Audio Engineering Society

Convention Paper 5446

Presented at the 111th Convention
2001 September 21–24 New York, NY, USA

Binaural Modeling of Multiple Sound Source Perception: Methodology and Coloration Experiments

Kazuho Ono[1], Ville Pulkki[2], and Matti Karjalainen[2]

1) NHK Science and Technical Research Laboratories,
1-10-11 Kinuta Setagaya-ku, Tokyo 157-8510, JAPAN
ono@strl.nhk.or.jp

2) Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing,
P. O. Box 3000, FIN-02150 HUT, FINLAND
Ville.Pulkki@hut.fi and Matti.Karjalainen@hut.fi

ABSTRACT

Binaural modeling of coloration perceived due to multiple coherent sources is studied under the condition that sounds arrive at a listener successively within a certain time delay. The model simulates the perception of coloration of two horizontally located sources under prominent precedence effect conditions. Listening experiments are conducted to ensure the validity of the modeling. A new methodology is adopted in the experiments to minimize the error owing to individuality of HRTFs and inaccuracy of HRTF measurements.

INTRODUCTION

This work has been carried out as an extension of a project in which the perceptual qualities of amplitude-panned virtual sources in different setups were studied [1][2]. Most of this previous work has been conducted on problems with localization and coloration of virtual sources, and it has already produced extensive knowledge on how directions of virtual sources are perceived in different setups.

The work so far has limited the situation to a simple case of simultaneously arriving sounds in an anechoic environment. In such case, the arrival time of sound from all the sound sources is within a period of 1 ms. In a real room, however, the time delays of reflections from walls usually exceed 1 ms and then the precedence effect should be taken into account to evaluate the perceptual quality.

There exist many experimental studies on the precedence effect [3]. The perceived direction is reported to depend generally on the time delays of primary and secondary sound. It is also reported to be related to human adaptation for the property of sound field [4]. However, guidelines for perfect prediction of the precedence effect for unknown sound signals still remain to be studied.

On the other hand, the precedence effect is often used practically in public address (PA) systems to reinforce the primary sound source by a secondary sound source located in a different direction, without giving the audience a disturbing sensation as sound is coming from the secondary source, by introducing a suitable time delay for the secondary sound source.

While the precedence effect deals with directional perception, the present work deals with multiple source perception from a viewpoint of coloration. The work proposes and validates a binaural model of coloration for two coherent sources under the condition that the sounds from these sources arrive at a listener successively within a certain time delay, particularly when sound sources are located in the horizontal plane. When the delay is within a few milliseconds, the precedence effect is expected to be effective, but the model assumes coloration to be a separate phenomenon, related to binaural loudness as a function of critical band.

Listening experiments are conducted in this study to ensure the validity of the modeling approach. These tests are conducted in an anechoic room with two loudspeakers reproducing the same sound signal with a certain time delay from one loudspeaker to another. The subjective assessment is done using the method of adjustment. During the experiment, sound signals close to the entrance of ear canals are recorded with miniature electret microphones for each subject. These signals are used for the input of the binaural model, instead of using sound signals generated by convolution of HRTFs and sounds from loudspeakers. In this methodology, the input of the binaural model is perfectly the same as in the listening condition, minimizing the error owing to individuality of HRTFs and that caused by HRTF measurements, which is often problematic in experiments using HRTFs.

Listening test results support the validity of the binaural model in general, since they were found out to be modeled quite accurately, at least for the case of two coherent sources in different directions and relative delays within 4 ms.

1. PERCEPTUAL ATTRIBUTES

1.1 Timbre

Timbre is defined by American Standards Association [5] as “that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar.” Timbre is characteristic to different instruments, the piano has a different timbre when compared with the harpsichord or the chain saw. It is assumed that timbre is primarily composed by the magnitude spectrum of sound and how it changes with time [6]. In this study the timbre of multiple sources is of interest. Coloration in sound reproduction is understood as the change of timbre, uncolored sound being the goal of perfect reproduction.

Plomp [7] has shown that perceptual differences between different sounds are closely related to the differences in the spectra of the sounds. He defined the spectrum as the levels in 18 1/3-octave frequency bands. A 1/3 octave bandwidth is close to the bandwidth of a critical band. This suggests that the loudness level at each critical band forms the spectrum that is used in timbre perception. Loudness is the subjective attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud [6].

Binaural effects of timbre detection have also been studied. It has been found that the perceived timbre of a sound source is dependent on the sum of decoded loudnesses of ear canal signals [8]. There might be, however, some deviations of this depending on the interaural level difference of a sound source [9].

This paper deals with multiple source perception under precedence effect conditions, which means the delay values between each source composing a multiple source are static. It does therefore not produce any time-varying modifications to sound objects. The scope of the study can therefore be narrowed only to concern static coloration of perceived timbre.

1.2 Directional attributes

Spatial and directional hearing have been studied intensively; for overviews, see for example [3] or [10]. The duplex theory of sound localization states that the two main cues of sound source localization are the *interaural time difference* and the *interaural level difference* which are caused respectively by the wave propagation time difference (primarily below 1.5 kHz) and the shadowing effect by the head (primarily above 1.5 kHz). The auditory system decodes the cues in a frequency-dependent manner. The cues resolve in which cone of confusion the sound source lies. A cone of confusion can be approximated by a cone having an axis of symmetry along a line passing through the listener's ears.

The precedence effect, however, should be taken into account in directional perception in a real room, in which the time delay of a reflection from a wall usually exceeds 1 ms, even if the room is not very reverberant as a concert hall. Blauert describes in his book [3] about the precedence effect that “When signals from two or multiple coherent sources, for example, a direct sound and successive reflections, reach a listener from different directions, the auditory event will nevertheless, often appear in a single direction only”. In this study, binaural perception of narrow-band loudness is studied under precedence effect conditions.

2. BINAURAL AUDITORY MODEL

In this study the perceived timbre is modeled as a binaural loudness level spectrum over frequency channels. A schematic diagram for the binaural model of neural decoding is presented in Fig. 1. It takes as input the sound signal arriving to the ear canals and computes the decoded frequency-dependent loudness level spectra. It models the middle ear, the cochlea, and the auditory nerve. The model is explained in more detail in [11] and [1].

The middle ear, cochlea, and auditory nerve models have been implemented based on the HUTear 2.0 software package [12]. The middle ear is modeled using a filter that approximates a response function derived from the minimum audible pressure curve [13]. The cochlear filtering of inner ear is modeled using a 42-band gammatone filter bank [14]. Center frequencies of the filter bank follow the ERB (equivalent rectangular bandwidth) scale [15]. The outputs from the auditory models are interpreted and compared with the listening test results.

