Audio Engineering Society

Convention Paper 6142

Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

Elicitation and Grading of Subjective Attributes of 2-channel Phantom Images

Hyun-Kook Lee and Francis Rumsey

Institute of Sound Recording, University of Surrey, Guildford, Surrey, GU2 7XH, England
Correspondence should be addressed to Hyun-Kook Lee: tonmeister1025@yahoo.co.uk

ABSTRACT

The subjective attributes of 2-channel phantom images of transient piano, continuous trumpet and male speech sources were elicited using pair-wise comparison between reference mono images and their phantom images. The attributes elicited included 'image focus', 'image width', 'image distance', 'brightness', 'hardness' and 'fullness'. The effect of interchannel time and intensity differences on the perceived difference between the real image and its phantom image was investigated for each sound source in respect of the elicited subjective attributes. Results show that the type of panning method (pure time, pure intensity and combination of the two) had a statistically significant effect on image focus and image width attributes. It was also found that the type of sound source had a significant effect on all the attributes.

1. INTRODUCTION

This paper is concerned with investigating the influences of interchannel time difference, interchannel intensity difference, and the type of sound source on the perceived auditory attributes of phantom images in two-channel stereophonic reproduction. The experiment described here is a part of pilot studies that were conducted prior to evaluating the effect of interchannel crosstalk in multichannel microphone techniques. Although subject to much debate, very little is known about the effect of interchannel crosstalk, or what kinds of auditory attributes listeners can perceive from the resulting phantom image and how those attributes affect the quality of the image. Therefore, it is not entirely clear what attributes are appropriate for the evaluation of the effect of interchannel crosstalk in various microphone techniques.

Reported studies on the perceived differences between phantom images created by the precedence effect and their corresponding mono images could be the basis for evaluating the attributes of 'crosstalk phantom image' (e.g. the phantom image having 'greater spatial extent' [1], 'image extended toward the echo source' [2] and 'fuller tonal colour' [3]). For experiments, there also might be a number of possible attribute scales that could be provided by the experimenter based on their own

experience and knowledge. However, as Berg and Rumsey [4] and Kjeldsen [5] point out, the use of the 'provided' attribute scales has a significant limitation in this kind of spatial subjective evaluation. Listeners are restricted to respond only in the experimenter's own terms, even if they find other valuable attributes that can also be evaluated. In this respect, a more reasonable method for subjective evaluation, especially of such a relatively undeveloped area as the effect of interchannel crosstalk, would be grading subjectively 'elicited' attribute scales.

It has been claimed by Rumsey [6] that the perceived auditory attributes of phantom images created from the interference of interchannel crosstalk signals between the adjacent microphones in multichannel microphone arrays depend on the combination of relevant time and intensity differences between the signals. Therefore, it will be appropriate to investigate the basic effect of time and intensity differences on the perception of phantom images.

Since a source emanating from a mono loudspeaker can be regarded as a 'real' source without any crosstalk interference, a way of investigating the basic effect of interchannel crosstalk would be to compare the perceptible differences between 'real' images and their corresponding phantom images, created with various combinations of time and intensity differences.

In this pilot experiment, a series of subjective listening tests were designed and conducted in order to elicit perceived attributes of phantom images in two-channel stereo when compared to the 'real' image provided by a mono loudspeaker, and to grade the magnitude of the elicited attributes depending on the panning method and the type of sound source. The scope of this experiment was limited to two-channel stereo since it would simplify the simulation of possible effects of the combination ratio of time and intensity differences.

2. DESCRIPTION OF THE LISTENING TESTS

2.1. Method

This experiment is based on the Quantitative Descriptive Analysis (QDA), which was originally developed for the evaluation of sensory attributes of products. Basically the QDA consists of three stages: elicitation process, grouping analysis and grading process [7]. Firstly, a group of qualified subjects are presented with stimuli and generate descriptive terms on the attributes of the product through discussion. Secondly, the elicited terms are grouped into a limited number of attribute scales through discussion based on the similarity of meaning of the terms. Finally, the stimuli are graded using the obtained attribute scales. An advantage of this method is that the subjects have an influence on the attribute scales that are to be used in grading. Therefore, it is possible to reduce a bias that might be caused when provided attribute scales were used. The test method used in this experiment was modified from the QDA in a way that there was no subject discussion in the elicitation process and the grouping analysis. It was thought that a bias might be caused among subjects in the course of discussion and the answers might be dependent on a few influential subjects' decisions. Therefore, only one subject was involved in each test, and provided his or her independent answers. The elicited terms were analysed only by the author without subject discussion, because in the grading process it was desired to use more commonly referred attribute scales based on literatures rather than subjective scales. Therefore, this experiment used something of a compromise between elicited and provided attribute scales, consisting of two listening tests: elicitation and grading tests.

2.2. Stimuli

Three sound sources were chosen with much consideration, including:

- Piano 'staccato' note of C3 (fo = 130Hz)
- Trumpet 'sustain' note of Bb3 (fo = 228Hz)
- Continuous speech signal

The sound sources of piano and trumpet were chosen in order to examine the perceived effect depending on the different natural characteristics of musical instruments, including transient and continuous natures (staccato vs. sustain). It was of interest to see how these characteristics affect the perception of image attributes. The importance of transient nature in sound for accurate localisation has been mentioned in many literatures such as [8], [9] and [10]. The authors generally agree that

continuous sound on its own is not able to provide a reliable cue for localisation. It was decided to use single notes played with the instruments instead of musical extracts in order to limit the variables strictly within the experimental scope. Ideally the sound sources should have been recorded anechoically, but this was not of practical possibility. The piano source was recorded using a single cardioid microphone of Schoeps CMC 5-U placed about 30cm over the hammers for the desired note. The piano was completely covered with thick cloth in order to reduce unwanted acoustic effects as much as possible. The trumpet sources were recorded in a small overdub booth of studio 3 of the University of Surrey, using a single cardioid microphone of AKG 414 B-ULS place about 1m away from the instrument. The recording space was acoustically isolated and had no audible reverberation. The onset and offset transients of the trumpet source were removed by fading in and out the beginning and ending for 1 second each, and the total duration of the stimulus was 4 seconds. The speech signal was chosen because it is a mixture of both transient and continuous natures with the wide rage of frequencies. Also speech signal has been one of the most popularly used stimuli in the classic localisation experiments and it was of interest to compare the results of the current experiment with those of the classic ones. The speech recording used was a Danish male speech that was anechoically made for the Bang and Olufsen's Archimedes project. An English speech recording was also available in the CD, but it was decided to use a foreign language like Danish rather than English in order to prevent the listener from paying attention to the language itself. The reference for comparison to the phantom image was the real image of a mono source radiated from the same direction as that phantom image direction. For each sound source, one mono stimulus and three stereo stimuli were produced using different panning methods including pure time difference, time-intensity combination and pure intensity difference. The sound pressure levels of all the stimuli were calibrated at 75dB. From an informal test that had been conducted before the main experiment, it was recognised that the variety of panning angles had a very small effect on the perceived attributes. Therefore, the test angle of this experiment was decided to be 20° only. The required time and intensity differences for 20° imaging were calculated using the combination function based on the author's own psychoacoustic values obtained from a localisation experiment using the above musical stimuli. The details of the localisation experiment and the development of the combination function will not be presented here, but it can be found in the author's internal report [11]. A composition of the test stimuli is shown in Table 1.

