A binaural hearing model has been developed over a number of years that predicts the perceived width and position of sounds, over frequency and over time. The most appropriate methods for applying this model to evaluations of spatial impression are considered, including suitable test signals. Examples of a range of measurements are shown in a range of situations.

INTRODUCTION

A measurement that can predict the perceived spatial impression of sounds and provide the results in an intuitive manner would be a useful tool to aid the development of sound recording and reproduction systems. In this paper the term ‘spatial impression’ is intended as an overall descriptor which refers to all perceivable attributes that can be described in terms of three physical dimensions, which includes the majority of the attributes laid out in Rumsey’s spatial scene analysis paradigm [1].

Currently, it is necessary to conduct expensive and time-consuming listening tests every time the spatial impression of a sound system needs to be evaluated. Whilst it may never be possible to entirely replace subjective tests with an objective measure, the use of an objective metric can be used to support subjective testing, such as pre-selecting algorithms for more rigorous subjective evaluation, or helping to bridge the gap between the physical and perceptual domains. In this way, a reliable objective measurement that predicts aspects of perceived spatial impression can save time and money when used in combination with subjective evaluation.

The majority of previous work that has been undertaken to develop such objective metrics has focused on perceived location (or localisation). This is a logical place to start when developing such measures, as it is a relatively straightforward phenomenon that can be studied relatively easily, and it can be easily implemented as a measurement. However, its perceptual importance and whether it matches the way that humans perceive sound is questionable.

For instance, a sound recording and reproduction system can be evaluated by placing sound sources at given positions around the recording system, and then using a measurement to predict the localisation accuracy of each of these when the sound is reproduced. This should give an accurate prediction of the perceived localisation of sound sources in these positions, but will not necessarily give an indication of the spatial impression of the acoustical environment. Whilst a completely accurate reproduction of the sound source localisation may indicate that the acoustical environment will similarly be reproduced accurately, it is difficult to interpret how any errors in the reproduction localisation will affect the perceived spatial impression. For instance, the direct sound and reflections of an acoustical environment combine to give rise to a single perceptual component with a modified spatial impression [2] – it would be complex to determine this interaction based on the intended and actual localisation of each of the components (i.e. the direct sound and each reflection).

Therefore, a measurement system that can evaluate the attributes that contribute to spatial impression as perceived by a listener would be more useful. To use Rumsey’s scene-based description structure [1], this can be interpreted as the location (azimuth, elevation, distance), and size (width, height, depth) of various
elements of a sound scene such as individual sources, groups of sources or the acoustical environment.

There is still a great deal of work to be done to develop measurements that predict the perceived effect of all of these attributes for a wide range of signals. The author, together with colleagues, has developed a binaural hearing model that can predict the perceived location and width of stimuli. This is described in this paper, together with related information about how this can be implemented in the analysis of sound recording and reproduction systems. Finally, example measurements of a range of systems are shown, giving an overview of the capability of such a measurement system.

1 BINAURAL HEARING 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 (or reverberation). In the future, the aim is to develop the model to be able to automatically separate these two components.

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 [3, 4]. Whereas recent research has suggested that it is not an accurate representation of the physiology of the binaural hearing process [5], the predictions of models of binaural perception based on the IACC have shown remarkable similarity to experimental data relating to lateralisation (e.g. [6, 7, 8]), binaural detection thresholds (e.g. [9, 10, 11]), and the perceived source width of an auditory stimulus (e.g. [12, 13, 14]).

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 2. 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 [15]
- the use of a ‘running’ measurement to uncover variations in the perceived parameters over time [16]
- the effect of the frequency and loudness of the input signal is taken into account when predicting the perceived width [15, 17]
- predictions of the perceived location and width are integrated into a single output [18]
- the results are output in terms of an angle rather than an unintuitive value such as a coefficient [16]
- the effect of the IACC on the perceived localisation is taken into account [18]
- the effect of double peaks in the IACC calculation are taken into account in the width and localisation prediction [18]

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 1) 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: 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 2: Block diagram of the main processing stages of the binaural model.

Figure 3: Example output of the interactive plot display showing the combined width and location 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 [16]. 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.

There are a number of ways in which the model can be improved, including integration of results across frequency, automated separation of sources and reverberation, and further development of temporal smoothing algorithm.

2 MEASUREMENT PROCESSES

Now that a suitable binaural model has been developed to predict some aspects of the perceived spatial impression of stimuli, the optimum method for applying this needs to be developed.