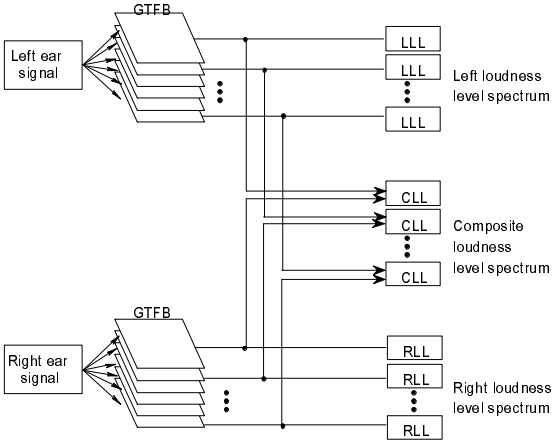


Fig. 1: Modeling of right-ear, left-ear, and composite loudness level spectra (RLL, LLL and CLL spectra). LL denotes the loudness level.

The loudness of each frequency band in each ear is computed using Zwicker's formulae [16]. Due to simplicity, this model is used instead of the more thorough model proposed in [6]. From the left-ear loudness level spectrum (LLL) and right-ear loudness level spectrum (RLL), composite loudness level spectrum (CLL) is computed by summing the loudness of each ear at each frequency band.

3. SIMULATION USING DUMMY HEAD HRTF'S

Coloration of a virtual source composed of two real sound sources is first simulated in anechoic condition with HRTFs and the auditory model. The HRTF data used for this simulation was that of KEMAR [17]. The model calculates the composite loudness level (CLL) spectra of the virtual source and a real reference source for 42 ERB channels as binaural loudness spectra. Coloration of a virtual source should be calculated by subtracting the single reference source's CLL spectrum from the virtual source's CLL spectrum, but to make it easy to compare with the listening test results described later, subtraction was done vice versa here to have different sign, which means the calculated values indicate how much gains are needed for virtual source to be equally loud to single reference source.

Calculation was conducted under certain delay conditions between the two sound sources. The SPL from each loudspeaker is the same at the listening position (without dummy head) for single reference source and virtual source. The source locations are limited to horizontal plane, as shown in Fig. 2. Fig. 3 shows the simulated results for a stereophonic setup (θ_1=-30°, θ_2=+30°), while Fig. 4 shows the results for a front-side setup (θ_1=0°, θ_2=+90°). In the front-side setup, delay was for the side loudspeaker (located at 90°).

In the stereophonic setup, large peak is around 1.7 kHz for 0 ms delay, which corresponds to the first comb filter notch in virtual source. The peak is found in lower frequencies for 2 ms and 4 ms delay, which corresponds to the shift of comb filter notch due to delays. In the front-side setup, similar peaks due to comb filter notch exist at similar frequencies except for a peak at 3 kHz for 0 ms delay.

4. LISTENING TESTS

Listening tests were carried out to ensure the validity of the modeling. All tests were conducted in an anechoic room with two loudspeakers reproducing the same sound signal with a certain time delay from one loudspeaker to another, which is the same setup as in the simulations described above.

A new methodology based on simultaneous recording of the sound signals at the entrances of the ear canals was adopted. This methodology enables perfect matching between the listening test results and the corresponding outputs of the auditory model, in which the individuality of HRTFs and the movement of listening positions are fully taken into account. The details are discussed below.

4.1.1 Coloration evaluation

In the test, the subjects adjusted the loudness of a narrow-band virtual source created by two real sources by level control to match as well as possible with a single real reference source. The reference single source was presented first, followed by a presentation of the virtual source, as shown in Fig. 6. The sound signal used for reference and that composing the virtual source are identical except for delay between loudspeakers. The sound signal was bandlimited noise for Listening test 1 and 2, and bandlimited pulse train for Listening test 3. The number of the bandlimited noise (or pulse train) bands is 19 whose lower and upper cutoff

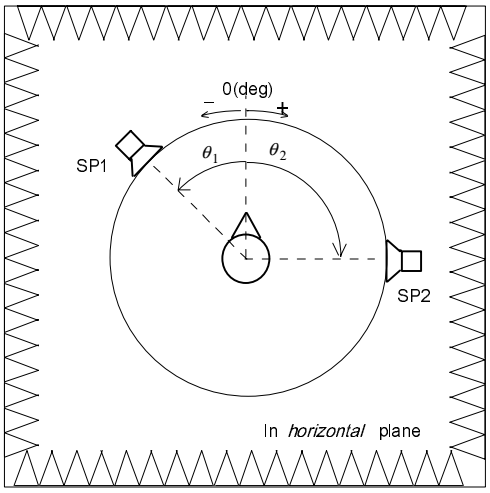


Fig. 2: Loudspeaker setup for simulation

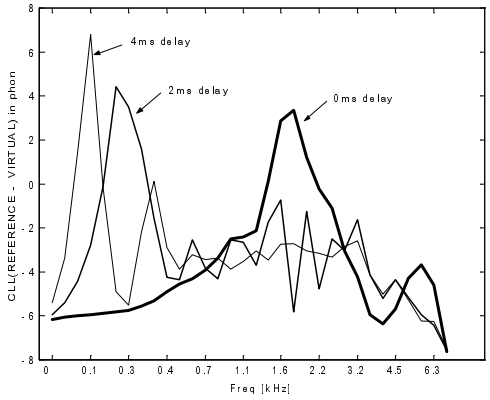


Fig. 3: Simulated coloration using KEMAR HRTFs for a stereophonic setup, for arrival time differences of 0, 2, and 4 ms.

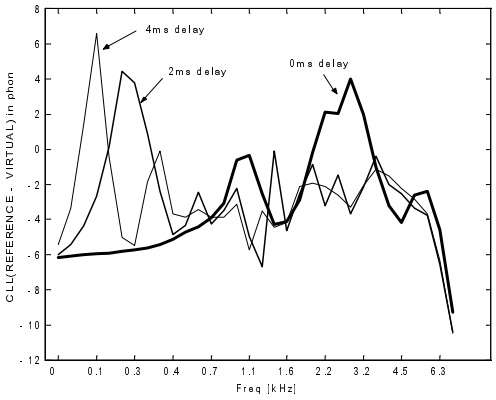


Fig. 4: Simulated coloration using KEMAR HRTFs for a front-side setup, for arrival time differences of 0, 2, and 4 ms.

frequencies are two adjacent Bark frequencies respectively, so that the frequency band for each signal starts from between 1 and 2 Bark, ending to between 19 and 20 Bark.

The sequence was repeated until the listener pressed 'enter' key to denote that he has selected the best matching level. The listener pressed up/down keys or left/right keys to change the loudness of a virtual source for test. The virtual source signal was controlled with a gain coefficient whose value was changed by 3 dB or 0.5 dB each time up/down keys or left/right keys were pressed, respectively.

The loudspeakers were in front of the listener at azimuth directions θ_1, θ_2 in the *horizontal* plane, and the sound pressure level from both loudspeakers generating the virtual source were equal. The listener was requested to keep his/her head immobile.

As a result a set of gain adjustment coefficients as a function of frequency channel were obtained. If a virtual source was not colored, the listeners' judgments would be constant. Colorations of virtual sources are shown as higher or lower adjustment values depending on frequency channel, corresponding to dips or peaks in virtual source coloration spectra.

