4aPPb22. Spectral integration of interaural time differences in auditory localization
Nicolas Le Goff*, Jorg M. Buchholz and Torsten Dau

*Corresponding author's address: Technical University of Denmark, Lyngby, 2800, -, Denmark, nlg@elektro.dtu.dk

This study investigates how the auditory system integrates spatial information across frequency. In experiment one, discrimination thresholds for interaural time differences (ITDs) were measured as a function of both reference ITD and center frequency (CF) of noises with bandwidth of one ERB. In addition, discrimination thresholds were also measured as a function of CF for different values of interaural coherence (IC) typical of sounds in realistic acoustic environments. For both high ICs and small reference ITDs, discrimination thresholds were lowest for CFs between 700 and 1000 Hz. For smaller ICs and larger reference ITDs, this dominance region shifted towards lower CFs. A conceptual localization model was developed that used the variance of the ITD thresholds to optimally weight the contribution of the individual frequency bands before spectral integration. In experiment two the model was tested by asking listeners to align a broadband noise signal with an ITD that was fixed across frequency onto a broadband noise target with different ITDs in individual 1 ERB-wide subbands. The results were consistent with both the model predictions and the shift of dominance range observed in experiment one.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Daily one-to-one verbal communication is challenging in the presence of interfering sound sources, background noise or reverberation. Despite an increasing interest on the topic, the ability for most listeners to communicate almost effortlessly in such adverse acoustical conditions remains unexplained. It has, however, been suggested that, among other aspects, the ability to localize the position of a talker by processing spatial cues plays a positive role (e.g., Bronkhorst, 2000). Speech is a broadband sound and the understanding of localization of broadband signals in realistic conditions is, therefore, crucial.

In recent binaural auditory models (e.g., Stern et al., 1988; Breebaart, 2001), it has been proposed that broadband signals are decomposed into individual frequency bands in the cochlea and that the binaural information present in each frequency band is spectrally integrated at higher stages along the auditory pathway. A number of studies have investigated this spectral integration process specifically for localization. The investigations of Raatgever (1980) have found experimental evidence that the salience of components in binaural stimuli was not equal for all frequencies. As a consequence, the inclusion of frequency weights has been considered, as for instance in the cross-correlation lateralization model (“weighted-image model”) proposed by Stern et al. (1988), and validated on different data sets (e.g., Trahiotis and Stern, 1989). A limitation of these studies, however, is that the investigations were restricted to controlled conditions that do not necessarily reflect realistic acoustical conditions.

For instance, it is known that reverberation in rooms causes a reduction in interaural correlation at the ears of a listener (Kutruff, 2000) and alter, to some degree, the interaural differences available to a listener. For a single sound source in an anechoic environment, a cross-correlation lateralization model would show that the peak of the cross-correlation function (coherence) reaches a value of 1 (full correlation). Furthermore, this peak would occur at a fixed interaural delay in each frequency band, corresponding to the acoustical interaural time difference (ITD). In contrast, in rooms, reflections of sound from the walls (reverberation) cause a decrease in coherence and a variation of the position of the peak in time. In addition, such disturbances of the cross-correlation functions are independent across frequencies and, hence, lead to possible misalignments of the cross-correlation patterns across frequencies.

Another aspect of realistic communication is that sound sources may occur at various azimuth angles, corresponding to ITD and interaural level difference (ILD) cues spread across the entire ecological range. It has been shown that the binaural system can effectively suppress interfering broadband sources at azimuth angles different from that of the talker (e.g., Peissig, 1997). While this type of directional binaural unmasking is important for the investigation of realistic communication, the possible influence of the azimuth angle on how frequencies are weighted remains unexplored.

The present study, therefore, aimed at extending previous results regarding spectral integration for localization in more realistic acoustical conditions. To this end, the spectral integration of ITDs was studied in three experiments where ITD detection thresholds were measured for narrowband and broadband signals. In the first experiment, to account for the decrease in interaural correlation occurring at the listener’s ears in rooms, ITD thresholds were obtained for fully and partially correlated signals. In the second experiment, ITD detection thresholds were measured as a function of the reference ITD in the detection task, corresponding to different azimuth angles.

