Effect of rear microphone spacing on spatial impression for omnidirectional surround sound microphone arrays

Francis Rumsey and Wyn Lewis

1 Institute of Sound Recording, University of Surrey, Guildford GU2 7XH, UK
Correspondence should be addressed to corresponding author (f.rumsey@surrey.ac.uk)

ABSTRACT
The effect of rear microphone spacing in a five-channel omni-directional array was evaluated in respect of the subjective attributes ‘envelopment’, ‘spaciousness’ and ‘naturalness’. Preference results were also obtained and a range of different programme material types was evaluated. Results suggest that, although large differences were not noticed, spacing had a statistically significant effect on envelopment and spaciousness, with spacings larger than the critical distance of the room giving rise to higher levels of these attributes. Distinct preference was shown for spacings of three and four metres as opposed to the extremes of two and five metres, in the particular acoustic environment tested, suggesting that there is an optimum range of microphone spacing outside which the reproduced spatial impression becomes less pleasing or natural.

INTRODUCTION
This paper is based on experiments conducted as part of an undergraduate project by Wyn Lewis. It is presented here because the results provide some useful insights into the subjective spatial effects of different lateral spacing of omnidirectional rear microphones in surround sound arrays. Anecdotal evidence from recording engineers has suggested that spaced arrays having rear microphone spacing of approximately three metres are successful in a number of recording environments, providing sufficient decorrelation to give a pleasing sense of spaciousness whilst maintaining good integration between the room impression and the front sound stage. It has also been suggested that such spacing is not so wide as to cause the rear loudspeaker signals to become ‘disconnected’ from the overall spatial impression, and thereby sounding unnatural or separately perceivable as sources. It seems reasonable to hypothesise that the extent to which two spaced microphone signals, derived from the reverberant field, can be decorrelated is limited (tending towards zero). Further physical separation may not result in an improvement in the spatial effect, and indeed the effect might even start to be perceived as worse as a result of other factors becoming prominent, such as excessively delayed crosstalk from front sources.

In order to test these otherwise anecdotal claims an experiment was conducted in which a number of different sources were recorded in a reverberant environment, under controlled conditions. The sources were recorded in five-channel surround sound using a fixed front microphone array and four different rear microphone spacings. These recordings were then replayed in a listening room to a number of subjects who judged attributes of the resulting spatial quality and indicated their preference for particular versions. Relationships could then be established between rear microphone spacing, spatial quality and listener preference, enabling some tentative conclusions to be drawn concerning the most effective microphone spacing.

RECORDING AND PREPARATION OF SOURCE MATERIAL
Microphone array and recording environment
The microphone array used for this experiment was configured as shown in Figure 1. The microphones concerned were Schoeps CMC-5U omnidirectional types. The front three channels were derived from a spaced ‘ABC’ array that remained constant throughout the experiment. It is not claimed that this array offers...
any particular imaging advantages over other types or spacings, it was simply a convenient compromise using spaced omni microphones that could be used to feed the front three channels in a consistent fashion. Since front imaging was not in question during this experiment, the choice of front array was not regarded as a crucial factor. (The choice of spacing could be challenged, for example.)

The distance between front and rear microphones needed to be reasonably large in order to reduce the level of direct sound present in the rear speakers, which can cause problems with front-back confusion. However the distance needed not to be so high that the delay caused by the distance was perceived as an echo. Therefore, and in accordance with experiential evidence from recording engineers, the distance of the rear microphones was fixed at 3m. The distance between the rear microphones was chosen to produce two signals that were decorrelated to differing degrees. This distance was one of the variables in the experiment and the separations chosen were 2m, 3m, 4m and 5m.

The height of the front microphones was 1.6m and the rear microphones were raised somewhat higher. The distance of the microphones from the sources was approximately equal to the critical distance of the room (~2.5m).

The recordings were made in a classical music recording studio, Studio 1 at the University of Surrey, that has an area of approximately 250 m² and a maximum reverberation time of approximately 2 seconds.

