Lee, Hyunkook, Mason, Russell and Rumsey, Francis

Perceptually modelled effects of interchannel crosstalk in multichannel microphone technique

Audio Engineering Society

Convention Paper 7200

Presented at the 123rd Convention
2007 October 5–8 New York, NY, USA

Perceptually Modelled Effects of Interchannel Crosstalk in Multichannel Microphone Technique

Hyun-Kook Lee[1], Russell Mason[2], and Francis Rumsey[3]

[1] *Digital Media Research Lab., LG electronics, Seoul, Korea*

[2, 3] *Institute of Sound Recording, University of Surrey, Guildford, Surrey, U.K.*

Correspondence should be addressed to Hyun-Kook Lee (hklworld@lge.com)

ABSTRACT

One of the most noticeable perceptual effects of interchannel crosstalk in multichannel microphone technique is an increase in perceived source width. The relationship between the perceived source-width-increasing effect and its physical causes was analysed using an IACC-based objective measurement model. A description of the measurement model is presented and the measured data obtained from stimuli created with crosstalk and those without crosstalk are analysed visually. In particular, frequency and envelope dependencies of the measured results and their relationship with the perceptual effect are discussed. The relationship between the delay time of the crosstalk signal and the effect of different frequency content on the perceived source width is also discussed in this paper.

1. INTRODUCTION

Since surround sound recordings became popular, interchannel crosstalk in multichannel microphone technique has become an issue of much discussion. Most current three-channel or five-channel main microphone techniques are designed so that phantom imaging of a sound source primarily relies on the time and intensity relationship between the signals from the two microphones covering the sector of the stereophonic recording angle in which the source lies. In those types of microphone techniques, therefore, there is the implicit assumption that signals from microphones other than the pair that is primarily responsible for phantom imaging can be treated as crosstalk. For instance, if a three-channel microphone array was used for recording a single sound source located in one recording sector of the array, signals from the microphone pair that covered the recording sector where the source lies would be considered to be primary while any signal from the contralateral microphone would be regarded as crosstalk. The crosstalk channel would have certain time and intensity relationships to the wanted channels depending on the distance and angle between microphones in the array and therefore the presence of the crosstalk would be likely to affect certain aspects of the perception of the phantom image, even if the location of that phantom image could be determined solely by the primary channels.

The perceptual effects of interchannel crosstalk have been investigated and reported by Lee and Rumsey [1]. That research was conducted with a series of subjective elicitation and grading experiments involving such independent variables as microphone array type, sound source type and acoustic condition. The main result of the elicitation experiment was that the most salient effects of interchannel crosstalk were increase in source width and decrease in locatedness (localisability). From the grading experiment, it was found that the magnitudes of both source width increase and locatedness decrease significantly depended on the ratio of interchannel time and intensity differences in the microphone arrays used. For both attributes, arrays employing a greater interchannel time difference (conversely, a greater intensity of the crosstalk signal due to the fact that the microphone arrays used for the experiment were near-coincident types) caused a greater effect. It was also found that sound source type was a significant factor for the source-width-increasing effect but not for the locatedness effect, while acoustic condition had a significant effect on the locatedness decrease, but not on the source width increase.

Additionally, [2] reported the result of an investigation into the crosstalk effects on subjective preference of the perceived sound quality. A subjective experiment was carried out to test the preference between practical recordings of musical performances made with 'OCT' [3] and those with 'ICA-3' [4], which are popular three-channel microphone techniques that differ in their characteristics regarding interchannel crosstalk (i.e. the OCT had a greater reduction of interchannel crosstalk than the ICA-3). Sound sources comprised solo piano, solo violin, string quartet and percussion ensemble. The result was that the ICA-3 was preferred to the OCT technique for the string quartet and solo piano recordings while the OCT was preferred to the ICA-3 for the solo violin and percussion recordings. This might suggest that the effects of interchannel crosstalk on perceived sound quality are not always regarded negatively but can be regarded positively, depending on the characteristics desired for recordings of different types of sound source.

This paper describes an investigation designed to extend the study summarised above. The effect of interchannel crosstalk was objectively measured using a perceptual model, and the relationship between the perceived results and their physical causes was analysed. The increase in perceived source width was of particular interest in the current research. In the context of concert hall acoustics, there have been numerous studies conducted on the relationship between the effect of acoustic reflections on the perceived source width and physical metrics for that effect. For example, it was reported [5, 6, 7] that low frequency reflections in particular have a significant influence on the perception of source width. Also, there has been research related to the effect of the temporal characteristics of a signal on the perceived source width [8]. However, in the context of sound recording and reproduction, such studies have rarely been conducted. Although interchannel crosstalk and acoustic reflections have a similar form (both represent delayed secondary arrivals of a primary signal), the former typically has a shorter delay time compared to the latter. In this regard, it would be meaningful to investigate the effects of physical parameters of a crosstalk signal on the perceived source width increase. It is expected that the findings from this investigation will contribute to extension of our understanding of the perception of the source width attribute in general.

The measurement model used in the investigation is introduced below, and the procedure of stimuli creation is described. Secondly, the results of the measurements are reported and compared with those of the subjective experiments described in [1]. Finally, the frequency and envelope dependencies of the measurement results and their relationship with the perceived effect are discussed.

2. DESCRIPTION OF THE MEASUREMENT MODEL

The binaural hearing model was developed to predict aspects of the perceived spatial impression of a wide range of signals. The main factors that are predicted are the location (in terms of an azimuth) and width (in terms of a subtended angle) of the signals, in a number of frequency bands, dynamically over time. In addition, there is a simple prediction of the perceived loudness (in phons) when a calibrated input signal is used.

Depending on the content of the measured sound, the predicted parameters are either of the perceived source or the perceived environment (i.e. reverberation).

The measurement was designed to be as widely applicable as possible, in that it should be able to accurately predict the perceived source or environment width of any binaural signal that is fed into it. This means that it should work for any situation where a human listener can perceive the source or environment width of a sound, be it concert hall acoustics, reproduced sound, virtual reality, or any other form of listening. In addition, it should give comparable results for any input signal rather than having to rely on a single test signal or audio extract.

The central calculation on which the measurement model is based is the interaural cross-correlation coefficient (IACC). This was shown to be inversely related to the perceived width of auditory stimuli as early as the 1960s [9, 10]. Whereas recent research has suggested that it is not an accurate representation of the physiology of the binaural hearing process [11], the predictions of models of binaural perception based on the IACC have shown remarkable similarity to experimental data relating to lateralisation (e.g. [7, 12, 13]), binaural detection thresholds (e.g. [14, 15, 16]), and the perceived source width of an auditory stimulus (e.g. [17, 18, 19]).

Measurements based on this calculation that attempt to predict aspects of the perceived spatial impression have been developed previously by a number of researchers, however there were problems with the implementation of these, as the results were not directly comparable for a wide range of situations. The model described in this paper was developed to solve a number of these problems, such as the dependence of the perceived width on the loudness and frequency of stimuli.

The basic block diagram of the binaural model is shown in Figure 1. The main features of this binaural hearing model are as follows:

- inclusion of half-wave rectification and low-pass filtering so that the correlation of high frequency signals is affected by the envelope of the signal rather than the fine temporal detail [20]

- the use of a 'running' measurement to uncover variations in the perceived parameters over time [21]

- the effect of the frequency and loudness of the input signal is taken into account when predicting the perceived width [20, 22]

- predictions of the perceived location and width are integrated into a single output [23]

- the results are output in terms of an angle rather than an unintuitive value such as a coefficient [21]

- the effect of the IACC on the perceived localisation is taken into account [23]

- the effect of double peaks in the IACC calculation are taken into account in the width and localisation prediction [23]

The results of this analysis can be output as an animation or as an interactive plot. The animated display (an example screenshot is shown in Figure 2) shows the width and location as an angle against loudness and frequency, which is animated over time. The interactive plot (an example is shown in Figure 3) shows either the width or location or a combination as an angle, together with loudness against time, for selected frequency bands and time segments.