<table>
<tr><th></th><th>Mono</th><th>Time</th><th>Combi</th><th>Intensity</th></tr>
<tr><td>Speech</td><td rowspan="3">Physical

Positioning</td><td rowspan="3">0.5 ms</td><td rowspan="3">0.25ms
+
4dB</td><td rowspan="3">8dB</td></tr>
<tr><td>Piano</td></tr>
<tr><td>Trumpet</td></tr>
</table>

Table 1 : Composition of the test stimuli: nine stimuli were produced using different panning methods.

2.3. Listening Room Arrangement

The test was conducted in the ITU-R BS.1116 listening room at the University of Surrey. Two Genelec 1032A loudspeakers were set up with the distance of 3m between each. An additional loudspeaker of the same model was placed at the 20° position so that its image would appear at the same (similar) direction as that of the phantom image. The loudspeaker arrangement for playback is shown in Figure 1. An acoustically transparent curtain was used in order to hide the nature of the test to the listener.

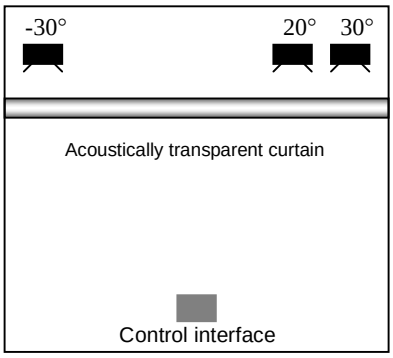

Figure 1 : Loudspeaker arrangements and listening positions for the pair-wise comparison of real and phantom sources

2.4. Test Subject

A total of eight subjects participated in the test. All were experienced listeners, selected from staff members, doctoral students and final year undergraduate students on the University of Surrey's Tonmeister course.

2.5. Elicitation Process

The listener's task was to compare two sound stimuli 'A' and 'B' using control interface provided, and describe perceived differences between them. Sound 'A' was to be the mono source and sound 'B' to be the phantom source. The nature of the stimuli was veiled to the listener. The control interface was designed using Cycling 74's "MSP" software as shown in Figure 2. The listeners were asked to answer a question written as 'B is ___ than A', using their own descriptive terms, and encouraged to spend as much time as they wanted in order to find all the perceptible differences. There were a total of nine trials, whose order was randomised.

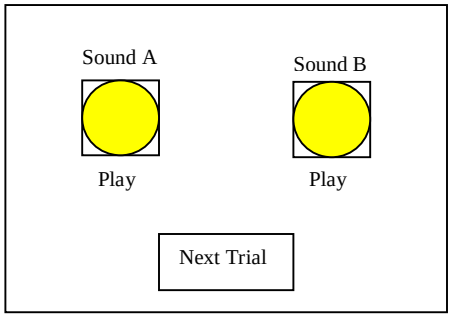

Figure 2 : Layout of the control interface used for comparing mono and phantom images

2.6. Grading Process

For each source, the listener was asked to grade three phantom sources whose images were created by pure time difference, pure intensity difference and the combination of the two, on a 10-point scale from –5 to 5 where 0 represents no difference compared to the mono. The six attribute scales obtained in the previous elicitation test were used for each sound source. The definitions on each scale were provided in the instruction in order to clarify the meanings of the terms. The listeners were instructed to compare all the phantom sources together on each scale in order to understand the basic differences among them first, and then grade each of them compared to the corresponding mono sources. This was in order to make sure the listener's systematic judgement for ranking of the phantom sources. The order of panning methods used for the phantom sources was randomised for each sound source. The control interface used in the test is shown in Figure 3.

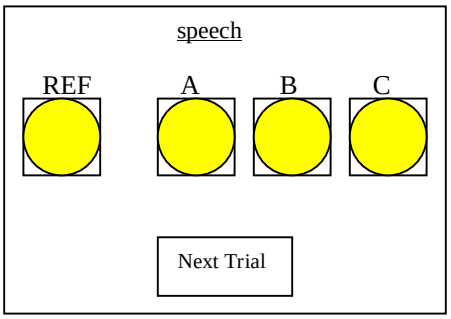

Figure 3 : Layout of the control interface used in the grading test

3. RESULTS AND DISCUSSION

3.1. Elicitation test

A number of descriptive terms were elicited from the subjects for each panning method of each sound source. The descriptive terms were combined for the corresponding sound source regardless of the panning method, in order to obtain general attribute scales for each sound source to be used for investigating the effects of the panning method in the grading test. All the descriptive terms obtained were classified into two broad groups: spatial and timbral attributes. The terms were then grouped into a few detailed attribute scales based on the similarity in meaning. The choice of the specific terms used for the spatial attribute scales is based on Rumsey [12]'s classification of spatial attributes, whereas that for the timbral ones is on Gabrielsson and Sjogren [13]'s classification of timbral attributes. Table 2 and 3 show the summaries of the elicited descriptive terms and the grouped attribute scales for each sound source. The number in the bracket represents the number of subjects who used the specific term. As can be seen, the most dominant attributes for all stimuli were the image focus and the

brightness. The fact that some of the scales were drawn from only one or two subjective descriptions did not matter at this stage since the purpose of this process was to obtain all the possible scales that are to be useful for grading the magnitudes of sound stimuli effects. It is interesting to note that all the subjective terms are generally grouped into the same attribute scales for each sound source in both spatial and timbral situations. The definitions of each attribute scale obtained are shown below:

- ***Image focus*** : how easy it is to determine the position of a source, i.e. focused / defocused
- ***Image width*** : the perceived width of the source, i.e. broad / narrow
- ***Image distance*** : the perceived distance from the listener to the sound source, i.e. close / distant
- ***Brightness*** : depends on the level of high frequencies, i.e. bright / dull
- ***Hardness*** : depends on the level of mid-high frequencies (e.g. 2000Hz to 4000Hz), i.e. hard/soft.
- ***Fullness*** : depends on the level of low frequencies, i.e. full/thin