4.1.2 Problems using HRTFs

The dummy head is a convenient tool and is used quite often in evaluating sound or in binaural recording. KEMAR, which was used for the simulation above to get coloration for virtual source, is one of the most popular dummy head, since it is regarded to have a shape corresponding to the average of human's head. But from the perceptual point of view, deviations from the average are often problematic especially in the perception of sound localization.

To avoid such problems, the ear canal signals are typically simulated using individually measured head-related transfer functions (HRTFs) and convolving them with an anechoic sound source. Although this is a much more exact method from the viewpoint that individuality of HRTFs are completely taken into account, the problems with listening condition still remain, because HRTFs is usually measured in an anechoic chamber while our listening situations are usually not anechoic. It has also the problem due to movements of human head in listening situations, especially when multiple coherent sources make a complicated sound field within a small area due to phase interference. A small movement of head can make large deviation of sound pressure at the entrance of the ear canal.

4.1.3 Simultaneous recording

We propose a new methodology for evaluating the binaural modeling of multiple source perception. It is based on simultaneous recording during the listening test, which makes it possible to analyze the ear canal signals that appeared in listening test, and to be free from the problem of individuality of HRTFs or that of listening position. The outline is shown in Fig. 5.

Two small electret microphones are set close to the entrance of the ear canals of the subjects (Fig. 7) to record the sound simultaneously with coloration evaluation. The microphones were close to the tragus in about 5 mm distance from the ear canal entrance so that they minimally changed the perceived sound, yet corresponded well to the ear canal signal [18]. For recording, the pair of reference and virtual sources were presented two extra times just after the subject pushed the 'enter' key to finish adjustment. The latter of the two extra sets was actually recorded for both the reference and the virtual sources. During the 'recording session', any key pushes were rejected to keep sound pressure the same as when the subjects decided the adjustment. By recording the sound sufficiently long after the adjustment it was

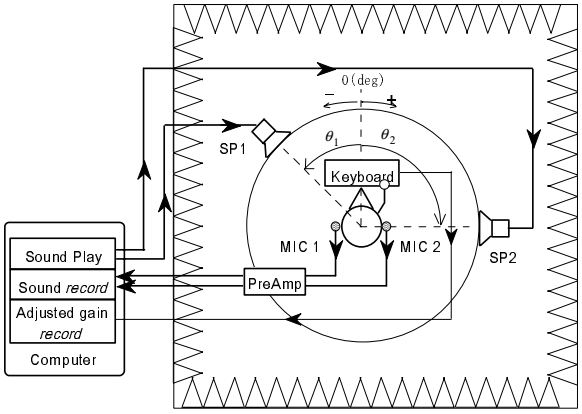


Fig. 5: Experimental setup for listening test

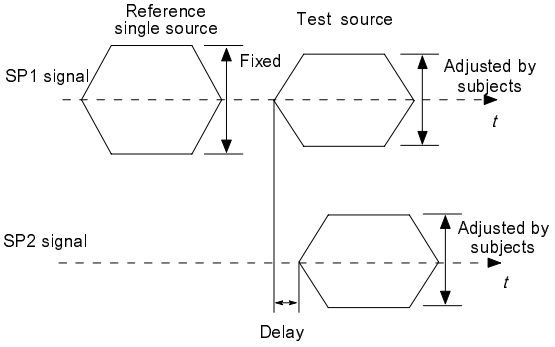


Fig. 6: Sound presentation

Fig. 7: Microphone(s) positioned close to the entrance of the ear canal

possible to avoid recording the sounds of keyboard clicks. The clicks could disturb the later analysis.

In this methodology, any sounds in an anechoic chamber during the recording session could also be recorded, so the subjects were required not to make any sounds or any movement, including swallowing during the recording period.

All the recorded sounds were checked if they had noise or disturbance by listening them carefully, and the sounds which were found to have extra noise were removed. The recorded sound was windowed by a hanning window for each reference source and virtual source sound respectively to pick out the corresponding section for both the reference source and the virtual source. The windowed sections were used as the inputs of the auditory model.

In the auditory model, all outputs from the auditory filter bank should in principle be used to compute total loudness, but in this study, in order to get a higher signal to noise ratio, only the outputs of 4 auditory filter banks whose center frequencies were closest to the center frequency of the input signals were used to calculate loudness values. Among the 4 center frequencies of the filter banks, 2 of them were just above the input signal freqency, and another 2, just below. By selecting specific outputs from all the filter banks, bandpass filtering of the input signals was quite easily achieved without design and change of bandpass filters corresponding to each input signals. Although the absolute values of total loudness calculated by this 'equivalent' filtering is not that of the input signal from microphones, the values are expected to be sufficiently useful as long as relative values, for example reference minus test, are concerned.

4.2 Listening test 1 (preliminary test)

A preliminary listening test was conducted for conventional stereophonic setup with 25 subjects to check the validity of the methodology and to find general tendencies of multiple sources perception. The number of loudspeakers used was two, which located at –30° and +30° respectively. The test was conducted in an anechoic chamber with two GENELEC 2029A loudspeakers located at 2 meters distance from the listener. The SPL of the single reference source was 65 dB, measured at the listening position. The test attendees were young males and females with normal hearing. They were all students at HUT, and in most cases with no prior experience in participating in a listening test panel.

The sound signals were 19 bandlimited noises whose center frequencies correspond to each Bark channel as described in 4.1.1. The reference source was presented from the left (-30°) loudspeaker and the virtual test source was presented from both loudspeakers with delay for right (+30°) loudspeaker. The delay values for right loudspeaker are 0 ms (no delay) and 4 ms. The perceived directions of virtual sources were expected to be different from the reference source, but to concentrate on loudness evaluation, subjects were not required to answer the perceived direction of the sounds but just to comment about the perceived direction voluntarily.

A boxplot and the mean values of results for adjusted gain are shown in Fig. 8, and in Fig. 9. Fig. 8 shows the results for the case with 0 ms delay, which corresponds to the coloration for conventional stereophonic setup. There exists a hump in the adjustments near 1.7 kHz where the data range is also widened, and the mean values decline slightly at higher frequencies. These results for stereophonic setup are similar to those shown in previous results, where coloration of an amplitude-panned virtual source was studied [19].

Fig. 9 shows the results for the case with 4 ms delay, where the precedence effect is expected to exist. In this case, the median value shows higher value only at 0.1 kHz, while in all other frequencies the value is quite stable in the range of –2 - –4 dB.

Figures 10 and 11 show the boxplot and the mean values of the difference (reference – test) of composite loudness level (CLL) calculated by the auditory model using the recorded sound as its input. In these figures, the mean values of the adjusted gain, shown already in Fig. 8 or Fig. 9, are also plotted again.

This difference of the loudnesses in the output of the auditory model describes the validity of the binaural auditory modeling. If the value shows 0 phon, it means that the model perfectly describes the binaural loudness, while a positive/negative value shows that the modeled composite loudness of the reference single source is more/less loud than the test source, although the subjects perceived them by adjustment to be equally loud.

