MODELING THE CONTRALATERAL HRTF

CARLOS AVENDANO[1], RICHARD O. DUDA[2] AND V. RALPH ALGAZI[1]

[1]CIPIC University of California at Davis, USA
carlos@ece.ucdavis.edu, algazi@ece.ucdavis.edu
[2]Department of Electrical Engineering, San Jose State University, USA
rod@duda.org

We show how the contralateral head-related transfer function (HRTF) can be modeled by a simple transformation of the ipsilateral HRTF. Components of the transformation are based on a spherical model of the listener's head. Listening tests reveal that the average localization error introduced by the model is approximately 5°.

INTRODUCTION

Current HRTF-based systems for synthesizing spatial sound typically employ experimentally measured head-related impulse responses (HRIRs). As a sound source moves away from the median plane, the response becomes much more complex for the contralateral ear than for the ipsilateral ear. The high-frequency shadowing of the direct sound by the head is the reason for much of this complexity, because it reveals the presence of secondary waves of small magnitude that would otherwise be difficult to observe. The contralateral ear frequently exhibits non-minimum-phase HRTFs, bright spots in the time domain, complex interference patterns in the frequency domain, and an interaural level difference (ILD) that is a strong function of elevation [1, 3]. If one wishes to model the HRTF, it would appear that models for the contralateral ear may have to be much more complex than models for the ipsilateral ear.

However, we present results that indicate that it is not necessary to reproduce all of the features of the contralateral response to obtain high-quality synthesized binaural sound.

Near the median plane, the contralateral and ipsilateral responses are quite similar. Away from the median plane, simple modifications of the ipsilateral response due to the head shadow may be adequate to approximate the contralateral response. The simple model for the contralateral ear that we propose and evaluate here consists of (a) low-pass filtering the ipsilateral response to account for head shadow and (b) introducing an appropriate interaural time difference (ITD)[1]. The resulting model for the contralateral ear has the interesting feature that, except for scale, it yields an ILD magnitude spectrum that is independent of elevation. In listening tests, localization results obtained with this model were within approximately 5° of those produced by the measured contralateral HRTFs.

1. THE CONTRALATERAL HRTF

The model proposed in this paper was motivated by certain general properties observed in the contralateral HRTF and by its relationship and similarity with its ipsilateral counterpart.

These properties are illustrated in Fig. 1, which shows both the HRIRs (left column) and the magnitude of the HRTFs (right column) for both the ipsilateral and contralateral ears of a human subject. Each image shows the data as a function of elevation and either time or frequency for a section of the cone of confusion defined by a constant azimuth angle. The coordinate system used in this work is an interaural polar coordinate system, where the azimuth angle θ is the angle between the vector to the sound source and the vertical median plane, and varies from $-\frac{\pi}{2}$ (left) to $\frac{\pi}{2}$ (right). The elevation

[1] Throughout this paper, by "ITD" we mean the frequency-independent difference between the onset times for the left and right HRIRs.

