FREQUENCY-DEPENDENT CHARACTERISTICS AND PERCEPTUAL VALIDATION OF THE INTERAURAL THRESHOLDED LEVEL DISTRIBUTION

Christian S. E. Cotton, Stephen G. Oxnard, Laurence J. Hobden and Ethan Stanhope

Meridian Audio Ltd.
11 Latham Road
Huntingdon, UK
PE29 6YE
chris.cotton@meridian.co.uk

ABSTRACT

The interaural thresholded level distribution (ITLD) is a novel metric of auditory source width (ASW), derived from the psychophysical processes and structures of the inner ear. While several of the ITLD's objective properties have been presented in previous work, its frequency-dependent characteristics and perceptual relationship with ASW have not been previously explored. This paper presents an investigation into these properties of the ITLD, which exhibits pronounced variation in band-limited behaviour as octave-band centre-frequency is increased. Additionally, a very strong correlation was found between [1 – ITLD] and normalised values of ASW, collected from a semantic differential listening test based on the Multiple Stimulus with Hidden Reference and Anchor (MUSHRA) framework. Perceptual relationships between various ITLD-derived quantities were also investigated, showing that the low-pass filter intrinsic to ITLD calculation strengthened the relationship between [1 – ITLD] and ASW. A subsequent test using transient stimuli, as well as investigations into other psychoacoustic properties of the metric such as its just-noticeable-difference, were outlined as subjects for future research, to gain a deeper understanding of the subjective properties of the ITLD.

1. INTRODUCTION

Auditory or apparent source width (ASW) is a subjective spatial listening quality which describes the perceived lateral extent of a sound source in an arbitrary sonic re/production system. During the design and validation of such a system, which may range from a live ensemble performance in a theatre, a pair of loudspeakers in a listening room, or a multichannel loudspeaker array in a vehicle cabin, the means to objectively measure and quantify ASW are strongly desired. A number of quantities used to measure or control ASW have been studied, such as correlation-based metrics like the interaural cross-correlation coefficient (IACC) and interchannel cross-correlation coefficient (ICCC), as well as energy-based metrics such as the early lateral energy fraction (LF_E) or the so-called "energy vector" [1–6]. In [7], a novel metric for the ASW of a stereophonic phantom image, the interaural thresholded level distribution (ITLD), was presented. ITLD is a quantity obtained by applying a binaural hearing model of the human auditory system, specifically the transduction mechanism of the inner hair cells within the cochlea, to a pair of binaural signals. A preliminary objective evaluation of ITLD as a metric of ASW showed not only that the metric had several advantages over IACC, but that it may also be a more perceptually relevant quantity. ITLD exhibits a stronger correlation with ICCC for a pink noise burst in anechoic and reverberant listening conditions, and displays greater sensitivity to room acoustics and source coherence compared to IACC, as well as delineating additional psychoacoustic cues.

It was outlined in [7] that further research into the psychoacoustic properties of the ITLD, as well as its behaviour in the frequency domain, was required for a sufficient assessment of the application of ITLD for evaluating ASW. This paper presents the findings of a subjective listening evaluation undertaken to ascertain a perceptual relationship between ITLD and ASW, and narrow-band characteristics of the metric in comparison to IACC. ASW data was collected inside a critical listening room from a semantic differential listening test using a set of decorrelated pink noise stimuli. ITLD values calculated from binaural measurements taken in the room using a dummy head and the same stimuli were then used to calculate R^2, r and p-values describing the relationships between [1 – ITLD] and ASW, as well as [1 – IACC] and a range of ITLD-derived quantities. A very strong correlation was found between wide-band [1 – ITLD] values, evaluated over the duration of the noise stimuli, and the averaged ASW data. It was also observed that the low-pass filter included in the ITLD formulation increased the correlation coefficient of all ITLD-derived quantities in relation to ASW. The frequency-dependent characteristics of the ITLD were measured in anechoic and reverberant conditions, the latter using the measurements from the subjective listening tests. Pronounced frequency-dependent behaviour was observed from these measurements, with a marked decrease in the ITLD of increasingly decorrelated stimuli between 125 Hz to 2 kHz in anechoic conditions, and between 500 Hz to 2 kHz in reverberant conditions. The low-pass filter stage in the ITLD calculation was shown to increase the range of values of wide-band and band-averaged data, which is reflected in the subjective listening tests by increased correlation coefficients.

2. MEASUREMENT OF AUDITORY SOURCE WIDTH

2.1. Cross-correlation Metrics

Cross-correlation metrics, such as the interaural cross-correlation coefficient (IACC), have been shown to be effective for evaluating ASW [1,4,5]. IACC, given by Eq. (1), is defined as the maximum absolute value of the interaural cross-correlation function (IACF), given by Eq. (2), evaluated over the time interval $[t_1, t_2]$ and ob-

tained from a range of time lags -1 ms $\leq \tau \leq$ 1 ms [8],

$$\mathrm{IACC} = \max_{-1\mathrm{ms}\leq\tau\leq1\mathrm{ms}} |\mathrm{IACF}(\tau)| \tag{1}$$

$$\mathrm{IACF}(\tau) = \frac{\int_{t_1}^{t_2} x_L(t)x_R(t+\tau)dt}{\sqrt{\left[\int_{t_1}^{t_2} x_L^2(t)dt\right]\left[\int_{t_1}^{t_2} x_R^2(t)dt\right]}} \tag{2}$$

where $x_{L,R}(t)$ represents a pair of binaural signals captured with a dummy head. The just-noticeable-difference (JND) of IACC is approximated to be 0.075 [8]. The early interaural cross-correlation coefficient ($\mathrm{IACC_E}$) is obtained by evaluating Eq. (1) and (2) over the interval [$t_1 = 0$, $t_2 = 80$ ms] and considers contributions from the direct sound and early reflections only. Similar to IACC is the interchannel cross-correlation coefficient (ICCC), which describes the cross-correlation of the input signals to an audio system $s_{L,R}(t)$. ICCC is obtained by taking the maximum absolute value of the interchannel cross-correlation function (ICCF), given by Eq. (3) and (4), respectively.

$$\mathrm{ICCC} = \max |\mathrm{ICCF}(\tau)| \tag{3}$$

$$\mathrm{ICCF}(\tau) = \frac{\int_{-\infty}^{+\infty} s_L(t)s_R(t+\tau)dt}{\sqrt{\left[\int_{-\infty}^{+\infty} s_L^2(t)dt\right]\left[\int_{-\infty}^{+\infty} s_R^2(t)dt\right]}} \tag{4}$$

Control over ASW was achieved in [4] using various algorithms and filter architectures to introduce decorrelation between the input signals, though this can also be achieved through the application of a relative phase shift [9]. In this sense, ICCC can be seen as a means to control ASW at the input stage, and IACC can be used to measure it at the ears of the listener [4].

