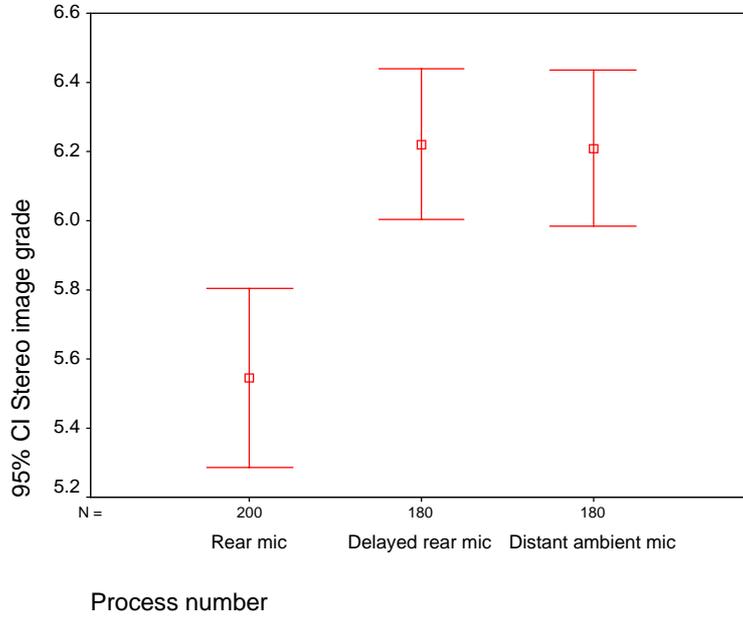

Figure 3: Mean value and the associated 95% confidence intervals of Spaciousness Grade for all listeners separated by microphone technique.
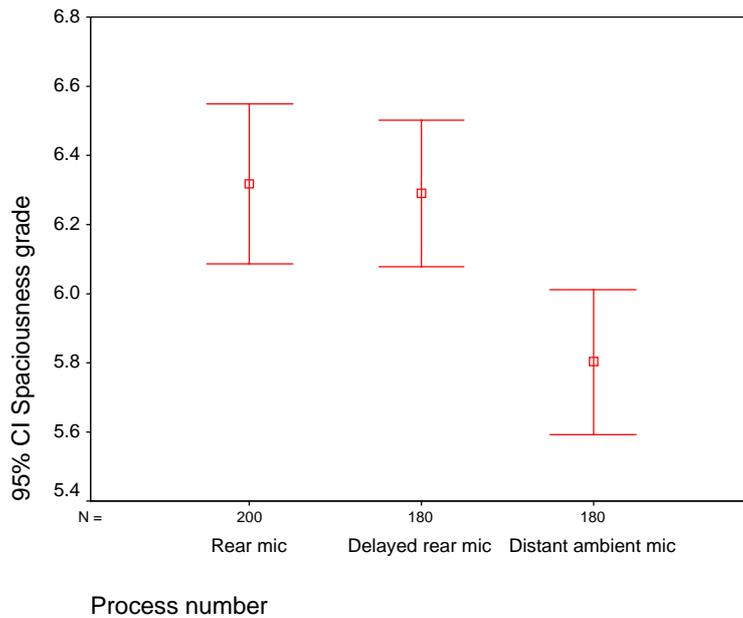
Figure 4: Mean value and the associated 95% confidence intervals of Stereo Image Grade for all listeners separated by microphone technique and musical excerpt.

Figure 5: Mean value and the associated 95% confidence intervals of Spaciousness Grade for all listeners separated by microphone technique and musical excerpt.
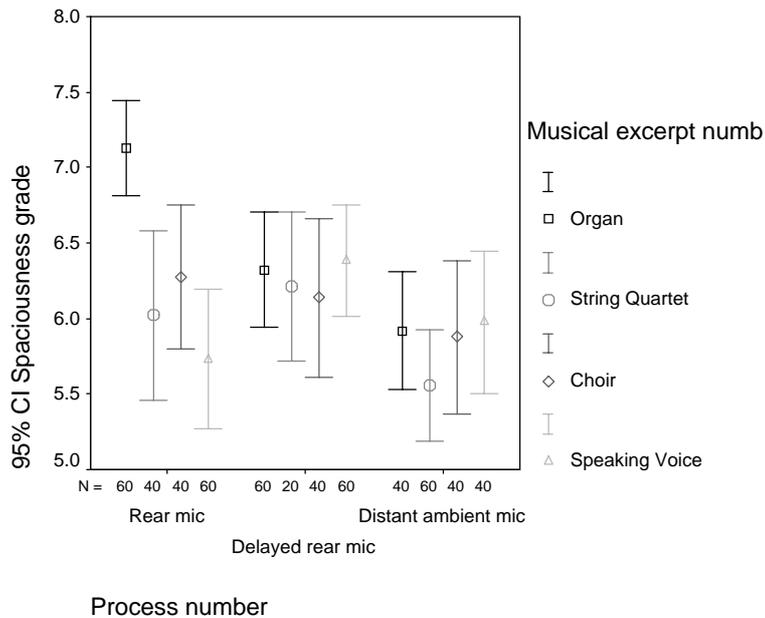
Figure 6: Sum of preference data for all listeners separated by microphone technique.
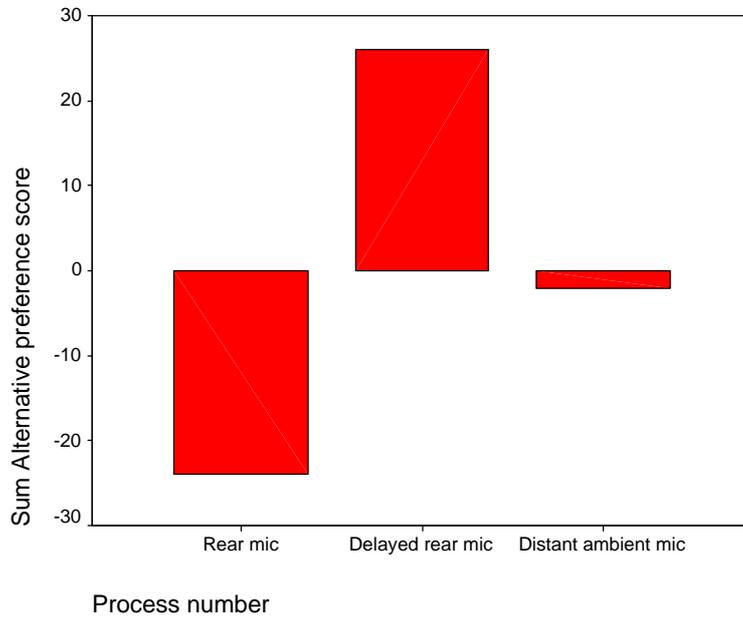
Figure 7: Sum of preference data for all listeners separated by microphone technique and musical excerpt.