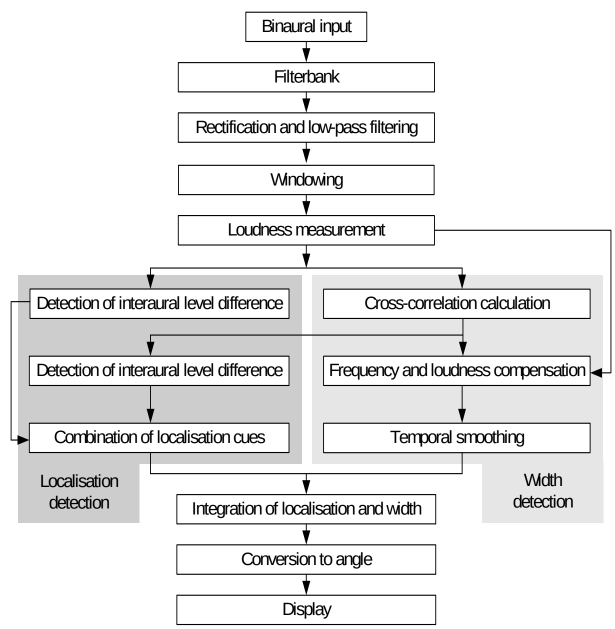


Figure 1: Block diagram of the main processing stages of the binaural model.

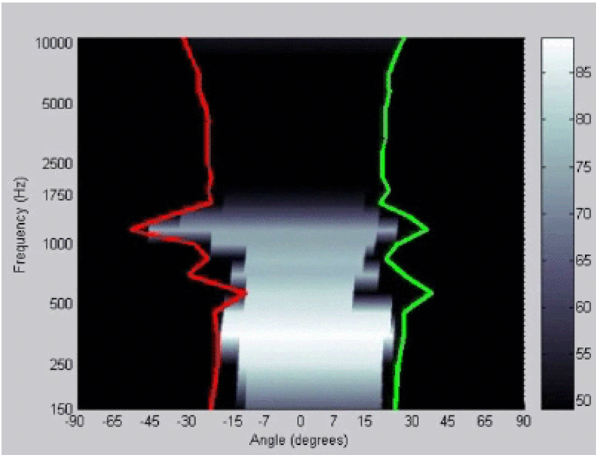


Figure 2: Example output of the animation display, showing the combined width and location in degrees on the x-axis, over frequency on the y-axis, with the loudness indicated by the brightness of the plot. The red and green lines indicate the 'peak hold' of the extreme left and right values respectively.

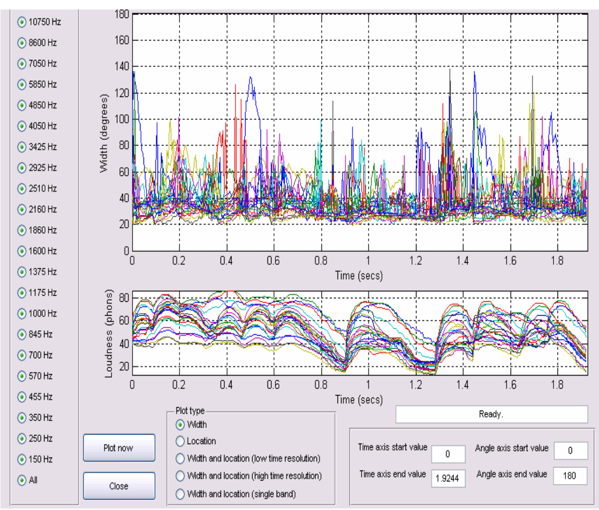


Figure 3: Example output of the interactive plot display showing the width of the sound on the y-axis over time on the x-axis in the upper plot, and the loudness of the sound on the y-axis over time on the x-axis in the lower plot, for the frequency bands selected on the left hand side.

A number of evaluations have been conducted on the resulting binaural model, including those reported in [21]. This has demonstrated that the model accurately predicts the perceived spatial impression of most simple stimuli (for example consisting of a single source in a reverberant environment), and with much greater accuracy than simple IACC measurement techniques.

3. STIMULI

A total of six sets of binaural stimuli were created for the current investigation. For this, six sets of multichannel stimuli used for the authors' previous subjective experiments [1], involving the combinations of two types of simulated microphone arrays and three types of sound source, were reproduced in University of Surrey's ITU-R BS.1116-compliant [24] listening room and the created soundfield was recorded using a dummy head placed in the listening position. Each set of the stimuli consisted of a pair of recordings made with and without interchannel crosstalk. The sound sources and microphone arrays involved in the stimuli are explained below.

The sound sources used for this investigation comprised cello, bongo and speech and they were chosen for their diverse temporal and spectral characteristics, with the cello being relatively continuous and having a complex harmonic structure, the bongo having a strong transient nature, and the speech having a mixture of transient and continuous sounds as well as a wide frequency range. The signal for each sound source was an anechoic mono recording of a performance excerpt taken from the Bang & Olufsen Archimedes project CD [25].

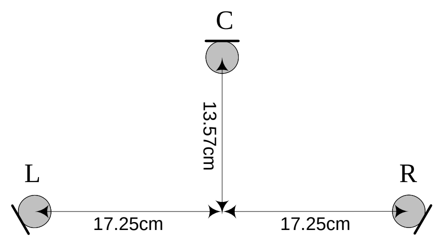
The microphone technique used was the so-called 'critical linking' three-channel microphone technique, proposed by Williams and Le Du [26]. The basic design concept of this technique aims to achieve a continuous distribution of phantom images across channels L (Left), C (Centre) and R (Right) by linking the stereophonic recording angles (SRAs) of each stereophonic segment C-L and C-R without overlap. Within one segment, the psychoacoustic laws for localisation in conventional two-channel stereophonic reproduction such as summing localisation or the precedence effect are applied independently without considering the influence of the other segment. For example, when a sound source is located at 45° in the recording segment C-R, localisation of the phantom image should be governed by the summing localisation effect between C and R only, and in this case L can be regarded as crosstalk to the channels C and R. Ideally, L should not be taken into account in the localisation process. In this regard, it is logical to examine the effect of interchannel crosstalk by comparing the image that is created with the crosstalk channel turned on (image formed by contributions from L-C-R) and that with the crosstalk channel turned off (C-R only). The critical linking technique supposedly enables one to create various array styles having different distances and angles between microphones while keeping the SRA across L, C and R constant. Therefore, the effect of the ratio of time to intensity differences between the crosstalk signal and the other channels can be investigated by comparing different microphone arrays sharing the same SRA.

For the current investigation, two critical linking arrays were simulated based on the relationship between interchannel time and intensity differences that are documented in [27]. The configurations for these arrays are shown in Figure 4. These particular arrays were chosen because the difference between each array in terms of the distance and angle between microphones was considered to be large enough to provide distinctive interchannel relationships for the crosstalk signals. The common SRA for these arrays was 180°, the simulated direction of the sound source was 45° from the centre line of the array and the distance from the centre point of the array was five metres. The interchannel time and intensity differences between L and C and between R and C calculated for each array for a source located at 45° from directly in front of the array are shown in Table 1.

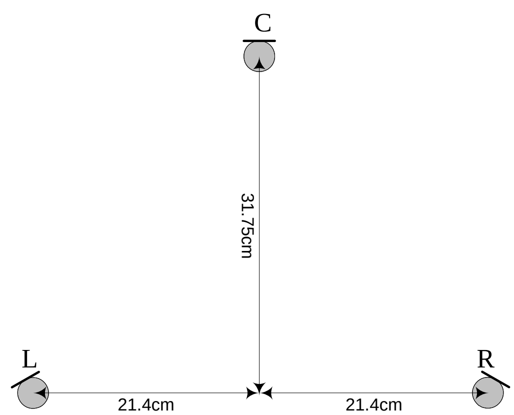
Examples of the waveforms of the short binaural stimuli created for each anechoic sound source are shown in Figure 5. The durations of the created stimuli were 1.6 – 2.0 seconds. The selection of the excerpts was made so that they included representative temporal characteristics of the sound sources (e.g. note and bow changes for the cello, syllable changes for the speech, and ongoing hits for the bongo).

	C to L delay	C to L intensity	C to R delay	C to L intensity
Array 1	0.64ms	- 20.5dB	- 0.08ms	- 0.7dB
Array 2	1.09ms	- 4.6dB	0.21ms	1.4 dB

Table 1: Time and intensity relationships between each microphone channel for two types of 'critical linking' array with a sound recording angle (SRA) of 180°, calculated for a source located at 45°.

