Acoustics 2008
Geelong, Victoria, Australia 24 to 26 November 2008

Acoustics and Sustainability:

How should acoustics adapt to meet future demands?

Evaluation of binaural loudness models with signals of different diffusivity

Luis Miranda and Densil Cabrera

Faculty of Architecture, Design and Planning, University of Sydney, NSW 2006, Australia

ABSTRACT

The effectiveness of three binaural loudness calculation procedures using dummy head recordings was tested to predict the results obtained from a listening test in which the sounds were broad noise bands with different degrees of diffusivity. The loudness calculation results obtained with the binaural recordings were also compared with the loudness calculation results obtained from an omnidirecitonal microphone. Using rms error as an indicator, one loudness calculation procedure proved to be more accurate than the other two for the binaural recordings while the loudness calculation results obtained with the omnidirecitonal microphone unexpectedly proved to be more reliable in predicting the results from the listening test than any of the binaural loudness calculation procedures tested.

INTRODUCTION

Loudness can be described as the level sensation of a sound; and, while it is highly related to physical intensity, it is also related to other characteristics of the sound, for example, bandwith, frequency content, length, etc. Loudness measurements were introduced as a method to correctly quantify loudness. Loudness measurements are usually carried out with a set of listeners which are subjected to a sound and in turn they most produce loudness estimation values through a common procedure; examples of this procedures are: magnitude estimation, magnitude production and intensity matching. Fletcher and Munson in their 1933 pioneering study (Fletcher and Musnon, 1933) compare the loudness pure tones of different frequencies; this work established the use of equal loudness curves for pure tones. The refinement of the loudness measrument procedures and advances in loudness theory in turn lead to the development of what now is a commonly used unit. The natural loudness unit used is the sone, introduced by Stevens in 1955 (Stevens, 1955), where he proposed a very simple mathematical formula to relate sound pressure level of mid-frequency pure tones to sones.

A year later Stevens publishes a series of papers (Stevens, 1956a; Stevens 1956b) in whih he describes the calculation of the loudness of a sound from the measurement of its spectral content; this became the base for the standard ISO 532-A. In the following decades various refinements have been made, resulting in more complex loudness calculation procedures or loudness models; just to mention a few we can name the models that have been standardized: the model proposed by Zwicker and Scharf in 1965 (Zwicker, 1965), standardized as ISO 532-B and the model proposed by Moore, Glasberg and Baer in 1997 (Moore et al., 1997), standardized as ANSI S3.4-2005. These and some other models have been proved to be reliable in predicting the loudness of a wide variety of sounds, but one of the main characteristics of these models is that they are designed to predict the loudness of frontally incident sounds, therefore all the sounds analyzed are assumed to be diotic (i.e. the same sound at each ear).

In the last few years there has been an increasing interest in modeling the loudness of dichotic sounds (i.e. different sound at each ear, including but not limited to binaural signals). The loudness model proposed by Moore, Glasberg and Baer has been revised in a few occasions, but the most relevant to this paper was published in 2007 (Moore and Glasberg 2007). In the older model the loudness of the sound analyzed for one ear was calculated and then multiplied by two. The principle behind this process is that the two ears perform a perfect summation of loudness; in the new revised version an interaction between the two ears is modeled and can be briefly described as a broadband inhibition of one ear to the other resulting in less than perfect summation.

The other binaural model tested in this paper is the one proposed by Sivonen and Ellermeier (2008). This model is the result of some research previously by Sivonen and colleagues (Sivonen, 2006 and 2007). In those preliminary studies the interaction of the two ears was analyzed in anechoic and reverberant sound fields and a 3 dB summation across ears was proposed. In the loudness model proposed in (Sivonen, 2008) the 3 dB summation is the first step of a process where the signal is then transformed to a diotic signal that can then be used as the input to any of the diotic loudness models, including the ones mentioned above.

There is one common factor about these two models; this is that they both base their assumptions in experiments conducted where the main dichotic feature was the from source direction. These models do not take into account the phase relation of the two signals arriving at the two ears or any time varying aspect of the signals; their input is only the spectral content of the signals.

The main difference between these two models is the way they sum the spectral content of the signals. In the Moore et. al. model the loudness summing across ears is the *last* step in the process, therefore summing in sones and the spectral content inhibition of one ear is broadly tuned; while in the Sivonen et. al. model the summing across ears is the *first* step and it occurs in physical sound units, not in sones. Furthermore, this summing is frequency specific and does not affect neighboring bands. The first Moore et. al. model (which assumes diotic signals) is also tested in this paper as a comparison of its performance in relation to the dichotic models; the loudness of each ear is calculated separately and then simply summed.