3.2. Grading Process

A multivariate ANOVA (MANOVA) test was carried out on the data obtained from the grading test, in order to investigate the effect of panning method and sound source on each image attribute. Dependent variables included all the attributes, and independent variables were panning method and sound source. It was recognised that the original data were dependent on the listeners' subjective interpretations of the five point interval scales. For instance, even if the ranking among the panning method was the same, the range of the scores given varied depending on the subject. Therefore, it was decided to normalise the original data based on ITU-R BS.1116 Rec. [14]. This certainly reduced variances among the subjects as a result. Figures from 4 to 9 show the profile plots of the normalised data with respect to both sound source and panning method. From an instant observation, it can be seen that the most effective attributes are 'image focus' and 'image width'. While the scores for those attributes appear to be quite large in the range of about –4 to +4, those for the other attributes are in the range of only about –1 to +1. Some of the attributes appear to have a similar tendency in their plots, e.g. image focus - image width (opposite direction) and brightness – hardness. This could be an indication that those attributes are strongly correlated to each other and can be subsumed in one effective attribute scale.

There were five main questions arising for the analysis of each attribute:

1. Is the type of sound source significant for the perception of a specific attribute?
2. Is the choice of panning method significant for the perception of a specific attribute?
3. For a specific sound source, how do different panning methods affect the magnitude of perceived effect?
4. For a specific panning method, how do different sound sources affect the magnitude of perceived effect?
5. Is there any correlation between attributes?

The answers for the first two questions can be found in the results of the MANOVA test shown in Table 4. If the MANOVA results showed there was a significant main effect in either sound source or panning method, the Post Hoc multiple comparison tests shown in Table 5 and 6 would examine the specific factors that caused the significance in the main effect. The third and fourth questions can be answered from the Post Hoc multiple comparison tests that are limited in one fixed factor at each test (either fixed source or fixed panning method). Since the whole results of these tests are rather too vast, the summary of significant levels for each pair-wise comparison is shown instead in Table 7 and 8. The last question can be considered by the 'bivariate correlation test' and the 'principal components analysis'.

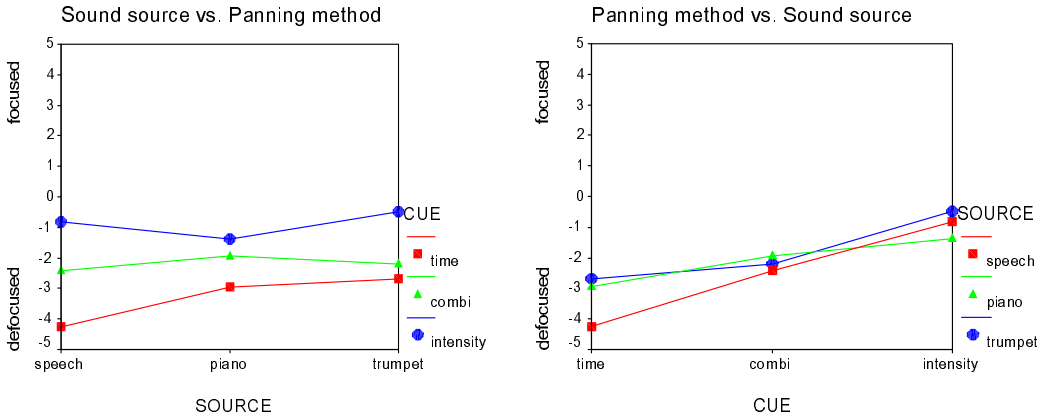

Figure 4 a and b : Mean plots for 'Image Focus' attribute: sound source vs. panning method and vice versa

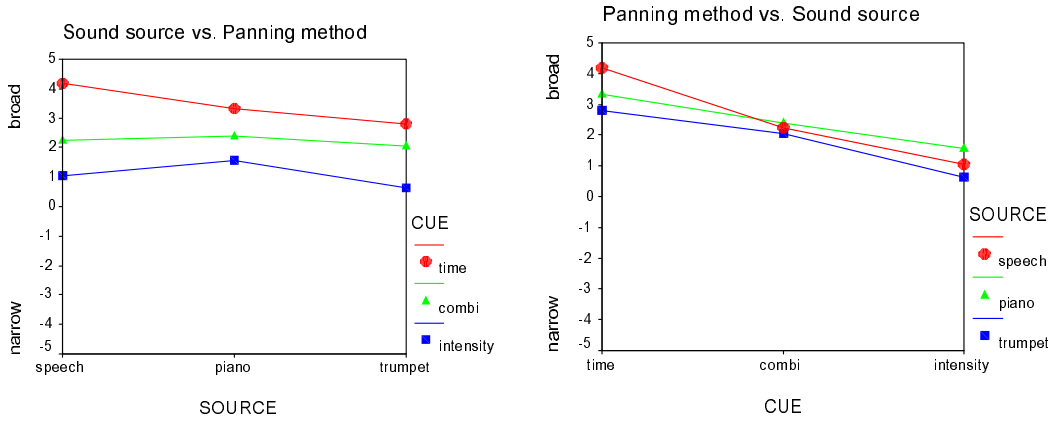

Figure 5 a and b : Mean plots for 'Image Width' attribute: sound source vs. panning method and vice versa

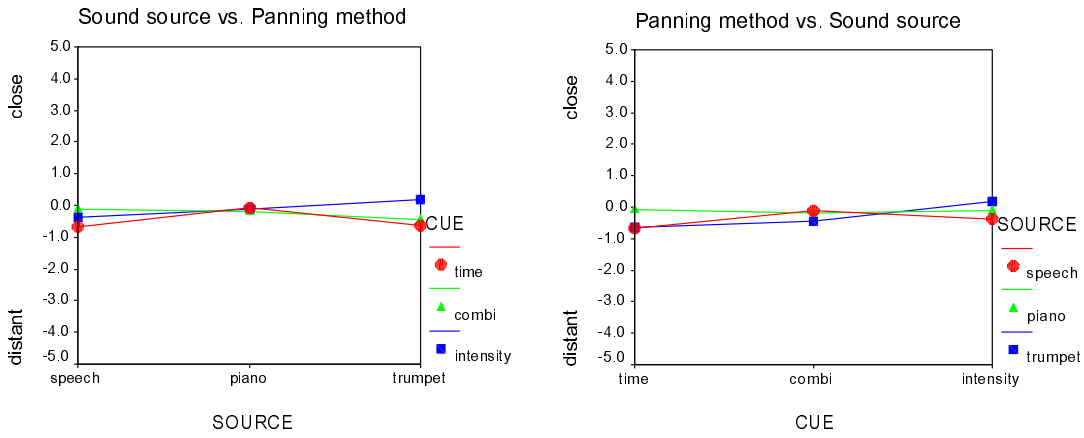

Figure 6 a and b : Mean plots for 'Image Distance' attribute: sound source vs. panning method and vice versa