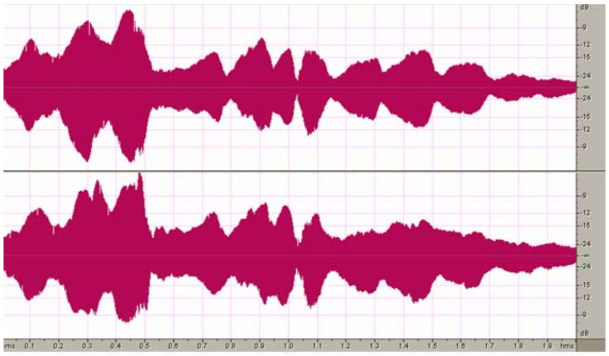


(a) Array 1: the L-R angle is 100°

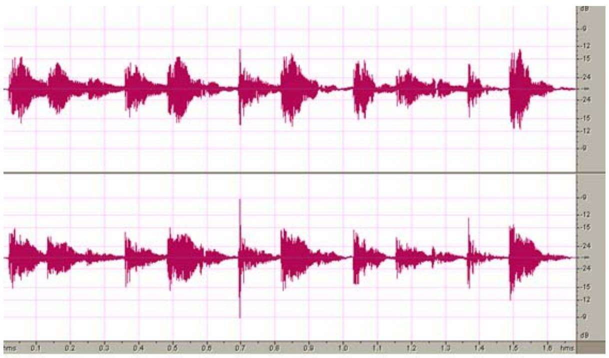


(b) Array 2: the L-R angle is 40°

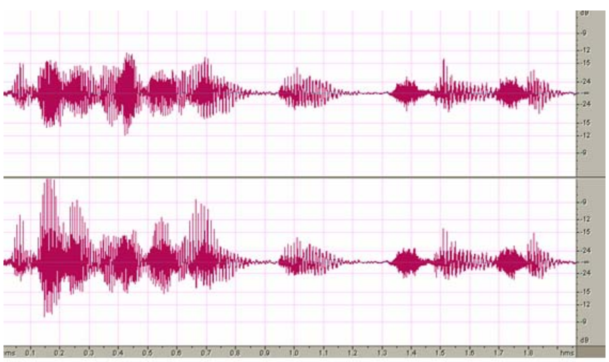
Figure 4: Configurations of 'critical linking' microphone arrays simulated for the measurement: the common SRA is 180°.



(a) Cello



(c) Bongo



(c) Speech

Figure 5: Waveforms of the stimuli used for the objective measurements, which were created from three-channel stereophonic recordings made with 'Array 2' (see Figure 4) in 'crosstalk-on' condition.

4. ANALYSES OF THE MEASURED RESULTS

4.1. Correspondence between the Measured Data and the Perceived Data

In order to validate the usability of the current prediction model as a tool for analysing the physical factors causing the perceived crosstalk effects, it would be first necessary to compare the measured results with the perceived results and to confirm whether the former could be used to predict the latter. If they did not match reasonably, it might be because the model does not implement all aspects of the complex cognitive process of spatial perception. Due to this model using a time-varying measurement of IACC in a number of frequency bands, there is no simple method of converting the magnitudes of measured differences between crosstalk-on stimuli and crosstalk-off stimuli into single numerical values for statistical analysis. Nonetheless, it is possible to measure the magnitudes of the visual changes indicated in the measured plots approximately, and compare the general trends between the perceived results and the measured results.

For comparison between the crosstalk-on stimuli (LCR) and crosstalk-off stimuli (CR), which were described in Section 3, with respect to the magnitude of change in measured source width, it was predicted that the LCR would appear to have a greater source width than the CR since the addition of the crosstalk signal would decrease the IACC. The plots in Figures 6 – 8 (see Pages 12 – 13) show the source width measurements made in 22 different octave frequency bands for each sound source. From the plots it can be seen for every source type that the source width measurements of CR and LCR have similar trends for microphone array 1 (see Figure 4), although there are some minor differences in the variation pattern of certain frequency bands. Microphone array 2 (see Figure 4), on the other hand, shows more obvious changes between CR and LCR in general. There are more frequency bands that produce large peaks in the LCR, and therefore the plot of LCR shows more erratic variations in source width measurement over time. This means that when the crosstalk signal has a higher ratio of time difference to intensity difference, it causes a higher degree of interaural decorrelation, leading to the perception of a greater source width. According to Mason [28] who suggests a close relationship between the interaural fluctuations over time and the IACC, this can be also explained as a more time-difference-based crosstalk signal producing a larger magnitude of fluctuations in ITD over time. These measurement results agree well with the perceived results that showed a statistically significant difference between microphone arrays.

For comparison between the magnitudes of source width change for each sound source, two cases can be considered separately depending on the behaviour of predicted values in different frequency bands. The first is for source widths determined by the frequency bands in which predicted values vary erratically, and the second is for those determined by the frequency bands crowded in the lower region of the plots in which predicted values remain relatively consistent. With respect to the former case, the cello appears to have more obvious changes between CR and LCR than the bongo and speech. However, for the latter case, it appears that the speech source gives rise to the most obvious change between CR and LCR. The cello does not seem to give rise to much difference between CR and LCR in this case. For example, it appears that most of the peaks for the cello source arise erratically in a single or a small number of frequency bands (Figure 6), while those for the speech source arise relatively regularly in a larger number of frequency bands (Figure 8). This might suggest that the source width changes for the speech source would have been more audible than those for the cello due to the wide range of frequency bands that gave rise to the changes. In this regard, the measurement results seem to agree with the perceived results showing that the speech source gave rise to the greatest change in source width.

4.2. Influences of Frequency Components and Signal Envelope

In order to investigate the frequency dependency of the source-width-increasing effect of interchannel crosstalk, the pattern of each frequency band was analysed in detail for each sound source. Only the stimuli for microphone array 2 were considered in this investigation because array 1 appeared to indicate no obvious changes between CR and LCR overall. The measured data for this investigation are presented in Figures 9 – 19 (see Pages 14 – 19).

From the observation of the measured differences between CR and LCR for each frequency band of the anechoic cello and bongo stimuli, it was possible to separate the frequency bands into four groups based on the variation patterns. The centre frequencies included in each group are shown in Table 2.

Group	Centre frequencies (Hz)
1	150, 250, 455, 570
2	700, 845, 1000, 1175, 1375
3	1600, 1860, 2160, 2510, 2925, 3425, 4050, 4850, 5850
4	7000, 8600, 10750

Table 2: Grouping of centre frequencies of the frequency bands that share similar time-varying patterns in the measured source width change for the cello and bongo sources.

Figure 9 shows the measurements made for the frequency bands of group 1. It can be firstly observed that there is no great difference between CR and LCR. The source width measured over time is relatively constant in both CR and LCR. However, for the frequency bands of group 2, the changes between CR and LCR become obvious in that the LCR has a greater magnitude of source width and a more erratic pattern of variations than the CR (Figure 10). In comparison with the loudness plot of the source signal shown in the figure, it can also be observed that the peaks of the measured source width over time for each frequency band in the LCR appear to correspond to the peaks of the loudness envelope of the frequency band signal. The overall plots for the frequency bands also appear to be largely related to the envelope of the overall waveform shown in Figure 5. Figure 11 shows the measurements of the frequency bands of group 3. Even though there are obvious changes observed between CR and LCR for these frequency bands, the patterns of source width increase for them appear to be different from those for the lower frequency bands shown above. The changes are mainly due to the random and sharp peaks, and therefore no envelope dependency is found. On the other hand, the measurements of the highest frequency bands of group 4 shown in Figure 12 have similar trends to those of the lowest frequency bands. That is, the source width is mostly constant over time, and no obvious change between CR and LCR is observed.

For the measurements of the bongo stimuli, the frequency bands can be separated into four groups in the same manner as shown for the cello stimuli (see Table 2). Firstly, it can be seen from Figure 13 that the measurements of the frequency bands of group 1 for the anechoic bongo stimuli change regularly over time depending on the envelope of the signal in both CR and LCR. The peaks appear to occur at the offset of each transient hit, and this is likely to be due to the influence of the decorrelated resonant sound during the decay. However, the differences between CR and LCR do not appear to be dominant for these frequencies. For the frequency bands of group 2, there are more noticeable differences between CR and LCR as can be observed in Figure 14. The LCR appears to have slightly greater widths in general and a greater number of large peaks than the CR, although these differences seem to be relatively small compared to the differences observed at the same frequency bands of the cello stimuli. It is interesting to note that the measurement of each frequency band in the LCR appears to be related to the signal envelope. Whereas the peaks of the measurements for the lower frequency bands occur at the dips of the signal envelope, the peaks for these frequency bands occur at the peaks of the signal envelope. Figure 15 indicates that the measurements change more randomly at the frequency bands of group 3, but the differences between CR and LCR are relatively small. The measurements for the highest frequency bands shown in Figure 16 (group 4) appear to be similar to those for the lowest bands in that the noticeable variations in the measurement occur at the dips of the signal envelope although their magnitudes are smaller. It can also be seen that the differences between CR and LCR are negligible.

