Lee, Hyunkook

A new time and intensity trade-off function for localisation of natural sound sources

Audio Engineering Society

Convention Paper 8149

Presented at the 128th Convention
2010 May 22–25 London, UK

A new time and intensity trade-off function for localisation of natural sound sources

Hyunkook Lee

LG Electronics, 221, Yangjae-Dong, Seocho-Gu, Seoul, Korea, 137-130
hklworld@lge.com

ABSTRACT

This paper firstly introduces a new set of psychoacoustic values of interchannel time difference (ICTD) and interchannel intensity difference (ICID) required for 10°, 20° and 30° localisation in the conventional stereophonic reproduction, which were obtained using natural sound sources of musical instruments and wideband speech representing different characteristics. Then it discusses the new concept of ICID and ICTD trade-off function developed based on the relationship of the psychoacoustic values. The result of the listening test conducted to verify the performance of the proposed method is also presented.

1. INTRODUCTION

This paper describes a series of subjective experiments carried out to investigate the localisation characteristics of natural sound sources in 2-0 stereophonic reproduction and to develop a novel interchannel time and intensity trade-off function that can be used for the design of stereophonic microphone techniques. Since 1940, a number of stereophonic localisation experiments have been conducted to investigate the independent influence of interchannel time difference (ICTD) or interchannel intensity difference (ICID) on the position of a phantom image perceived between two loudspeakers[1, 2, 3, 4, 5]. The data obtained from these experiments could become the basis for the design of stereophonic microphone techniques since the localisation of phantom images and the relevant stereophonic recording angle (SRA) rely on the Interchannel relationship between the recorded signals. However, the results of those experiments are divergent depending on the type of sound source used (e.g. noise [1], wide-band speech [2, 3, 4], speech and maracas [5]). Arguably, the localisation data that have been most widely quoted for microphone technique design are Simonsen [5]'s, which were obtained using speech and maracas as sources. Williams [6] used Simonsen's data for developing a ICTD-ICID trade-off relationship for the phantom image locations of 10°, 20° and 30° and this relationship was used for the analysis of SRAs for existing stereophonic microphone techniques and the design of his own multichannel microphone technique. However, Simonsen's data differ largely in ICID values from the data obtained by Wittek [4]. This seems to suggest that data obtained in a specific experimental condition might not be directly applied to localisation in a different condition.

Even though sound recordings made with microphone techniques deal with musical sources in most cases, to date experimental data related to the localisation characteristics of musical sources have not been presented apart from those of Simonsen's using maracas. This seems to be due to the complex nature of musical sources, which would make it difficult to strictly control experimental variables. However, it seems more valid to use the data obtained with musical sources for the design of microphone techniques since they are most likely to be encountered in practical situations. From this background, the current localisation experiment was conducted using musical sound sources with different temporal and spectral characteristics. A wideband speech source was also included for comparison with other classic data obtained using speech.

In the current study the independent influences of ICTD and ICID on the localisation of phantom images at the locations of 10°, 20° and 30° were first investigated. Then the significances of the differences between the results obtained with different sound sources were statistically analysed. Finally, it was attempted to develop a novel ICTD and ICID trade-off function and verify its performance through listening test.

2. EXPERIMENTAL DESIGN

2.1. Test method

In most localisation tests, the listener is presented with sound stimuli created with various interchannel differences in regular intervals and asked to judge the locations of the perceived phantom images. This type of method is useful if it is desired to obtain a continuous localisation curve and error bars for perceived angles. However, it was not deemed appropriate for the current listening test because the purpose of this experiment was to obtain useful values of ICTD and ICID required for specific phantom source locations of 10°, 20° and 30°, rather than perceived angles for certain interchannel values. Therefore, this test was designed so that the listener adjusted ICTD or ICID using a slider provided in a control interface to match the positions of the phantom images to those of the markers indicated at +10°, +20° and +30° between the loudspeakers. Time delay or intensity attenuation was applied only to the left channel so that the phantom image appeared only in the centre-right region. In this way it was expected to obtain more accurate values of ICTD and ICID that worked specifically for the desired angles. The listeners were allowed to listen to the stimuli repeatedly until they were completely sure about their decisions. The listeners were asked to face the front consistently while listening to the sounds in order to avoid any effect of head movement.

The control interface was developed using Cycling 74's 'MSP' software shown in **Figure 1**. The range of ICTD that could be applied on the left channel was from 0 to 5ms with the interval of 0.1ms. However, the scale shown in the slider was presented with the representative numbers of 0 to 50 in order to prevent the listener from being biased by their experience and knowledge about the influence of ICTD. The range of ICID scale was from 0 to –100, where 0 represents zero difference and –100 represents -∞dB, and the resulting values were later transformed into the corresponding decibel values. When the listener adjusted ICTD, ICID was maintained at 0, and vice versa. The order of the angles to be judged was randomised for each stimulus in order to avoid a psychological order effect.