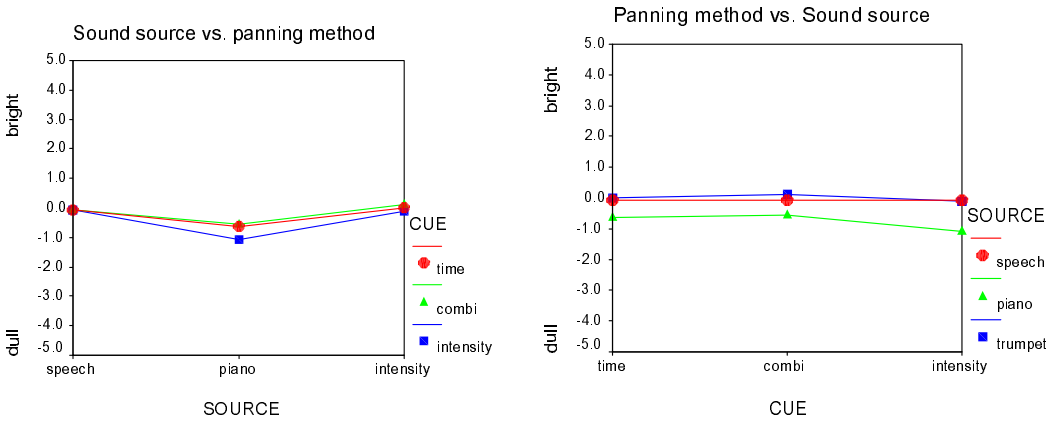

Figure 7 a and b : Mean plots for 'Brightness' attribute: sound source vs. panning method and vice versa

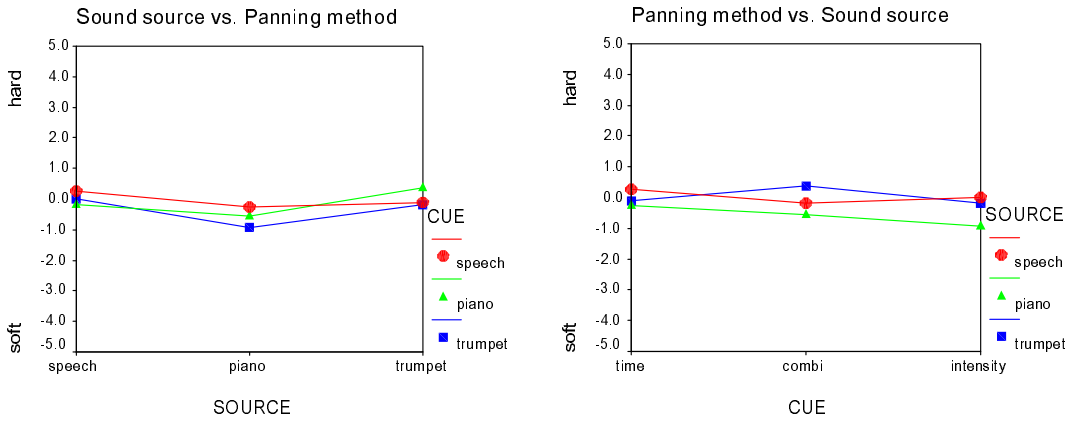

Figure 8 a and b : Mean plots for 'Hardness' attribute: sound source vs. panning method and vice versa

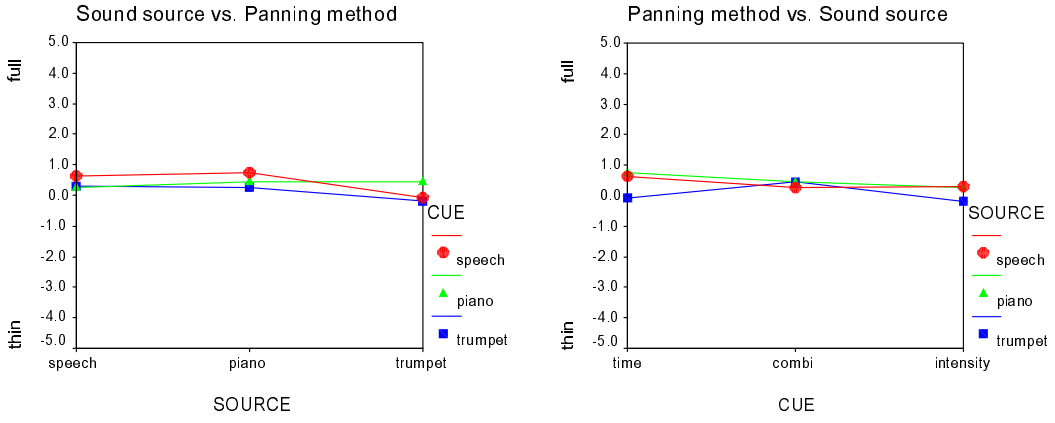

Figure 9 a and b : Mean plots for 'Fullness' attribute: sound source vs. panning method and vice versa

3.2.1. The main effect of sound source and panning method

Table 4 shows that the main effect of sound source (ignoring the type of panning method) was significant for all image attributes (p<0.05) except for image distance (p=0.510). This means that at least a pair of sound sources had a significant difference apart from the image distance attribute, and the detailed effect can be examined by looking at the p values for each pair of sound sources in Table 5. It can be seen that for image focus attribute, speech and trumpet had a significant difference (p=0.010) whereas the other pairs were not significant. For image width, significant difference can be found between speech and trumpet (p=0.014) and between piano and trumpet (p=0.045). For both brightness and hardness, piano was significantly different from speech (p=0) and trumpet (p=0). For fullness, only piano and trumpet had a significant effect (p=0.039). The significant effect found between sound sources for the timbral attributes seem to be a natural result to a degree because each sound source has different timbral characteristics. However, this result cannot be generalised since only low frequency piano and trumpet sources were used for this test. If high frequency sources had been used, the result might have differed. It can be seen in Table 4 that only image focus and image width had significant effects of panning method when sound source is ignored. All pairs of panning methods for those attributes have highly significant effects (p=0) as shown in Table 6. The result suggests that panning method has a main effect on spatial perception rather than timbral one.

3.2.2. Panning method effect for each sound source

Figure 4a clearly shows that for every source type, the magnitude of perceived effect for the image focus attribute increases in the order of intensity, combination and time panning (Intensity<Combination<Time), which agrees with the literatures suggesting the difficulty of accurate localisation with pure time difference. This also confirms that coincident stereophonic microphone technique has the advantage of spaced omni microphone technique in the ability of creating stable image. However, the effect between each panning method depends on the sound source. From Table 7, the effect of panning method on the image focus appears to be most significant for the speech source since each pair has significance at 0%. For piano source, only the difference between time and intensity pannings was significant (p=0.04). For trumpet, the difference between time and combination pannings could be neglected since it is insignificant (p=1.000).