The model could be used for either comparative or absolute evaluations of a reproduced sound field. This sound field may be affected by any and all components in the signal chain which, depending on the individual situation, may include the original source signal, an original acoustical environment (including the source, environment and capturing system), additional signal processing, some form of storage or distribution medium, reproduction signal processing, and the reproduction system (including the loudspeakers and the acoustical environment) and finally the binaural capture system. An absolute measurement will include the effect of each of these components, and to evaluate the effect of one of these requires the quantification and standardisation of all the other factors in the signal path.

Comparative evaluations are somewhat simpler, and are the basis of other perceptual quality meters, such as PEAQ [19]. As an example, if the aim of the recording and reproduction was to accurately recreate the perception of the original acoustical environment, then comparable measurements could be undertaken of both the original sound field and the reproduced version. Such an approach was taken by Furlong [20], and it is a useful and straightforward technique for evaluating a complete system. However, it is rare that all the parameters of a complete recording and reproduction system are in the control of one individual or organisation.

A more common type of test would be to evaluate the effect of one component in the recording and reproduction chain, and this can be done by exchanging or altering the given component, which will enable the evaluation of differences caused by this quantifiable change. This is also a relatively simple task, however it is complicated by the possibility that the other components in the signal chain will act to mask changes. For instance, a change to a spatial processing algorithm may have one effect with a certain reproduction system, but may produce a different effect with a different reproduction system. For this reason, the comparisons need to be conducted with care, and if possible in as many situations as possible.

Another important consideration in this type of test is the source signal that is used. Ideally, this would be representative of a wide range of stimuli that will be reproduced through the system under test. It should also include the extreme range of values of all the relevant physical parameters that may be present in typical stimuli in order to fully test the capabilities of the reproduction system. This is considered in more detail in the next section.

In order to limit complexity, this paper will focus on the problem of evaluating some aspect of a sound reproduction system, be it some form of processing algorithm, the loudspeakers and their setup, the acoustical environment, or the listening position. In this context, it is necessary to consider what subjective attributes are to be evaluated, as this will help to structure the application of the binaural model.

The overall characteristic that will be investigated could be described as the ‘quality of the spatial impression’. However, as discussed above, this is a multi-dimensional term, made up of many individual attributes. In addition, quality is recognised to be a higher level cognitive attribute, which includes the detection of lower level quantitative judgments together with an evaluation of the suitability of these towards a certain aim or expectation. The binaural model predicts the perception of lower level sensory attributes, so it is logical to limit the application to these attributes, at least until further work has been undertaken to map the relationship between these and quality for a wide range of situations and stimulus types.

Therefore, it would be simpler to measure low level attributes, based on the factors that the binaural model currently predicts. The predictions of the perceived width can be used to determine the accuracy and range of widths that can be reproduced by a system. Likewise, the predictions of perceived location can be used to determine the accuracy and range of localisation that can be reproduced by a system. Measurements can be made of the consistency of the results across time, position, frequency, or other variables. Finally the prediction of the perceived loudness can be used to evaluate whether there is any emphasis or suppression
of the reverberation, which may affect the perceived envelopment [21] 1.

3 TEST SIGNALS

As discussed in the previous section, one of the most important and challenging issues in testing sound reproduction systems is the selection of suitable test signals. These should be representative of the type of stimuli that will be reproduced through the system under test, and should cover a wide range of each relevant physical parameter in a relatively short time.

The potential experiment questions posed in the previous section dictate the properties of the test signals to a certain extent. To evaluate the range of locations and widths that can be reproduced, the test signal will have to contain a range of interchannel correlation values and level / time differences 2. To evaluate the accuracy of these factors, the ‘correct’ perceived effect of these parameters will have to be pre-determined. To evaluate the suppression or enhancement of reverberation, the test signal will have to include a reverberation-like decay in order to monitor any level compression or expansion effects that may affect how the reverberation is perceived relative to the direct sound.

3.1 Signal spectrum

One important factor in the choice of test signal is the effect of the signal spectrum on the result. The interaural cross-correlation coefficient (IACC) has been shown to be related to the perceived width of a sound, and is basically a measure of the similarity of the signals arriving at each ear. This can be affected by differing notches in the transfer response from the source to each ear when combined with the reflections from the acoustical environment. For instance, the plots in Figure 4 show a fast Fourier transform of a binaural sound field consisting of a direct sound and a single lateral reflection. It can be seen that this causes comb filters that are different at the two ears due to the differing path length of the reflection to each ear. The signal overlaid in the first plot will have a relatively high IACC as the spectral components are at frequencies where the response of the sound field is relatively similar at the two ears. However, the signal overlaid in the second plot contains spectral components at a frequency where there is a large difference between the sound field responses at the two ears; therefore this will cause a relatively low IACC.