In this paper these models are tested with stimuli where not only the direction of the signals is controlled but also the cross-correlation between the signals. A listening test was conducted to obtain results that could be used to test against the results obtained from the loudness models.

LISTENING TEST: METHOD

A loudness matching listening test was conducted to quantify the sound pressure level adjustment required for equal loudness for signals played from various directions and with various degrees of diffusivity. For this test an eight channel electroacoustic system was used in an anechoic chamber and the test signals used were broad noise bands together encompassing much of the audible spectrum divided in low, mid and high, plus a broadband pink noise.

Subjects

Nineteen listeners participated in this experiment. Their hearing thresholds were measured according to ANSI S3.21-2004, “Methods for manual pure threshold audiometry”, and the results of the listeners whose threshold levels exceeded 20 dB in any of the bands tested (500 Hz to 8 kHz) were excluded from the analysis, leaving sixteen participants. The age of the participants ranged from twenty-one to thirty-three years, with a mean of twenty-seven. The group was composed of four female and twelve male subjects. Ten out of these sixteeen listeners had previous experience in similar listening tests.

Loudspeaker setup and listening room

The experiment was conducted an anechoic chamber designed to be anechoic for the 200 Hz 1/3-octave band and above. The background noise in this anechoic chamber is below the hearing threshold.

The loudspeaker setup consisted of eight Yamaha MSP5 loudspeakers. The loudspeakers were set at 1.2 m above the floor and 1.5 m away from the measuring position. The centre loudspeaker was positioned in front of the measuring position, that is, with as azimuth of 0°. The other loudspeakers were positioned at azimuth angles of 45°, 90°, 135° and 180°, and the last three loudspeakers were positioned symmetrically to the loudspeakers angled at 45°, 90° and 135°of azimuth.

The subject was seated at the centre of the loudspeaker arrangement. A screen was positioned just below the 0° (or ‘centre’) loudspeaker. The subject was asked to face this loudspeaker at all times and this screen provided a visual reinforcement to make the subject look straight ahead.

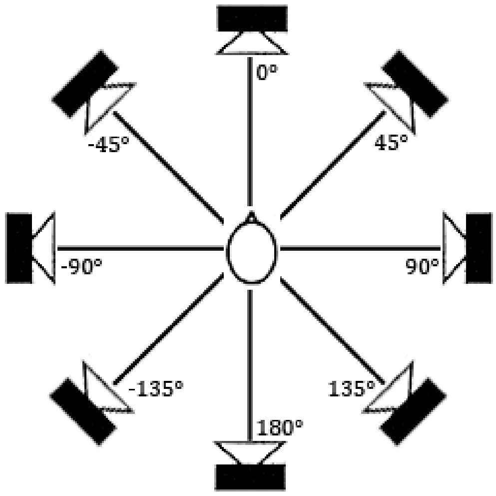

Figure 1. System setup.

Signal playback and response data collection

A computer located outside the anechoic room was used to provide playback and to record the response of the subjects. This computer was connected to a digital audio interface (MOTU 896HD), which provided individual playback for the eight channels. Individual signals for each loudspeaker were necessary to ensure proper calibration between channels, and to for the creation of a horizontally diffuse soundfield at the listening position.

The experiment was run using a program developed in Max/MSP. This program controlled the playback and playback level of the signals, provided level changes for the signals being adjusted and recorded the frequency band, diffusivity and predominant channel of the signals being tested and the level change made by the subject.

Stimuli

For this experiment three parameters of the signals were tested: spectral content, diffusivity, source direction. These parameters are individually explained in the following subsections. The resulting stimuli are a combination of all values of these parameters.

Spectral content

The stimuli used for the listening experiment were four different steady state noise signals. Using FFT filtering in Max/MSP noise files were created with different spectral content parting from pink noise. The spectrum was divided in three, the broadband noises were: ‘low’, encompassing the 63, 125 and 250 Hz octave bands; ‘mid’, encompassing the 500, 1000 and 2000 Hz octave bands; and, ‘high’, encompassing the 4000, 8000 and 16000 Hz octave bands with a cutoff point at 20000 Hz. Pink noise was also used in this experiment, this was low pass filtered below the 63Hz octave band due to the loudspeaker system’s frequency response

The differences in transfer function for each loudspeaker to the listening position were corrected by measuring the impulse response of each loudspeaker separately and then, from the impulse response, generating inverse filters. The inverse filter for each loudspeaker was convolved with each stimulus resulting in particular noise files for each loudspeaker. The files were played back at a sampling rate of 44.1 kHz and a 16-bit resolution.