Fig. 10 depicts the results for 0 ms delay. It shows that the model-based composite loudness level difference is within ±2 phon from ideal for almost all frequency channels and has quite stable value of ±1 phon for low and mid frequencies. We can also see a small hump in the frequency range of 1.7 kHz. A possible reason for this hump could be related to the spatial instability of sound field around dips due to comb filtering, but it may be for a further study to test this.

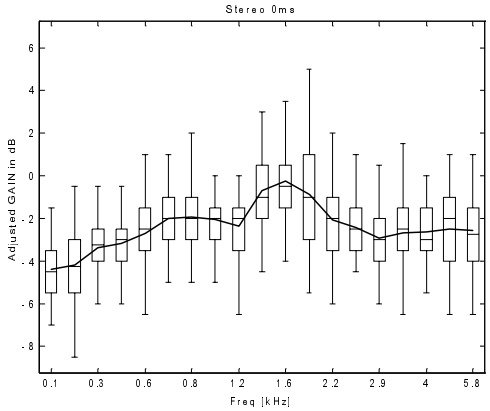


Fig. 8: Adjusted gain of virtual source to match the loudness of reference source. Stereophonic setup, no delay between channels.

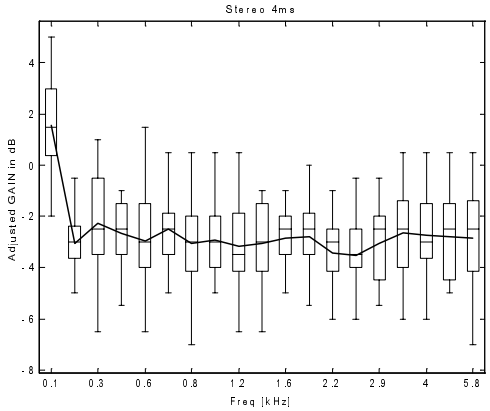


Fig. 9: Adjusted gain of virtual source to match the loudness of reference source. Stereophonic setup, 4 ms delay between channels.

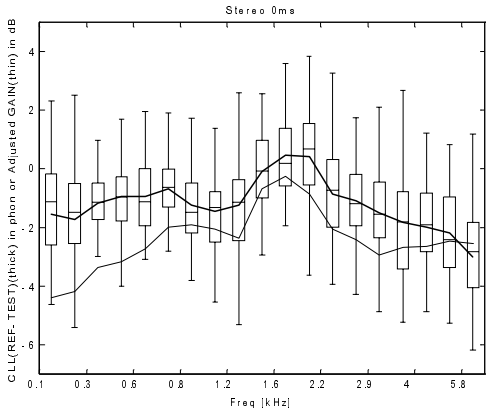


Fig. 10: Composite loudness level (CLL) difference calculated using binaural auditory model. Plotted with adjusted gain as in Fig. 8. Stereophonic setup, no delay between channels.

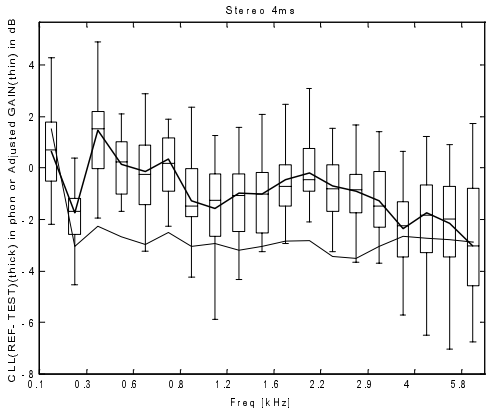


Fig. 11: Composite loudness level (CLL) difference calculated using binaural auditory model. Plotted with adjusted gain as in Fig. 9. Stereophonic setup, 4 ms delay between channels.

Fig. 11 depicts the results for 4 ms delay. It shows that the difference of the composite loudness level is ±2 phon for all frequencies and exhibits quite stable value of ±1 phon for low and mid frequencies. These results suggest the general validity of using CLL to represent binaural loudness for the stereophonic setup, especially for low and mid frequencies.

4.3 Listening test 2

Further experiments for more conditions and parameters were conducted by a smaller number of listeners. In the next experiment, two loudspeaker setups which are the stereophonic setup (θ_1=–30°, θ_2=+30°) and a front-side setup (θ_1=0°, θ_2=+90°), and 3 delay values which are 0 ms, 2 ms, and 4 ms, were adopted. Delay was added to the right loudspeaker (located at +30°) for stereophonic setup and side loudspeaker (located at +90°) for front-side setup. The number of subjects was 3 and they repeated each condition twice. They were all well trained listeners.

Test signals were 19 bandlimited noises whose center frequencies correspond to Bark channels as described in 4.1.1 (same as the Listening test 1). The reference source was presented from the left loudspeaker (located at –30°) for stereophonic setup and from the front loudspeaker (located at 0°) for front-side setup.

The test was conducted in anechoic chamber with 2 GENELEC 2029A loudspeakers located at 2 meters distance from the subjects for stereophonic setup, and 1.6 meters for front-side setup. The SPL of the single source was 65 dB at 1kHz, measured at the listening position, and it was changed with frequency according to the equal loudness curve at 60 phon, since in Listening test 1 it was found that the variation in perceived loudness was too large when using constant SPL

The perceived directions of the virtual sources were expected to be different from the reference single source, but to concentrate on loudness evaluation, subjects were not required to answer the perceived direction of the sounds, but were asked to comment about the perceived direction voluntarily.

Figure 12 depicts the boxplot and the mean values of adjusted gain for the stereophonic setup, while Fig. 13 shows the results for front-side setup. All the data show similarity to the simulation results by KEMAR calculated by CLL, which implies that CLL is expected to be a valid measure for perceived loudness in this condition.

Figures 12-(a) and 12-(c) show the results for the same conditions as in Listening test 1 conducted by 25 subjects. These two results show similarity to each other in general. This indicates that the 3 subjects adopted in Experiment 2 show the same tendency as the average of 25 subjects adopted in Listening test 1, which suggests the validity of other results in Listening test 2.

Figures 14 and 15 illustrate the boxplot and the mean values of the difference of composite loudness level (CLL) calculated by auditory model using the recorded sound as its input. In these figures, the mean values of the adjusted gain, shown already in Fig. 12 and Fig. 13 are also plotted again.

In the stereophonic setup it is shown that the difference of the composite loudness level is ±2 phon for all frequencies and delays except when the frequency is around 0.3 kHz and the delay is 2 ms, and that it is stable within ±1 phon in almost all conditions. In the front-side setup it is shown that the difference of the composite loudness level is within ±2 phon for all frequencies and delays except for the highest frequency with no delay, and is stable between 0 phon and +2phon in almost all conditions. These results suggest that CLL is also valid for binaural loudness in front-side setup or in different delay values.