The frequency bands of the speech stimuli can also be separated into several groups, the number of groups and the range of frequencies belonging to each group differ from the cello and bongo. There are a total of three groups of frequency bands that show different trends in the measurements, as indicated in Table 3.

Group	Centre frequencies (Hz)
1	150, 250, 455, 570
2	700, 845, 1000
3	1175, 1375, 1600, 1860, 2160, 2510, 2925, 3425, 4050, 4850, 5850, 7000, 8600, 10750

Table 3: Grouping of centre frequencies of the frequency bands that share similar time-varying patterns in the measured source width change for the speech source.

The measurements made for the group 1 frequency bands of the anechoic speech stimuli are shown in Figure 17. It can be seen that for both CR and LCR, the measurements increase at the dips of the signal envelope, in other words in the space between each syllable. However, the difference between CR and LCR appears to be very small. For the frequency bands of group 2, as shown in Figure 18, there are more obvious differences in the measurements between CR and LCR. The temporal variations in the measurements for the LCR appear to occur at the peaks of the signal envelopes of the corresponding frequency bands. Finally, Figure 19 indicates that the measurements made for the frequency bands of group 3 have more erratic temporal variations compared to those for the lower frequency bands. Due to the sharp and random peaks, none of the frequency bands appear to exhibit envelope-dependency. However, it is interesting to note that with all the frequency bands of this group considered together, there are certain temporal regions where the large peaks become crowded (e.g. in the region around 1.3 - 1.4 seconds), and the envelope of the crowded peaks appear to be related to the signal envelope.

5. DISCUSSIONS

From the measurement results, relationships between the source width increasing effect of interchannel crosstalk and the related physical factors can be discussed.

The results showed that for the frequency bands up to 570Hz there were no obvious differences between the crosstalk-on and crosstalk-off stimuli of all source types in the measured source width. This means that the source width increasing effect of interchannel crosstalk was small at low frequencies. At the lower-middle frequencies in general, the crosstalk appeared to have the most obvious effect on the increase in the measured width. It was also found that the source width increase caused by crosstalk at the lower-middle frequencies mainly occurred around the onsets of the signal envelope rather than the offsets, and this seems to be related to the delay time of the crosstalk signal and the onset duration of the primary signals. Griesinger [8] hypothesises that source width is perceived only when the secondary delayed sound arrives within the onset duration of the primary sound. From this point of view, the range of the delay time of the crosstalk signal used in the current studies (0.5 – 1.1ms) is small enough to contribute to the perception of source width since the onset durations of the current sound sources are greater than 10ms.

Furthermore, the dominance of the middle frequency effect seems to be due to the relationship between the pattern of interaural fluctuation and the delay time of the crosstalk signal. The range of delay times involved in the crosstalk signal corresponds to approximately the cycle duration of signals around 1,000 Hz, and therefore strong interaural time and intensity fluctuations (thus low IACC) might have occurred at the middle frequencies. However, in the case of acoustic reflections in rooms, which typically have longer delay times, the effect of interaural fluctuation on source width increase might occur at lower frequencies. In fact, the ranges of delay times of the reflected signals used for studies that reported the importance of low frequencies in source width perception [5, 6] were much greater than those used in the current study. Based on the above discussion, it is generally suggested that the measured effect of the frequency component of the secondary signal on source width perception might be dependent on the range of delay time of the secondary signal.

It was observed from the results that the upper-middle frequencies generally produced very erratic and random variations in the measurements across all frequency bands in this range. From this, it might be proposed that some kind of cognitive process (or rapid decision making process in the brain) was involved in the perception of source width change. That is, plausibility of the change in each frequency band might have been taken into account in the detection of audible changes. In the context of localisation studies, the 'plausibility hypothesis' of Rakerd and Hartmann [29] suggests that unreasonably large ITD (Interaural Time Difference) cues produced by the interaction between direct sound and room reflections are ignored by the brain in the process of localisation and only plausible ITD cues are used. Similarly, the rapid and large variations in IACC (or rapid and large ITD fluctuations) shown in the measurement results for the upper-middle frequencies might have been recognised as implausible cues by the brain, and therefore disregarded in the process of source width perception.

It was shown in the results that the patterns of the temporal variations at the highest frequencies became very similar to those at the low frequencies. This is likely to be due to the simulation of the breakdown of phase locking that is included in the process of the current model. The human hearing system fails to detect fine temporal details at high frequencies due to the breakdown of phase locking, and the perceived width becomes dependent on the IACC of the signal envelope rather than the signal itself [30]. From this, it is considered that the source widths of those higher frequency bands were determined by the envelope of the signal, having a relatively low frequency.

6. SUMMARY AND CONCLUSIONS

This paper described an investigation into the measured effects of interchannel crosstalk, using a perceptual model. The measurement model chosen for this investigation was an IACC-based width and location prediction model, which was described in Section 2. This model was particularly useful for the current study since it is intended to measure any audible signal rather than specific test signals. Firstly, the correspondences between the measured data and the perceived data were examined. For this, the measurements of the crosstalk-on stimuli and crosstalk-off stimuli were compared with respect to the independent variables, which were the type of microphone array and the type of sound source.

It was found that the measurements for the source width attribute matched the perceptual data reasonably well. The influences of the different frequency components and their relationship with the signal envelope were investigated with regard to the source width increasing effect of interchannel crosstalk. It was found that at low frequencies up to the centre frequency of 570Hz there was no obvious crosstalk effect. At the middle frequencies up to around 1000Hz, the source width increasing effect of crosstalk was most dominant, having a positive correlation with the onsets of the signal envelope. At the higher frequencies, the measurements became largely erratic and the envelope dependency disappeared.

It is concluded that the perception of increased source width due to interchannel crosstalk, within the context of the current experimental study, is mainly a middle frequency phenomenon and dependent on the temporal characteristics of signal. It is also suggested that the effect of the frequency content of the secondary signal on perceived source width might be closely related to the delay time of the secondary signal.

7. REFERENCES

[1] Lee, H.K., Rumsey, F. (2005): 'Investigation into the effect of interchannel crosstalk in multichannel microphone technique', Audio Engineering Society 118th Convention, Preprint 6374.

[2] Lee, H.K. (2006): 'Effects of Interchannel Crosstalk in Multichannel Microphone Technique', *Ph.D thesis*, University of Surrey, England.

[3] Theile, G. (2001): 'Multichannel natural recording based on psychoacoustic principles', In *Proceedings of the Audio Engineering Society 19th International Conference*, pp.201-229.

[4] Herrmann, U. and Henkels, V. (1998): 'Main microphone techniques for the 3/2-stereo-standard', www.hhton.de

[5] Morimoto, M. and Maekawa, Z. (1988): 'Effects of low frequency components on auditory spaciousness', *Acustica*, 66, pp.190-196.

[6] Hidaka, T., Beranek, L. and Okano, T. (1995): 'Interaural cross-correlation lateral fraction, and low- and high- frequency sound levels as measures of acoustical quality in concert halls', *Journal of the Acoustical Society of America*, 98, pp.988-1007.

[7] Mason, R., Brookes, T. and Rumsey, F. (2005b): 'Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli', *Journal of the Acoustical Society of America*, 117, pp.1337-1350.

[8] Griesinger, D. (1997): 'The psychoacoustics of apparent source width, spaciousness & envelopment in performance spaces', *Acta Acustica*, 83, pp.721-731.

[9] Chernyak, R. I. and Dubrovsky, N. A. 1968: 'Pattern of the noise images and the binaural summation of loudness for the different interaural correlation of noise', *Proceedings of the 6th International Congress on Acoustics*, Tokyo, pp. A53-A56.

[10] Keet, W. de V. (1968): 'The influence of early lateral reflections on the spatial impression', *Proceedings of the 6th International Congress on Acoustics*, Tokyo, pp. E53-E56.

[11] Boehnke, S. E., Hall, S. E., and Marquardt, T. (2002): 'Detection of static and dynamic changes in interaural correlation', *Journal of the Acoustical Society of America*, 112, pp. 1617-1626.

[12] Sayers, B. M. (1964): 'Acoustic-image lateralization judgments with binaural tones', *Journal of the Acoustical Society of America*, 36, pp. 923-926.