Therefore, it is apparent that the spectral content of a signal can have a large effect on the resulting interaural cross-correlation. Based on this, if a test signal is to successfully detect any such constructive and destructive interferences, it needs to be spectrally complex in order to excite these differences. However, on the other hand if the signal is spectrally too complex, any differences that occur will be a small proportion of the overall signal. For example, a significant destructive interference across a narrow frequency range would have a large effect on the IACC if a sound consisted of two tonal components, one at the interference frequency and one elsewhere. However, if there were a large number of tonal components across a wide frequency range, the fact that the destructive interference is only a small proportion of the signal would mean that the IACC would remain close to 1.

---

1 This is not necessarily an attribute of spatial impression as defined in the introduction, though it is regularly considered to be similar and could be included in a measurement as the relevant information is required as part of the binaural hearing model.

2 This assumes that the system under test uses fairly simple time or level based panning. If the recording / reproduction system involves a more complex panning system then the localisation performance can be tested using that method of sound source positioning.
A wide-band test signal, such as noise, has the advantage that all components and potential interactions can be evaluated at once, and there is less risk of missing an interaction as would happen with the signal in the upper plot of Figure 4. It is also simple to vary the interchannel cross-correlation (ICCC) of a wide-band test signal, and amplitude envelopes that are representative of a reverberant decay can be introduced in a consistent manner. On the other hand, the disadvantage of a wide-band signal is that these interactions could be a small proportion of the overall result as discussed above, and noise is dissimilar to the music or speech signals that are most likely to be played through a sound reproduction system.

A tonal or harmonic signal would be more musically representative, and therefore the results of any measurements made of these would be likely to be more externally valid, especially when considering the effect of the spectral complexity on the result as discussed above. However, in order to detect all significant interactions, either a number of these signals would be required with differing pitches, or the signal will need to include a frequency sweep of some form – this will mean that the signal (and therefore the measurement and analysis time) will be longer. In addition, it would be difficult to vary the ICCC of such a signal in a meaningful way, and any amplitude envelope that is introduced will have to be repeated for each pitch – resulting in a longer signal duration.

As there is no way to compromise between these two signal types and still retain the requirements set out at the start of this section, it was decided to use two signals, one wide-band and one tonal.

### 3.2 Wide-band signal

The wide-band signal that was chosen is a noise-like signal made up of a large number of sine tones – a spectrogram of this is shown in Figure 5. The sine tones are spaced 5 Hz apart, based on the fact that the bandwidth of strong interactions and resonances such as modes in most listening environments are equal to or greater than this [22], meaning that this signal should excite any significant irregularities in the response. The starting phase of the sine tones is pseudo-random, in order to minimise the crest factor of the noise.

The ICCC can be swept from 0 to 1 or vice versa, and is varied by altering the arrangement of the tonal components in each channel. The number of output channels is theoretically unlimited, though a smaller spacing of sine tones may be required if there are a large number of channels.

The amplitude envelope of this signal is designed to test the dynamic response of the reproduction system. It has a transient attack, followed by a sustained segment that is 12dB lower in level than the attack. The sustain segment lasts for 3 seconds, during which the ICCC sweeps from 0 to 1 or 1 to 0, as specified. Following this, there is an exponential decay, which simulates a reverberant decay with a reverb time (RT60) of 2 seconds.

The perceived location of the sound can be manipulated by introducing interchannel time or level differences, based on previous experimental data such as those listed in [23]. Finally, the signal is filtered to give a frequency response similar to pink-noise.

![Figure 5: A spectrogram of the wide-band test signal showing the spectral content of the signal on the y-axis against time on the x-axis, with the magnitude shown on the colour scale in dBFS.](image)

This signal can be used to evaluate the following factors, for wide-band signals:

- the range of widths (which may be otherwise described as the ‘focus’ and ‘spaciousness’ of the reproduced sound) that can be reproduced by a system
- any enhancement or suppression of the reverberation
- the accuracy of the reproduced localisation.