Recording setup and alignment
The microphone gains were adjusted for each configuration of rear microphones so that a pink noise source replayed through a loudspeaker in the intended source position gave rise to equal output level from all five microphones. This pink noise was recorded at each microphone spacing so as to enable subsequent replay level alignment.

Programme extracts were recorded onto a Sony PCM-800 eight-track digital audio recorder at standard recording levels.

Source material
The sources were single live musical instruments chosen to span a range of pitch, dynamic and timbral characteristics, as follows:

1. Solo Cello – the chosen extract was a typical piece written for solo cello from the Romantic period consisting of slow melodic lines spanning a wide pitch and dynamic range.
2. Solo Clarinet – the chosen extract was a piece composed in the 20th century and consisted of virtuosic passages, combined with slow ‘melodic’ passages, again spanning the range of the instrument, but mainly in its upper register, and being played moderately loud (mf), to loud (f).
3. Solo Violin – the chosen extract was a piece composed in the Romantic period, playing slow melodic lines with some pauses in between, and again spanning the range of the instrument whilst being played loudly.
4. Solo Piano – the extract was composed in the Classical period, consisting of short notes, being played in the mid-register of the instrument, and moderately loud.
5. Speech – a male voice, recognisable to the subjects of the test, expressively reading a passage from a book with rises and falls in the sound of his voice, pauses in between sentences, and talking clearly. The loudness of his voice was raised from normal conversational volume to that of a speaker at a presentation in front of an audience so that he could be clearly understood a distance away.

Practical considerations required that each extract was repeated four times, once for each microphone spacing. This was not ideal as small variations were encountered in performance, but this was regarded as a minor issue in relation to the aims of the experiment.

The duration of each extract was between 30 seconds and one minute, longer than recommended in standards for sequential comparison tests, but a suitable compromise in view of the number of attributes to be judged in each case.

REPLAY CONDITIONS
The listening test was conducted in the ITU-R BS.1116 listening room at the University of Surrey. The monitors used were five Genelec 1032A loudspeakers which were set up for multichannel surround sound reproduction, conforming to the layout specified in ITU-R BS.775. All the loudspeakers were set at the height of 1.2m, with the base distance (between the front left and right loudspeakers) set at 2.5m and all the loudspeakers were placed at least 1m away from the walls of the room.

The loudspeakers were each aligned to produce an acoustic level of 70dBA at the listening position, using pink noise at a level equivalent to recorded reference level. This listening level was judged comfortable yet loud enough to be revealed on the screen. All the lights in the room were dimmed to create a comfortable environment that reduced the importance of visual cues and enabled subjects to concentrate on aural cues relating to spatial impression.

EXPERIMENTAL DESIGN
The experiment was based on a modified, blind, single stimulus rating (SSR) paradigm owing to the use of tape-based replay that made direct inter-stimulus comparisons difficult. This method relies strongly on the listener’s memory of previous stimuli to make reliable comparisons, and is likely to result in the stimuli being rated on an absolute rather than comparative basis. In order to make the task somewhat more straightforward and comparative for subjects, a reference item (the 2m microphone spacing, labelled A) was repeated between each item to be rated in order to provide an anchor point for judgements. The order of items presented therefore followed the pattern A-B-A-C-A-D, where B, C and D were examples of different (larger) microphone spacings, presented in pseudo-random order. Stimulus A was rated on the first audition and simply used as a reminder on subsequent occasions.

Clearly this approach has benefits and difficulties: a benefit is that the repeated reference enables more reliable comparisons to be made between sequential presentations of the wider microphone spacings and the reference spacing, but the difficulty is that the reference spacing is rated first on all occasions and repeated more often than the other versions. This might have given rise to some experimental bias and/or order effects, but was chosen as a compromise solution in view of limited time and facilities.
SCALES
The problems involved with varying interpretations of perceived attributes of spatial impression have been discussed before [1, 2]. It is not proposed to enter into a long discussion of this issue here. In this experiment the following questions were of interest:

- How is the sensation of being immersed in reverberant sound affected?
- How is the naturalness of the reproduced sound field affected?
- How is the sense of space in the reproduced sound field affected?