The variation of CLL difference is relatively large for the front-side setup, although the mean values for CLL difference are quite stable. The data for each subject (not shown in this paper) show that it is mainly due to differences between subjects, and the data were relatively consistent within each subject.

4.4 Listening test 3

Listening tests 1 and 2 were conducted for stationary noise-like sound, which is not usually used for precedence effect experiments. At the tests, subjects' informal (voluntarily) comments reported that they had seldom clear perception of sound localization for the virtual source except when there was no delay between two loudspeakers, while they had clear localization for the reference single source.

In Listening test 3, only the sound signal was changed to a bandlimited pulse train, which is expected to have more prominent precedence effect because of its transient properties. The pulse train signals were made from 10 Hz impulses filtered by 6th-order-Butterworth bandpass filters, whose center frequencies correspond to Bark bands as described in 4.1.1. The SPL of the single

reference source was 60 dB at 1kHz, measured at the listening position, and it was changed with frequency according to the equal loudness curve at 60 phon. Other conditions were the same as in Listening test 2. In this test, two of the three subjects were the same as in Listening test 2, and one subject was different, but all of them were well trained listeners.

In these tests, subjects' informal (voluntarily) comments reported that all listeners felt that most of the virtual sources were located at a similar direction to the single reference source, when there was delay between two loudspeakers. This shows that the precedence effect was prominent in this condition.

Figure 16 depicts the boxplot and the mean values of adjusted gain for the stereophonic setup, while Fig. 17 is for the front-side setup. All data show similarity to the simulation results by KEMAR, calculated by CLL, and also to the results in Listening test 2, in general. This implies that CLL is expected to be valid for perceived loudness in this condition, also for pulse trains.

Figures 18 and 19 show the boxplot and the mean values of the difference of composite loudness level (CLL) calculated by the auditory model using the recorded sound as its input. The mean values of the adjusted gain, showed already in Fig. 15 and Fig. 16, are also plotted again.

The results show that the difference of the composite loudness level is ±2 phon for all frequencies and delays, except when the frequency is around 0.3 kHz and the delay is 2 ms for both stereophonic and front-side setup, and that it is stable within ±1 phon in almost all conditions. These results suggest that CLL is also valid for binaural loudness for transient sounds.

The large hump around 0.3 kHz for both stereophonic and front-side setup shows that CLL does not represent the binaural loudness at this frequency range. At this frequency range, variation of CLL difference is also large in Listening test 2 although the mean value is around 0 phon. Further studies may be necessary for more precise modeling for specific frequencies in which CLL difference shows large variation, including around 0.3 kHz.

5. Discussion

In this study the model used to simulate the timbre of sound sources did not include any kind of precedence effect. However, the results for stationary bandpass noise and bandpass pulse trains show that computed CLL difference between virtual source and reference single source had generally stable values around 0 phon when subjects judged the reference and test signals equally loud. This suggests that CLL can be used to represent binaural loudness even for multiple sources without including any precedence effect. The similarity of results between Listening test 2 and Listening test 3 shows that binaural loudness perception can be modeled independently of directional perception for a virtual source created by two sources which have less than 4 ms delay difference.

The subjects of our experiments reported several informal findings from listening to the test signals. The direction of the virtual source varied widely, as expected, especially for the stationary noise case. For the pulse train it was more stable due to stronger precedence effect. An interesting finding was that the perceived distance of the virtual source was often different from the reference real source. Also, the timbre varied noticeably although a critical band width bandpass noise was used. One possible explanation could be that this is based on off-band listening due to spectral spreading of auditory excitation. The most surprising percept was, however, that in few cases the real source and the virtual source for bandpass noise were found to have different pitches. This multitude of attributes to simple bandpass noises shows that auditory perception, even in this simplified case, is a complicated process where different perceptual factors such as loudness, timbre, direction, and pitch may be present in various combinations.

Further studies are necessary to generalize the validity of this modeling and to make it more precise. The studies will include multiple sources consisting of more than two real sources, sound sources in the median or lateral plane, various level conditions and incoherence between sound sources. Step by step this methodology should be extended to cases of real-world complexity. This modeling approach, in order to test independency between loudness perception and directional perception, will be one of the important keys to decompose the perception of spatial sound, which is usually a multi-dimensional problem.

The relationship of this work to the study for binaural loudness by coherent signals presented using heaphones [8] has also to be mentioned. The study showed that the binaural loudness was found to be the summation of the loudnesses of each ear, when the interaural level difference is less than 10dB. Our preliminary analysis showed that in the frequency range adopted in our study, the interaural differences are below 10dB in general except for very high frequencies above 4kHz, which shows good accordance with the study by headphones. From this point of view, the present work can be situated as an extension of the work by headphone to a multiple source problem with a few milliseconds delay in which signals for both ears are partially incoherent. On the other hand, for high frequencies, loudness level for each ear is strongly affected by the shadowing effect of the human head, which means binaural difference of loudness gets larger, typically above 10 dB. In this case, binaural difference will also have to be taken into account to represent binaural loudness.

The new methodology adopted in this study, based on simultaneous recording and adjustment, was found useful for our study, because it can record for model-based analysis an identical state to the listening test condition. It works especially well when the sound field by multiple source makes listening-position-dependent deviation of sound pressure level due to phase interference, which was sometimes perceived also in our listening tests. By recording simultaneously it is possible to compare the data precisely with the gain adjustment by subjects. It also makes it possible to judge whether a large deviation within a subject's data is from his/her inconsistency or from the movement of listening position.

We found this methodology critical to noise generated by subjects. The subjects should be well informed about this noise problem and should be well trained to be quiet during the recording session when applying this methodology.

6. Conclusion

Binaural modeling of virtual source coloration was studied under precedence effect conditions. New methodology was adopted to obtain complete matching between the listening test conditions and binaural modeling. Our experiments showed that this methodology highly useful. It was found that CLL (composite loudness level) obtained by summing up the loudnesses of each ear is an accurate measure for binaural loudness and perceived coloration of virtual sources both for stationary and transient sounds.

References

[1] V. Pulkki and M. Karjalainen. Localization of Amplitude-Panned Virtual Sources I: Stereophonic panning, Accepted for publication in *JAES*.

[2] V. Pulkki. Localization of Amplitude-Panned Virtual Sources II: Two and three-dimensional panning, Accepted for publication in *JAES*.

[3] J. Blauert. *Spatial Hearing*, Revised edition, The MIT Press, 1997, Cambridge, MA, USA.

[4] R. K. Clifton and R. L. Freyman, The precedence effect: Beyond echo suppression, In R. H. Gilkey and T. R. Anderson, editors. *Binaural and Spatial Hearing in Real and Virtual Environments*, Lawrence Erlbaum Assoc., 1997.

[5] American Standards Association. *Acoustical Terminology SI*, American Standards Association, New York, 1960.

[6] B. C. J. Moore. A Model for the Prediction of Thresholds, Loudness, and Partial Loudness, *J. Audio Eng. Soc*., 45(4):224-240, 1997.

[7] R. Plomp. *Aspects of Tone Sensation*, Academic, London, 1976.