### 3.3 Tonal signal

The tonal signal that was chosen is based on the research of Neher et al [24], who investigated the properties of musical signals and their effect on the IACC when reproduced in a room.

In order to ensure that there are at least two tonal components in each of the frequency bands in the binaural hearing model (so that there is an interaction to
measure), two harmonic complexes are used, which are harmonically related to each other. The first has a fundamental frequency of 55 Hz with 34 harmonics, the second a fundamental frequency of 220 Hz with 42 harmonics. This results in a signal with a frequency range of 55 Hz to 9240 Hz. The frequency components are held constant for half a second, sweep up an octave over approximately 3.5 seconds, then rest at the final pitch for half a second. Therefore, the frequency range at the end of the signal is 110 Hz to 18480 Hz. The sweep rate was derived from experimentation – it was chosen as a compromise of being able to see all significant deviations in the result whilst keeping the stimulus duration as short as possible. A spectrogram of the resulting signal is shown in Figure 6.

Figure 6: A spectrogram of the tonal test signal showing the spectral content of the signal on the y-axis against time on the x-axis, with the magnitude shown on the colour scale in dBFS.

The level of each harmonic in the tone complex is equal, though it is filtered after generation to give a frequency response similar to pink-noise. It is not possible to vary the ICCC in a meaningful way using this stimulus, so this is kept at a value of 1. As for the wide-band stimulus, the perceived location of the sound can be manipulated by introducing interchannel time or level differences.

This signal can be used to evaluate the following factors, for tonal signals:
- the narrowest width (which may be otherwise described as the ‘focus’) that can be reproduced by a system
- the accuracy of the reproduced localisation
- the variation or consistency of the reproduced localisation and width over frequency.

4 EXAMPLE MEASUREMENTS

A number of 2-channel stereo reproduction systems in a range of acoustical environments were tested by reproducing both test signals through each and capturing the results with a binaural head and torso simulator (HATS). The resulting binaural signals were then passed through the binaural hearing model. A subset of these results is discussed below.

4.1 Listening room

The first reproduction system consisted of two monitor loudspeakers positioned at a height so that the tweeters were at ear level, a distance 2m from the HATS, and ±30° from the median plane, in an ITU-R BS 1116 standard listening room. It was expected that this system would be representative of a fairly high quality reference system. The results of the measurements of this system are shown in Figure 7 to Figure 11 below.

Figure 7 shows the combined width and location results for the wide-band stimulus, averaged over all the measured frequency bands, together with the loudness in each frequency band. An ideal width and location result for this test signal would be the reproduction of wide range of widths with a smooth transition in between, with a location centred on 0° throughout. In other words, the stimulus should start narrow, and then should gradually increase in width to at least ±30° (the loudspeaker spacing) as the ICCC of the signal gradually moves from 1 to 0 between 0.5 and 3 seconds. An ideal loudness result for this test signal would be a
short transient attack that is approximately 12 phons above the loudness of the sustain section, followed by a decay of approximately 30 phons per second.

It can be seen from Figure 7 that the overall trend follows the expectation that the sound should increase in width over the duration of the signal, but the area of paler colour at the start of the stimulus indicates that not all frequency bands are in agreement. Further analysis indicated that some of the high frequency bands had a significant interaural level difference, as shown in Figure 8, which caused the prediction of the localisation of these frequency bands to be to one side or the other. This was found to be due to the fact that the loudspeakers used in the system under test had non-matching tweeters, with differing level and phase responses.

Figure 8: Output of the binaural hearing model interactive plot display showing the interaural level difference 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 above 1 kHz, for the wide-band stimulus reproduced in the listening room.

The results for the tonal stimulus are shown in Figure 10 and Figure 11. Again, there is a large amount of variation at higher frequencies caused by the mismatched tweeters.

Analysis of the loudness measurements indicates that this reproduction system has a relatively linear amplitude response. This means that this system will not emphasise or suppress the perceived level of the reverberation with respect to the direct sound.

Figure 9: Output of the binaural hearing model interactive plot display showing the combined width and location 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 up to 1 kHz, for the wide-band stimulus reproduced in the listening room.

Figure 10: Output of the binaural hearing model interactive plot display showing the combined width and location 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 all frequency bands, for the tonal stimulus reproduced in the listening room.
The measured results in Figure 11 show that below 1 kHz there is little variation in width and location as the tonal signal sweeps across frequency. This is to be expected as the reproduction room is acoustically treated to reduce modal and reflection-based artefacts, therefore minimising frequency-dependent interactions.