[13] Okano, T. (2000): 'Image shift caused by strong lateral reflections, and its relation to inter aural cross correlation', *Journal of the Acoustical Society of America*, 108, pp. 2219-2230.

[14] Constan, Z. A., and Hartmann, W. M. (2001): 'Sound localization by interaural time differences at high frequencies', *Journal of the Acoustical Society of America*, 109, pp. 2485.

[15] Sayers, B. M., and Cherry, E. C. (1957): 'Mechanism of binaural fusion in the hearing of speech', *Journal of the Acoustical Society of America*, 29, pp. 973-987.

[16] Robinson, D. E., and Jeffress, L. A. (1963): 'Effect of varying the interaural noise correlation on the detectability of tonal signals', *Journal of the Acoustical Society of America*, 35, pp. 1947-1952.

[17] Bernstein, L. R., and Trahiotis, C. (1992): 'Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies', *Journal of the Acoustical Society of America*, 91, pp. 306-316.

[18] Plenge, G. (1972), 'Über das problem der im-kopf-lokalisation (On the problem of in-head localisation)', *Acustica*, 26, pp. 241-252.

[19] Kurozumi, K., and Ohgushi, K. (1983): 'The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality', *Journal of the Acoustical Society of America*, 74, pp. 1726-1733.

[20] Blauert, J., and Lindemann, W. (1986): 'Spatial mapping of intracranial auditory events for various

degrees of interaural coherence', *Journal of the Acoustical Society of America*, 79, pp. 806-813.

[21] Mason, R., Brookes, T. and Rumsey, F. (2004): 'Evaluation of an auditory source width prediction model based on the interaural cross-correlation coefficient', 148th Meeting of the Acoustical Society of America, *Journal of the Acoustical Society of America*, 116, no. 4 (October), pp. 2475.

[22] Mason, R. and Brookes, T. (2008): 'Loudness dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli', in preparation.

[23] Mason, R., Brookes, T. and Rumsey, F. (2004): 'Integration of measurements of interaural cross-correlation coefficient and interaural time difference within a single model of perceived source width', Audio Engineering Society 117th Convention, Preprint 6317.

[24] ITU-R (1994): 'Recommendations ITU-R BS.1116: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems', *International Telecommunications Union*.

[25] Hansen, V. and Munch, G. (1991): 'Making recordings for simulation tests in the Archimedes project', *Journal of the Audio Engineering Society*, 39, pp.768-774.

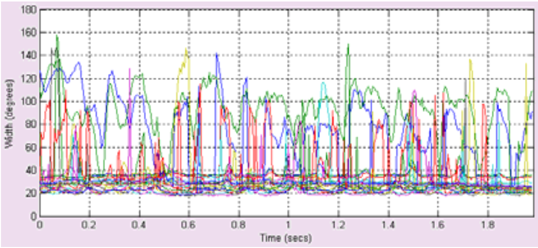
[26] Williams, M. and Le Du, G. (1999): 'Microphone array analysis for multichannel sound recording', Audio Engineering Society 107th Convention, Preprint 4997.

[27] Williams, M. (2004): *Microphone arrays for stereo multichannel sound recording. Vol.1* (Segrate: Editrice Il Rostro).

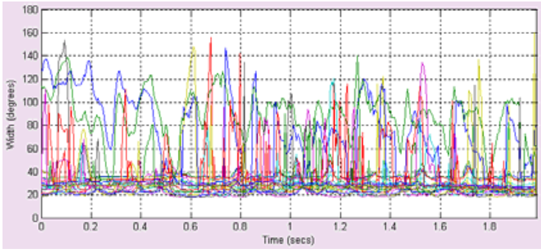
[28] Mason, R. (2002): 'Elicitation and measurement of auditory spatial attributes in reproduced sound', *Ph.D thesis*, University of Surrey, England.

[29] 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.

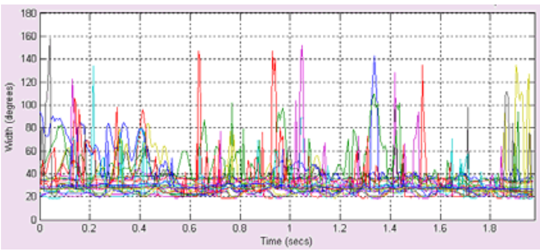
[30] Mason, R., Brookes, T. and Rumsey, F. (2004): 'Development of the interaural cross-correlation coefficient into a more complete auditory width prediction model', *Proceedings of the 18th International Congress on Acoustics*, pp. 2453-2456.



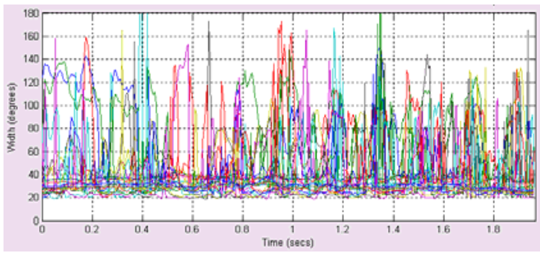
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

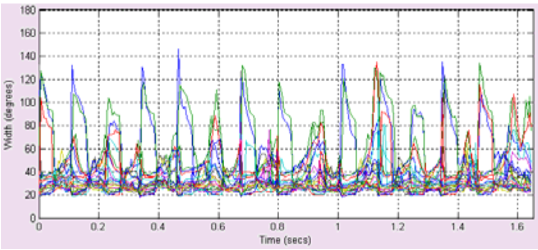


(c) Array 2: crosstalk-off (CR)

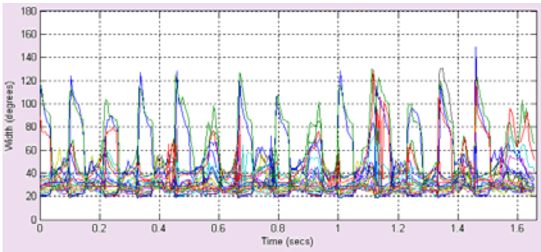


(d) Array 2: crosstalk-on (LCR)

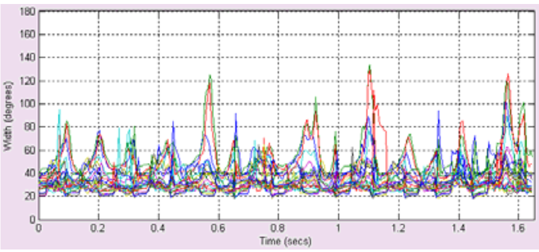
Figure 6: Comparisons of the plots of width measurements for the ‘cello’ stimuli.



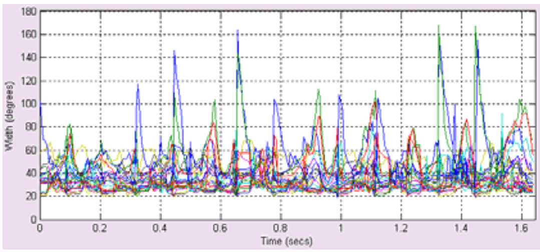
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

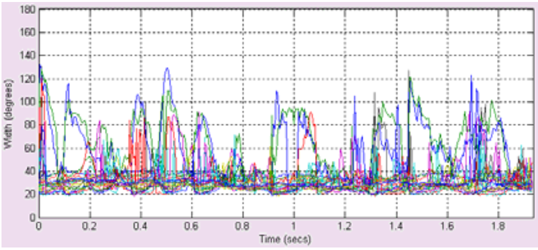


(c) Array 2: crosstalk-off (CR)

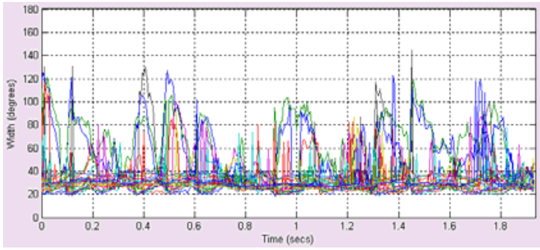


(d) Array 2: crosstalk-on (LCR)

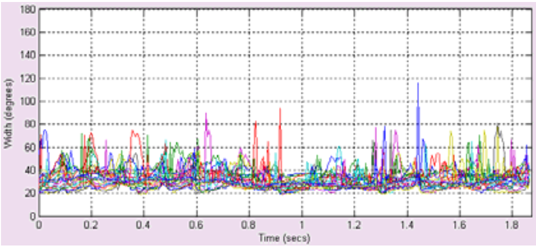
Figure 7: Comparisons of the plots of width measurements for the ‘bongo’ stimuli