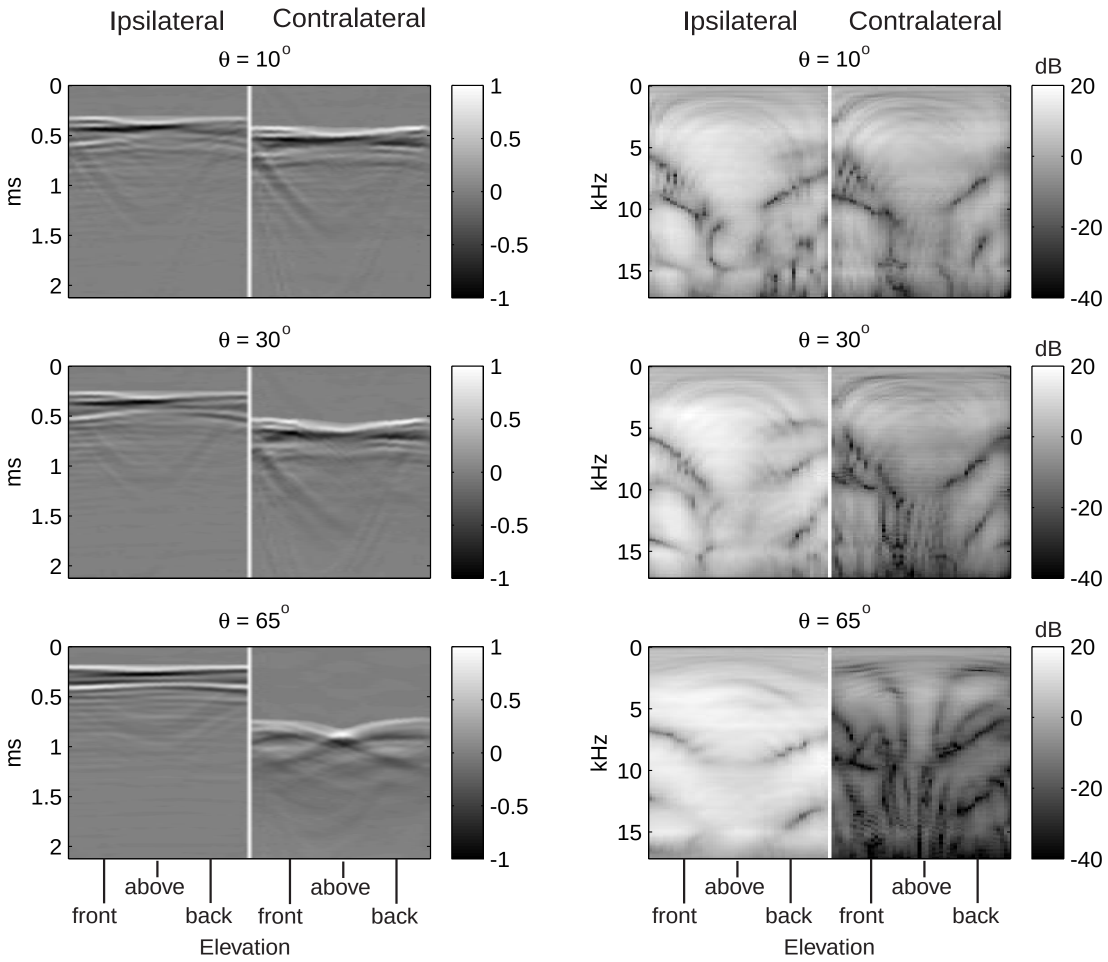

Figure 1: HRIRs (left) and corresponding HRTFs (right) as functions of elevation for three azimuth angles, $\theta = 10^\circ, 30^\circ, 65^\circ$. The magnitudes of the HRIRs have been normalized to make apparent the time domain structure. The HRTF magnitudes have not been normalized and show the energy level of the original data.

angle ϕ is the angle of the horizontal plane with the projection of the source into the median plane, and, starting directly below the subject, varies from $-\frac{\pi}{2}$ to $\frac{3}{2}\pi$.

Because the human head is basically symmetrical, the HRTF for both ears reaches its maximum similarity in the median plane. For a spherical head model [1, 8], the head shadow at the two ears is identical, the frequency response is essentially flat, and the ITD is exactly zero.

As azimuth increases, the contralateral HRTF continue to exhibit the same fine-structure features as the ipsilateral HRTF, but high frequencies start to be attenuated because of the head shadow (see Fig. 1, $\theta = 10^\circ$). On the ipsilateral side, the energy at high frequencies slowly increases, reaching a maximum boost of 6 dB for normal incidence ($\theta \approx 90^\circ$). To a first approximation, the difference between ipsilateral and contralateral HRTFs for small azimuth angles ($\theta < 20^\circ$) is an elevation-independent low-pass transfer function. In the time domain, the most apparent difference is a slight increase in ITD.

For larger azimuth angles ($25^\circ < \theta < 55^\circ$), the features of the contralateral ear do not match the ipsilateral features as closely (Fig. 1, $\theta = 30^\circ$). Above $3 - 4$ kHz, the energy level differences between both sides increase to more than 10 dB. In the time domain, an elevation-dependent variation of the onset of the contralateral HRIR is evident. There are two reasons for this occurrence on a “cone of confusion”: (a) the fact that the ears are not diametrically opposed, but are offset downward and backward, and (b) the fact that the head shape is more nearly ellipsoidal than spherical [5].