It can be seen in Figure 4 and 5 that the image width attribute has a similar tendency of the panning method effect to the image focus. The perceived width of image increases in the order of intensity, combination and time panning for every sound source. This can be interpreted as an indication that images created by spaced omni microphone techniques would appear wider than that created by coincident techniques. Near-coincident techniques would be placed in between those. For the speech source, each pair of panning methods has a significant difference as shown in Table 7. The difference between combination and intensity pannings for the piano source can be neglected (p=0.079) while the other pairs appear to be significant. Combination and time pannings for the trumpet source also have a small difference in their effects (p=0.531), unlike the other pairs. Table 7 shows that there are no significant effect between any panning methods for any sound source for the image distance and the brightness attributes. The only significant difference for the hardness attribute was observed between time and intensity for piano source (p=0.030). For the fullness attribute, only the difference between combination and intensity for the trumpet source appeared to be significant (p=0.023).

3.2.3. Sound source effect for each panning method

As can be checked in Table 8, the image focus and width attributes had a similar sound source effect for each panning method. That is, the effect of the speech source was significant compared to the other sources for the time panning, while there was no considerable sound source effect found for the other panning methods. Figure 4b and 5b show that the trumpet source had the smallest effect for the image focus and width. This might look rather contradictory to the literature pointing out the difficulty of localisation with purely continuous source. However, the nature of this test was not to compare the three different sound sources directly with each other, but to compare those with each corresponding mono source. If the trumpet source had been originally difficult to localise, there

might have not been much difference detected between the mono source and the phantom source. On the other hand, it was relatively easier to listen to the speech source and the image of the mono source would have been much more distinctive than the phantom source. In this respect, the result obtained here appears to be promising.

It is interesting to observe that the brightness and hardness attributes had a similar sound source effect as well for each panning method. It can be seen from Table 8 that both attributes showed significant effects for the piano source when the intensity panning was used. This means that the image became significantly duller or softer when the piano source was panned with pure intensity difference cue. This might be due to a possible comb filter effect especially at upper harmonics, caused when the intensity difference between the loudspeakers is transmitted to both ears with acoustic crosstalk of time delay. However, it is difficult to generalise the above result because the piano source used in this test was only a single C3 note, which has a low fundamental frequency. The result might have been differed if a piano source of a higher note had been used. For the image distance and fullness attributes, there was no considerable sound source effect found for any panning method.

3.2.4. Correlation between each attributes

The similarities that were found between some attributes in the above sections gave rise to an expectation that those attributes are strongly related to each other. A bivariate correlation test was performed in order to examine the relationship between each attribute. The result of the correlation analysis, shown in Table 9 reports that the image focus and image width attributes are negatively correlated at a high level (correlation coefficient = -0.893). This means that more defocused images had a linear relationship with wider images. From the similar results obtained for the brightness and hardness in the above section, it was expected that those attributes would be also strongly correlated, but they appear to be only moderately correlated (0.494). The strong correlation between the image focus and width suggests that they could be simultaneously considered in a combined scale if the same experimental conditions are applied.

3.2.5. Principal Components Analysis (PCA)

The result of a principal component analysis that was carried out in order to confirm if the image focus and width attributes could be subsummed in one scale is shown in Table 10. As can be seen, there are only five effective components. This means that at least two attribute were considered to be principally one component. Figure 10 clearly shows that the image focus and width attributes can be considered as one component. The strong relationship between those two attributes might be simply a natural phenomenon at least for the sound sources used in this experiment. However, it might give rise to a doubt if there was any bias occurred in understanding the definitions of those two attributes. It is possible that the meanings of the terms used for the scales were not clear enough for the subjects and caused some confusion for judgement. This needs to be further examined.

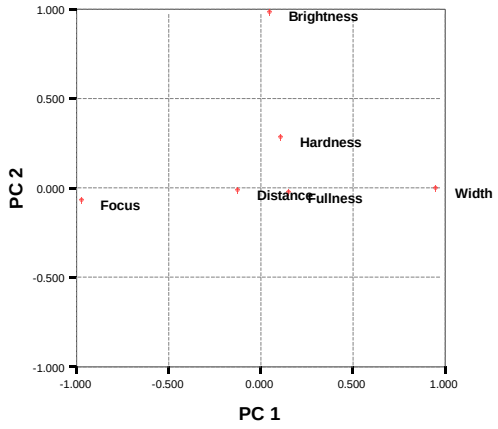

Figure 10 : Scatter plot of the principal component analysis: PC1 represents a hidden dimension that is constituted by image focus and image width.

4. SUMMARY AND CONCLUSIONS

A set of subjective listening tests was conducted in order to elicit and grade the attributes of phantom images in a two channel stereophonic listening environment. There were three different sound sources: speech, transient piano and continuous trumpet. The directions of the phantom images were manipulated by three different panning methods: pure time panning, pure intensity panning and combination of the two. Firstly, the subjects described perceived differences between the phantom image and the real image using their own terms. The subjective terms were then categorised into six groups, which were to be used as attribute scales for the grading test. Finally, the same subjects graded the magnitude of perceived differences between the phantom images and the real images using the obtained attribute scales. The results obtained from this experiment are summarised below:

1. Six elicited attributes were common for all sounds. Three of them were spatial attributes (image focus, image width and image distance) and three were timbral attributes (brightness, hardness and fullness).

2. Image focus and image width were the most dominant attributes in general. The magnitudes of effect of the other attributes were relatively small.

3. Image focus and image width attributes were strongly correlated.

4. Regardless of panning method, the type of sound source had a significant effect on the perceived difference between real images and phantom images for all attributes except image distance.

5. Regardless of sound source type, only image focus and image width attributes had significant panning method effect on the perceived difference between real images and phantom images.

6. For image focus and image width scales, the magnitude of the difference between real image and phantom image was greatest for time difference panning, and smallest for intensity difference panning for every source type.

7. For brightness and hardness attributes, the largest difference between real and phantom images was caused by intensity panning on the piano source.

8. For fullness attribute, the difference between real and phantom images was the largest with time panning on speech and piano sources. The other conditions caused insignificant difference.

In conclusion, the results from the pilot experiment summarised above suggest that the perception of 2-channel phantom image attributes depend not only on the combination of time and intensity differences, but also on the type of sound source. The results form the basis for further experiments to evaluate the effect of interchannel crosstalk resulting from different designs of multichannel microphone techniques and recording conditions.