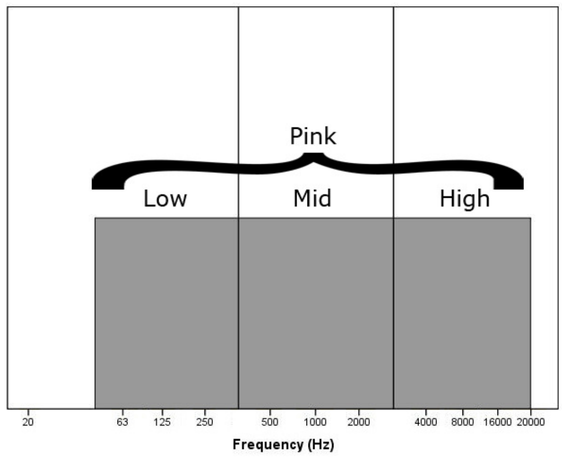

Figure 2. Frequency content of the stimuli used.

Diffusivity

Each of the eight loudspeakers had independent (uncorrelated) noise files for the low, mid, high and pink signals. This was done to ensure minimal coherence between the noise files played back from each loudspeaker. The correlation function in Matlab was used to check that the coherence between each noise file was kept to a minimum.

Four degrees of horizontal diffusivity were tested in this experiment. The diffusivity levels were created by changing the mix between the most powerful loudspeaker signal and the signals from the rest of the loudspeakers. Having different noise files with no correlation playing from each of the loudspeakers let us create an approximately diffuse sound field. In theory perhaps we could have created the impression of a diffuse field with just the two lateral loudspeakers playing uncorrelated noise files with the same power to the listening position, but if the listener were to move his/her head by a minimum amount the effect might change appreciably. The rear loudspeakers were only used in the experiment for aiding in the creation of this diffuse field and thereby giving the subject a certain degree of freedom for head movements.

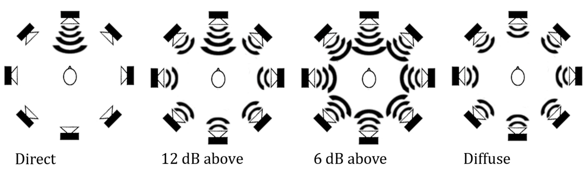

Figure 3. Diffuse conditions tested in the experiment.

The four different degrees of diffusivity tested were: (i) completely diffuse field (where all the eight loudspeakers play independent noise signals at the same level); (ii) direct sound (where only one loudspeaker plays noise); (iii) direct sound 6 dB above the levels of each of the loudspeakers creating a diffuse field (in this all the loudspeakers play noise at a certain level except that one loudspeaker that plays noise 6 dB above the level of each of the other ones; i.e. the main loudspeaker makes a contribution 2.5 dB below the summed contributions of the remaining seven loudspeakers); and, (iv) the same condition as (iii) but with the dominant loudspeaker 12 dB above the level of each of the other ones (or 3.5 dB above the summed contributions of the other loudspeakers). The 6 and 12 dB levels were chosen after listening to 3, 6, 9 and 12 dB above the individual loudspeakers forming the otherwise diffuse field in the setup and deciding that these conditions were discernible enough from the other conditions tested and also from each other.

Source direction

For this experiment only five directions were tested: 0°, 45°, 90°, -45° and -90°. Assuming left-right symmetry, this was reduced to three directions in the analysis. The loudspeakers behind the 90° line were only used to provide a more stable diffuse field and were not used as dominant sound sources. For this directions all the diffuse conditions explained above were tested. It should be noted that the fully diffuse condition has no direction.

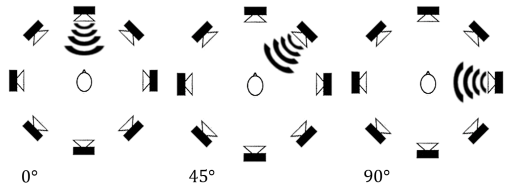

Figure 4. Three of the five directions tested in the experiment. The other two directions -45° and -90° were combined with 45° and 90° respectively in the analysis.

Playback level

The playback level used was 3 sones, this level was chosen because it was the highest level that still gave enough headroom for the lowest frequencies to be played back without audible distortion (for the highest conceivable subjective magnitude adjustment). The system setup was calibrated using a Bruel & Kjaer 4190 microphone placed at the centre of the listening position at a height of 1.2 m. The sound pressure level to loudness (sone) modeling was performed in PsySound3 (Cabrera, 2008a) (available as a pre-release version in www.psysound.org) according to Moore, Glasberg and Baer's steady state loudness model (Moore, 1997). The system was calibrated with an accuracy of ±0.2 dB.