Detection experiments were followed by a localization experiment, which represented more realistically everyday conditions. The listener’s task was to align a broadband pointer signal carrying a single ITD with a broadband target signal that carried a different ITD in each frequency band. The results of the three experiments were used to specify and validate a conceptual localization model in which the information in each frequency band was optimally weighted by the variance of the internal noise limiting the ITD detection process.

METHOD

This study consists of two ITD detection experiments, a localization experiment and the development of a conceptual localization model. All experiments were conducted in an insulated sound booth and all stimuli were presented over headphones.
Exp. 1: ITD Thresholds as a Function of Frequency and Interaural Correlation

The purpose of this experiment was to investigate how reverberation can influence the ability to detect ITDs. ITD thresholds were, therefore, obtained for different interaural correlations, for both narrowband and broadband stimuli.

A 3-interval, 3-alternative forced-choice method was combined with a 2-down 1-up tracking of the 70.7% point of the psychometric function (Levitt, 1971). At the beginning of the experiment, the ITD was varied with a step size of 1.584 (2 dB). After the first two reversals, the step size was reduced to 1.122 (0.5 dB). Thresholds were measured as the average of the last eight reversals. The task of the listeners was to detect the target interval, which contained a stimulus lateralized by means of an ITD. The stimuli in the two reference intervals were diotic. Each interval contained a new noise token that was 300-ms long and was gated with 5-ms onset and offset ramps. Stimuli were presented at a fixed total power of 70 dB SPL and had a bandwidth of 1 ERB (Glasberg and Moore, 1990) or were broadband with a bandwidth of 3 kHz and centered at 1500 Hz. The ITDs were applied by filtering the right-ear signals with a non-causal all-pass filter that had a constant group delay corresponding to the desired ITD. The filter was realized by applying a phase shift specific to the ITD and to each frequency bin in the spectral domain. Partially correlated stimuli were generated by mixing two independent noise realizations (Hartmann, 2011). Four listeners (including the first author), who had normal hearing and who were familiar with psychoacoustic tasks, participated in the experiment. Three repetitions of each measurement were carried out.

Exp. 2: ITD Thresholds as a Function of Frequency and Reference ITD

In natural listening condition, while one’s attention is predominantly focused ahead in the horizontal plane, sound sources also occur at larger azimuth angles. The purpose of the second experiment was to measure the dependence of the ITD detection thresholds as function of the reference ITD, corresponding to different azimuth angles.

The method and apparatus used in this experiment were similar to those of the first detection experiment. ITD detection thresholds were measured for the same narrowband and broadband noises. The interaural correlation of the signals was, however, fixed to 1 for the whole experiment. The ITD carried by the stimuli in the two reference intervals (“reference ITD”) was varied. ITD detection thresholds were obtained for reference ITDs of 200, 400 and 600 µs, spanning most of the ITD ecological range of ITDs. Three paid subjects and the first author participated in the experiment. All participants received training and were familiar with psychoacoustic experiments. Four repetitions of each measurement were carried out.

Exp. 3: Localization and Spectral Integration of ITDs

A limitation of the two previous experiments is that ITD detection is a basic task that is not directly related to even the simplest spatial localization tasks that occur in everyday acoustic scenarios. In this experiment, the lateralization percept of broadband noise stimuli carrying multiple ITDs was investigated in a direction-matching experiment.

The task of the listener was to align a signal (pointer) carrying a single ITD with a signal carrying multiple ITDs (target). Both pointer and target signals consisted of nine 1-ERB-wide bands distributed between 148 and 1572 Hz and had a total power of 70 dB SPL. Signals were 250-ms long and had 5-ms onset and offset ramps. Frequency bands were separated by 2-ERB to allow an independent adjustment of the ITD in each frequency band. The target signals carried a different ITD in each frequency band that was arranged in four different distributions, shown by the circles, squares, upward and downward triangles on the left side of Fig. 3b. The target signals were grouped in 2 categories: low-ITDs (squares and circles) and high-ITDs (upward and downward triangles). In each category, the 2 types of ITD distributions had the same median value, 200 or 400 µs for the low- and high-ITD categories, respectively. In the low-ITD category, ITDs ranged between 100 and 300 µs and in the high-ITD category, ITDs ranged between 300 and 500 µs. Within one category, the distribution of the ITD along frequency bands was, however, symmetrical around the median ITD. ITDs were present in both target and pointer signals. ITDs were applied with the same technique as in the detection experiments. The pointer signals carried a single ITD that was adjusted by the listeners using a graphical interface and using one of 3 step sizes: 150, 50 or 20 µs. The listeners could play the target or pointer signals repeatedly at their convenience. The initial position of the pointer signal was...
randomly chosen between -700 and 700 µs. A measurement ended when the listener decided that the target and pointer signals matched each other in terms of lateralization. The same four listeners who participated in the first detection experiment participated in this experiment. Four repetitions of each condition were carried out.