2.2. Directional Energy Metrics

Further to the cross-correlation metrics described above are quantities which consider the contribution of directional energy when evaluating ASW, such as the early lateral energy fraction ($\mathrm{LF_E}$) and the so-called "energy vector" presented in [6] and evaluated in [5]. $\mathrm{LF_E}$ describes the ratio of lateral energy to omnidirectional energy at the ears of the listener, evaluated over a period of 80 ms as shown in Eq. (5),

$$\mathrm{LF_E} = \frac{\int_{0.005}^{0.080} p_L^2(t)}{\int_{0}^{0.080} p^2(t)} \tag{5}$$

where and p and the p_L are the instantaneous sound pressure of an impulse response captured from an omnidirectional and figure-of-eight microphone, respectively [8]. $\mathrm{LF_E}$ was shown to be a fair predictor of ASW in [5], while other works have shown the metric to be unsuitable [1, 2, 10], due to the lower integration limit in the numerator of Eq. (5) which omits the direct sound component from the $\mathrm{LF_E}$ calculation [10].

2.3. Interaural Thresholded Level Distribution

ITLD is a quantity obtained from a binaural signal pair using a mathematical hearing model of the inner ear section of the human auditory system (HAS) [7]. Specifically, low-pass filtering and half-wave rectification is applied to the binaural signal pair in order to model the transduction of acoustic pressure signals as cochlear nerve signals. The inner ear, comprised of the cochlea and vestibular system, is the terminal section of the HAS. The cochlea consists of three coiled, fluid-filled ducts separated by several membranes, the most significant for this work being the basilar membrane (BM). Upon the BM sits the organ of Corti which contains inner and outer hair cells (IHC and OHC, respectively). These are sensitive to displacement of the fluid in the cochlear ducts caused by vibrations from the ossicular chain of the middle ear reaching the cochlea, giving rise to the transmission of electrical signals propagated via the auditory nerve [11]. The action of the IHC has been accounted for in many binaural hearing models [12–15]. A simple approximation, referred to in this work as a "cochlear nerve transduction model" (CNTM) can be implemented using low-pass filtering and half-wave rectification to mimic the inability of the IHC to effectively resolve high frequencies and respond to rarefactive signal components [11, 16–18].

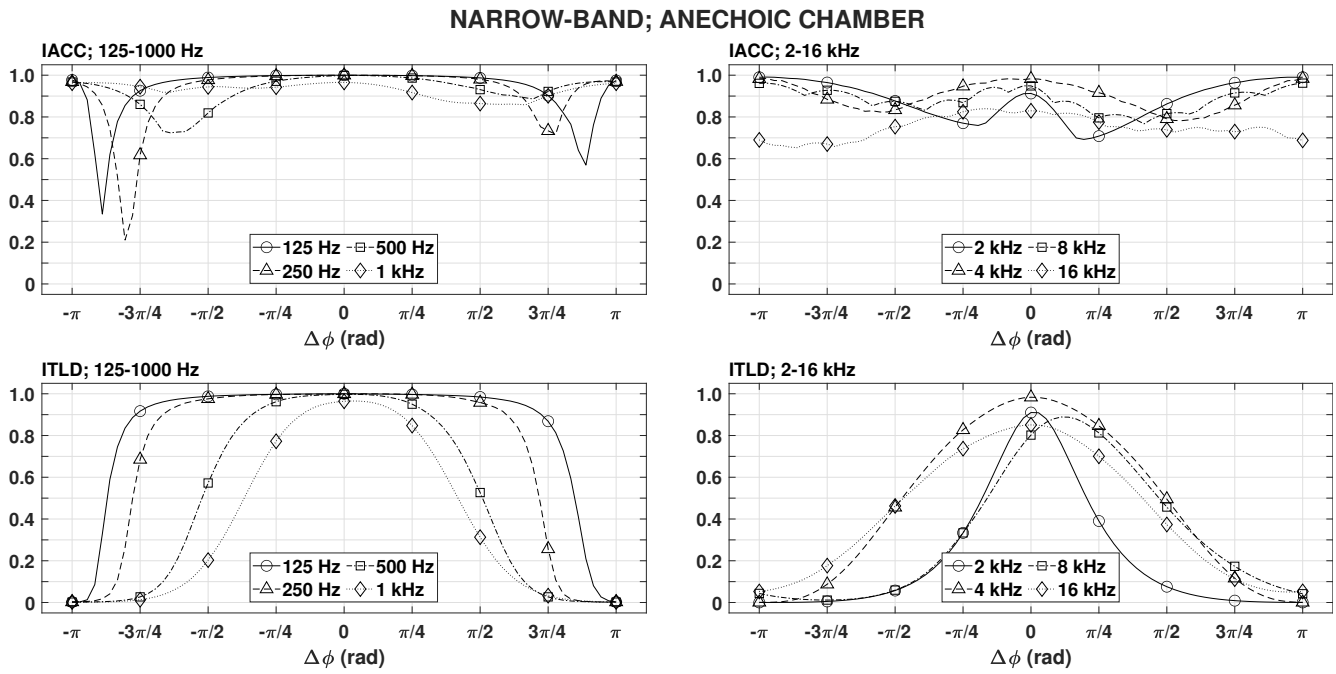
ITLD is obtained by applying these processes to a binaural signal pair $x_{L,R}(t)$, using a 6th-order low-pass filter at 1 kHz, after [16], and taking the normalised integrated sum of their product as shown in Eq. (6),

$$\mathrm{ITLD} = \frac{\int_{t_1}^{t_2} \Lambda_L(t)\Lambda_R(t)dt}{\sqrt{\left[\int_{t_1}^{t_2} \Lambda_L^2(t)dt\right]\left[\int_{t_1}^{t_2} \Lambda_R^2(t)dt\right]}} \tag{6}$$