5. LIMITATIONS AND IMPROVEMENTS

The frequency range of stimuli was limited to low frequencies for musical sound sources. There were also high frequency sources used in the localisation test, but they were excluded in the elicitation and grading tests as the time allowed for the experiment was limited. The small frequency range of stimuli might have affected the results of the elicitation and grading tests, especially for the timbral attribute perceptions. If higher frequency sources had been also used, there might have been different gradings for the given attributes.

The stimuli used in this experiment were considered to be rather difficult to listen to. Some subjects found it was tiring and tedious to concentrate on listening to them for long periods, which might have affected the results to some degree. The nature of the sources was originally chosen in order to strictly separate the transient and continuous natures of sound. However, it was recognised in the course of the experiment that it would be more realistic to use musical extracts rather than single notes.

It was found that even though written definitions for the attributes were provided in the instructions, extra verbal explanations were required for some subjects. Some subjects did not seem to be fully familiar with the given attributes, therefore a training session might have helped to familiarise the subject with the task. It is supposed

that the strong correlation between the image focus and width attributes could have been caused by the lack of training session.

A larger number of subjects would be required for further experiments for more statistically reliable results.

6. FUTURE WORK

Future experiments will be based on three frontal channels and involve musical sources. They will include a similar form of elicitation with grading tests, and will investigate the effect of various combinations of time and intensity differences across three-channel. Another consideration might be listener's preference between the qualities of mono image without any interchannel crosstalk, and crosstalk phantom image. The results of these future experiments should aid the design of multichannel microphone techniques that take into account the interchannel crosstalk.

7. ACKNOWLEDGEMENTS

The authors wish to thank the staff and students at University of Surrey's Tonmeister course for having participated in the subjective listening tests, and Dr.Slawomir Zielinski for helping with statistical analysis.

8. REFERENCES

[1] Freyman, R.L., Clifton, R.K., Litovski, R.Y. (1991) Dynamic process in the precedence effect. *Journal of the Acoustical society of America*, 90, pp.874-884.

[2] Perrott, D.R., Marlborough, K. and Merrill, P. (1988) Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate. *Journal of the Acoustical Society of America*, 85, pp.282-288.

[3] Streicher, R. and Everest, F. A. (1998) *The New Stereo Soundbook, 2nd Ed.* (CA: TAB Books)

[4] Berg, J. and Rumsey, F. (1999) Spatial attribute identification and scaling by Repertory Grid Technique and other methods. In Proceedings of the AES *15th International Conference, Rovaniemi*, pp.51-66. Audio Engineering Society

[5] Kjeldsen, A. (1998) The measurement of personal peference by Repertory Grid Technique. *104th Audio Engineering Society Convention*, Preprint 4685.

[6] Rumsey, F. (2001) *Spatial Audio* (Oxford: Focal Press)

[7] Bech, S. (1999) Methods for subjective evaluation of spatial characteristics of sound. In *Proceedings of the AES 16th International Conference, Rovaniemi*, Finland, pp.487-504. Audio Engineering Society

[8] Wallach, H., Newman, E.B. and Rosenzweig, M.R. (1949) The precedence effect in sound localisation. *American Journal of Psychology*, 52, pp.315-336.

[9] Rackerd, B. and Hartmann, W.M. (1986) Localisation of sound in rooms, III: Onset and duration effects. *Journal of the Acoustical Society of America*, 80, pp.1695-1706.

[10] Yost, W.A., Wightman, F.L. and Green M.D. (1971) Lateralisation of Filtered Clicks. *Journal of the Acoustical Society of America*, 50, pp. 1526-1531.

[11] Lee, H.K. (2004) M.Phil Thesis, University of Surrey (www.surrey.ac.uk/soundrec)

[12] Rumsey, F. (2002) Spatial quality evaluation for reproduced sound: Terminology, Meaning, and a Scene-based paradigm. *Journal of the Audio Engineering Society*, 50, pp.651-666.

[13] Gabrielsson, A. and Sjogren, H. (1979) Perceived sound quality of sound-reproducing systems. *Journal of the Acoustical Society of America*, 65, pp.1019-1033.

[14] ITU-R BS.1116. (1994) Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. *International Telecommunications Union, Reccommendation ITU-R 1116*, pp.276-297

Sound source	*Spatial Attributes*	
	Group	**Descriptive terms**
SPEECH	*Image focus*	Less localised (4) Less focused (2) Less present (1) Less stable (1) Phasier (2) Less Coherent (1)
	Image width	Wider (7)
	Image distance	More distant (1) Further away (1)
PIANO	*Image focus*	Harder to locate (2) Less defined (2) Less focused (1) More elevated (1) More reverberant (2)
	Image width	Wider (6)
	Image distance	More distant (1) Closer (1)
TRUMPET	*Image focus*	Harder to locate (2) Less focused (2) Less solid (1) More diffused (1)
	Image width	Wider (1)
	Image distance	More distant (1) Further away (1) Closer (1)

Table 2 : Summary of spatial attribute scales drawn from the elicited descriptive terms

Sound source	*Timbral Attributes*	
	Group	**Descriptive terms**
SPEECH	*Brightness*	Less bright (2) More cloudy (1) Duller (1) Muddier (1) Less breathy (1)
	Hardness	Softer (1)
	Fullness	Fuller (1) Bassier (1) Less bassy (1) Less body (1)
PIANO	*Brightness*	Brighter (1) Duller (2) Less dark (1) Less bright (1) Less toppy (1) Less harsh (1) Less bassy (1)
	Hardness	Softer (1)
	Fullness	Less attack (2) Less punch (1) Bassier (1) Fuller (1)
TRUMPET	*Brightness*	Brighter (3) Duller (1) More present (1) More nasal (1)
	Hardness	Stronger (1) Harsher (1)
	Fullness	Fuller (1) Less bassy (2)

Table 3 : Summary of timbral attribute scales drawn from the elicited descriptive terms