As we approach the interaural pole, e.g. $\theta > 60^\circ$, the complexity of the contralateral ear response, both in time and frequency, is far greater than the relatively simpler structure observed on the ipsilateral side (see third row in Fig. 1, $\theta = 65°$). In the time domain we can observe waves reaching the ear from the front and back of the head, creating complex interference patterns. The elevation dependence of the onset time of the contralateral HRIRs is even more pronounced, reaching a maximum deviation from the constant onset predicted by the spherical head model. However, above 4 kHz the average high-frequency ILD is now greater than 15 dB and the exact response at the shadowed ear becomes less critical.

2. THE MODEL

We now use the observations in the previous section to develop a model for the contralateral HRTF. We base the model on the ipsilateral response, and write it in the frequency domain as:

$$\widehat{H_c}(\omega, \theta, \phi) = F(\omega, \theta, \phi)\, H_i(\omega, \theta, \phi), \qquad (1)$$

where ω is angular frequency, $H_i(\omega, \theta, \phi)$ is the ipsilateral HRTF, $\widehat{H_c}(\omega, \theta, \phi)$ is the model of the contralateral HRTF, and $F(\omega, \theta, \phi)$ is a linear transfer function to be determined.[2] For any ipsilateral response with finite non-zero energy in the bandwidth of interest, we can find a transfer function that will approximate the contralateral response arbitrarily closely. Thus, the general form in (1) does not represent an interesting model of the contralateral HRTF.

Based on our previous observations we can greatly simplify the transformation and obtain an efficient model as follows. The transformation in (1) can be factored as

$$F(\omega, \theta, \phi) = e^{-jD(\theta,\phi)}\, S(\omega, \theta). \qquad (2)$$

The first factor in the right-hand side of (2) corresponds to the time delay necessary to preserve the ITD. The second factor in (2) is a function of frequency and azimuth and accounts for head shadow effects and other individual characteristics. The advantage of this decomposition is that each component can be computed independently based on simple geometrical models, whether derived from anthropometric measurements or otherwise approximated.

A simple expression for the head-shadow transformation can be derived from a spherical head model. If $S_i(\omega, \theta)$ and $S_c(\omega, \theta)$ are the sphere HRTFs for the ipsilateral and contralateral sides respectively, then the head shadow transformation can be written as

$$S(\omega, \theta) = \frac{S_c(\omega, \theta)}{S_i(\omega, \theta)}. \qquad (3)$$

The spherical head model used in our implementation is based on an infinite series solution to the diffraction of sound around a sphere [4]. The model is a function of source range and sphere radius, and can be scaled according to the listener's anatomy.

The function $D(\theta, \phi)$ introduces the appropriate onset time to the contralateral HRTF model. Notice that it is a function of both azimuth and elevation according to the pattern observed in the data. This onset time correction can be simply computed as

$$D(\theta, \phi) = \widehat{T}_c(\theta, \phi) - T_i(\theta, \phi), \qquad (4)$$

where $T_i(\theta, \phi)$ is the onset time of the ipsilateral response and $\widehat{T}_c(\theta, \phi)$ is a model for the onset time on the contralateral side. In this model, it is important to control the ITD to avoid the introduction of azimuth errors. The best results are obtained by using the actual, measured ITD. However, for most purposes, adequate results may be obtained using an ITD based on a spherical or elliptical-head model customized to individual listeners. Such a geometric model for the contralateral onset time has recently been proposed in [5], where an elliptical head with offset ears is used to approximate the elevation-dependent variation of ITD on a cone of confusion that is observed in human subjects.

An additional gain correction was introduced to correct for the fact that the HRIRs were measured at a range of 1 m from the center of the head, which is close enough so that there is a small but noticeable difference between the distances to each ear. The exact low-frequency ILD is known for a source located at a given distance from an ideal sphere [4]. This theoretical solution was used to compute an empirical gain correction function $\alpha(\theta)$, dependent only on azimuth, and based on the power in the frequency interval between 200 Hz and 600 Hz. The correction of the spherical head results was within 2 dB near the poles, decreasing to zero towards the median plane.

The model that we have developed for the contralateral HRTF is therefore completely based on a transformation of the experimentally measured ipsilateral HRTF. The transformation (2) involves two functions that are determined by simple geometric models based on anthropometry.

[2] Note that we have assumed that the HRTF is independent of the range r to the source. This is a good approximation for a point sound source in the far field. However, we subsequently discovered that there is a small range-dependent effect in our experimental data, which was measured at a range of 1 m. This is accounted for ahead.