**Localization Model**

As part of the investigation, a conceptual localization model was designed. It aimed to predict the spectral integration used for localization. One important aspect of the model is the weighting applied to the individual frequency channels, which is derived from the data obtained in the ITD detection experiments. Here the focus was to predict the localization of signals carrying ITDs at different values of the ecological range. The model was validated using the data obtained in the localization experiment.

As part of the model design, it is assumed that the uncertainties involved in the ITD threshold detection task in Experiments 1 and 2 are equal to the uncertainties involved in the localization task of Experiment 3. Applying the principles of detection theory (Green and Swets, 1966), it is, furthermore, assumed that these uncertainties can be described by an internal auditory (decision) noise with a variance \( \sigma^2 \) that is frequency \(( f)\) and interaural delay \(( \tau)\) dependent, using the following equation:

\[
\sigma^2(f, \tau) \approx \left( \frac{\tau_{\text{JND}}(f, \tau)}{d'} \right)^2 \tag{1}
\]

Here, \( d' \) is the sensitivity index (deprime) and \( \tau_{\text{JND}} \) the ITD threshold or just noticeable difference (JND) as, for instance, measured in Experiment 2, and shown in Fig 2. Since a 3-AFC task was applied in the detection experiments to measure the 70.7% correct point on the psychometric function the deprime is given by \( d' \approx 1.28 \) (Hacker and Ratcliff, 1979). Moreover, since the ITD thresholds \( (\tau_{\text{JND}}) \) were measured at different reference ITDs \( (\tau_{\text{ref}}) \), using a pointer method, the following mapping onto the interaural delay \(( \tau)\) axis was applied:

\[
\tau = \tau_{\text{ref}} + \frac{\tau_{\text{JND}}}{2} \tag{2}
\]

Hence, by applying the narrowband ITD data shown in figure 2, the variance of the auditory internal noise can be derived as a function of frequency and interaural delay using equations 1 and 2. However, the experimental data only defines the internal noise for a number of discrete interaural delays and frequency points, which do not necessarily coincide with the ones applied in the localization experiment. Hence, the data needs to be approximated (or fitted) by an analytical function, which similar to eq. 1, is as follows:

\[
\tau_{\text{JND}}(f, \tau) = \sqrt{\sigma^2(f, \tau) \cdot d'} \tag{3}
\]

With the variance \( \sigma^2(f, \tau) \) arbitrarily defined by:

\[
\sigma^2(f, \tau) = \frac{g_0(\tau)}{B(f) \cdot f} \cdot \left( 1 + \left( \frac{f}{f_{\text{up}}(\tau)} \right)^{N(\tau)} \right) \tag{4}
\]

Where \( B(f) \) is the equivalent rectangular bandwidth (ERB) of an auditory filter at center frequency \(( f)\), which according to Patterson et al. (1988) is given by:

\[
B(f) = 24.7 \cdot (0.00437 \cdot f) + 1 \tag{5}
\]

And the interaural delay-dependent fitting parameters:
\[ g_0(\tau) = 0.06 \cdot \tau + 19 \cdot 10^{-6} \]
\[ N(\tau) = -\left(2000 \cdot \tau \right)^2 + 1 \]
\[ f_{up}(\tau) = -0.85 \cdot \tau \cdot 10^{-6} + 1000 \]

Here the frequency \( f \) and bandwidth \( B \) are given in Hertz and the interaural delay \( \tau \) is given in microseconds. Equation 4 describes a band-stop shaped variance with an overall sensitivity \( g_0 \) that is interaural delay-dependent. The low-frequency roll-off mainly reflects an observer with constant-phase sensitivity as described by the term \( 1/f \). The term \( 1/B(f) \) refers to the fact that the variance of a (auditory) bandpass filtered noise is inversely proportional to its bandwidth. The remaining terms in eq. 4 describe a highpass filter whose cut-off frequency \( f_{up} \) and order \( N \) are dependent on the interaural delay. The analytical approximations given in eq. 4 are shown in Fig. 2 by the solid lines and are in good agreement with the corresponding narrowband ITD thresholds (dashed lines).