[8] U. T. Zwicker and E Zwicker. Effects of binaural loudness summation and their approximation in objective loudness summation, *Proc. Inter-noise 90*, 1990.

[9] P. M. Zurek. Measurements of binaural echo suppression, *J. Acoust. Soc. Am*., 1979, 66(6), 1750-1757, December, 1979.

[10] R. H. Gilkey and T. R. Anderson, editors. *Binaural and Spatial Hearing in Real and Virtual Environments*, Lawrence Erlbaum Assoc., 1997.

[11] V. Pulkki, M. Karjalainen and J. Huopaniemi. Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model, *J. Audio Eng. Soc*., 1999, 47(4):203-217, April 1999.

[12] A. Härmä and K. Palomäki. HUTear -- a free Matlab toolbox for modeling of auditory system, 96-99,*Proc. Matlab DSP Conference*, 1999,Comsol Ltd., Espoo, Finland, November, http://www.acoustics.hut.fi/software/HUTear/

[13] B. C. J. Moore. *An introduction to the psychology of hearing*, Academic Press, San Diego, fourth edition, 1997

[14] R. Patterson, K. Robinson, J. Holdsworth, D. Mckeown, C. Zhang and M. H. Allerhand. Complex sounds and auditory images. In L. Demany, Y. Cazals and K. Horner, editors, *Auditory Physiology and Perception*, pages 429-446. Pergamon, Oxford, 1992.

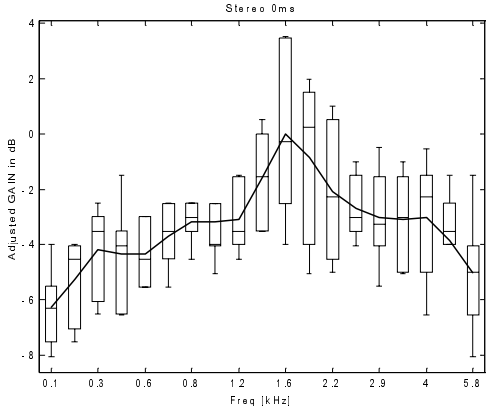
[15] B. C. J. Moore, R. W. Peters and B. R. Glasberg. Auditory filter shapes at low center frequencies. *J. Acoust. Soc. Am*., 88(1):132-140, July 1990.

[16] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer-Verlag, Heidelberg, Germany, 1990.

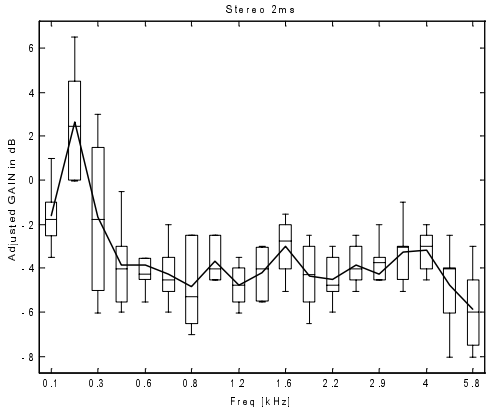
[17] B. Gardner and K. Martin, HRTF Measurements of a KEMAR Dummy-Head Microphone, *MIT Media Lab Perceptual Computing*,#280, MA, USA, 1994 (The HRTF data is retrievable from WWW URL: http://sound.media.mit.edu/KEMAR.html).

[18] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, Evaluation of Artificial Heads in Listening Tests, *J. Audio Eng. Soc*., 47(3):83-100, March 1999.

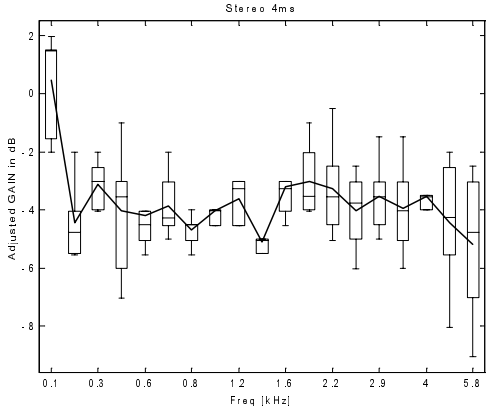
[19] V. Pulkki, Coloration of Amplitude-Panned virtual Sources, In *Proceedings of the 110th Convention of the Audio Engineering Society*, Amsterdam, May. 2001, Preprint 5402.



(a) Stereophonic setup, no delay.

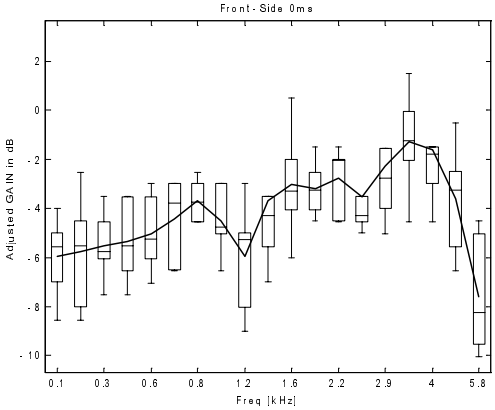


(b) Stereophonic setup, 2 ms delay.

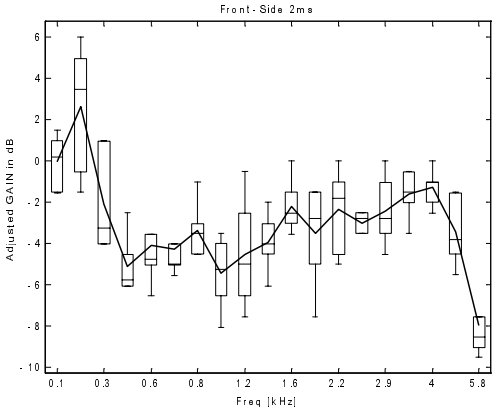


(c) Stereophonic setup, 4 ms delay.

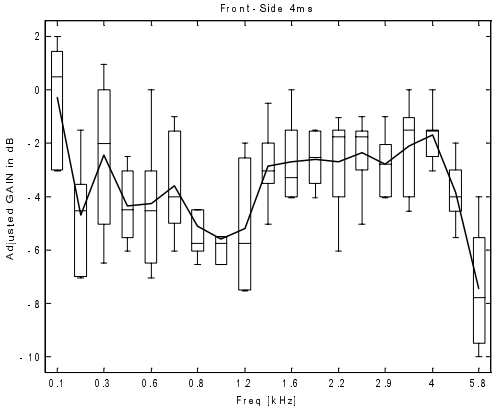
Fig. 12: Adjusted gain of virtual source to match the loudness of reference source by bandlimited noise. (Results of Listening test 2)



(a) Front-side setup, no delay.