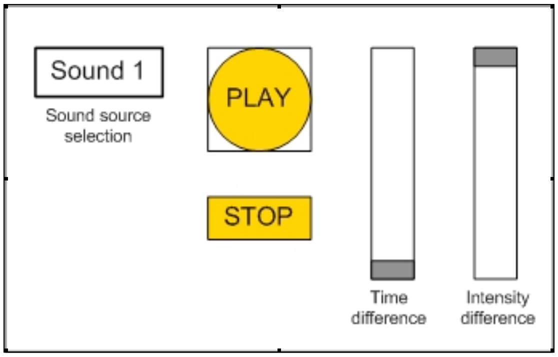

Figure 1 Control interface for the localisation test developed using Cycling 74's MSP software

2.2. Sound stimuli

Five sound stimuli were chosen for this experiment, comprising:

- Piano 'staccato' note of C3 (f_0 = 130Hz)
- Piano 'staccato' note of C6 (f_0 = 1046Hz)
- Trumpet 'sustain' note of Bflat3 (f_0 = 228Hz)
- Trumpet 'sustain' note of Bflat5 (f_0 = 922Hz)
- Continuous wideband speech

The piano and trumpet were chosen in order to examine the effect of temporal characteristics of different musical instruments (i.e. transient vs. continuous). For each musical source, low and high notes were chosen to investigate the effect of spectral characteristics. The speech source was included for its broadband frequency spectrum as well as complex temporal characteristics. Since a number of earlier localisation tests used speech sources, the speech source used in this test was considered to be a useful reference for comparison. It was decided to use single notes instead of performance extracts in order to control the variables strictly. Ideally all the sound sources would have been recorded under an anechoic condition, but this was unavailable. Alternatively, the piano sources were recorded in a small recording booth of studio B at the Metropolis recording studios, using a single cardioid microphone (Schoeps CMC 5-U) placed about 30cm over the hammers for the desired notes. 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 (AKG 414 B-ULS) placed about 1m away from the instrument. The recording space was acoustically isolated and had no audible reverberation. In order to investigate the continuous nature of the trumpet strictly, the onset and offset transients of the trumpet sources were removed by fading in and out the beginning and ending for one second each, and the total duration of the stimulus was four seconds. The speech signal was chosen because it is a mixture of both transient and continuous natures with the wide range of frequencies. The speech recording used was Danish male speech that was anechoically recorded 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 rather than English in order to prevent the listener from paying attention to the language itself.

2.3. Physical setup

The listening test was conducted in the ITU-R BS.1116-compliant listening room at the University of Surrey. Two Genelec 1032A loudspeakers were arranged in the standard 60° configuration, with a distance of 2.4m between them.

2.4. Test subjects

A total of five listeners took part in the test. All were critical and experienced listeners, including research staff and doctoral students at the Institute of Sound Recording of the University of Surrey. Because of the nature of the test requiring highly critical listening skill, it was decided to employ a relatively small number of experienced listeners rather than a large number of inexperienced listeners, and repeat the test three times for each listener in order to ensure a sufficient amount of data for analysis.

3. RESULTS AND DISCUSSIONS

3.1. Localisation characteristics

Figure A.2 shows the results of the localisation test using pure ICTD cue. The plots represent the median values and associated 25th and 75th percentile bars for the subjective data obtained. All the subjects found that it was almost impossible to localise the high note trumpet. The low note trumpet, on the other hand, was reasonably localisable but the subjects still found it difficult to localise easily because the positions of phantom images randomly changed even with a very small head movement. For the transient piano sources, both low and high notes are relatively well localised.

The localisation difficulty for the continuous trumpet notes with pure ICTD seems to confirm the literature reporting the importance of transient component in localisation relying on the time difference between two sounds [7, 8, 9, 10]. This result might also be explained by Rakerd and Hartmann [8]'s 'plausibility hypothesis'. This hypothesis suggests that transient cue is plausible as it wins in a competition with reflections and its ITD becomes apparently detectable, while the steady-state ongoing ITD cue is implausible because it conflicts with room effect.

It appears that the low piano note was localised slightly more certainly than the high piano note. Bank and Green [11] found that for transient noise signals low frequency components below about 2000Hz were essential for accurate localisation in stereophonic reproduction. Based on Yost *et al* [12], this is because low frequency transients excite more space in the cochlear partition than high frequency ones and excite more fibres, thus producing more substantial positional displacement. The high note piano used in this experiment has a complex tonal nature containing lower harmonics. However, the low note piano has richer low frequency components by its nature and this would have led to a better localisation certainty.

The speech source appears to have the best localisation certainty in general. This seems to be due to the fact that the continuous speech source has consecutive transients at every syllable change as well as wide frequency range with the fundamental frequency of about 100Hz.