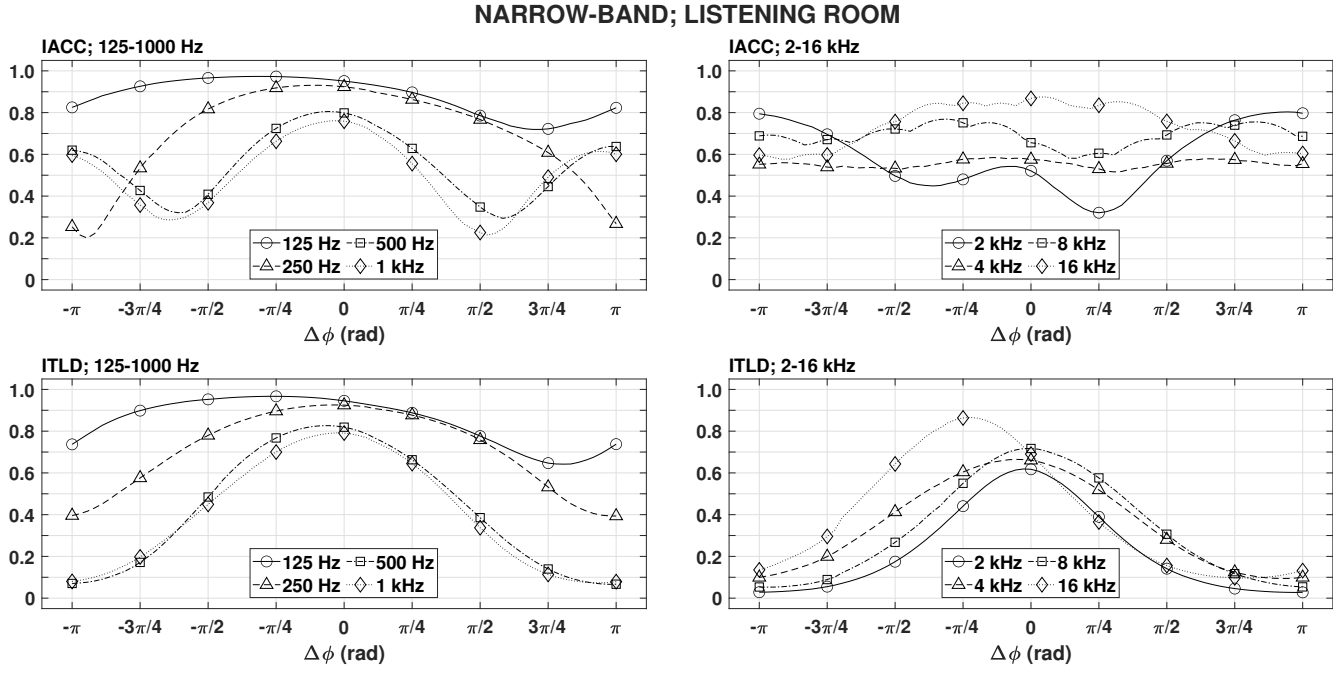
where $\Lambda_{L,R}(t)$ represents the output of the CNTM processing applied to $x_{L,R}(t)$, and t_1 and t_2 are the time integration limits for calculating the metric. In [7], the interval [$t_1 = 0$ ms, $t_2 = 80$ ms] was used in order to consider only the initial portion of recorded signals and for direct comparison with $\mathrm{IACC_E}$. In Sec. 3 the time interval is defined as [t_1 = 0, $t_2 = \infty$] for both IACC and ITLD. Eq. (6) yields an ITLD of 0 when $x_{L,R}(t)$ are in anti-phase and a value of 1 when they are exactly phase-coherent, and hence when using ITLD to evaluate ASW, [1 – ITLD] should be used to obtain a positive perceptual correlation between ITLD and ASW.

3. FREQUENCY-DOMAIN PROPERTIES OF THE ITLD

The variability of IACC and $\mathrm{IACC_E}$ over different frequency bands has been widely studied, with $\mathrm{IACC_{E3}}$, the average value of $\mathrm{IACC_E}$ over 500 Hz, 1 kHz and 2 kHz octave-bands, displaying a stronger correlation with ASW [1,3,5]. To investigate the frequency-dependent properties of the ITLD, binaural measurements were taken of a set of stereophonic decorrelated pink noise bursts, windowed with a 15% tapered-cosine function, using a Neumann KU-100 dummy head in an anechoic chamber of dimensions 3.05 m × 3.40 m × 5.50 m, and a critical listening room of dimensions 4.40 m × 2.67 m × 6.30 m and reverberation time of 0.25 s. The noise bursts were decorrelated using a relative phase shift approach after [9], increasing from -π to +π rad in intervals of $\pi/36$ (5 cyclic degrees). Playback was rendered using a pair of loudspeakers with full-range drive units in enclosed cabinets. The loudspeakers were spaced 2 m apart at a distance of 1.73 m from the dummy head in the anechoic chamber, and a 2.9 m spacing at a distance of 2.51 m in the listening room to produce an azimuth of 60° between the loudspeakers at the central listening position in both environments. Band-pass filtering, using 512-tap 96 kHz octave-band FIR filters at centre-frequencies f_c increasing from 125 Hz to 16 kHz, was then applied to the recordings before IACC and ITLD values were calculated, shown in Fig. 1. Band-averaging of values over 500



(a)



(b)

Figure 1: *Narrow-band comparison of IACC and ITLD values measured in an anechoic chamber (a) and a reverberant listening room (b).*

Hz, 1 kHz and 2 kHz was also performed for both metrics, shown in Fig. 2 alongside wide-band values and the effects of omitting the 6th-order low-pass filter on ITLD values. Observations made in the following are discussed in Sec. 5.

3.1. Narrow-band Behaviour

Under anechoic conditions, ITLD displays pronounced behaviour between 125 Hz and 2 kHz, which begins to break down at 4 kHz and above. For any $|\Delta\phi| \leq 3\pi/4$ rad, ITLD will decrease as the f_c of the octave-band is increased, with the extent of the reduction