In this experiment the term spaciousness is distinguished from envelopment, and given its more literal meaning pertaining to the perception of being in a large space. Recent work [3] indicates that it may be possible to perceive the boundaries of a reproduced environment separately from the perception of envelopment (defined here as relating to a sense of immersion in reverberant sound). This does not discount the fact that the perception of spaciousness in a recording might improve as the perception of envelopment increases, but both perceptions are not regarded as one and the same.

As one of the main objectives of multichannel surround sound systems is arguably to recreate a more natural-sounding acoustic environment, it seems crucial to find out if the reproduced sound field appears to be natural. This scale is clearly context dependent and multidimensional but helps to highlight the existence of unnatural effects such as discrete echoes, image distortion, phasiness and so forth, that might otherwise remain hidden.

The following definitions were provided to subjects (they were all familiar with the natural environment termed ‘Studio 1’):

**Envelopment.** How immersed do you feel in the reverberant sound field? How enveloping is it?

**Naturalness.** How natural is the soundfield? All the recordings were made in Studio 1. How similar is it to being in the recording environment?

**Spaciousness.** The degree to which the sound is perceived as open, not constrained to the locations of the loudspeakers.

Subjects were also asked to denote a preference for one of the four items in each sequence. Scales ranging from 0–10 for each attribute were provided on paper, with no intermediate anchor points or descriptors. Grades were quantised to one decimal place. Subjects were instructed that the end points of the scales corresponded to the lowest and highest possible quantities of the attribute in question.

SUBJECTS
Fourteen subjects took part in the listening test, all of which were students on the Music and Sound Recording degree course at the University of Surrey. Ten were final year students, two were second year and two were first year, and all were unpaid volunteers. Whilst the subjects could be considered ‘expert’ listeners for many types of experiments concerning sound recording, their familiarity with listening to and evaluating surround sound recordings was limited.

Owing to time limitations, the subjects were not pre-screened, neither were they formally trained in relation to the test method or the attributes concerned. However the attributes to be graded were explained to them in detail and an opportunity for questions was given.

RESULTS
In order to eliminate differences between subjects in their use of the scale and to normalise results, the raw responses were z-transformed as recommended in ITU-R BS.1116 for scales in which no intermediate anchor points are defined, while retaining the original 0-10 scale. A multifactor ANOVA (MANOVA) test was then performed on the transformed data to determine significant effects, having first checked appropriate assumptions regarding the suitability of the data. The results of the ANOVA analysis is shown in the Appendix. For some attributes, it will be noted that the model is only significant at the 90% level, which is not ideal but may be adequate to make some observations in a relatively non-critical experiment.

Preference data was scored and analysed as follows: the preferred item in each group of observations was scored +1 and the remainder were scored −1. A simple count could then be made of the preference scores.

**Effect of microphone spacing**
Alterations in rear microphone spacing had a significant but not large effect on the subjective attributes in question. The effect on spaciousness was only significant at the 90% level (which is arguably adequate for a comparison of microphone techniques). The effect on envelopment was significant with greater than 99% confidence, and the effect on naturalness was significant at 95% in respect of an interaction effect between spacing and programme item (in other words the effect of spacing was not the same for all programme items).

Means and associated confidence intervals for envelopment, spaciousness and naturalness (by programme item) are shown in Figures 2–4.
Preference
A clear preference resulted for spacings of 3 and 4 metres, with 5 metres being least preferred. The results are shown in Figure 5.

Effects of programme
The programme item was a significant factor (99%) in the results for envelopment but not for either of the other dependent variables (spaciousness, naturalness). This is shown in Figure 6.