Figure 3 shows the results of the localisation test using a pure ICID cue. The effect of transient characteristics appears to be less dominant in the case of ICID in that the continuous trumpets of both low and high notes were localised reasonably well. This seems to suggest that the continuous nature of a sound is plausible when ICID cue is used for localisation. In general, however, the results of the ICID localisation have a similar tendency to the results of the ICTD localisation. That is, the speech and piano sources were more certainly localised than the trumpet sources.

From the above results, it can be generally seen that the localisation using pure ICID was more stable than that using pure ICTD, which supports the literature. It is interesting to observe that for both ICTD and ICID pannings the size of the error bar becomes greater as the localisation angle moves from 10° to 30°. This seems to be related to the findings of the minimum audible angle (MAA) of Mills [13]. Mills carried out a subjective experiment to measure the smallest angular change of sound source that the listener could just detect, which is the so-called 'minimum audible angle (MAA)', using pure tones and it was found that the MAA became larger as the loudspeaker pair moved away to the side of the listener. In addition, it can also be observed from the results that the shift factors of ICTD or ICID required for the phantom source positions of 10° and 20° have almost constant relationships. However, the shift factor for 30° appears to be much greater that those for 10° and 20°. This phenomenon can also be observed in the classic localisation curves, being almost linear up to 75% of the shift region and becoming exponential as the angle increases further up to 100% [14]. The results of both Mills and the author seem to suggest that in stereophonic reproduction the listener's sensitivity for localising a phantom source decreases as the direction of the source moves from the front to the side.