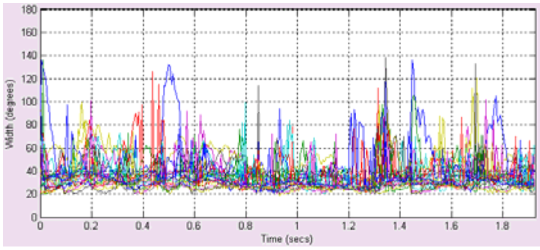
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

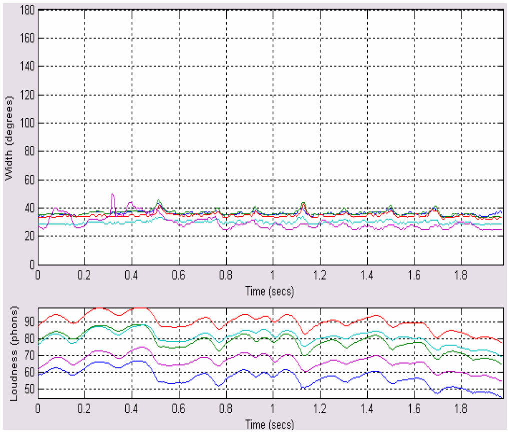


(c) Array 2: crosstalk-off (CR)

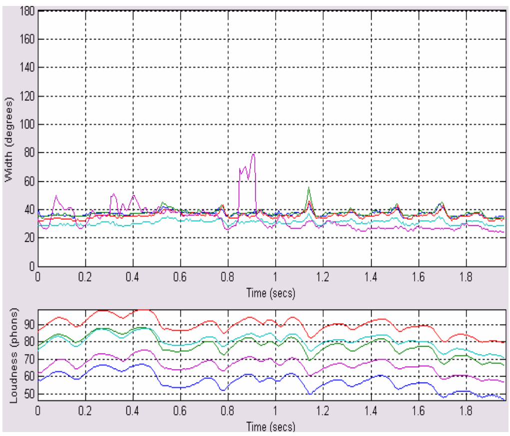


(d) Array 2: crosstalk-on (LCR)

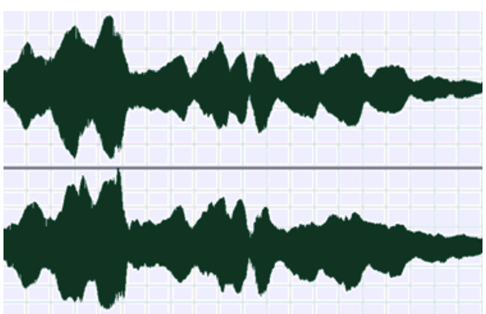
Figure 8: Comparisons of the plots of width measurements for the ‘speech’ stimuli



(a) Crosstalk-off

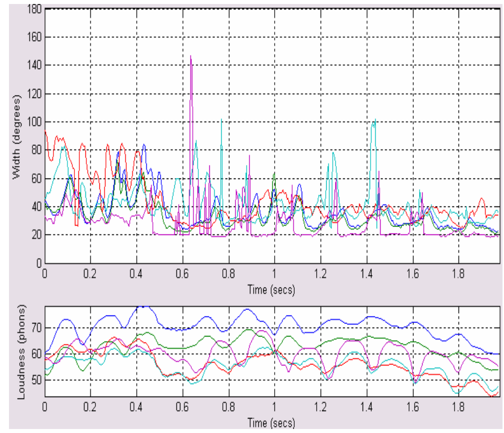


(b) Crosstalk-on

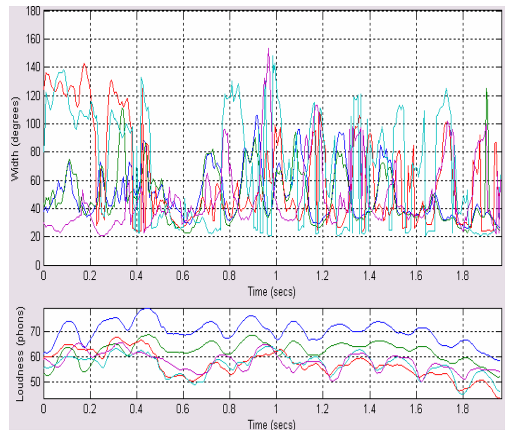


(c) Waveform

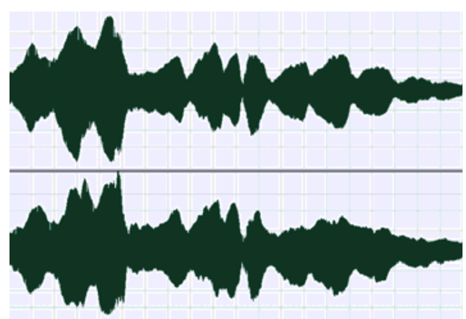
Figure 9: Plots of the width measurement made for the cello stimuli of microphone array 2, with centre frequencies of 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



(a) Crosstalk-off

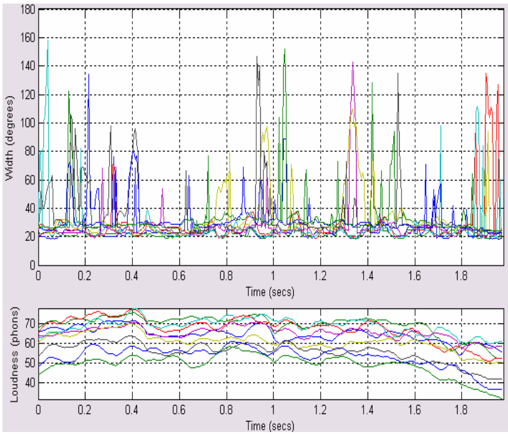


(b) Crosstalk-on

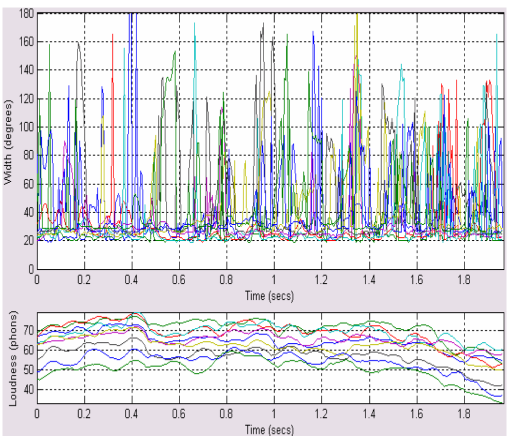


(c) Waveform

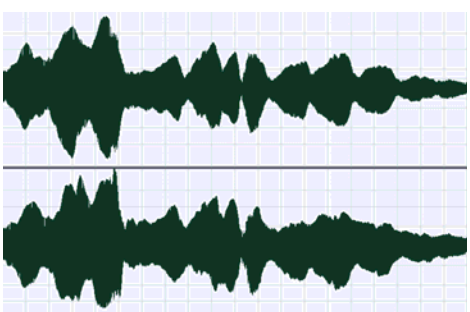
Figure 10: Plots of the width measurement made for the cello stimuli of microphone array 2, with centre frequencies of 700, 845, 1000, 1175, 1375 Hz, and waveform of the binaural signal



(a) Crosstalk-off

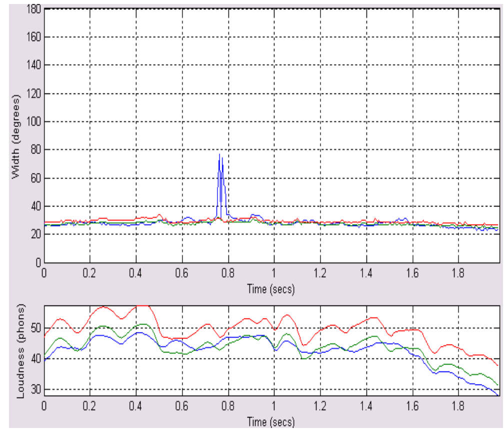


(b) Crosstalk-on

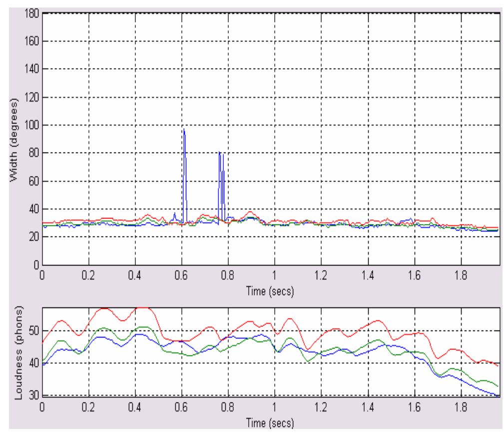


(c) Waveform

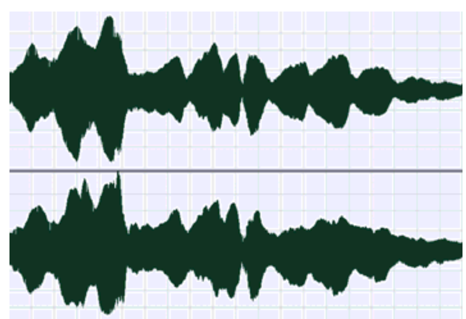
Figure 11: Plots of the width measurement made for the cello stimuli of microphone array 2, with centre frequencies of 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850 Hz, and waveform of the binaural signal