DISCUSSION
Although the differences are not pronounced and only trends can be observed, the 2m spacing gives the lowest values for spaciousness and envelopment. This spacing is smaller than the critical distance or reverberation radius of the hall in question (CD = about 2.5m) and is likely to result in relatively high inter-channel correlation. If 90% confidence is considered justifiable (which it arguably is for comparisons between microphone techniques) the 3m spacing produces higher spaciousness than the 2m spacing, as determined by a Tukey HSD multiple comparison between means. Other spacings are not significantly different, but the trend is for them to give slightly higher spaciousness than the 2m spacing. The inconclusiveness of this result is possibly due to the difficulty of the task for the subjects and it is probable that greater attribute-specific training would have improved their sensitivity to changes in spaciousness. Nonetheless it is interesting to note the relatively small effect as far as recording engineers with substantial listening experience are concerned.

In general there is a trend towards rising envelopment with increased spacing, but with a dip at the 4m spacing that is unexplained. Interestingly, this dip at 4m also arises in the spaciousness results. Envelopment at a spacing of 5m is significantly higher than that at 2m with 95% confidence, and the difference between 2 and 3m is significant at the 90% level (Tukey HSD).

Averaged over all programme items, naturalness is not significantly different at any spacing, but the trend is slightly rising to 4m then falling to 5m. This suggests that excessive spacing may give rise to less natural spatial impression. A larger number of observations and more consistent subjects would be needed for greater certainty. Separated by programme item (with which microphone spacing shows a significant interaction), we notice that some items (e.g. violin, piano) have a rising naturalness trend from 2-3-4m that then falls significantly at 5m. The other items do not significantly differ. This may be related to the audibility of the background reverberant information stream, as the items showing this effect involved more staccato passages enabling the hearing of reverberation between notes, and thus the more reliable and distinctive judgement of the naturalness of the spatial impression. This hypothesis is possibly borne out by looking at the influence of programme material on perceived envelopment (Figure 5). The same two items (violin and piano)
are significantly different (higher, P=0.02) from the cello which was more legato (speech is somewhere in between).

The trend towards a preference for spacings of 3 and 4m, looked at alongside the other results, suggests that there is a preferred optimum level of envelopment and spaciousness above which results begin to seem unnatural for some types of programme material, and where other artefacts such as audible delays or echoes become apparent. This is clearly supposition, and further tests would be required to determine the reasons for the lack of naturalness in the 5m-spaced version.

A discriminant function analysis, conducted to discover the attribute having the greatest influence on preference, shows significant differences between attributes and suggests that naturalness is most highly correlated with preference ratings, followed by spaciousness, followed by envelopment. This not entirely surprising since the term ‘naturalness’ itself contains strong connotations of ‘liking’ (especially if the bottom of the scale is taken to mean that the recording sounded ‘unnatural’). It nonetheless serves to demonstrate that spaciousness and envelopment ratings on their own are not sufficient to determine whether or not the result will be pleasant to listen to. For example, here it is shown that the highest envelopment does not necessarily correspond to the greatest preference.

Overall the results are not wholly conclusive and a future experiment could be conducted to compare such subjective results with measured results of the reproduced soundfield (such as interchannel or interaural correlation) as well as involving a greater degree of training and listener selection. Nonetheless interesting trends are noticed.

CONCLUSION

Rear microphone spacing of an omnidirectional surround sound array has been shown to have a small but significant effect on some aspects of subjective spatial impression and on listener preference. The results are not wholly conclusive, can only be said to apply to this particular recording/reproduction context and there is a need for better listener training, but the trend was for listeners to prefer recordings with spacings of three or four metres, as opposed to the very wide spacing of five metres. The spacing of two metres gave rise to lower spaciousness and envelopment than some wider spacings. There is some evidence that although envelopment is highest at the widest spacing, this does not correspond to optimum naturalness or listener preference.