Experiment design

In order to find how loudness sensitivity is affected by the experiment parameters, the subjects were asked to match the loudness of two sounds; a reference sound was played first and marked as A, the sound to match was marked as B. The sound matching procedure would produce level difference results (in decibels) from the reference value to the adjusted value; and this level difference can be easily analyzed. The pairs of sounds used were of the same frequency content. The reference in this experiment would always be the direct (non-diffuse) sound coming from the centre loudspeaker; this was chosen as the reference sound since, the model used for calibration is for frontally incident free field (Moore, 1997), and so other field conditions were not defined by the model. The matching sounds were of various levels of diffusivity and predominant direction; as stated before, only the front five loudspeakers were tested for the loudness match, the back three only aiding in creating the diffuse field. This gave a total of 64 different pairs: (5 (match loudspeakers) x 3 (levels of diffusivity) x 4 (signal spectra)) + (1 (fully diffuse level) x 4 (signal spectra)). The whole test comprised three repetitions of these pairs, giving a total of 192 pairs.

The stimuli were played back in a continuous loop so that the subject could switch between the reference and the sound to match as necessary (but the two sounds were not present at the same time). The subjects were given a keypad to make adjustments in level using up and down arrows. The steps of these adjustments were 1 dB, although the subjects did not know the exact level being changed. When the subject made a decision a key would take him/her to the next pair of stimuli.

The order of the 192 stimulus pairs was divided in sets of three 64 pairs as indicated above. Each set was run one after the other and the pairs within each sets were completely randomized, and was different for the three sets and for all subjects. The initial level of the sound to match was also randomized from -20 to +20 dB above or below the reference level. All the subjects completed the test in one session ranging from 45 minutes to close to 2 hours. The subjects were advised to take a break at the middle of the session, indicated by a counter on the computer screen. This was done to prevent fatigue and unreliable results.

LISTENING TEST: RESULTS

Consistency checks

The individual subject results were analyzed in order to exclude inconsistent subjects. The results were first checked for inconsistencies in left-right symmetry within the subject's own results. To perform this analysis the response between the centre loudspeaker and the loudspeaker located at 45° azimuth and the response between the 0° loudspeaker and the loudspeaker located at -45° azimuth were averaged. These averages were then subtracted which would indicate a bias towards one of the ears. The same was done with the results for the loudspeakers located at 90° and -90° azimuth and the two results averaged. The results for the direct comparison between the 45° and -45° loudspeakers and 90° and -90° loudspeakers were also analyzed. None of the subjects showed a substantial bias towards a particular side, therefore none of the subjects were excluded on this basis.

Combined results of the subjects were analyzed to identify outliers prior to the final analysis. Histograms of the results for each combination of centre to 45°, centre to -45°, centre to 90° and centre to -90° were plotted with normal distribution curves. The results tended to follow a normal distribution curve and the outliers were easily identified. The subjects that consistently lay outside the normal distribution curve were excluded. None of the subjects were excluded on this basis.