2.1. Implementation

The transformation used in the model was performed in the time domain. The head-shadow correction was implemented as a minimum-phase FIR filter derived from the magnitude responses obtained from the spherical head model. The two parameters of this model — source range and head radius — were obtained from measurements of our data collection apparatus and by estimating the average radius of the head for each subject.

The fractional time delay in $D(\theta, \phi)$ was implemented by upsampling the ipsilateral response by a factor of ten, shifting by the closest integer at the higher sampling rate, and downsampling to the original sampling rate, thus obtaining a resolution of a tenth of a sample. For the present evaluations the original ITD of the measured data was used.

3. MODEL EVALUATION

The model was evaluated in two different ways. First, an objective measure was used to determine the spatial distribution of errors between the measured and modeled contralateral HRTF as a function of source location. The second evaluation involved perceptual comparisons of localization errors between the measured HRTF and the model. We first describe the data measurement procedure, and then consider each evaluation in turn.

3.1. Measurement Procedure

The evaluation of the model made use of the experimentally measured HRIRs of three different subjects. The data were measured at 1250 locations in space, with elevation increments of $\Delta\phi = 5.625°$ for a range $-45° \leq \phi \leq 231°$ and at 25 different azimuth angles with a 5° spacing in the front, increasing towards the interaural poles.

The HRIRs were measured using the blocked-ear-canal technique by attaching the probe tubes of two Etymōtic ER-7C microphones to plastic ear plugs, which were then inserted into the subjects' ear canals. The impulse responses were obtained using Golay codes (Crystal River Snapshot^{TM} system), played through Bose Acoustimass Cube speakers mounted on 1-m-radius hoop that was rotated about the subject's interaural axis. The sampling rate of the measurements was 44.1 kHz. The resulting impulse responses were windowed and truncated to remove room reflections, and free-field equalized to compensate for the speaker and microphone transfer functions.

Headphone compensation filters were also computed for each subject. The same set of headphones (AKG K240-DF) was used during all listening sessions, and the individualized compensation filters (implemented digitally) were always applied to the stimuli before presentation.

3.2. Objective Error Measure

An objective measure based on the frequency selectivity of the hearing mechanism was used to evaluate the model. The data and the model magnitude responses were smoothed using critical band shaped filters [9]. The normalized mean squared error between the smoothed spectra was computed for all azimuths and elevations. The error was computed as

$$E(\theta, \phi) = \frac{\sum_\omega [B_c(\omega, \theta, \phi) - \widehat{B_c}(\omega, \theta, \phi)]^2}{\sum_\omega B_c(\omega, \theta, \phi)^2}, \qquad (5)$$

where $B_c(\omega, \theta, \phi)$ is the smoothed spectrum of the measured contralateral HRTF, $\widehat{B_c}(\omega, \theta, \phi)$ is the model, and the summation is over the discrete frequency points from 0 Hz to 22.05 kHz.

Fig. 2 shows the error on a dB scale, computed as

$$E(\theta, \phi)_{dB} = 10 \log_{10}[1 + E(\theta, \phi)], \qquad (6)$$

such that a perfect model would result in a 0-dB error. The error is shown as a function of location for each of the three subjects. Notice that the model introduces more error toward the interaural poles, and low (below front) and high (below back) elevations. Overall, the error is lower for subject S2 and higher for subject S1. As expected, the error is lowest near the median plane in all cases.

This objective error is based solely on the spectral mismatch between measurements and model. Since it includes a perceptual weighting, it may indicate the regions of space where the localization errors caused by the model may occur. No evaluation of such a correlation has yet been performed. Note that since the error was computed in the spectral domain, time domain artifacts (in particular, ITD errors) may also result in localization errors. Because we used the onset time of the measured HRTF in the present evaluation, there was no need for an objective evaluation in the time domain.

3.3. Perceptual Tests

There are two ways in which models may differ from the measured HRTF: (a) they may produce perceptible changes in timber, and (b) they may introduce localization errors.