REFERENCES


APPENDIX

Statistical Tables

Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent Variable</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>Z/Envelopment</td>
<td>99.371</td>
<td>19</td>
<td>5.230</td>
<td>2.484</td>
<td>.00</td>
</tr>
<tr>
<td>Z/Spaciousness</td>
<td>64.031</td>
<td>2.370</td>
<td>19</td>
<td>1.492</td>
<td>.086</td>
<td></td>
</tr>
<tr>
<td>z/Naturalness</td>
<td>56.677</td>
<td>2.983</td>
<td>19</td>
<td>1.499</td>
<td>.086</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>Z/Envelopment</td>
<td>10090.803</td>
<td>1</td>
<td>10090.803</td>
<td>4793.24</td>
<td>.00</td>
</tr>
<tr>
<td>Z/Spaciousness</td>
<td>9842.076</td>
<td>4356.956</td>
<td>1</td>
<td>4356.956</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>z/Naturalness</td>
<td>8975.629</td>
<td>4493.436</td>
<td>1</td>
<td>4493.436</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>SPACING</td>
<td>Z/Envelopment</td>
<td>31.940</td>
<td>3</td>
<td>10.641</td>
<td>5.057</td>
<td>.00</td>
</tr>
<tr>
<td>Z/Spaciousness</td>
<td>14.501</td>
<td>4.854</td>
<td>3</td>
<td>2.144</td>
<td>.098</td>
<td></td>
</tr>
<tr>
<td>z/Naturalness</td>
<td>7.613</td>
<td>2.538</td>
<td>3</td>
<td>1.271</td>
<td>.284</td>
<td></td>
</tr>
<tr>
<td>PROGRAM</td>
<td>Z/Envelopment</td>
<td>27.873</td>
<td>4</td>
<td>6.966</td>
<td>3.310</td>
<td>.01</td>
</tr>
<tr>
<td>Z/Spaciousness</td>
<td>9.886</td>
<td>2.471</td>
<td>4</td>
<td>1.094</td>
<td>.366</td>
<td></td>
</tr>
<tr>
<td>z/Naturalness</td>
<td>4.661</td>
<td>1.166</td>
<td>4</td>
<td>0.584</td>
<td>.674</td>
<td></td>
</tr>
<tr>
<td>SPACING * PROGRAM</td>
<td>Z/Envelopment</td>
<td>39.558</td>
<td>12</td>
<td>3.296</td>
<td>1.566</td>
<td>.10</td>
</tr>
<tr>
<td>Z/Spaciousness</td>
<td>39.650</td>
<td>3.304</td>
<td>12</td>
<td>1.463</td>
<td>.138</td>
<td></td>
</tr>
<tr>
<td>z/Naturalness</td>
<td>44.399</td>
<td>3.700</td>
<td>12</td>
<td>1.854</td>
<td>.043</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>Z/Envelopment</td>
<td>547.356</td>
<td>22</td>
<td>2.475</td>
<td>1.199</td>
<td></td>
</tr>
<tr>
<td>Z/Spaciousness</td>
<td>587.322</td>
<td>2.596</td>
<td>22</td>
<td>1.199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z/Naturalness</td>
<td>518.922</td>
<td>2.366</td>
<td>22</td>
<td>1.199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>Z/Envelopment</td>
<td>10737.530</td>
<td>22</td>
<td>4.877</td>
<td>276.271</td>
<td>.00</td>
</tr>
<tr>
<td>Z/Spaciousness</td>
<td>10433.438</td>
<td>486.157</td>
<td>22</td>
<td>1.872</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z/Naturalness</td>
<td>9551.233</td>
<td>2.366</td>
<td>22</td>
<td>1.199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Z/Envelopment</td>
<td>10737.530</td>
<td>22</td>
<td>4.877</td>
<td>276.271</td>
<td>.00</td>
</tr>
<tr>
<td>Z/Spaciousness</td>
<td>10433.438</td>
<td>486.157</td>
<td>22</td>
<td>1.872</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z/Naturalness</td>
<td>9551.233</td>
<td>2.366</td>
<td>22</td>
<td>1.199</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standardized canonical discriminant function coefficients for preference

<table>
<thead>
<tr>
<th>Function</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z/Envelopment</td>
<td>0.30</td>
</tr>
<tr>
<td>z/Naturalness</td>
<td>0.64</td>
</tr>
<tr>
<td>Z/Spaciousness</td>
<td>0.42</td>
</tr>
</tbody>
</table>

AES 112TH CONVENTION, MUNICH, GERMANY, 2002 MAY 10–13