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	image focus	84.423	8	10.553	16.290	.000	.674
	image width	80.162	8	10.020	15.296	.000	.660
	image distance	5.133	8	.642	.967	.470	.109
	brightness	9.779	8	1.222	3.745	.001	.322
	hardness	10.028	8	1.253	4.273	.000	.352
	fullness	5.765	8	.721	2.103	.048	.211
Intercept	image focus	324.998	1	324.998	501.676	.000	.888
	image width	366.483	1	366.483	559.430	.000	.899
	image distance	5.346	1	5.346	8.056	.006	.113
	brightness	5.298	1	5.298	16.230	.000	.205
	hardness	2.347	1	2.347	8.002	.006	.113
	fullness	7.069	1	7.069	20.633	.000	.247
Sound source	image focus	6.069	2	3.034	4.684	.013	.129
	image width	6.600	2	3.300	5.037	.009	.138
	image distance	.903	2	.452	.681	.510	.021
	brightness	8.341	2	4.171	12.778	.000	.289
	hardness	5.840	2	2.920	9.955	.000	.240
	fullness	2.348	2	1.174	3.427	.039	.098
Panning method	image focus	69.021	2	34.510	53.271	.000	.628
	image width	67.715	2	33.858	51.683	.000	.621
	image distance	1.521	2	.760	1.146	.325	.035
	brightness	.813	2	.406	1.245	.295	.038
	hardness	1.444	2	.722	2.462	.093	.072
	fullness	1.313	2	.656	1.915	.156	.057
Sound source * Panning method	image focus	9.333	4	2.333	3.602	.010	.186
	image width	5.847	4	1.462	2.231	.076	.124
	image distance	2.708	4	.677	1.020	.404	.061
	brightness	.625	4	.156	.479	.751	.029
	hardness	2.743	4	.686	2.338	.065	.129
	fullness	2.104	4	.526	1.535	.203	.089
Error	image focus	40.813	63	.648			
	image width	41.271	63	.655			
	image distance	41.811	63	.664			
	brightness	20.563	63	.326			
	hardness	18.480	63	.293			
	fullness	21.584	63	.343			
Total	image focus	450.233	72				
	image width	487.916	72				
	image distance	52.290	72				
	brightness	35.639	72				
	hardness	30.855	72				
	fullness	34.418	72				
Corrected Total	image focus	125.236	71				
	image width	121.434	71				
	image distance	46.944	71				
	brightness	30.342	71				
	hardness	28.507	71				
	fullness	27.349	71				

Table 4 : Multivariate ANOVA results table with Image focus, Image width, Image distance, Brightness, Hardness and Fullness as dependent variables and Panning method and Sound source as independent variables

Multiple Comparisons

Bonferroni

Dependent Variab	(I) Sound sourc	(J) Sound sourc	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
image focus	speech	piano	-.4192	.23235	.228	-.9906	.1523
		trumpet	-.7071*	.23235	.010	-1.2786	-.1356
	piano	speech	.4192	.23235	.228	-.1523	.9906
		trumpet	-.2879	.23235	.660	-.8594	.2836
	trumpet	speech	.7071*	.23235	.010	.1356	1.2786
		piano	.2879	.23235	.660	-.2836	.8594
image width	speech	piano	.1029	.23365	1.000	-.4718	.6776
		trumpet	.6875*	.23365	.014	.1128	1.2622
	piano	speech	-.1029	.23365	1.000	-.6776	.4718
		trumpet	.5846*	.23365	.045	.0099	1.1593
	trumpet	speech	-.6875*	.23365	.014	-1.2622	-.1128
		piano	-.5846*	.23365	.045	-1.1593	-.0099
image distance	speech	piano	-.2721	.23517	.755	-.8505	.3063
		trumpet	-.1054	.23517	1.000	-.6838	.4730
	piano	speech	.2721	.23517	.755	-.3063	.8505
		trumpet	.1667	.23517	1.000	-.4118	.7451
	trumpet	speech	.1054	.23517	1.000	-.4730	.6838
		piano	-.1667	.23517	1.000	-.7451	.4118
brightness	speech	piano	.6888*	.16492	.000	.2831	1.0944
		trumpet	-.0625	.16492	1.000	-.4681	.3431
	piano	speech	-.6888*	.16492	.000	-1.0944	-.2831
		trumpet	-.7513*	.16492	.000	-1.1569	-.3456
	trumpet	speech	.0625	.16492	1.000	-.3431	.4681
		piano	.7513*	.16492	.000	.3456	1.1569
hardness	speech	piano	.6042*	.15635	.001	.2196	.9887
		trumpet	.0000	.15635	1.000	-.3845	.3845
	piano	speech	-.6042*	.15635	.001	-.9887	-.2196
		trumpet	-.6042*	.15635	.001	-.9887	-.2196
	trumpet	speech	.0000	.15635	1.000	-.3845	.3845
		piano	.6042*	.15635	.001	.2196	.9887
fullness	speech	piano	-.0833	.16897	1.000	-.4989	.3323
		trumpet	.3346	.16897	.156	-.0810	.7502
	piano	speech	.0833	.16897	1.000	-.3323	.4989
		trumpet	.4179*	.16897	.048	.0023	.8335
	trumpet	speech	-.3346	.16897	.156	-.7502	.0810
		piano	-.4179*	.16897	.048	-.8335	-.0023

Based on observed means.

*. The mean difference is significant at the .05 level.

Table 5 : Multivariate Post Hoc multiple comparison test for sound source

Multiple Comparisons

Bonferroni

Dependent Varia	(I) panning meth	(J) panning meth	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
image focus	time	combi	-1.1042*	.23235	.000	-1.6756	-.5327
		intensity	-2.3958*	.23235	.000	-2.9673	-1.8244
	combi	time	1.1042*	.23235	.000	.5327	1.6756
		intensity	-1.2917*	.23235	.000	-1.8631	-.7202
	intensity	time	2.3958*	.23235	.000	1.8244	2.9673
		combi	1.2917*	.23235	.000	.7202	1.8631
image width	time	combi	1.2292*	.23365	.000	.6545	1.8038
		intensity	2.3750*	.23365	.000	1.8003	2.9497
	combi	time	-1.2292*	.23365	.000	-1.8038	-.6545
		intensity	1.1458*	.23365	.000	.5712	1.7205
	intensity	time	-2.3750*	.23365	.000	-2.9497	-1.8003
		combi	-1.1458*	.23365	.000	-1.7205	-.5712
image distance	time	combi	-.2083	.23517	1.000	-.7868	.3701
		intensity	-.3542	.23517	.411	-.9326	.2243
	combi	time	.2083	.23517	1.000	-.3701	.7868
		intensity	-.1458	.23517	1.000	-.7243	.4326
	intensity	time	.3542	.23517	.411	-.2243	.9326
		combi	.1458	.23517	1.000	-.4326	.7243
brightness	time	combi	-.0625	.16492	1.000	-.4681	.3431
		intensity	.1875	.16492	.780	-.2181	.5931
	combi	time	.0625	.16492	1.000	-.3431	.4681
		intensity	.2500	.16492	.404	-.1556	.6556
	intensity	time	-.1875	.16492	.780	-.5931	.2181
		combi	-.2500	.16492	.404	-.6556	.1556
hardness	time	combi	.0833	.15635	1.000	-.3012	.4679
		intensity	.3333	.15635	.111	-.0512	.7179
	combi	time	-.0833	.15635	1.000	-.4679	.3012
		intensity	.2500	.15635	.344	-.1345	.6345
	intensity	time	-.3333	.15635	.111	-.7179	.0512
		combi	-.2500	.15635	.344	-.6345	.1345
fullness	time	combi	.0625	.16897	1.000	-.3531	.4781
		intensity	.3125	.16897	.207	-.1031	.7281
	combi	time	-.0625	.16897	1.000	-.4781	.3531
		intensity	.2500	.16897	.432	-.1656	.6656
	intensity	time	-.3125	.16897	.207	-.7281	.1031
		combi	-.2500	.16897	.432	-.6656	.1656

Based on observed means.