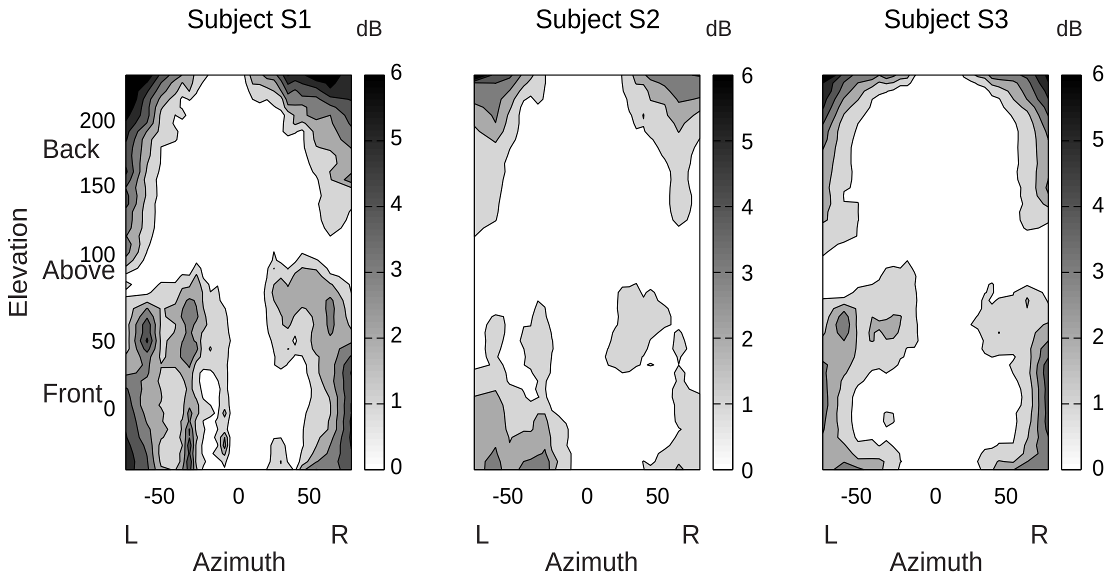

Figure 2: Energy-normalized mean squared error as a function of location for the three subjects.

Preliminary listening sessions with sound sources moving in space showed that the model resulted in a good sound quality, very similar to the sound perceived for the measured HRTF. The aim of the following perceptual tests was to evaluate more rigorously the effectiveness of the model in providing good sound localization. A static localization test was performed for sound sources widely distributed in space.

A two-alternative forced-choice AB-CD paradigm was used to quantify this localization error. The task was to compare the perceived "distance" between the measured and the modeled HRTF to the distance between the measured HRTF and a perturbed version of itself. The perturbation applied was a differential change in the azimuth coordinate, $\Delta\theta$. The difference between the measured HRTF and the perturbed HRTF was used to set a reference for the definition of "distance" with which the model was evaluated.

The measured HRTF was presented as stimulus A and C. The model under test was presented as stimulus B, and the a perturbed version of the measured HRTF was presented as D. The order of presentation was randomized to obtain four different sequences, i.e. AB-CD, CD-AB, BA-DC and DC-BA. Subjects were forced to choose the pair (first or second) that was "closest", but were allowed to listen to the same stimulus sequence for an unlimited number of repetitions before deciding.

3.4. Experimental Conditions

A total of $N = 100$ randomly selected locations were tested for each differential azimuth change and for each subject. Five perturbation angles were tested, $\Delta\theta = 0^\circ, 2.5^\circ, 5^\circ,\ 7.5^\circ$, and 10°. Azimuth perturbations were randomly assigned positive or negative values. When the azimuth of the perturbed stimulus did not coincide with our data sampling grid, $\Delta\theta$ was approximated by increasing or decreasing the ITD of the closest HRTF pair. Three subjects with no significant hearing impairment participated on the test. Separate listening sessions were arranged for each perturbation angle.

The auditory source stimulus was a 500-ms white Gaussian noise sequence, 100% amplitude modulated at 40 Hz, with a 500-ms gap between successive stimuli. Amplitude modulation was employed to increase the transient cues that are important for accurate localization [7]; the 40-Hz modulation frequency was chosen so that even the low-frequency auditory channels, which have longer time constants than the high-frequency channels, would receive significant onset information [10]. This stimulus was convolved with the left and right HRIRs to produce the binaural stimulus, which was presented to the subject through compensated headphones.