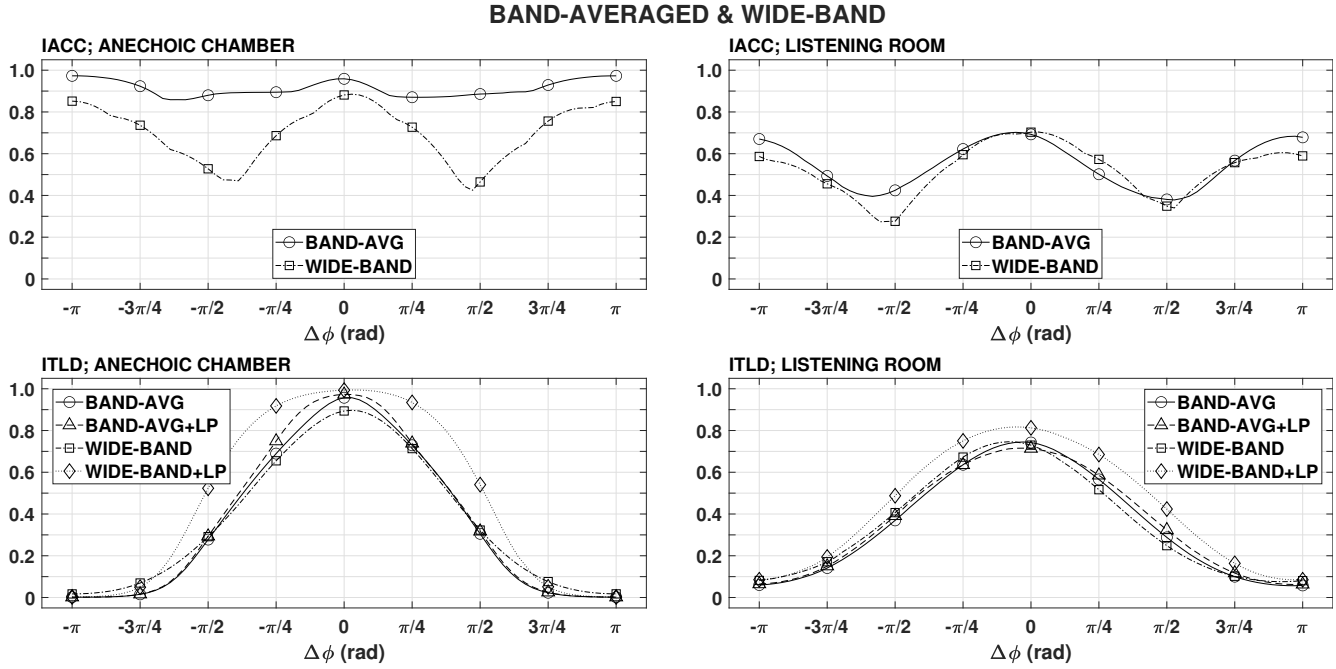


Figure 2: *Band-averaged and wide-band comparison of IACC and ITLD values in an anechoic chamber (left column) and a reverberant listening room (right column). The effects of omitting the 6th-order low-pass filter from ITLD calculations are marked in both listening environments.*

in value diminishing as $|\Delta\phi| \rightarrow 0$. This behaviour is illustrated in Fig. 1(a), where the trends of ITLD values resemble tapered-cosine windows with the cosine fraction increasing with f_c. It is also observed that for 125 Hz $\leq f_c \leq$ 2 kHz, ITLD values at $|\Delta\phi| = \pi$ rad are largely invariant to the changing f_c. For higher frequencies, 4 kHz $\leq f_c \leq$ 16 kHz, the inverse-relationship characteristic seen in the lower octave-bands breaks down, although the tapered-cosine trend is still observed: values in the 4 kHz band are increased over the 2 kHz band and are similar to the 500 Hz and 1 kHz bands. The 8 kHz band follows the trend of the 2 kHz band for $\Delta\phi < 0$, and the 4 kHz band for $\Delta\phi > 0$, displacing the symmetry seen in the plots to about $\Delta\phi = \pi/8$ rad. Additionally, ITLD values in the 16 kHz band are increased over the 2 kHz band for all $|\Delta\phi| > 0$, though good symmetry about $\Delta\phi = 0$ is still observed. The range of values (ROV) for bands 125 Hz $\leq f_c \leq$ 1 kHz, as well as 4 kHz, range appropriately from $\rightarrow 0$ at $|\Delta\phi| = \pi$ rad to $\rightarrow 1$ at $\Delta\phi = 0$, while the 2 kHz, 8 kHz and 16 kHz bands see a small reduction at $\Delta\phi = 0$, no larger than 0.2. Under reverberant conditions, the inverse-relationship and tapered-cosine shapes observed in the anechoic case are also present, but limited to the 500 Hz, 1 kHz and 2 kHz bands and to a smaller extent. The tapered-cosine trend is not observed in the 125 Hz band, and the ROV of the 125 Hz and 250 Hz bands are reduced significantly to 0.75-0.95 and 0.4-0.9, respectively, though they both approach an appropriate maximum value as $|\Delta\phi| \rightarrow 0$. ITLD values decrease for the 500 Hz, 1 kHz and 2 kHz bands similarly to the anechoic case, though the decrease in value between these bands is far smaller in the reverberant case, and their ROV is reduced to 0.1-0.8 (500 Hz and 1 kHz) and 0.05-0.6 (2 kHz). As with the anechoic case, ITLD values increase at 4 kHz and decrease at 8 kHz, with both bands exhibiting an ROV of around 0.1-0.7. ITLD in the 16 kHz band also exhibits the same behaviour as the 8 kHz band in the anechoic recordings, characterised by asymmetry about $\Delta\phi$ = 0 and the trend closely following the 2 kHz band for $\Delta\phi > 0$.

Narrow-band IACC values exhibit significantly different behaviour in both the anechoic and reverberant cases compared to ITLD. Under anechoic conditions, IACC displays a greatly reduced ROV for octave-bands $f_c \geq 500$ Hz, of which the 2 kHz band exhibits the largest ROV of 0.8-1.0. As such, the pronounced "W" trend seen in the wide-band data in Fig. 2 is largely absent or drastically less pronounced. The 125 Hz and 250 Hz bands display a greater ROV (0.35-0.95 and 0.2-0.95, respectively), however this is localised to $-\pi \leq \Delta\phi \leq -\pi/2$ rad and not reflected for the same positive values of $\Delta\phi$. A shallow "W" trend is seen in the 2 kHz, 4 kHz and 8 kHz bands with a reduced ROV as described above, while the 16 kHz band is relatively flat, increasing from 0.7 at $|\Delta\phi| = 3\pi/4$ rad to 0.85 at $\Delta\phi = 0$, corresponding to approximately 2 JNDs. In contrast to the anechoic measurements, the data captured in the listening room suggests that IACC and ITLD exhibit some similarities below 4 kHz. The 125 Hz and 250 Hz bands exhibit a similar trend over the range of $\Delta\phi$ with largely comparable ROV, and a pronounced "W" shape is seen in IACC for the 500 Hz, 1 kHz and 2 kHz bands. Of these bands, ITLD exhibits an ROV approximately 33% greater than IACC for 500 Hz and 1 kHz, and 25% greater than IACC for 2 kHz. For bands $f_c \geq$ 4 kHz, no "W" trend is observed in the IACC values, and the 4 kHz band varies by less than 2 JNDs over the range of $\Delta\phi$.

3.2. Band-averaged and Wide-band Behaviour

Fig. 2 compares wide-band values of ITLD and IACC with band-averaged values calculated over the 500 Hz, 1 kHz and 2 kHz

bands. For ITLD, band-averaged values are slightly increased in the anechoic case at $\Delta\phi = 0$ compared to wide-band data, but otherwise they are largely comparable and exhibit a strong tapered-cosine trend for both listening environments. The minimum ITLD ROV is increased by approximately 28% in the anechoic case compared to the listening room, but this is to be expected on account of the inherent nature of an anechoic chamber in contrast to a more reverberant listening room. Band-averaged IACC values are restricted to a range of approximately 2 JNDs in the anechoic chamber, whereas a much larger ROV is observed in the listening room, corresponding to approximately 20 JNDs. IACC values calculated from the listening room measurements show the wide-band and band-averaged data to be broadly comparable, both in terms of general trend and ROV.