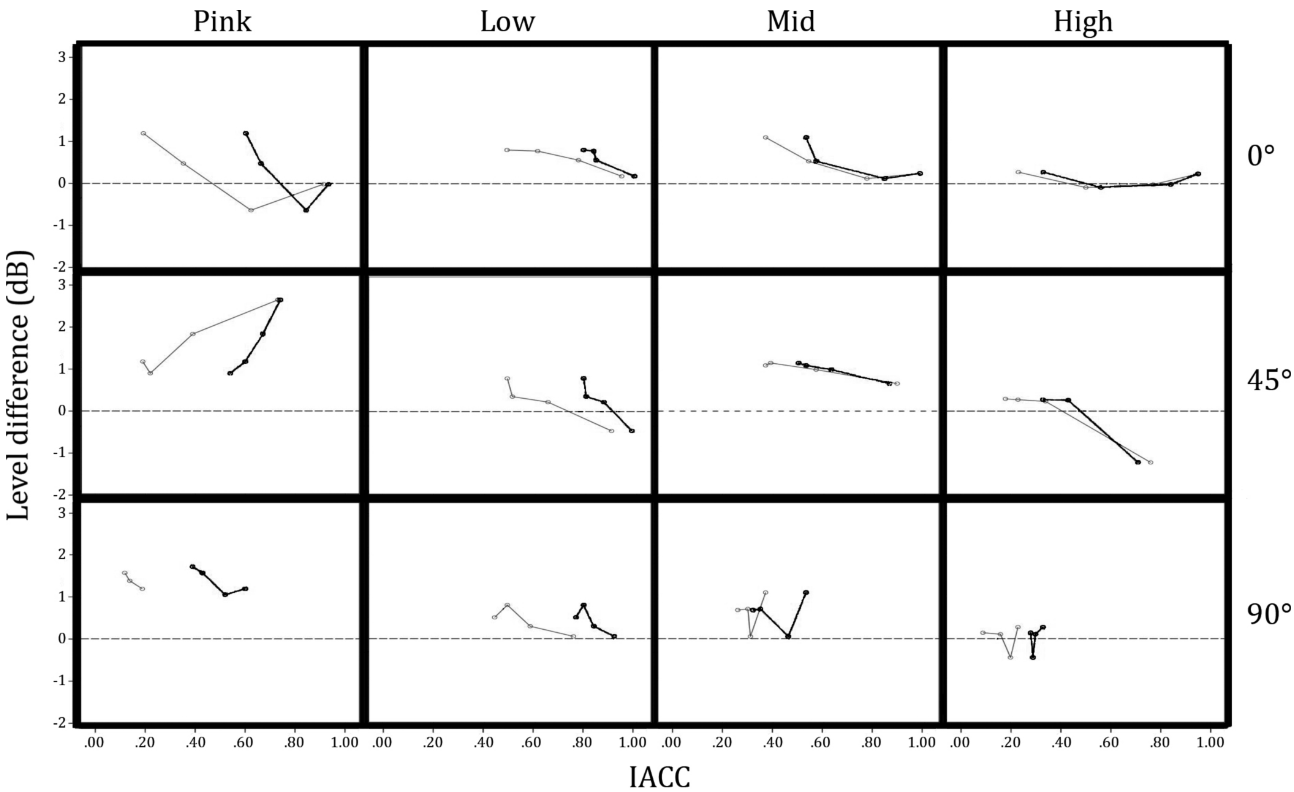

Figure 5. Level difference plotted against IACC. The light line corresponds to the A-weighted value; the dark line corresponds to the un-weighted value. A positive value indicates a greater sensitivity to the sound being compared to the frontal reference.

Results analysis

Results were analysed in terms of the strength and significance of effects of the independent variables (spectral content, diffusivity and direction) using factorial analysis of variance (ANOVA). The comparison between the levels obtained and the frequency content and direction returned significant values, the diffusivity did not. The frequency content had values of F=32.673 and p<0.001; the direction had values of F=6.910 and p<0.001; finally the diffuse condition returned the least significant values, F=2.848 and p=0.36. This tells us that the frequency content and loudspeaker direction have a more decisive influence on loudness than the diffusivity of the signal heard. There is a strong dependence of loudness on the direction x frequency content combination. The values returned being F=11.324 and p<0.001. This means loudness depends highly on the directivity of the source and the frequency content at the same time. Another combination that shows influence on loudness is the frequency content x diffusivity combination. This had values of F=4.702 and p=0.001. This means frequency and diffusivity have strong dependence on loudness, but the important conclusion we can draw from this is that not all the diffuse conditions might have the same effect on loudness and that this might change with frequency content. On this point the direction x diffusivity didn't show a significant influence, which means that the effect of diffusivity on loudness isn't direction dependent.

A physical analysis method that should characterize the subjective diffusivity of the soundfields is be the Interaural Cross Correlation. In order to do this, each signal combination was played and recorded with a head and torso simulator (HATS – Bruel & Kjaer type 4128C). From these binaural recordings, IACC was calculated, which yields one simple number that describes the relationship between the signals arriving at the two ears. This was done in PsySound3 following the method explained in by Ando (1998). In Figure 5 the subjective results will be plotted against IACC (which varies from 1, or maximally correlated, to 0, or no correlation). Theoretically the values from the 45° and -45°, and 90° and -90° should be the same, and the values measured were within ±0.03. Due to their long wave periods, lower frequencies tend to yield higher IACC values. The lower IACC values that are seen for 90°, even for direct sound, are a reflection of the relatively broad bandwidth of the noise signals

The method explained by Ando (1998) applies an A-weighting filter to the signal prior to performing the IACC. The result is an IACC that is weighted according to the most typical loudness filter deemphasizing the lower part of the spectrum. As an alternative method results without the weighting are also presented in Figure 5.