*. The mean difference is significant at the .05 level.

Table 6 : Multivariate Post Hoc multiple comparison test for Panning method

			SOUND SOURCE		
Dependent Variable	**Panning method**		**Speech**	**Piano**	**Trumpet**
Image Focus	Time	Combi	.000	.087	1.000
		Intensity	.000	.004	.001
	Combi	Time	.000	.087	1.000
		Intensity	.000	.606	.011
	Intensity	Time	.000	.004	.001
		Combi	.000	.606	.011
Image Width	Time	Combi	.000	.036	.531
		Intensity	.000	.000	.002
	Combi	Time	.000	.036	.531
		Intensity	.002	.079	.042
	Intensity	Time	.000	.000	.002
		Combi	.002	.079	.042
Image Distance	Time	Combi	.366	1.000	1.000
		Intensity	1.000	1.000	.134
	Combi	Time	.366	1.000	1.000
		Intensity	1.000	1.000	.345
	Intensity	Time	1.000	1.000	.134
		Combi	1.000	1.000	.345
Brightness	Time	Combi	1.000	1.000	1.000
		Intensity	1.000	.498	1.000
	Combi	Time	1.000	1.000	1.000
		Intensity	1.000	.348	1.000
	Intensity	Time	1.000	.498	1.000
		Combi	1.000	.348	1.000
Hardness	Time	Combi	.165	.634	.465
		Intensity	.776	.030	1.000
	Combi	Time	.165	.634	.465
		Intensity	1.000	.411	.335
	Intensity	Time	.776	.030	1.000
		Combi	1.000	.411	.335
Fullness	Time	Combi	.937	.855	.083
		Intensity	1.000	.281	1.000
	Combi	Time	.937	.855	.083
		Intensity	1.000	1.000	.023
	Intensity	Time	1.000	.281	1.000
		Combi	1.000	1.000	.023

Table 7 : Multiple comparison between panning methods against each sound source: the numerical value represents significance level *p*.

			PANNING METHOD		
Dependent Variable	**Sound Source**		**Time cue**	**Combi cue**	**Intensity cue**
Image Focus	Speech	Piano	.017	.294	.734
		Trumpet	.004	1.000	1.000
	Piano	Speech	.017	.294	.734
		Trumpet	1.000	1.000	.230
	Trumpet	Speech	.004	1.000	1.000
		Piano	1.000	1.000	.230
Image Width	Speech	Piano	.133	1.000	.818
		Trumpet	.011	1.000	1.000
	Piano	Speech	.133	1.000	.818
		Trumpet	.795	1.000	.141
	Trumpet	Speech	.011	1.000	1.000
		Piano	.795	1.000	.141
Image Distance	Speech	Piano	.294	1.000	1.000
		Trumpet	1.000	1.000	.865
	Piano	Speech	.294	1.000	1.000
		Trumpet	.404	1.000	1.000
	Trumpet	Speech	1.000	1.000	.865
		Piano	.404	1.000	1.000
Brightness	Speech	Piano	.178	.158	.017
		Trumpet	1.000	1.000	1.000
	Piano	Speech	.178	.158	.017
		Trumpet	.114	.031	.026
	Trumpet	Speech	1.000	1.000	1.000
		Piano	.114	.031	.026
Hardness	Speech	Piano	.368	.330	.007
		Trumpet	.724	.062	1.000
	Piano	Speech	.368	.330	.007
		Trumpet	1.000	.001	.034
	Trumpet	Speech	.724	.062	1.000
		Piano	1.000	.001	.034
Fullness	Speech	Piano	1.000	.982	1.000
		Trumpet	.196	.992	.365
	Piano	Speech	1.000	.982	1.000
		Trumpet	.096	1.000	.518
	Trumpet	Speech	.196	.992	.365
		Piano	.096	1.000	.518

Table 8 : Multiple comparison between sound sources against each panning method: the numerical value represents significance level *p*.

Correlations

		FOCUS	WIDTH	DISTANCE	BRIGHT	HARDNESS	FULLNESS
FOCUS	Pearson Correlation	1	-.893**	.180	-.143	-.204	-.239*
	Sig. (2-tailed)	.	.000	.130	.232	.085	.043
	N	72	72	72	72	72	72
WIDTH	Pearson Correlation	-.893**	1	-.271*	.058	.211	.306**
	Sig. (2-tailed)	.000	.	.021	.628	.075	.009
	N	72	72	72	72	72	72
DISTANCE	Pearson Correlation	.180	-.271*	1	-.078	-.240*	-.341**
	Sig. (2-tailed)	.130	.021	.	.516	.042	.003
	N	72	72	72	72	72	72
BRIGHT	Pearson Correlation	-.143	.058	-.078	1	.494**	-.051
	Sig. (2-tailed)	.232	.628	.516	.	.000	.670
	N	72	72	72	72	72	72
HARDNESS	Pearson Correlation	-.204	.211	-.240*	.494**	1	.253*
	Sig. (2-tailed)	.085	.075	.042	.000	.	.032
	N	72	72	72	72	72	72
FULLNESS	Pearson Correlation	-.239*	.306**	-.341**	-.051	.253*	1
	Sig. (2-tailed)	.043	.009	.003	.670	.032	.
	N	72	72	72	72	72	72

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 9 : Pearson correlation analysis between the image attributes

Total Variance Explained

	Initial Eigenvalues			xtraction Sums of Squared Loading			Rotation Sums of Squared Loadings		
Componer	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.389	39.816	39.816	2.389	39.816	39.816	1.888	31.471	31.471
2	1.341	22.345	62.161	1.341	22.345	62.161	1.014	16.905	48.376
3	1.081	18.023	80.184	1.081	18.023	80.184	1.009	16.821	65.198
4	.669	11.142	91.326	.669	11.142	91.326	1.004	16.736	81.934
5	.424	7.064	98.389	.424	7.064	98.389	.986	16.426	98.360
6	.097	1.611	100.000	.097	1.611	100.000	.098	1.640	100.000

Extraction Method: Principal Component Analysis.

Table 10 : Result of the principal component analysis