3.2.1. ITLD Low-pass Filter Effects

As discussed in Sec. 2.3, the 6th-order 1 kHz low-pass filter (LPF) is included in the formulation of the ITLD in order to account for the inability of the IHC to resolve high frequencies [11] in the CNTM. The filter was omitted in the narrow-band measurements to accommodate a direct comparison between IACC and ITLD in each octave-band, but included in Fig. 2 to demonstrate the effect on measured values of ITLD. In both listening environments, the LPF increases the band-averaged ITLD by a small margin, at most 0.05, while wide-band measurements are shown to be affected more. In the anechoic case, the LPF increases wide-band ITLD values by as much as 0.25 for $\pi/2 \leq |\Delta\phi| \leq \pi/4$ rad and decreases ITLD for $|\Delta\phi| > 2\pi/3$ rad. In the reverberant case, ITLD values are increased to a smaller extent for all $\Delta\phi$ with the LPF included. Omitting the LPF increases the contribution of direct and reflected high frequencies to the ITLD calculation, which adds both rectified and non-zero energy from shorter wavelengths and therefore alters the number of non-zero elements in $\Lambda_{L,R}(t)$ at different $\Delta\phi$. This means that at larger phase shifts, $|\Delta\phi| > 2\pi/3$ rad, the rectified regions in $\Lambda_{L,R}(t)$, which decrease ITLD, are reduced while at smaller $\Delta\phi$ the non-zero regions, which increase ITLD, are also reduced as more negative cycles are introduced. These effects are diminished with the LPF applied, causing the reduction in ITLD for $|\Delta\phi| > 2\pi/3$ rad seen in the anechoic case, and the increase for both the anechoic and reverberant cases at $|\Delta\phi| < 2\pi/3$ rad. The absence of a decrease in ITLD at larger $\Delta\phi$ in the listening room may be caused by the inability of the LPF to overcome the combined effects of room asymmetry and the reverberant nature of the soundfield adding reflections which significantly affect the ITLD calculation.

4. PERCEPTUAL VALIDATION OF THE ITLD

While the objective characteristics of the ITLD have been detailed at length in [7] and furthered in this work, ascertaining the nature of the perceptual relationship between ITLD and ASW is of foremost importance when considering its application as a metric of ASW.

4.1. Experimental Design

A semantic differential listening test based on the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) framework [19] was designed in this work. 14 participants ranked the perceived ASW of decorrelated pink noise stimuli from 0 (narrowest) to 100 (widest). A set of 7 2-channel pink noise sources were generated using an interchannel relative phase-shift $\Delta\phi$ increasing from 0 to π rad in steps of $\pi/6$, with a 15% tapered-cosine window applied, as in Sec. 3. The order of the test stimuli was randomised for each participant, with the known and hidden references corresponding to $\Delta\phi = \pi/2$ rad, and the anchor corresponding to $\Delta\phi = 0$ (no decorrelation). The tests were performed in the critical listening room used in Sec. 3, with the same loudspeaker placement and listening position. A MATLAB app was designed for the test to facilitate playback and collect the ASW data, with the user interface shown in Fig. 3. Participants were able to listen to each of the stimuli as many times as they required, and as such the maximum output amplitude of the stimuli was set to a safe listening level. The collected perceptual data was then normalised and averaged before Pearson's correlation analysis was performed using wide-band and band-averaged values of ITLD and IACC measured in Sec. 3 to obtain R^2, r and p-values for both metrics.

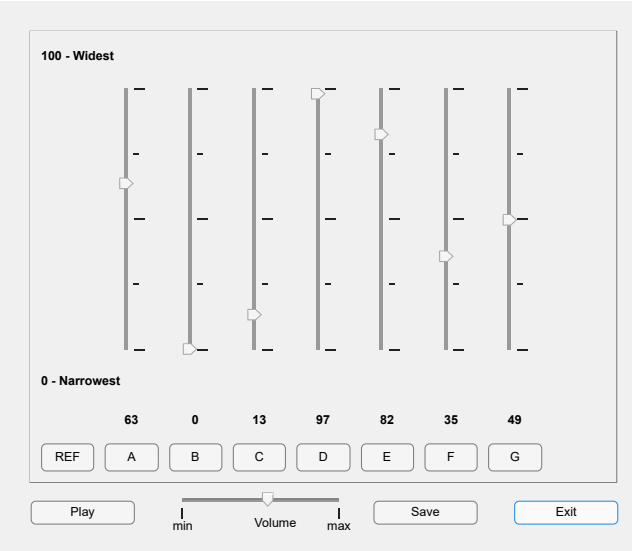


Figure 3: *User interface for evaluating the perceptual correlation between ITLD and ASW using the MUSHRA framework.*

4.2. Results

Fig. 4 compares the relationships between wide-band and band-averaged [1 – ITLD] data, with the time integration interval set to $[t_1 = 0, t_2 = \infty]$ and the LPF both omitted and applied, against the averaged ASW scores collected from the perceptual listening test. A very strong correlation is observed for all ITLD datasets, with the largest R^2 values exhibited when the LPF is included as part of the ITLD calculation. The wide-band LPF case displays the largest R^2 of 0.992, with a p-value $\ll 0.001$, while the band-averaged LPF case performs marginally poorer with an R^2 of 0.983 and p-value $\ll 0.001$. Without the LPF applied, the wide-band and band-averaged cases yield R^2 values of 0.940 and 0.969 and p-values < 0.001 and $\ll 0.001$, respectively.