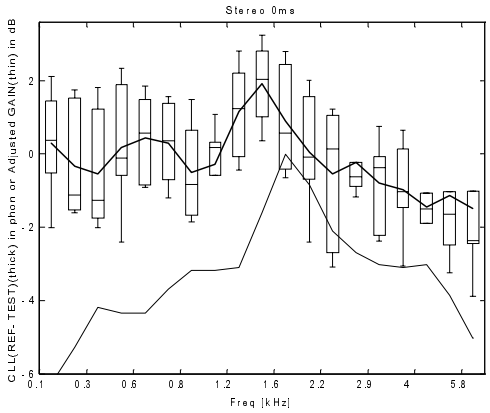


(b) Front-side setup, 2 ms delay.

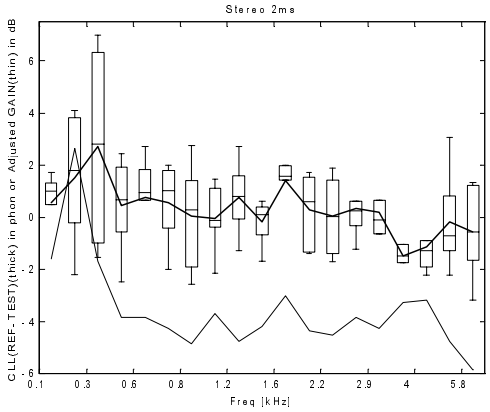


(c) Front-side setup, 4 ms delay.

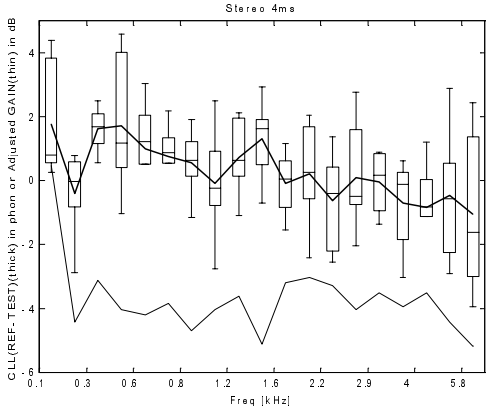
Fig. 13: Adjusted gain of virtual source to match the loudness of reference source by bandlimited noise. (Results of Listening test 2)



(a) Stereophonic setup, no delay.

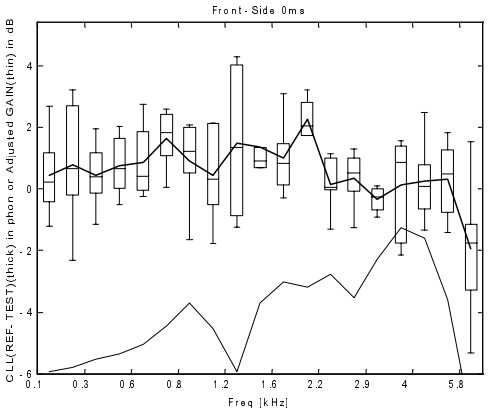


(b) Stereophonic setup, 2 ms delay.

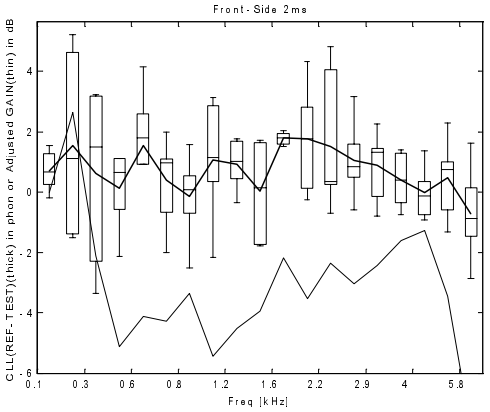


(c) Stereophonic setup, 4 ms delay.

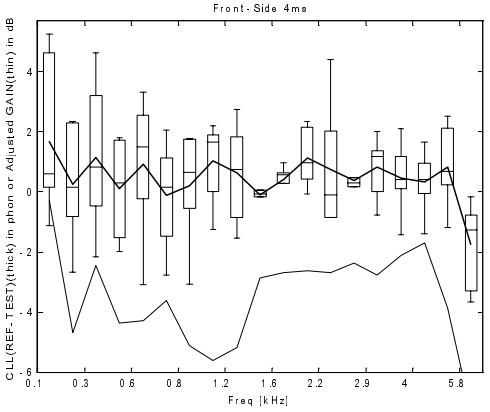
Fig. 14: Composite loudness level (CLL) difference by bandlimited noise calculated using binaural auditory model. Plotted with adjusted gain. (Results of Listening test 2)



(a) Front-side setup, no delay.

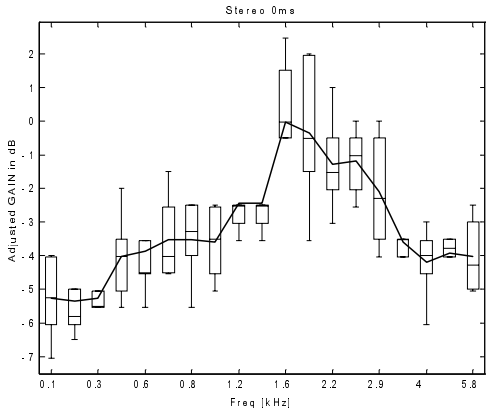


(b) Front-side setup, 2 ms delay.

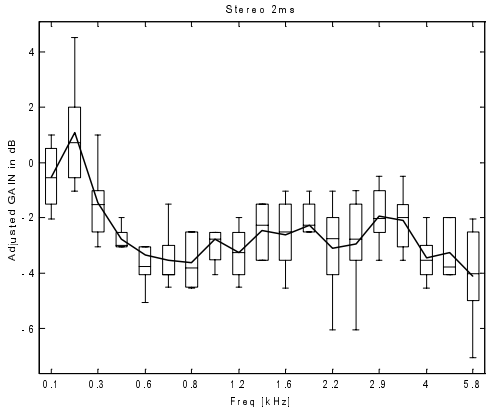


(c) Front-side setup, 4 ms delay.

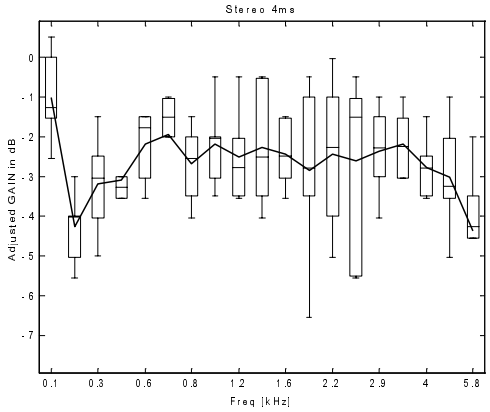
Fig. 15: Composite loudness level (CLL) difference by bandlimited noise calculated using binaural auditory model. Plotted with adjusted gain. (Results of Listening test 2)