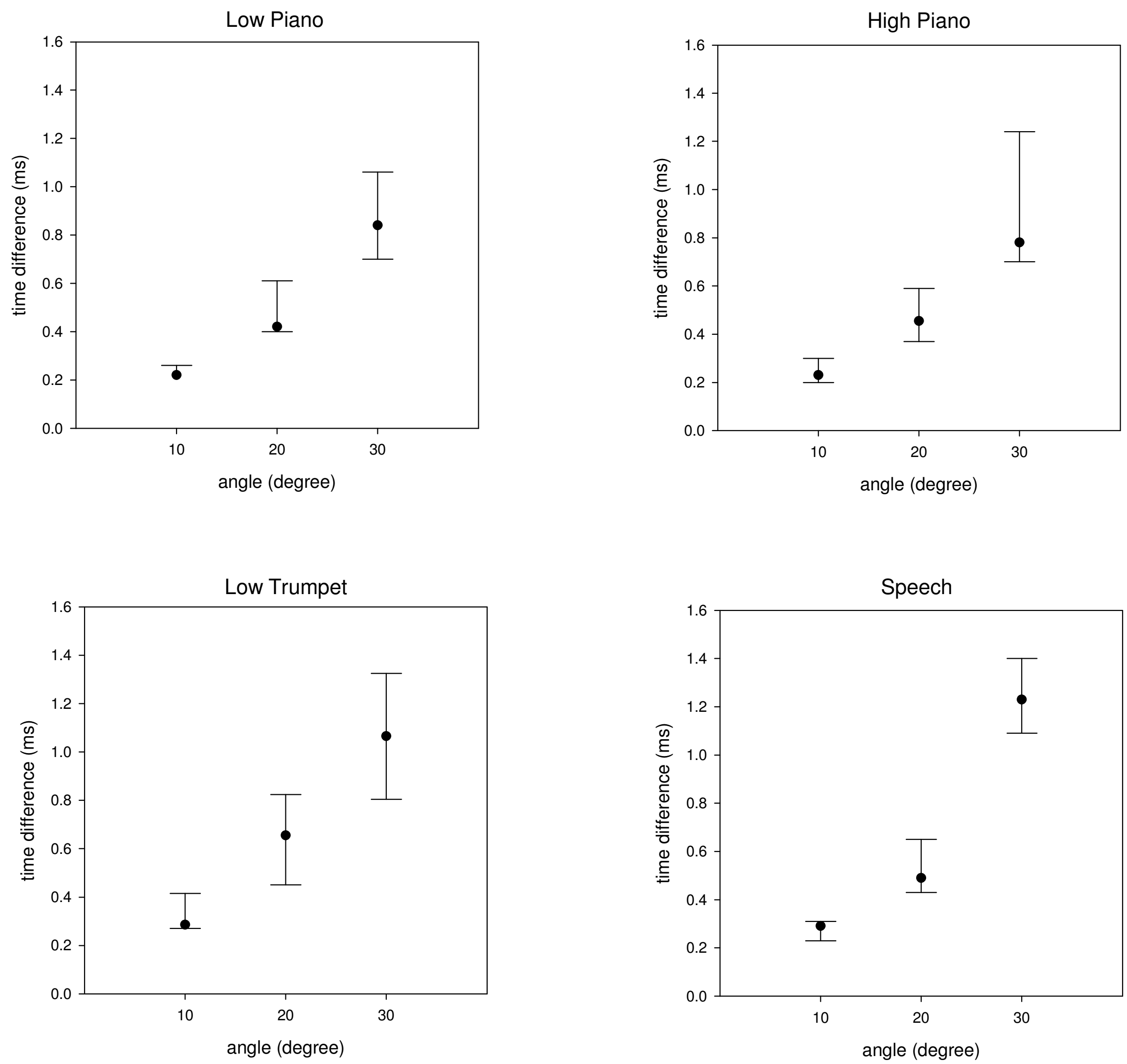

Figure 2 Localisation by pure interchannel time difference (ICTD): Median values and associated 25th to 75th percentile

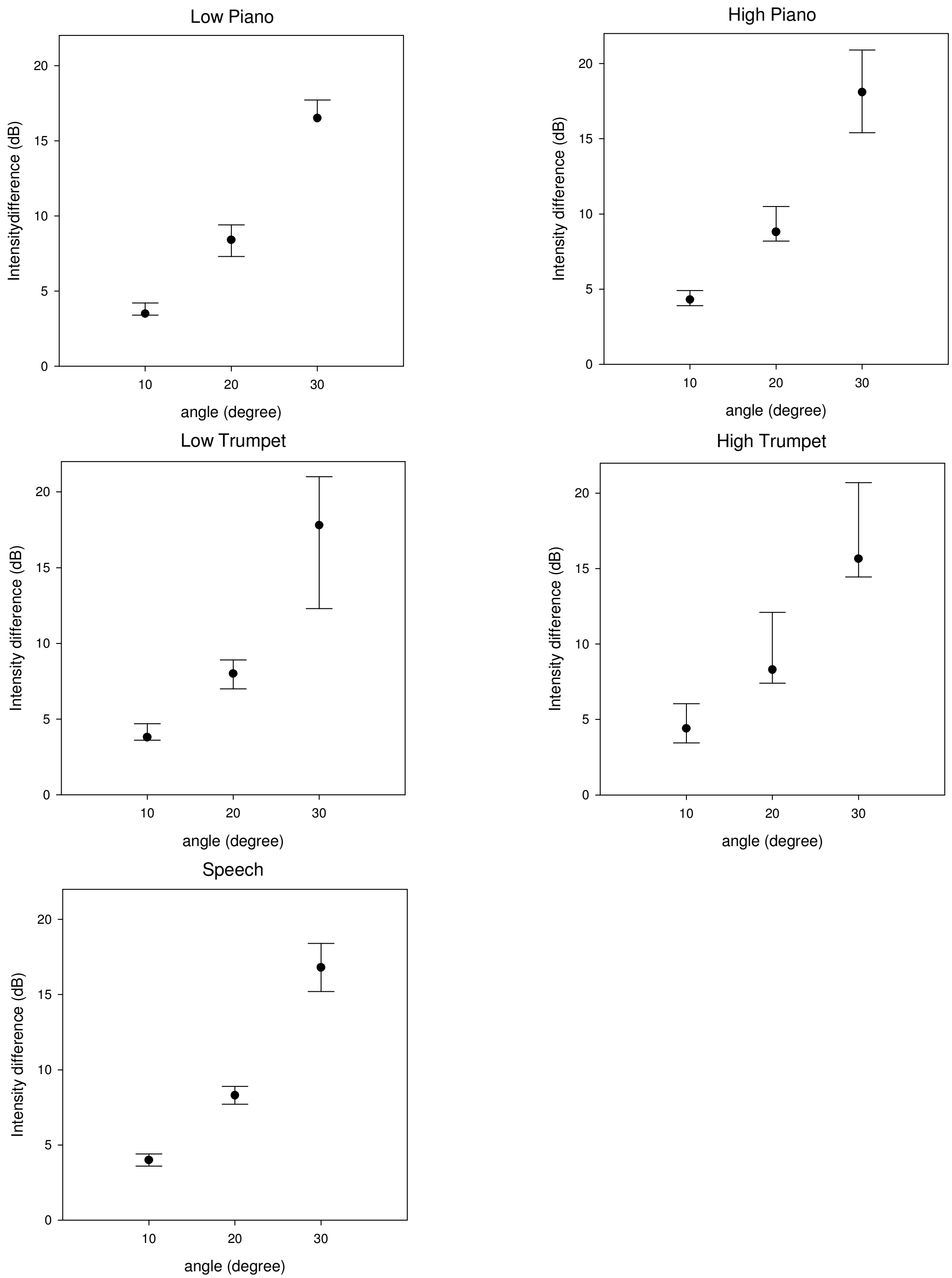

Figure 3 Localisation by pure interchannel intensity difference (ICID): Median values and associated 25th to 75th percentile