The relationship between [1 – IACC] and the averaged data is difficult to interpret owing to the nature of the swing of IACC values over the range $0 \leq \Delta\phi \leq \pi$ rad. [1 – ITLD] values exhibit a single minimum and maximum value at $\Delta\phi = 0$ and $\Delta\phi = \pi$ rad,

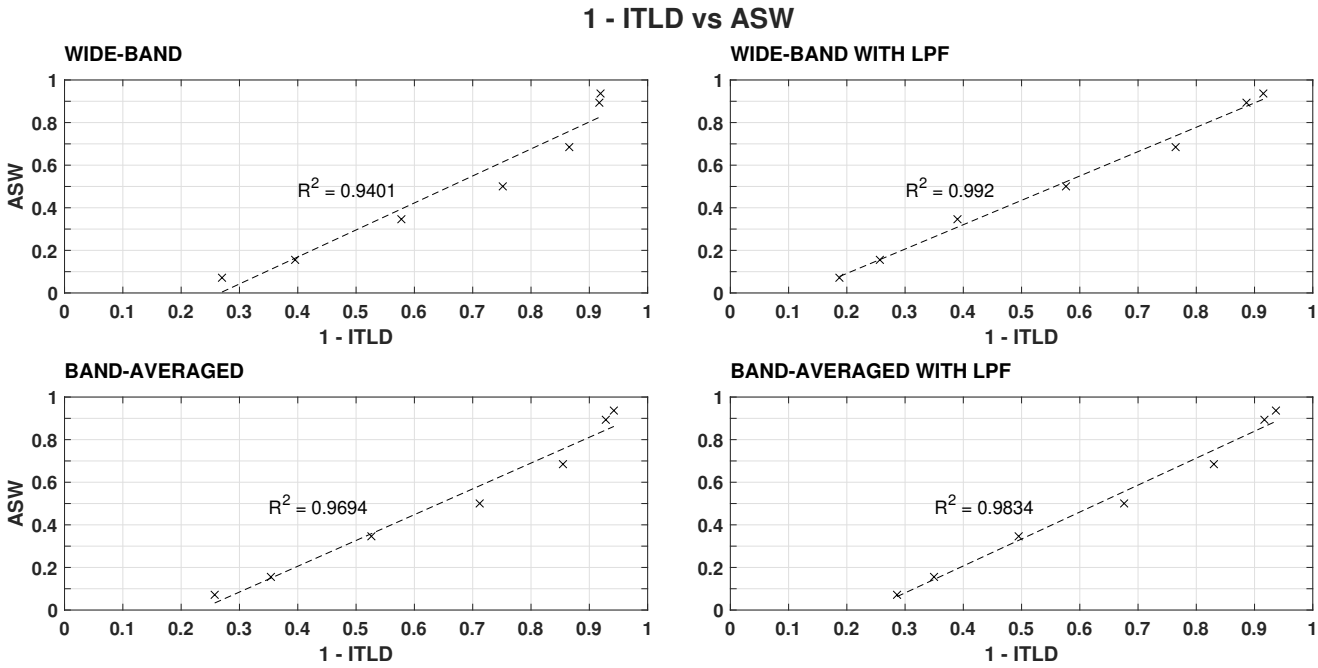


Figure 4: *[1 – ITLD] vs ASW for wide-band (top left), band-averaged (bottom left), wide-band low-pass filtered (top right) and band-averaged low-pass filtered (bottom right) ITLD values.*

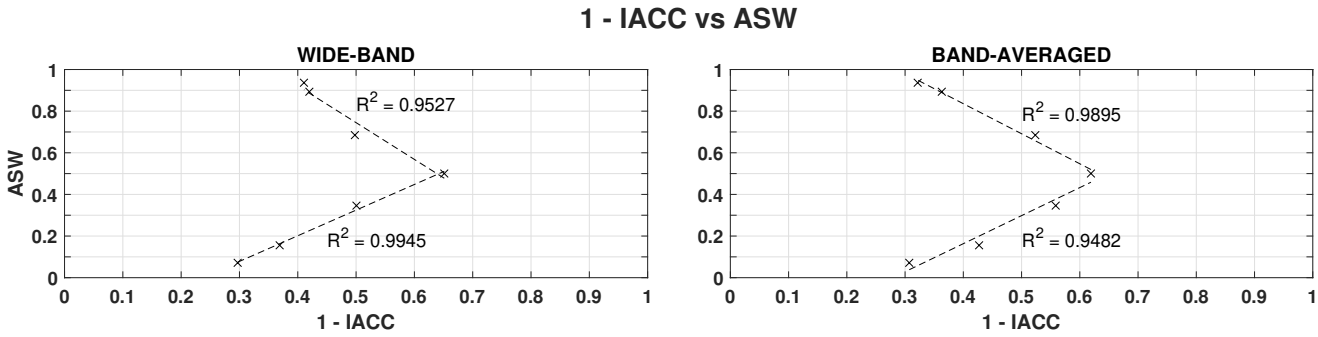


Figure 5: *[1 – IACC] vs ASW for wide-band (left) and band-averaged (right) IACC values. Symmetry in IACC values about $\Delta\phi = \pi/2$ rad means that a regression line across the entire range of ASW values is not possible, and is instead shown in two linear sections.*

respectively, whereas [1 – IACC] values display a pair of minima at these $\Delta\phi$ which is not reflected in the collected ASW data, and therefore meaningful R^2 and *p*-values cannot be obtained over the range $0 \leq \Delta\phi \leq \pi$ rad. This is readily observed from Fig. 5, which compares the relationships between wide-band and band-averaged [1 – IACC] data against the averaged ASW data. R^2 values are shown for the first and last 4 data points in both cases, corresponding to $0 \leq \Delta\phi \leq \pi/2$ and $\pi/2 \leq \Delta\phi \leq \pi$ rad, respectively, referred to henceforth as $\Delta\phi_\alpha$ and $\Delta\phi_\beta$. In the wide-band case, the relationship over $\Delta\phi_\alpha$ is characterised by a very strong R^2 of 0.995, and a *p*-value < 0.01, while the band-averaged case exhibits a smaller R^2 of 0.948 and an increased *p*-value of $p < 0.03$ over the same interval. Conversely, over the interval $\Delta\phi_\beta$, the band-averaged [1 – IACC] data exhibits a larger R^2 than the wide-band case, 0.990 compared to 0.953, as well as a smaller *p*-value, $p < 0.01$ compared to $p < 0.03$.