(a) Crosstalk-off

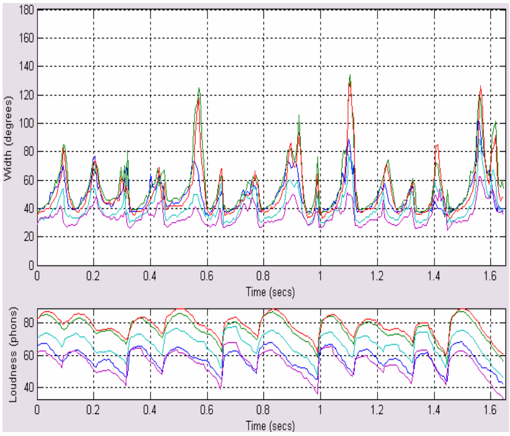


(b) Crosstalk-on

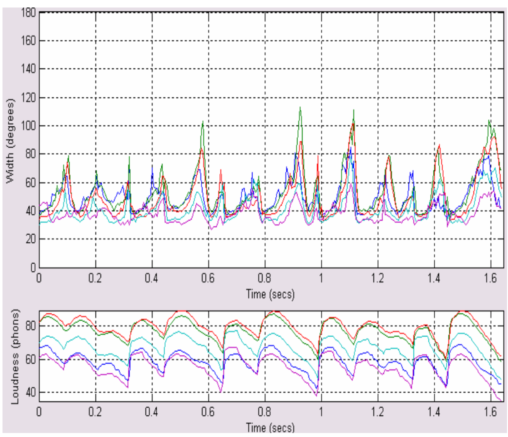


(c) Waveform

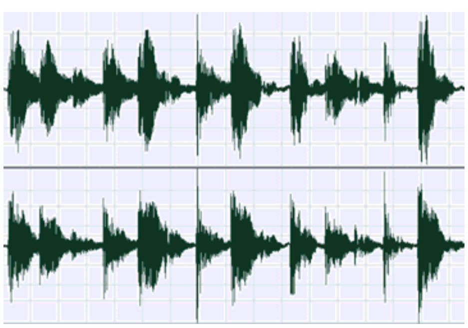
Figure 12: Plots of the width measurement made for the cello stimuli of microphone array 2, with centre frequencies of 7050, 8600, 10750 Hz, and waveform of the binaural signal



(a) Crosstalk-off

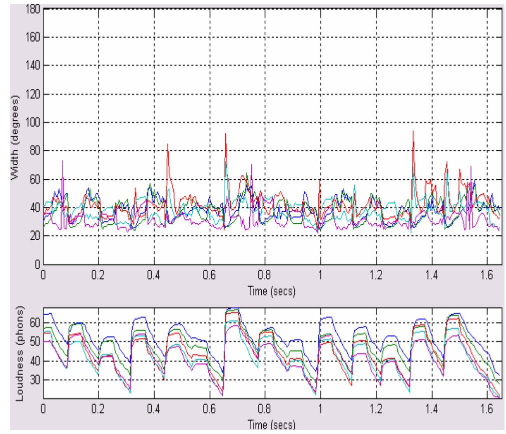


(b) Crosstalk-on

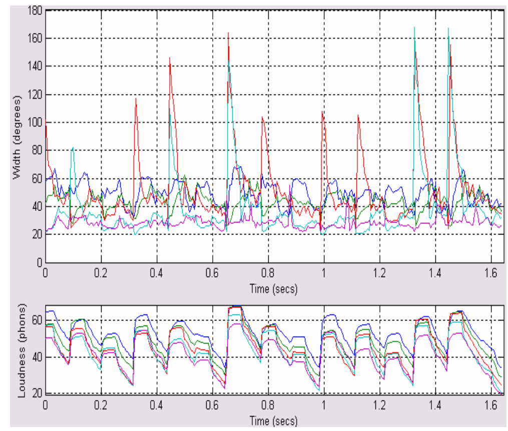


(c) Waveform

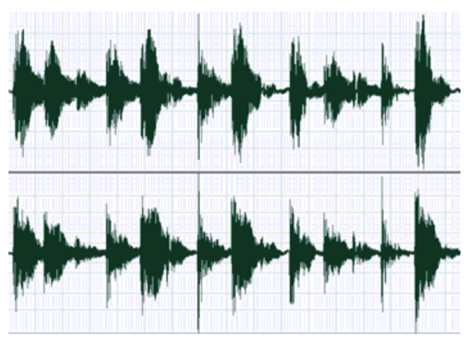
Figure 13: Plots of the width measurement made for the anechoic bongo stimuli of microphone array 2, with centre frequencies of 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



(a) Crosstalk-off

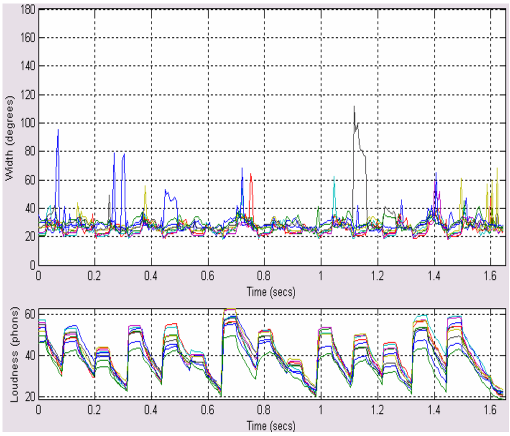


(b) Crosstalk-on

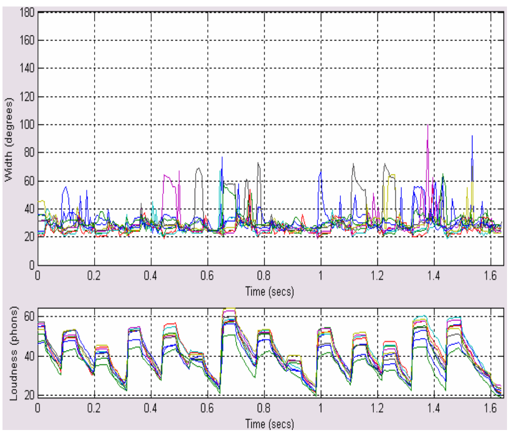


(c) Waveform

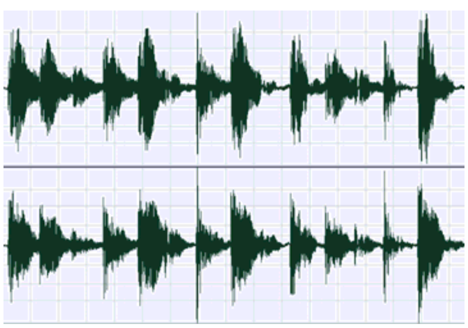
Figure 14: Plots of the width measurement made for the anechoic bongo stimuli of microphone array 2, with centre frequencies of 700, 845, 1000, 1175, 1375 Hz, and waveform of the binaural signal