3.2. Statistical analysis

In order to examine the significance of the differences observed between the sound sources, the 'Friedman' test, which is a non-parametric statistical test, was carried out. The 'ANOVA' test, which is a parametric test, was not appropriate for this experiment since the panning angle scale (10°, 20° and 30°) had an ordinal nature and the homogeneity of variance required for the ANOVA test was not met in this case. The results of this test shown in **Table 1** indicate that the differences between sound sources were not significant for both ICTD and ICID localisations ($p > 0.05$). Since there is no significant difference between sound sources, it is possible to combine the data for all sound sources. **Table 2** shows the overall median values and associated 25th to 75th percentiles, and these data are plotted in **Figure 4** and **5**. It can be noted again in the unified plots that the localisation certainty tends to become worse as the angle increases. The increase of median value is almost constant up to 20° but becomes steep from 20° to 30°.

	ICTD localisation	ICID localisation
N	36	45
Chi-Square	3.765	5.721
Df	3	4
Asymp. Sig.	0.288	0.221

Table 1 Effects of sound stimuli in time and intensity pannings, analysed using the Friedman test

Panning method	Angle	25th percentile	median	75th percentile
ICID (dB)	10	3.5	**4.0**	4.4
	20	7.6	**8.4**	9.25
	30	15.4	**17.1**	19.6
ICTD (ms)	10	0.22	**0.27**	0.32
	20	0.41	**0.50**	0.72
	30	0.75	**1.1**	1.36

Table 2 Overall median values and 25th to 75th percentiles

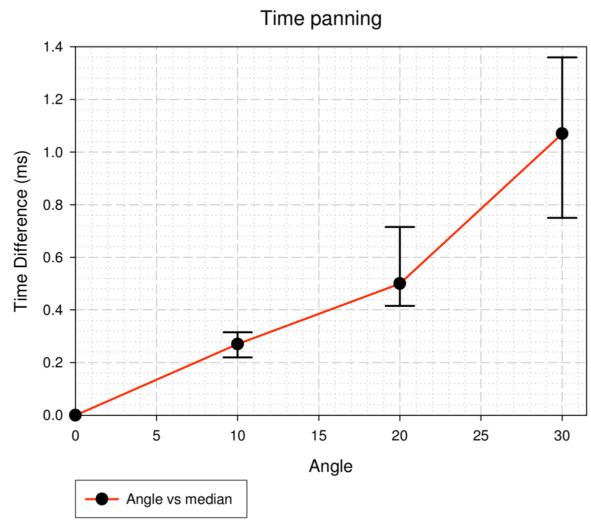

Figure 4 Plots of overall median values and 25th to 75th percentiles for the ICTD localisation

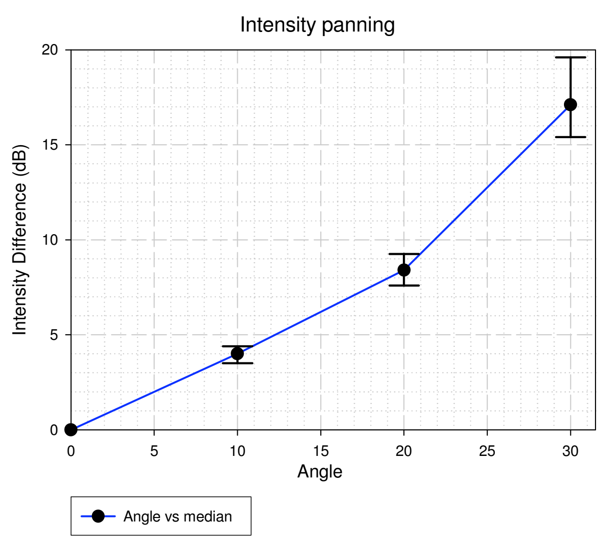

Figure 5 Plots of overall median values and 25th to 75th percentiles for the ICID localisation

The obtained results can be compared with others from similar experiments. As can be seen in **Table 3**, the differences among Simonsen, Wittek and the author's results are very small regarding the ICTD values. Regarding the ICID values, however, the author's results appear to be very different from Simonsen's while they are very similar to Wittek's. Generally Simonsen's ICID values are 2-3dB less than Wittek and the author's, and this is considered to be significant in that this range of intensity differences could cause noticeable angular shifts of phantom images. In fact, Simonsen's values did not satisfy the supposed angular shifts in the informal listening test carried out in the ITU-R BS.1116 listening room at the University of Surrey. The differences between the different authors' results seem to have resulted from the different experimental conditions, such as the acoustic condition of the listening room, the type of sound source used and the number of subjects.