4.2.1. ITLD Time-integration Limits

The results shown in Fig. 4 use ITLD values calculated over the duration of the listening test stimuli, $[\mathrm{t}_1 = 0, \mathrm{t}_2 = \infty]$. It is common to see ASW metrics calculated over the early portion of a response, generally the first 80 ms, such as with $\mathrm{IACC_E}$ and $\mathrm{LF_E}$. To observe the effects of varying the length of the integration window on the strength of the correlation between [1 – ITLD] and the averaged ASW values, 6 ITLD sub-quantities were formed for both wide-band and band-averaged ITLD values, the latter indicated in by a "3" subscript: ITLD and $\mathrm{ITLD_3}$ represents the wide-band and band-averaged ITLD evaluated over the interval $[\mathrm{t}_1 = 0, \mathrm{t}_2 = \infty]$, $\mathrm{ITLD_E}$ and $\mathrm{ITLD_{E3}}$ represent evaluation over $[\mathrm{t}_1 = 0, \mathrm{t}_2 = 80$ ms$]$, following [7], and $\mathrm{ITLD_L}$ and $\mathrm{ITLD_{L3}}$ are evaluated over $[\mathrm{t}_1 = 80$ ms $, \mathrm{t}_2 = \infty]$. R^2, *r* and *p*-values describing the relationships between these quantities and the averaged ASW data, with comparison with and without the application

of the LPF, are given in Table 1. The data confirms the observation made in Sec. 4.2 that the application of the LPF increases the strength of the correlation between the different ITLD quantities and the averaged ASW data in every case, as well as decreasing the *p*-value. All quantities with the LPF applied have an R^2 value greater than 0.95, and *p*-value far below 0.05, in some cases by up to 4 orders of magnitude. The strongest-performing quantities are the wide-band [1 – ITLD] and [1 – $ITLD_L$] with the LPF applied, which have the same R^2 and *p*-values listed in Table 1. [1 – ITLD] has an exceptionally small increase in R^2 value over [1 – $ITLD_L$] of 3.920×10^{-8}, with the *p*-value decreased by an even smaller margin of 2.402×10^{-11}. The additional 80 ms in the [1 – ITLD] integration interval corresponds to a small section of the initial rise of the tapered-cosine window applied to the noise bursts, and therefore a subsequent evaluation of these quantities with transient stimuli may be the subject of further research.

Table 1: *R^2, r and p-values describing relationships between [1 – ITLD] sub-quantities and averaged ASW data.*

Quantity	LPF	R^2	*r*	*p*
[1 – ITLD]	n	0.940	0.970	3.044×10^{-4}
	y	0.992	0.996	1.969×10^{-6}
[1 – $ITLD_3$]	n	0.969	0.985	5.608×10^{-5}
	y	0.983	0.992	1.217×10^{-5}
[1 – $ITLD_E$]	n	0.895	0.946	1.264×10^{-3}
	y	0.982	0.991	1.440×10^{-5}
[1 – $ITLD_{E3}$]	n	0.957	0.978	1.303×10^{-4}
	y	0.958	0.979	1.283×10^{-4}
[1 – $ITLD_L$]	n	0.940	0.970	3.044×10^{-4}
	y	0.992	0.996	1.969×10^{-6}
[1 – $ITLD_{L3}$]	n	0.969	0.985	5.608×10^{-5}
	y	0.983	0.992	1.217×10^{-5}

5. DISCUSSION

ITLD exhibits marked frequency-dependent behaviour under anechoic conditions between 125 Hz and 2 kHz, characterised by a tapered-cosine shape in the swing of its values over the full range of relative phase shifts $-\pi \leq \Delta\phi \leq \pi$ rad, with the cosine-fraction increasing with the centre-frequency f_c of the octave-band under observation. The trend breaks down above 2 kHz, where the swing in ITLD values over the range of $\Delta\phi$ exhibits a broader bell-curve shape, and asymmetries about $\Delta\phi = 0$ become apparent at higher frequencies. This behaviour can be explained by considering that as the f_c of the band of interest is increased, a greater number of negative cycles which are rectified to 0 as part of the ITLD calculation are introduced into the CNTM signals $\Lambda_{L,R}(t)$, which increases the size of the region where $\Lambda_L(t) \cdot \Lambda_R(t)$ is equal to 0. This is why ITLD reduces more at larger $\Delta\phi$, as the regions where $\Lambda_{L,R}(t)$ are both non-zero are already reduced. While the same phenomena is seen in the listening room measurements, the effects of room reflections on ITLD calculation are marked, reducing the ROV of each band relative to anechoic measurements and reducing the tapered-cosine trend. A bell-curve trend in the swing of ITLD values over $\Delta\phi$, seen at higher frequencies in the anechoic data, is observed for octave-bands $f_c \geq 500$ Hz, with asymmetries seen at high frequencies again. Room reflections reduce the coherence of the signals measured at the ears of the dummy head and increase loudspeaker cross-talk effects which, like with increasing f_c, introduce additional negative cycles into $\Lambda_{L,R}(t)$, again reducing regions of non-zero values in the product of the CNTM signals and thus the ITLD. The low-pass filter used in the ITLD calculation was shown to increase the range of values measured over the range of $\Delta\phi$, by decreasing the ITLD at the largest phase shifts while increasing it between $-\pi/2 \leq \Delta\phi \leq \pi/2$ rad. The decrease in ITLD at $|\Delta\phi| \geq 3\pi/4$ rad can be attributed to the same behaviour observed for the narrow-band data described above, and the increase in ITLD at smaller $\Delta\phi$ is also a result of this phenomena. At smaller $\Delta\phi$, the regions in $\Lambda_{L,R}(t)$ that are simultaneously non-zero are increased. Applying the LPF reduces the contribution of high frequencies to the ITLD calculation, which are observed in the narrow-band data to reduce the ITLD. This means that a greater number of negative cycles are removed from regions in $\Lambda_{L,R}(t)$ which would otherwise have contributed to a non-zero value, increasing the measured ITLD. These observations are marked in the anechoic data, whereas the listening room measurements show the extent of the LPF effects to be reduced. This is due to the effects of room reflections on the ITLD calculation as described above, which the LPF is evidently unable to overcome.

The results of the semantic differential listening test outlined in Sec. 4 show a very strong correlation between 6 wide-band and band-averaged sub-quantities of ITLD and averaged ASW data for a steady-state pink noise burst. The largest R^2 value and smallest *p*-value between the ITLD sub-quantities and ASW were obtained from [1 – ITLD] (the wide-band ITLD data evaluated over the interval $[t_1 = 0, t_2 = \infty]$) yielding values of R^2 = 0.992 and $p \ll 0.001$. The difficulty in obtaining the relationships between wide-band and band-averaged [1 – IACC] values and ASW over the full range of $\Delta\phi$ was outlined in Sec. 4.2. Although some portions of the [1 – IACC] data showed a marginally stronger correlation with ASW compared to [1 – ITLD], [1 – ITLD] *p*-values are consistently smaller compared to [1 – IACC], by up to 4 orders of magnitude. For all ITLD sub-quantities, applying the LPF to the [1 – ITLD] calculation strengthened the correlation with ASW and reduced *p*-values, in some cases by up to 2 orders of magnitude. The largest increase in the R^2 value was observed for the wide-band [1 – $ITLD_E$] data, increasing from 0.895 to 0.982. Despite this, the R^2 value for [1 – $ITLD_E$] and the band-averaged [1 – $ITLD_{E3}$] were among the lowest of the sub-quantities, which may be attributed to integration interval $[t_1 = 0, t_2 = 80$ ms] corresponding to a small portion of the initial transient of the tapered-cosine window applied to the test stimuli. It is also observed from the data in Table 1 that band-averaging [1 – ITLD] values yields a weaker correlation with ASW, which is contrary to what is typically observed for IACC [1]. Objective comparisons of wide-band and band-averaged values of both metrics in Fig. 2 support these observations, as band-averaged ITLD values exhibit a smaller ROV compared to wide-band data, while band-averaged IACC values are more symmetrical about the y-axis compared to wide-band IACC.