(a) Stereophonic setup, no delay

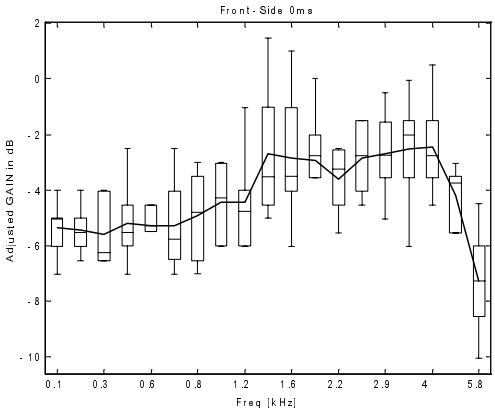


(b) Stereophonic setup, 2 ms delay

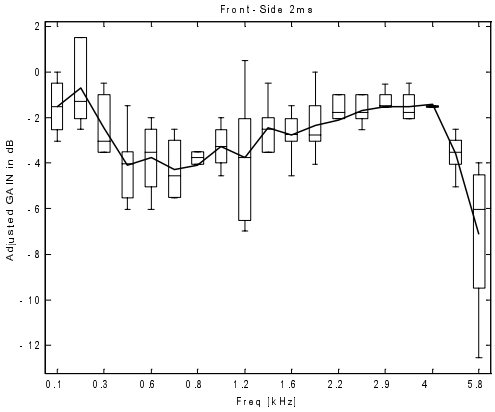


(c) Stereophonic setup, 4 ms delay

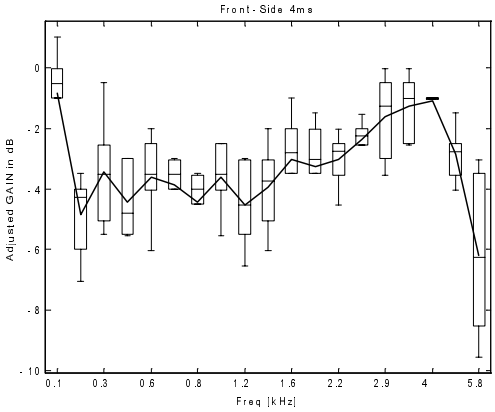
Fig. 16: Adjusted gain of virtual source to match the loudness of reference source by bandlimited noise. (Results of Listening test 3)



(a) Front-side setup, no delay.

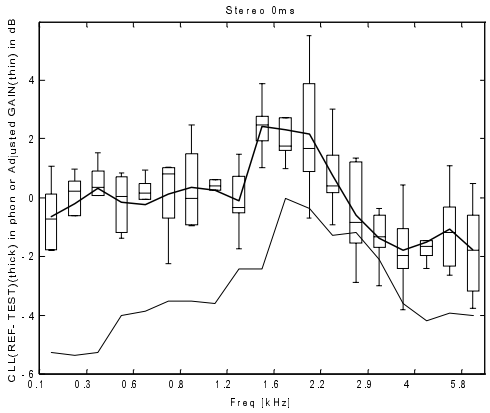


(b) Front-side setup, 2 ms delay.

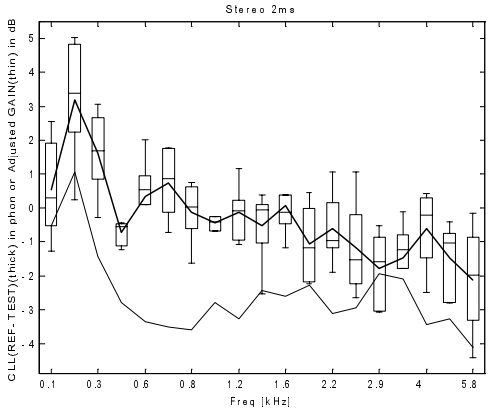


(c) Front-side setup, 4 ms delay.

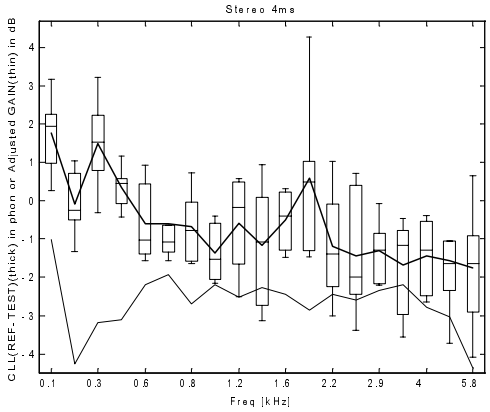
Fig. 17: Adjusted gain of virtual source to match the loudness of reference source by bandlimited noise. (Results of Listening test 3)



(a) Stereophonic setup, no delay.

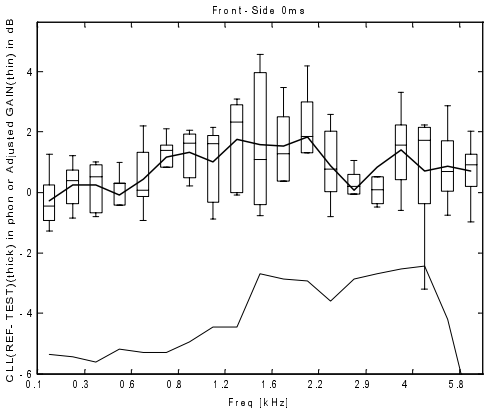


(b) Stereophonic setup, 2 ms delay.

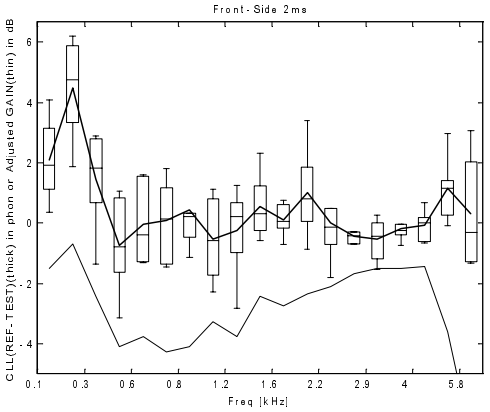


(c) Stereophonic setup, 4 ms delay.

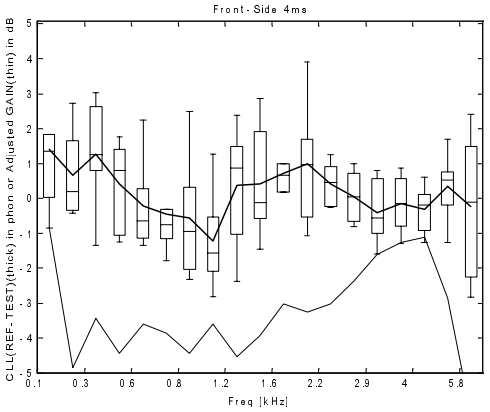
Fig. 18: Composite loudness level (CLL) difference by bandlimited pulse train calculated using binaural auditory model. Plotted with adjusted gain. (Results of Listening test 3)



(a) Front-side setup, no delay.



(b) Front-side setup, 2 ms delay.



(c) Front-side setup, 4 ms delay.

Fig. 19: Composite loudness level (CLL) difference by bandlimited pulse train calculated using binaural auditory model. Plotted with adjusted gain. (Results of Listening test 3)