Researcher		De Boer [2]	Simonsen [5]	Wittek [4]	Lee (author)
Sound source		Speech	Speech / maracas	speech	Speech / various
ICID	10°	5dB	2.5dB	4.4dB	4.0dB
	20°	11dB	5.5dB	8.8dB	8.4dB
	30°	not indicated	15dB	18dB	17.1dB
ICTD	10°	0.7ms	0.20ms	0.23ms	0.27ms
	20°	1.7ms	0.44ms	0.45ms	0.50ms
	30°	not indicated	1.12ms	1.0ms	1.1ms

Table 3 Comparisons of psychoacoustic values required for the localisation of 10°, 20° and 30° angles

4. DEVELOPMENT OF ICTD-ICID TRADE-OFF FUNCTION

4.1. Method

Using the unified localisation data obtained from the current experiment, it was attempted to develop ICTD-ICID trade-off functions for the phantom image locations of 10°, 20° and 30°. The basic combination method used was based on Theile's hypothesis, which suggests that the degree of total angular shift of phantom image can be calculated simply by summing the angular shifts by individual time and intensity differences, provided the individual shift is linear. The simple equation for this hypothesis is shown below.

$$\vartheta\,(\Delta t, \Delta I) = \vartheta\,(\Delta t) + \vartheta\,(\Delta I)$$

The unified data plots in **Figure 4** and **5** show that the psychoacoustic values required for 10° and 20° shifts are almost linearly increased in both time and intensity pannings. In other words, the increasing factors of the 0° - 10° and 10° - 20° shift regions are almost constant and therefore it was possible to apply the above combination function in this case. In an informal listening test conducted by the author and two colleagues who are critical listeners, this combination function was found to be valid for the interchannel data of up to the 20° shift region. However, this function could not be directly applied for the 30° shift because the shift factor of the 20° - 30° region is much greater than those of the lower regions. For example, a simple combination of individual shifts by time and intensity such as $\Delta t(10°) + \Delta I(20°)$ will not complete the desired 30° shift if $\Delta I(20°)$ is from the 0°-20° region. Even if it is based on the region of 10°-30°, there will be two different shift factors to be considered. Therefore, for the 30° shift, it was decided to divide the whole shift region into three effective regions and consider each separately as shown below.

$$\vartheta(30°) = \vartheta\,(0° - 10°) + \vartheta(10° - 20°) + \vartheta(20° - 30°)$$

Shift factors of ICTD and ICID required for each phantom image shift region were obtained by simplifying the results of the localisation experiments shown in **Figures 4** and **5** within the error ranges of 25th to 75th percentiles in such a way that the shift regions up to 20° have constant shift factors, as shown in **Table 4**.

Region shift	ICTD	ICID
0° - 10°	0.25ms	4dB
10° - 20°	0.25ms	4dB
20° - 30°	0.60ms	9dB

Table 4 Phantom image shift factors of ICTD and ICID for the shift region regions of 0° - 10°, 10° - 20° and 20° - 30°

4.2. Result

Using the proposed shift factors shown in **Table 4**, various combinations of ICTD and ICID were calculated. **Figure 6** shows the obtained combination curves for each localisation angle. It can be seen that the curves for the 10° and 20° shifts are completely linear and the calculated curve for the 30° shift is almost linear. As a result, three linear ICTD-ICID trade-off functions were developed for 10°, 20° and 30° shifts. The proposed linear trade-off curves could be advantageous to Williams [6]'s trade-off curves shown in **Figure 7** in that it would be much easier to calculate the required ICTD and ICID for trade-off with the linear curves.

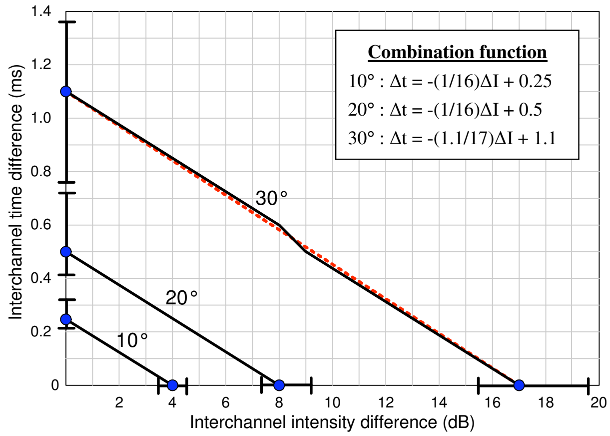

Figure 6 Proposed ICTD and ICID trade-off curves for 10°, 20° and 30° images, based on the psychoacoustic values obtained from the localisation test (see **Table 2**); Plots show the simplified median values and 25th to 75th percentiles.