6. CONCLUSIONS AND FURTHER WORK

This paper has presented a perceptual validation of the interaural thresholded level distribution to subjectively evaluate its application as a quantity of auditory source width, and an inves-

tigation into the frequency-dependent characteristics of the metric. A semantic differential listening test based on the MUSHRA framework was designed to obtain relative ASW scores of decorrelated pink noise bursts inside a critical listening room, which were then compared against measured values of [1 – ITLD], ITLD-derived quantities and [1 – IACC] to ascertain a perceptual relationship between the metric and ASW. Results from the listening test showed a very strong correlation between 12 ITLD-derived quantities and ASW with excellent statistical significance, with the best-performing metric being the wide-band [1 – ITLD] evaluated over the duration of the test stimuli with a 6th-order 1 kHz low-pass applied. A very strong correlation, comparable to [1 – ITLD], between [1 – IACC] and ASW was also identified with some limitations owing to the nature of IACC values in response to extreme decorrelation. Frequency-dependent characteristics of the ITLD were ascertained by applying an octave-band FIR filter bank to the pink noise bursts used in the perceptual validation test, from which a clear and consistent trend was observed in anechoic conditions between 125 Hz and 2 kHz, and between 500 Hz and 2 kHz in reverberant conditions. ITLD values measured from binaural signals with large amounts of decorrelation were observed to gradually decrease as the centre frequency of the band of interest was increased, with the application of the 1 kHz low-pass filter shown to increase the range of values in both wide-band and band-averaged cases. Further research into the perceptual correlation between [1 – ITLD] and ASW when using transient stimuli will provide a deeper understanding of the subjective relationships presented here, as well as exploring other psychoacoustic properties of the metric such as its just-noticeable-difference.

7. REFERENCES

[1] Takayuki Hidaka, Leo L Beranek, and Toshiyuki Okano, "Interaural cross-correlation, lateral fraction, and low-and high-frequency sound levels as measures of acoustical quality in concert halls," *J. Acoust. Soc. Am.*, vol. 98, no. 2, pp. 988–1007, 1995.

[2] Toshiyuki Okano, Leo L Beranek, and Takayuki Hidaka, "Relations among interaural cross-correlation coefficient (iacc e), lateral fraction (lf e), and apparent source width (asw) in concert halls," *J. Acoust. Soc. Am.*, vol. 104, no. 1, pp. 255–265, 1998.

[3] Masayuki Morimoto and Kazuhiro Iida, "Appropriate frequency bandwidth in measuring interaural cross-correlation as a physical measure of auditory source width," *Acoust. sci. tech.*, vol. 26, no. 2, pp. 179–184, 2005.

[4] Franz Zotter and Matthias Frank, "Efficient phantom source widening," *Archives of Acoustics*, vol. 38, 2013.

[5] Matthias Frank and Franz Zotter, "Simple technical prediction of phantom source widening," *AIA/DAGA, Fortschritte der Akustik, Meran*, 2013.

[6] Michael A Gerzon, "General metatheory of auditory localisation," in *Audio Engineering Society Convention 92*. Audio Engineering Society, 1992.

[7] Christian SE Cotton, James Blake, Ethan Stanhope, Stephen G Oxnard, and Laurence J Hobden, "Evaluating the interaural thresholded level distribution as a psychophysically-motivated metric of auditory source width," in *Audio Eng. Soc. Conf.: AES 2023 Int. Conf. on Spatial and Immersive Audio*. Audio Eng. Soc., 2023.

[8] "Acoustics — Measurement of room acoustic parameters — Part 1: Performance spaces," Standard, Int. Org. for Standardization, Geneva, CH, June 2009.

[9] Laurence J. Hobden and Christopher Gribben, "Two-channel sine sweep stimuli: a case study evaluating 2-n channel upmixers," in *Audio Eng. Soc. Conv. 146*. Audio Eng. Soc., 2019.

[10] Matthias Frank, "Source width of frontal phantom sources: Perception, measurement, and modeling," *Archives of Acoustics*, vol. 38, no. 3, pp. 311–319, 2013.

[11] Jens Blauert, *Communication acoustics*, vol. 2, Springer, 2005.

[12] Russell Mason, "Implementation and application of a binaural hearing model to the objective evaluation of spatial impression," in *Audio Eng. Soc. Conf.: 28th Int. Conf.: The Future of Audio Technology–Surround and Beyond*. Audio Eng. Soc., 2006.

[13] Richard Lyon, "A computational model of binaural localization and separation," in *ICASSP'83. IEEE Int. Conf. Acoust., Speech, and Sig. Processing*. IEEE, 1983, vol. 8, pp. 1148–1151.

[14] Matti Karajalainen, "A binaural auditory model for sound quality measurements and spatial hearing studies," in *1996 IEEE Int. Conf. Acoust., Speech, and Sig. Processing. Conf. Proc.* IEEE, 1996, vol. 2, pp. 985–988.

[15] Russell R Pfeiffer and DO Kim, "Response patterns of single cochlear nerve fibers to click stimuli: descriptions for cat," *J. Acoust. Soc. Am.*, vol. 52, no. 6B, pp. 1669–1677, 1972.

[16] Russell Mason, Tim Brookes, and Francis Rumsey, "Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli," *J. Acoust. Soc. Am.*, vol. 117, no. 3, pp. 1337–1350, 2005.

[17] Richard Lyon, "A computational model of filtering, detection, and compression in the cochlea," in *ICASSP'82. IEEE Int. Conf. Acoust., Speech, Sig. Processing*. IEEE, 1982, vol. 7, pp. 1282–1285.

[18] Ray Meddis, "Simulation of mechanical to neural transduction in the auditory receptor," *J. Acoust. Soc. Am.*, vol. 79, no. 3, pp. 702–711, 1986.

[19] B Series, "Method for the subjective assessment of intermediate quality level of audio systems," *International Telecommunication Union Radiocommunication Assembly*, 2014.

[20] Toshiyuki Okano, "Image shift caused by strong lateral reflections, and its relation to inter-aural cross correlation," *J. Acoust. Soc. Am.*, vol. 108, no. 5, pp. 2219–2230, 2000.