This analysis has indicated that the spatial response of this reproduction system is relatively neutral across some of the frequency range, with consistent localisation and width cues, a wide range of reproduced widths, and no suppression or enhancement of the reverberation. However, it has indicated problems at higher frequencies, which were subsequently found to be due to mis-matched high frequency drivers in the loudspeakers.

4.2 Automobile

As a comparison, measurements were made of an in-car audio system where the 2-channel stereo is reproduced over multiple loudspeaker drivers. This is a very different situation to the listening room measurements above, both in terms of the loudspeaker arrangement and the acoustical environment. The HATS was set up in the driver’s seat, and binaural signals were recorded for both test stimulus types. The results of the measurements of this system are shown in Figure 12 to Figure 15 below.

Figure 12 shows the combined width and location results for the wide-band stimulus, averaged over all the measured frequency bands. As before, an ideal result for this test signal would be to start narrow and centred on 0°, then increase in width. It can be seen in Figure 12 that the perceived width prediction does increase as the ICCC reduces, though not as clearly, and to a lesser extent, than the listening room measurement.

In the previous set of measurements, the results at low frequencies showed a much clearer trend. Similar analysis (including the frequency range below 1 kHz) was made of the automobile measurements, and the results are shown in Figure 13.

It can be seen in Figure 13 that the variation in width is still not clear even in this frequency range. Further examination of individual frequency bands indicates that the variation in IACC (and therefore the width) is relatively small as the ICCC is varied, especially compared to the results in the listening room. In addition, in a number of frequency bands the location is displaced to one side or the other. These factors are likely to be a result of the off-centre listening position and the use of multiple drivers in different locations for each channel.
The results for the loudness measurement show that, as for the listening room system, the automobile audio system relatively accurately reproduces the amplitude envelope of the wide-band stimulus. This indicates that the level of the reverberation will not be enhanced or suppressed compared to the direct sound. It can be seen from the end of the decay that the dynamic range of this system is less than the listening room, but this is caused by the increased background noise of this measurement situation.

The measurements of the tonal stimulus (shown in Figure 14 and Figure 15) show a similar trend to those of the wide-band stimulus. The tonal stimulus has an ICCC of 1, and yet the prediction of the perceived width is still relatively wide. In addition, the width and location of the sound in each frequency band varies as the pitch of the stimulus changes. This indicates that the imaging of this system is inconsistent across frequency.

The measurements of this system indicate that its spatial performance is less neutral than the listening room system. This is to be expected due to the inherent problems of an automobile system, such as the off-centre listening position, the use of multiple drivers in a range of positions, and the non-optimal acoustical environment that is imposed. Specific problems that were found using this analysis were inconsistent location and width over frequency and a limited range of widths that can be reproduced. Further use of this analysis technique could help to suggest solutions to these problems, and to test updated systems.

5 FURTHER WORK

There are number of ways in which this work can be developed further. Firstly, the binaural hearing model and its display could be improved by revising the temporal smoothing algorithms that are employed, developing an automated method for separating source-related and environment-related aspects, and by investigating the optimum method to integrate the results across the frequency bands rather than taking a
loudness-weighted average. Secondly, the measurement method could be developed further, by taking into account head movements made when listening, and by making developing a database of evaluations so that the predictions of the low-level perceptual attributes can be converted to a quality or preference scale.

6 CONCLUSIONS

This paper has briefly summarised a binaural hearing model that has been developed to predict the perceived localisation and width of a binaural recording and display the results in an intuitive manner.

The potential problems in employing such a binaural hearing model to testing sound recording and reproduction systems have been considered. Based on this, a method for testing the spatial response of a sound reproduction system has been proposed, together with the type of attributes that can be tested.

It was discussed that one of the most important factors in testing sound reproduction systems using this method is the selection of appropriate test signals. Two test signals were proposed, one wide-band and one tonal, which aim to uncover the maximum information about the spatial performance of a reproduction system whilst keeping the measurement time to a minimum.

Measurements were made of a range of sound reproduction systems using this technique. Examples were shown of two systems, and it was found that the measurement technique gave useful information about the spatial performance of these.

Finally, a number of methods for further developing the measurement technique were suggested.

ACKNOWLEDGEMENTS

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK – grant GR/R55528/01 and EP/D049253.

REFERENCES