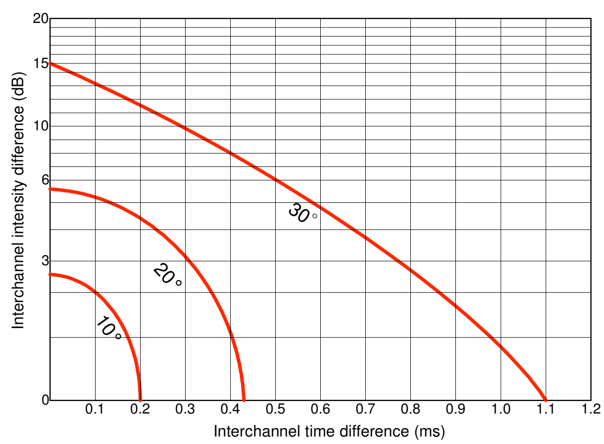

Figure 7 Williams' ICTD and ICID trade-off curves for 10°, 20° and 30° images, based on the psychoacoustic values obtained by Simonsen [5] (after Williams [6])

4.3. Verification of the proposed trade-off function

In order to verify the feasibility of the proposed combination functions, an additional subjective listening test was carried out with the identical subjects using the speech source in the same listening condition. A total of 17 test stimuli were created with various combinations of ICTD and ICID based on the proposed trade-off function, as listed in **Table 5**, and the subjects were asked to indicate the perceived locations of phantom images using reference markers placed with 5° intervals between the loudspeakers. Stimuli A to C were created aiming for 10° imaging, D to H for 20° and I to Q for 30°. The stimuli were recorded onto computer hard disk and played back to the subjects in a random order. Each stimulus was 30 second long, which gave the subjects enough time for judgment. Two subjects repeated the test three times and three repeated twice.

The result of the test is shown in **Figures 8 to 10**. It can be firstly noticed that the localisation accuracy for 30° images is generally worse than those for 10° and 20° ones. All the 25th to 75th percentile intervals in the 10° and 20° results cover the desired image positions although the medians do not always match the exact positions. However, in the 30° case some of the images with more ICTD do not appear to be localised at the desired position at all. In general it can be suggested that the images with more ICTD tends to have worse accuracy compared to those with more ICID and this is particularly true for 30° images. The deviation between the median angles and the desired angles is within the range of only 2°-3° and in this sense the proposed combination function could be accepted to be used.

Stimuli	Combination	Stimuli	Combination
A	0dB+0.25ms	J	2dB+0.97ms
B	2dB+0.13ms	K	4dB+0.84ms
C	4dB+0ms	L	6dB+0.71ms
D	0dB+0.5ms	M	8dB+0.58ms
E	2dB+0.38ms	N	10dB+0.45ms
F	4dB+0.25ms	O	12dB+0.32ms
G	6dB+0.13ms	P	14dB+0.19ms
H	8dB+0ms	Q	17dB+0ms
I	0dB+1.1ms		

Table A.5 Sound stimuli of various time and intensity combinations, based on the linear combination functions : A – C for 10°, D – H for 20° and I – Q for 30°

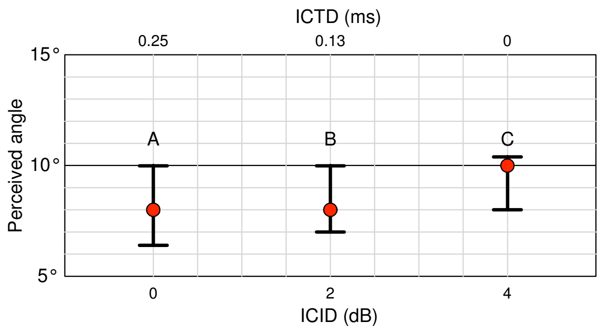

Figure 8 Data plots of perceived phantom image angles for the stimuli A, B and C indicated in Table 5: Median values and associated 25th to 75th percentiles

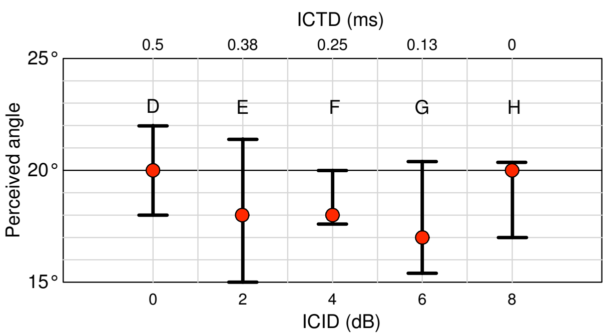

Figure 9 Data plots of perceived phantom image angles for the stimuli D, E, F, G and H indicated in Table 5: Median values and associated 25th to 75th percentiles