In order to model the measured localization data described in Experiment 3, it is assumed that the auditory system combines ITD information (or estimates) from different (discrete) frequency channels, \( i \), as follows:

\[ \hat{S} = \sum_{i=1}^{N} a_i \cdot \hat{S}_i \]

Where \( N \) is the number of frequency channels considered, \( S_i \) the ITD estimated at frequency channel \( i \), \( S \) the frequency-integrated (or overall) ITD estimate, and the weights \( a_i \) given by:

\[ a_i = \frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \frac{1}{\sigma_i^2} \]

In Experiment 3, the stimuli consisted of \( N \) bandpass filtered noises with center frequencies \( f_i \) whereby each bandpass filtered noise carried a different ITD \( \tau_i \). Within the localization model it is assumed that the ITD estimates \( (S_i) \) directly refer to the frequency channel-dependent ITDs \( (\tau_i) \) that are applied in the stimulus. Moreover, it is assumed that the involved variance \( \sigma_i^2 \) can be derived from eq. 4 by setting \( f = f_i \) and \( \tau = \tau_i \). The resulting weights \( a_i \) for the four different stimulus conditions used in Experiment 2 and shown in Fig. 2 are shown in Fig. 3a.

**RESULTS**

In this section, experimental data as well as analytic ITD threshold approximations and model predictions are presented for the three experiments. In all figures, open symbols denote experimental data and filled black symbols denote data obtained with the model.

The results of the first ITD detection experiment in which the effect of a reduction in interaural correlation was studied are shown in Fig. 1. The ITD thresholds for the 1-ERB-wide stimuli are shown by the connected symbols as a function of the center frequency. The corresponding thresholds for the broadband stimuli are shown on the right.
FIGURE 1: ITD thresholds as a function of frequency and interaural correlation for narrowband stimuli are shown by connected symbols on the left. Isolated thresholds on the right were measured with 3-kHz-wide stimuli centered at 1.5 kHz. Error bars represent the standard error of the mean.

For all center frequencies and bandwidths, the thresholds increase with decreasing interaural correlation. The stimuli are perceived as increasingly “wider” for smaller interaural correlations, which results in an increasing uncertainty of their location. Similar to the literature data (e.g., Klumpp and Eday, 1956; Zwislocky and Feldman, 1956), for all interaural correlations, ITD thresholds measured with narrowband stimuli decrease from low center frequencies up to about 1000 Hz and increase for higher center frequencies. For all interaural correlation values, the ITD thresholds measured for the wideband noise are just slightly below the lowest thresholds measured for the narrowband stimuli. This difference is, however, only statistically significant (single tail t-test, 5% significant level, unequal variances) for the data obtained with an interaural correlation of 0.97.

The increase in thresholds resulting from the decrease in interaural correlation is consistent with the result of the progressive increase in external noise that corrupts the temporal fine structure of the internal representation of the signals. Overall, the frequency dependency of the ITD thresholds, which, according to localization model developed in this paper, reflects the spectral weighting of the ITDs, appears to be fairly independent of the interaural correlation so that the ITDs carried by signal components between 750 and 1000 Hz are dominant for all interaural correlations.

The results of the second ITD detection experiment in which the effect of the reference ITD was studied are shown in Fig. 2 (open symbols) together with analytical approximations (filled symbols). The layout of the figure is similar to that of Fig.1 with narrowband data on the left and broadband data isolated on the right. Each symbol represents data obtained with a different reference ITD. Experimental thresholds shown for a reference ITD of 0 µs are those of the first experiment shown in Fig. 1 for fully correlated stimuli.

FIGURE 2: Experimental ITD detection thresholds (dashed lines) and corresponding analytical approximations (solid lines). Data are shown as a function of frequency and reference ITD for narrowband stimuli by connected symbols on the left. Isolated
thresholds on the right were measured with 3-kHz-wide stimuli centered at 1.5 kHz. Error bars represent the standard error of the mean.