(a) Crosstalk-off

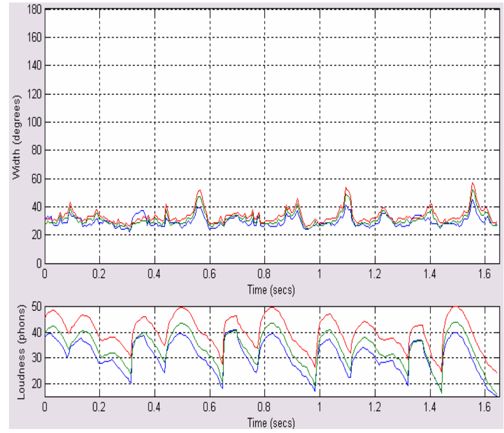


(b) Crosstalk-on

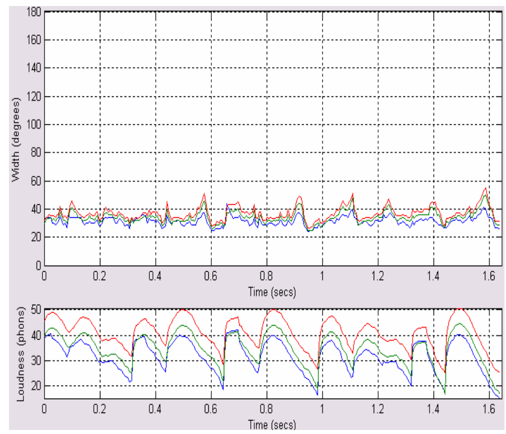


(c) Waveform

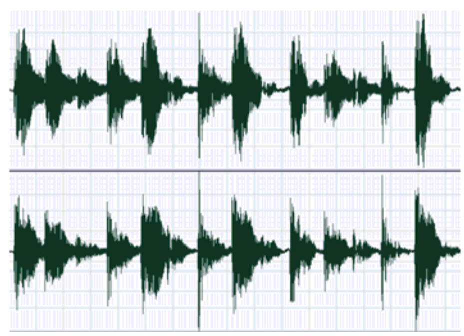
Figure 15: Plots of the width measurement made for the anechoic bongo stimuli of microphone array 2, with centre frequencies of 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850 Hz, and waveform of the binaural signal



(a) Crosstalk-off

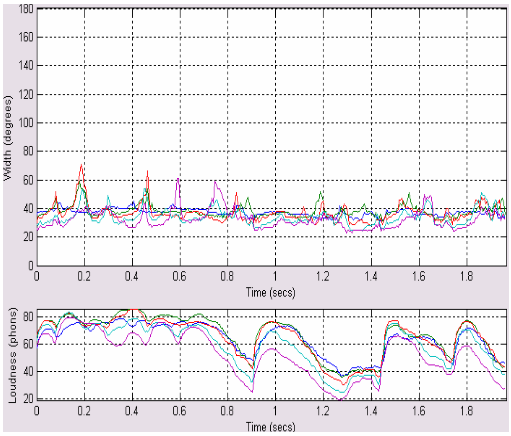


(b) Crosstalk-on

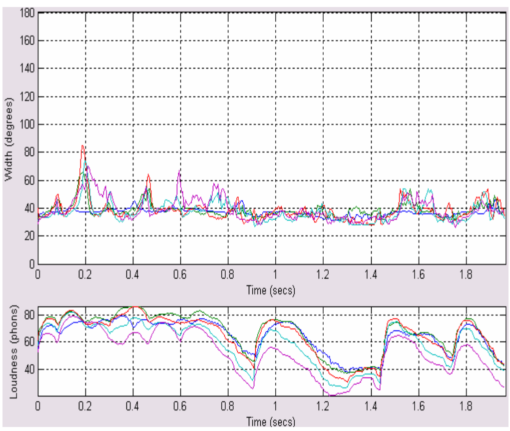


(c) Waveform

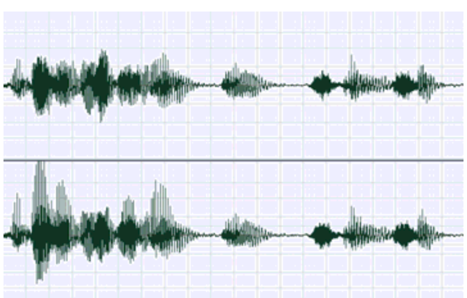
Figure 16: Plots of the width measurement made for the anechoic bongo stimuli of microphone array 2, with centre frequencies of 7050, 8600, 10750 Hz, and waveform of the binaural signal



(a) Crosstalk-off

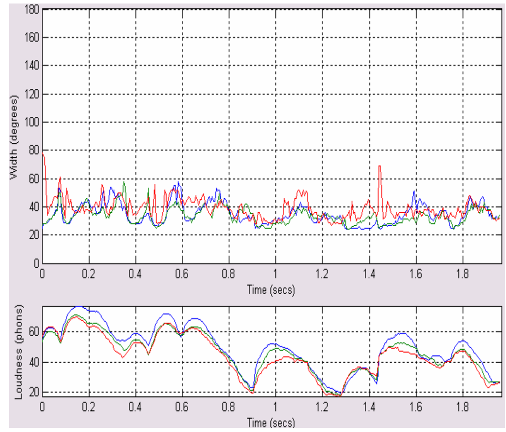


(b) Crosstalk-on

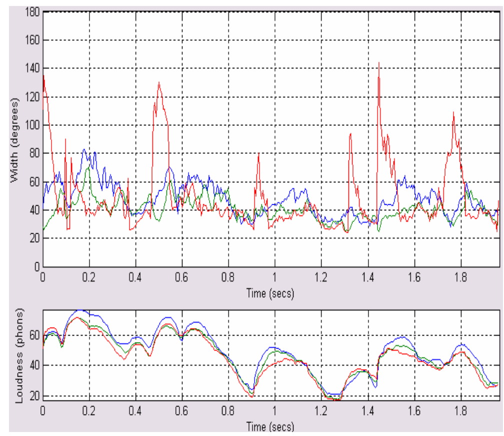


(c) Waveform

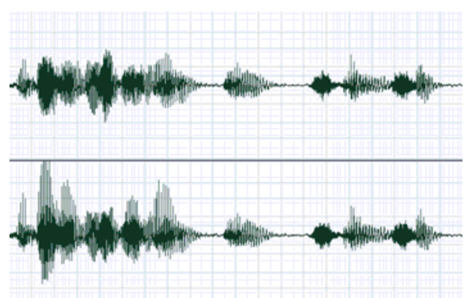
Figure 17: Plots of the width measurement made for the anechoic speech stimuli of microphone array 2, with centre frequencies of 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



(a) Crosstalk-off

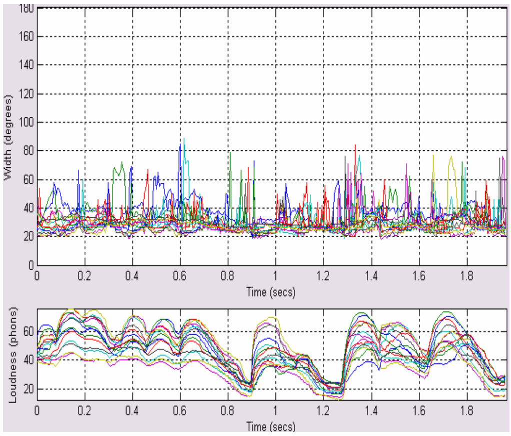


(b) Crosstalk-on

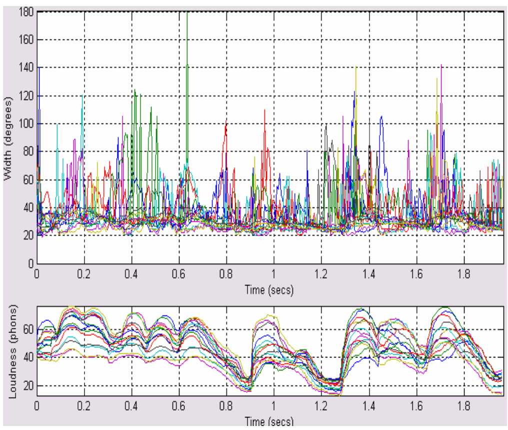


(c) Waveform

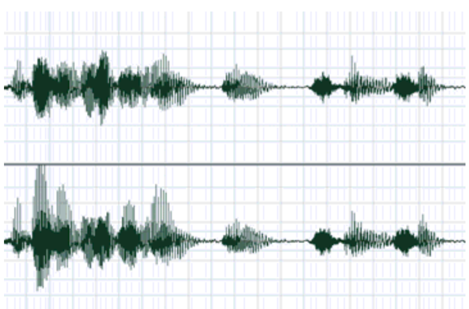
Figure 18: Plots of the width measurement made for the anechoic speech stimuli of microphone array 2, with centre frequencies of 700, 845, 1000 Hz, and waveform of the binaural signal



(a) Crosstalk-off



(b) Crosstalk-on



(c) Waveform

Figure 19: Plots of the width measurement made for the anechoic speech stimuli of microphone array 2, with centre frequencies of 1175, 1375, 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850, 7050, 8600, 10750 Hz, and waveform of the binaural signal