BINAURAL LOUDNESS MODEL EVALUATION

In this section the results of the experiment are evaluated using some of the loudness models referred to in the introduction. The models chosen to be evaluated were the original model of Moore et. al. (1997), the binaural model Moore et al. model (2007) and the most recent binaural model proposed by Sivonen et. al. (2008).

The procedure to evaluate the loudness models is the same for all of them. The models are evaluated using the signals recorded with the dummy head. These signals were played back and recorded at the calibration level at the centre of the head position and then adjusted according to the results obtained from the listening test. Based on the subjective data of the experiment, all of these adjusted signals should have a loudness of 3 sones (which was the modelled loudness of the reference stimuli). The modelled loudness of these adjusted recordings was then calculated. A model would be more accurate when the difference between the loudness of the reference stimuli and that of the adjusted stimuli is close to zero. Hence for multiple stimuli, the accuracy of a model can be assessed through the root mean square (rms) deviation from 3 sones, with a low value indicating accuracy.

Figure 6 show the results obtained from these calculations. The topmost graph shows results for the Moore et al. (1997) model. Although this was presented as a diotic model, in this evaluation it is implemented using input from binaural recordings made with the HATS. The outer ear transfer function was that of the HATS instead of the free field response given in the paper – that is, the filtering was done acoustically as part of the measurement process. The other small difference between the model as originally proposed and this implementation is the summation of monaural loudness of the two ears to generate an estimate of binaural loudness (by contrast, the original model assumes that the signals at the two ears are identical, and so multiplies the monaural loud ness result by 2).

The middle and bottom graphs included in Figure 6 show the results for the other two binaural models. Since the Sivonen and Ellemeier (2007) model performs binaural summation before spectral values are input into an arbitrary diotic loudness model, in this case we were able to use the Moore et al. (1997) model subsequent to binaural summation, meaning that all three evaluations are using the same model with the difference being where and how summation is performed. From these analyses, it appears that the Sivonen and Ellemeier (2007) produces somewhat better estimates of binaural loudness than the other binaural loudness models for the stimuli tested, and evidenced by a lower rms deviation from 3 sones.

LOUDNESS MODEL EVALUATION WITH OMNIDIRECTIONAL VERSUS BINAURAL RECORDINGS

Traditionally, loudness models were designed to predict the loudness of frontal incidence sounds and the input to such loudness models would come from an omnidirectional microphone positioned in the place where the centre of the listener's head would be; and the effect of the head would be taken into account with the implementation of a transfer function through signal processing. As opposed to this, in the previous section, we used binaural recordings made with a dummy head to calculate the loudness of stimuli, thereby implementing the transfer functions acoustically. It is obvious that the omnidirectional method is simpler and more cost effective, but it does not take into account the directional attributes of the stimuli as it normally assumes they are all frontally incident. Therefore in this section the two methods (omnidirectional versus binaural) are compared to determine if there is a clear advantage of using one method or the other.

For this section the stimuli used in the listening test were adjusted in level according to the results from the listening test. These adjusted stimuli were played back through the loudspeaker setup and recorded using a omnidirectional microphone situated in the centre of the setup, in the place where the centre of the listener's head would be positioned. The recorded stimuli were analyzed using the Moore et. al. 1997 (Moore, 1997) loudness model (with the frontal free field outer ear transfer function included in the calculation) and the results compared. Figure 7 shows the results from the loudness model calculations, in terms of deviation from the reference stimulus (which is 3 sones). Remarkably, the rms deviation is substantially less than those of the models implemented using dummy head measurements.

DISCUSSION

The rms error of these three models was plotted against IACC as this gives us a better way of characterizing the diffusivity of the signal as received by the two ears. The results were chosen to be graphed separated by frequency content to see if some trend existed. Linear fit lines were graphed, even if their fit was poor, in order to give us an idea of the trend of the results.

A common factor we can see from the results of the three models tested is that the high frequency stimuli have a tendency to be over estimated. This means that the loudness predicted by the model should be in general lower than the results that are being obtained at the moment. In a similar comparison done by the authors in a listening experiment with narrower bands of noise (Cabrera et al. 2008b) the results for an 8 kHz octave band had a tendency to be overestimated by the loudness models tested. This band is included in the high frequency content signal. A possible explanation for this overestimation is that the head-related transfer functions of the HATS differ in magnitude from those of the average subject in the experiment within the 8 kHz octave band. Hence, the results obtained in the present experiment may be influenced by this error tendency in the physical model model more than the diffusivity of the signal. The largest differences in loudness for these models are for signals with low IACC, which occurs particularly for signals coming from 90°.