In general, ITD thresholds increase with increasing reference ITD. Thresholds obtained with broadband stimuli are, as in the previous experiment, slightly below the lowest thresholds obtained with narrowband stimuli for the corresponding reference ITD. For the narrowband stimuli, the threshold increase is, however, not equal for all center frequencies. As a consequence, the range where the lowest thresholds lie shifts from between 750 and 1000 Hz for a reference ITD of 0 µs to between 300 and 500 Hz for a reference ITD of 600 µs. In other words, there is a change in the spectral weighting as a function of the reference ITD. Such a change is reflected in the frequency weights applied to the stimuli used in the modeling of the localization experiment that are shown in Fig. 3a.

**FIGURE 3**: (a) Weights $a_i$ applied in the localization model for the four localization conditions. (b) Representation of the stimuli as a function of frequency (left), matched ITDs (open symbols on the right, “data”) and model predictions (dark filled symbols on the right, “model”). Dotted lines represent the median value of the ITDs carried by the stimuli. Error bars represent the 95% confidence interval of the mean. Markers correspond to the same stimulus conditions in the two panels.

Figure 3a shows the weights ($a_i$) applied in the localization model (equation 8) and derived from the ITD detection thresholds measured as a function of the reference ITD in Experiment 2. As suggested by the thresholds in Fig. 2, the overall dependency of the weights on frequency is shifted towards lower frequencies for the largest ITDs.

Figure 3b shows data from Experiment 3. The four types of ITD distributions are shown on the left by the connected symbols, and the average matched ITDs are shown on the right by corresponding open symbols (data). Error bars represent the 95% confidence interval of the mean of the sampled population for each condition. The horizontal dotted lines represent the median ITD for each two groups of stimuli. Model predictions are shown on the right by the corresponding filled dark symbols (model). Markers correspond to the same stimulus conditions in the panels (a) and (b).

Considering the two high-ITD stimuli (upward and downward triangles), the two matched ITDs are very close to the median ITD, i.e., 400 µs. A two-tailed t-test shows that two matched ITDs are not significantly different ($\alpha=0.05$, $t = 1.326<2.131$). The model predictions (391 and 394 µs) account well for the proximity of the matched ITDs. This is mostly due to the fact that the peak of the weights $a_i$ for the two high-ITD configurations occur at around the same center frequency (598 Hz), for which the ITD carried by the stimuli is identical, i.e., 400 µs.

For the two low-ITD stimuli (square and circle), the two matched ITDs differ from one another and are roughly symmetrically positioned around the median ITD, i.e., 200 µs. A two-tailed t-test shows that the two matched ITDs are significantly different ($\alpha =0.05$, $t = 4.639>2.131$). The model predictions (177 and 204 µs) are, as the experimental matched ITDs, different from one another. This is mostly due to the fact that the weights $a_i$ for the two low-ITD configurations are peaking at different center frequencies. It should be noted that although the ITD in the center frequency with the largest weight contributes most, all frequency channels contribute to some extent in this model approach.
DISCUSSION AND CONCLUSION

A conceptual localization model accounting for the spectral integration of ITD information was proposed and has successfully predicted the localization of complex stimuli carrying multiple ITDs over a wide frequency range. The proposed model suggests that the localization of broadband stimuli is the result of an optimal weighting of the contribution of all frequency bands based on the variance of the internal noise that was assumed to limit both the detection and localization of ITDs.

One could, however, consider the alternative situation, where no spectral integration is assumed and only the frequency band in which the binaural information is most salient is used, and effectively ignoring all other bands. This alternative was tested with the proposed model. To assuming a single channel model, the weight of the most salient channel was set 1 and all other channel weights were set to 0. The predictions of such a single channel model on the lateralization data reported here were, for the two high-ITD distributions exactly equal to 400 µs, and for the two low-ITD distributions equal to 175 and 200 µs. Predictions were, therefore, reasonable though coarser when only one channel was taken into account.

As shown in the first ITD detection experiment, the reduction in interaural correlation had an effect on the detectability of the ITDs, although the dependency on frequency of the detection thresholds did not change significantly. One question left open is whether the combination of the effects of a reduction in interaural correlation and an increase in reference ITD could be accounted for by the present model. Further work could include the measurement of ITD detection thresholds as a function of the reference ITD with partially correlated stimuli. The current model could be extended if necessary and tested on an extension of the localization experiment reported here with partially correlated stimuli.

REFERENCES