3.5. Results

If the model were perfect and the perturbation were $\Delta\theta = 0^\circ$, one would expect a percentage of responses favorable to the model (success rate) of 50%. Since the model introduces errors, the success rate is less than 50%, but increases as $\Delta\theta$ increases. The accuracy of the model is measured by the perturbation needed to achieve a 50% success rate. The results of the perceptual evaluation are summarized in Fig. 3. For all three subjects we observe that the 50% success rate point indicates that on average the model performs within 5°-6° of the real HRIR.

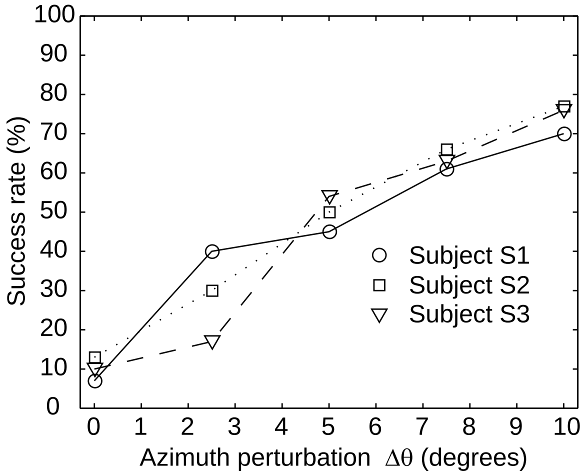

Figure 3: Results of perceptual tests. The ordinate is the percentage of responses favorable to the model. The abscissa is the differential azimuth change used as perturbation.

4. DISCUSSION AND CONCLUSIONS

The proposed contralateral model is a simple transformation of the ipsilateral response based on a spherical head model plus appropriate additional time delay. We have shown that this model introduces an average localization error of about 5°. If this error is acceptable, it allows modeling efforts to focus on the ipsilateral ear, which has relatively simpler characteristics.

As we observed earlier, it is known that the true ILD varies with elevation [3]. Because the ILD using the model is independent of elevation, spectral error is inevitable, and, as Fig. 2 illustrates, the objective error increases as the source moves away from the median plane. Additional perceptual tests are needed to quantify the spatial sensitivity of our results. Informal observations indicate that even the largest spatial displacements introduced by the model are on the order of 10° to 15°. While this is fairly large for a differential test, it is within normal limits of absolute localization accuracy [2].

5. ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under grant NSF IRI-96-19339. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the National Science Foundation. We would also like to thank Dennis Thompson for his help in collecting the data used in this work, and Pierre Divenyi for his guidance on perceptual testing procedures.

REFERENCES

[1] Brown C.P. and Duda R.O. 1998. A Structural Model for Binaural Sound Synthesis. IEEE Transactions on Speech and Audio Processing, 6, pp. 476-488.

[2] Carlile S. 1996. Auditory Space. In Virtual Auditory Space: Generation and Applications, Carlile S. ed. R. G. Landes, Austin, TX.

[3] Duda R.O. 1997. Elevation Dependence of the Interaural Transfer Function. In Binaural and Spatial Hearing in Real and Virtual Environments, Gilkey R.H. and Anderson T.R. eds. Lawrence Erlbaum Associates, New Jersey.

[4] Duda R.O. and Martens W. 1998. Range Dependence of the Response of a Spherical Head Model, J. Acoust. Soc. Am., 104, 5, pp. 3048-3058.

[5] Duda R.O., Avendano C. and Algazi V.R. 1999. An Adaptable Ellipsoidal Head Model for the Interaural Time Difference. Proc. 1999 ICASSP, Phoenix, Arizona, March 15-19, 1999.

[6] Green D.M. and Swets J.A. 1966. Signal Detection and Psychophysics. R. E. Krieger Pub. Co., Huntington, NY.

[7] Hafter E.R. and Buell T.N. 1990. Restarting the Adapted Binaural System. J. Acoust. Soc. Am., 88, pp. 806-812.

[8] Kuhn G. F. 1977. Model for the Interaural Time Difference in the Azimuthal Plane, J. Acoust. Soc. Am., 62, pp. 157-167.

[9] Schroeder M.R. 1977. Recognition of Complex Acoustic Signals. Life Sciences Research Report 5, p. 324, Bullock T. H. ed. Abakon Verlag, Berlin.

[10] van den Brink W.A. and Houtgast T. 1990. Spectro-Temporal Integration in Signal Detection. J. Acoust. Soc. Am., 88, pp. 1703-1711.