Another feature we can see for the models is that the pink noise signal has a tendency to be underestimated. This means that the loudness result for this signal is lower than the loudness result expected. This is common to all models and has the biggest differences compared to the low and mid frequency content signals. The other two signals, low and mid frequency content, are slightly different for the different models. In the Sivonen and Ellermeier model we can see that the low frequency content signal tends to have a smaller error than the mid frequency content signal. This situation is the opposite for the other two models.

With the rms error as an indicator we can see that the Sivonen and Ellermeier model is the one that is best at predicting the results from the listening test using head and torso simulator measurements. There is little difference in performance between the Moore et al. (1997) model and the Moore and Glasberg (2007) model when applied to binaural signals. Yet surprisingly, the older Moore et. al. (1997) model without a dummy head, and with the identical (diotic) partial loudnesses of the two ears simply summed, is better at predicting the results than the more sophisticated binaural models. We would have expected that if we take into account the direction-dependent head related transfer functions and the binaural effects of soundfield diffusivity (by using a head and torso simulator) in the loudness calculation these results would be better than a model that ignores these influences.

Although this result appears to raise the question of whether binaural loudness models work, in fact the question is whether such models need further refinement to account for soundfields of variable diffusivity. For the directional loudness experiment described by Cabrera et al. (2008b) which did not involve any soundfield diffusivity, the authors have since performed the same rms error calculations using binaural and omidirectional microphone recordings. In that experiment, the model using the omnidirectional microphone results was the poorest. An rms deviation of 1.11 sones was found for the omnidirectional implementation of Moore et al. (1997), compared with: 0.42 sones for the same model applied to binaural recordings; 0.62 sones for Moore and Glasberg (2007); and 0.37 sones for Sivonen and Ellermeier (2008). The problem appears to be, then, that none of the existing binaural loudness models are based on subjective data that involve substantial variation in soundfield diffusivity.

CONCLUSION

Three binaural loudness models were tested for this paper. The model proposed by Sivonen et. al. (2008) proved to be the best at predicting the results from the loudness matching listening test conducted. In a surprising result, the loudness model that assumes that the signals tested were diotic was even better at predicting the results from the listening test. These results show that there are still some improvements that should be done to the existing binaural loudness models. Improving these binaural loudness models can provide better understanding of binaural hearing that can be used to develop applications, for example in multichannel loudness metering.

With regard to the effect of diffusivity on loudness, the experiment results show effects that differ between the signal spectra, and do not show any effect independent of the signal spectrum. With the present data, the main conclusion on the effect of diffusivity is that further experimental studies are required to provide greater clarity on this issue.