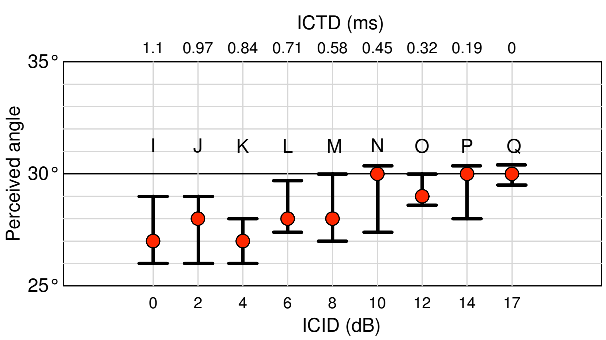

Figure 9 Data plots of perceived phantom image angles for the stimuli I, J, K, L, M, N, O, P and Q indicated in Table 5: Median values and associated 25th to 75th percentiles

5. CONCLUSION

A localisation experiment was carried out to obtain the individual localisation values of interchannel time difference (ICTD) and interchannel intensity difference (ICID) for 10°, 20° and 30°. The sound sources used for this experiment were wideband speech and musical sources of staccato piano notes and sustained trumpet notes. They were chosen to represent various spectral and temporal characteristics of sound (low frequency vs. high frequency, transient vs. continuous, wideband with ongoing transients). The use of various natural sources was expected to lead to a more reliable and practical result.

It was found that sources with more low frequency energy were generally more easily localised than those with more high frequency energy. It was also noticed that transient energy was necessary for ICTD to be plausible. Although the localisation characteristics for different sources vary, the data could be combined since the statistical analysis suggested that there was no significant differences between the data. The resulting localisation values of ICTD and ICID differed from the traditional values especially in ICID.

Based on the unified localisation values of ICTD and ICID obtained, the relationship between each localisation segment in the conventional stereo reproduction was examined. In consideration of this relationship, new trade-off functions for calculating the ratio between interchannel time and intensity differences required for the localisation of 10°, 20° and 30° images was developed. The proposed function has a linear characteristics and it differs from the conventional trade-off function [6].

The performances of these functions were verified through subjective listening tests. The result suggested that the accuracies of the proposed functions are reasonably good. For the 10° and 20° image positioning, any combination ratio between ICID and ICTD gave good results. For the 30° positioning the maximum performance could be achieved with the use of more ICID compared to ICTD, although the amount of error due to the use of more ICTD was in the negligible range. It is expected that the linear natures of the proposed functions obtained using natural sound sources could be useful for recording engineers to calculate the stereophonic recording angle of certain microphone technique or to design their own techniques.

6. ACKNOWLEDGEMENTS

The author would like to thank Dr.Francis Rumsey for his help with experimental design, all the excellent listeners from IoSR who participated in the listening tests, Dr.Byeongmoon Jeon, Dr.Sungyong Yoon, Dr.Tacksung Choi at LG Electronics for their encouragements on this publication.

7. REFERENCES

[1] Mertens, H. (1965): 'Directional hearing in stereophony: Theory and experimental verification', *European Broadcasting Union Review, Part A, 92, pp.1-14.*

[2] de Boer, K. (1940): 'Stereophonic Sound Reproduction', *Philips Technical Review*, 5, pp.107-114.

[3] Leakey, M. (1959): 'Some measurement of the effects of interchannel intensity and time differences in two channel sound systems', *Journal of the Acoustical Society of America*, 31, pp.977-986.

[4] Wittek, H. (2000): 'Untersuchungen zur richtungsabbildung mit L-C-R hauptmikrofonen', *Master's Thesis*, Institut für Rundfunktechnik, Germany.

[5] Simonsen, G. (1984): Master's Thesis. Technical University of Lyngby, Denmark.

[6] Williams, M. (1987): 'Unified theory of microphone systems for stereophonic sound recording', Audio Engineering Society 82nd Convention, Preprint 2466.

[7] Rakerd, B. and Hartmann, W. (1985): 'Localisation of sound in rooms, II: The effect of a single reflecting surface', *Journal of the Acoustical Society of America*, 78, pp.524-533.

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

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

[10] Zurek, P. (1980): 'The Precedence effect and its possible role in the avoidance of interaural ambiguities', *Journal of the Acoustical Society of America*, 67, (March), pp.952-964.

[11] Banks, S. and Green M. (1973): 'Localisation of High- and Low-Frequency Transients', *Journal of the Acoustical Society of America*, 53, pp.1432-1433.

[12] Yost, W., Wightman, F. and Green M. (1971): 'Lateralisation of filtered clicks', *Journal of the Acoustical Society of America*, 50, pp. 1526-1531.

[13] Mills, A. (1958): 'On the minimum audible angle', *Journal of the Acoustical Society of America*, 30, pp.237-246.

[14] Wittek, H. and Theile, G. (2002): 'The Recording Angles: based on localisation curves', Audio Engineering Society 112th Convention, Preprint 5568.