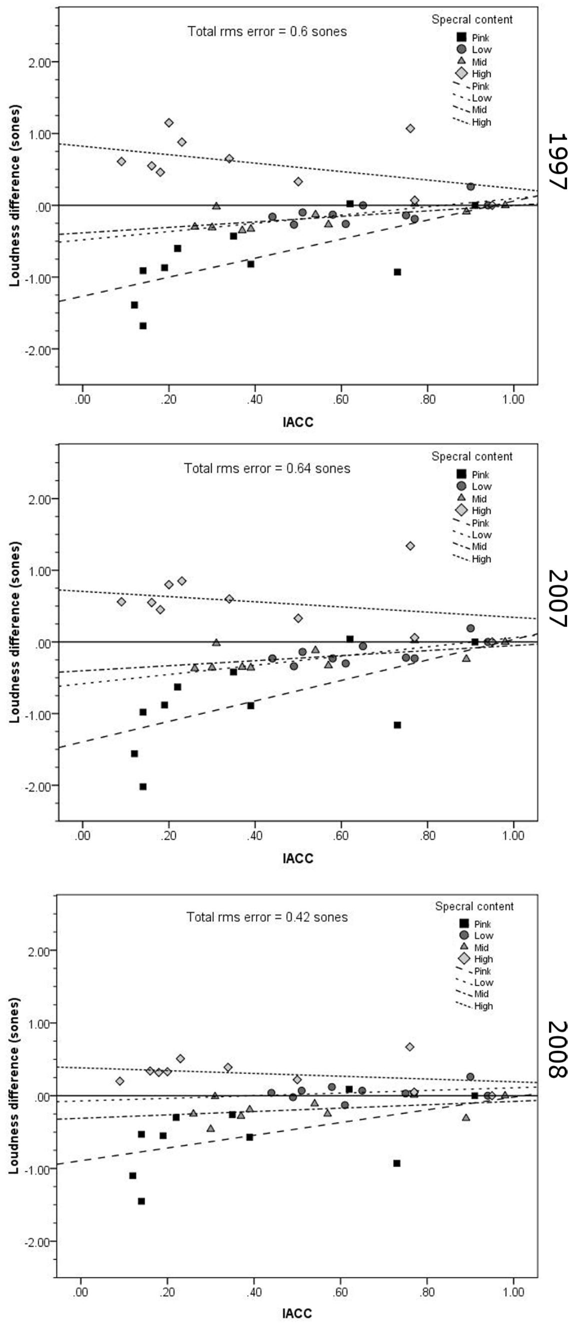

Figure 6. Difference in sones from the loudness of the reference stimuli and the adjusted stimuli according to the listening test results, plotted against IACC (A-weighted). The results are separated by the spectral content of the stimuli. A positive value indicates an overestimation of loudness from the model tested. Linear fit lines for each of the different spectral content stimuli and are also displayed. The models used are the Moore et. al. (1997) model using binaural input signals, the Moore and Glasberg (2007) model and the Sivonen and Ellermeier (2008) model.

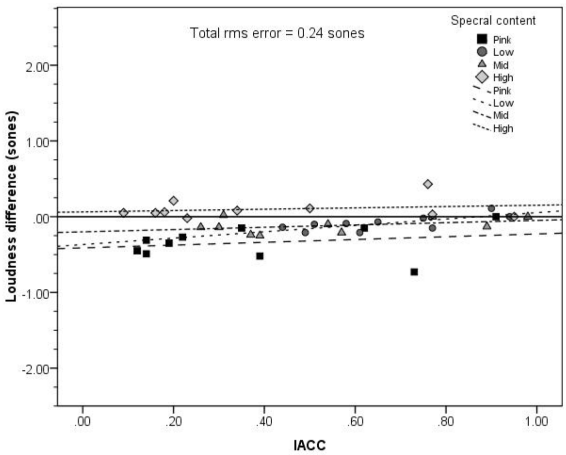

Figure 7. Difference in sones from the loudness of the reference stimuli and the adjusted stimuli according to the listening test results, plotted against IACC. The results are separated by the spectral content of the stimuli. A positive value indicates an overestimation of loudness from the model tested. Linear fit lines for each of the different spectral content stimuli are also displayed. The model used is the Moore et. al. 1997 model (Moore, 1997).

REFERENCES

Ando, Y. 1998, *Architectural Acoustics*, Springer-Verlag, New York, USA.

ANSI S3.4-2007 American National Standard - Procedure for the computation of loudness of steady sounds.

Cabrera, D., Ferguson, S., Rizwi, F. and Schubert, E. 2008a, “PsySound3: a program for the analysis of sound recordings,” *Acoustics 2008*, Paris, France.

Cabrera, D., Miranda, L. and Dash, I. 2008b, “Directional loudness measurements for a multichannel system,” *Acoustics 2008*, Paris, France.

H. Fletcher, W. A. Munson, 1933, “Loudness, Its definition, measurement and calculation”, *J. Acoust. Soc. Am.,* Vol. 5, 82-108.

ISO 532:1975 Acoustics -- Method for calculating loudness level.

Moore, B.C.J., Glasberg, B.R., and Baer, T. 1997, “A model for the prediction of thresholds, loudness, and partial loudness,” *J. Audio. Eng. Soc.* 45(4), 224-240.

Moore, B.C.J. and Glasberg, B.R. 2007, “Modeling binaural loudness”, *J. Acoust. Soc. Am.* 121(3), 1604-1612.

Sivonen, V.P. and Ellermeier, W. 2006, “Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation,” *J. Acoust. Soc. Am.,* 119(5) 2965-2980.

Sivonen, V.P. 2007 “Directional loudness and binaural summation for wideband and reverberant sounds”, *J. Acoust. Soc. Am.*, 121(5) 2852-2861.

Sivonen, V.P. and Ellermeier, W. 2008 “Binaural loudness for artificial-head measurements in directional sound fields,” *J. Audio Eng. Soc.* 56(6) 452-461.

Stevens, S.S. 1955, “The measurement of loudness,” *J. Acoust. Soc. Am.* 27(5) 815-829.

S. S. Stevens, 1956a, “The direct estimation of sensory magnitudes: loudness”, *The American Journal of Psychology*, 69(1) 1-25.

S.S. Stevens, 1956b, “Calculation of the loudness of complex noise”, *J. Acoust. Soc. Am.,* 28(5) 807-832.

Zwicker, E. and Scharf, B. 1965, “A model of loudness summation,” *Psych.l Rev.* 72(1) 3-26.