THE SENSITIVITY MATRIX FOR A SPECTRO-TEMPORAL AUDITORY MODEL

Jan H. Plasberg, David Y. Zhao and W. Bastiaan Kleijn

KTH (Royal Institute of Technology)
Department of Signals, Sensors and Systems
10044 Stockholm, Sweden
{ Jan.Plasberg, David.Zhao, Bastiaan.Kleijn }@s3.kth.se

ABSTRACT

Perceptually optimal processing of speech and audio signals demands distortion measures that are based on sophisticated auditory models. High-rate theory can simplify these models by means of a sensitivity matrix. We present a method to derive the sensitivity matrix for distortion measures based on spectro-temporal auditory models under the assumption of small errors. This method is applied to an example auditory model and the region of validity of the approximation as well as a way to analyze the characteristics of the model with subspace methods are discussed.

1. INTRODUCTION

In recent years, the development and usage of perceptual models within the various fields of speech and audio processing has become increasingly important. Not only speech/audio coding, but also the fields of speech enhancement, noise reduction, speech recognition, audio restauration profit if properties of human hearing are taken into account.

In signal processing we usually try to approach the problem of close-to perceptually optimal speech and audio processing by the development of simplified and sometimes ad hoc solutions, most of them only approximating the behavior of simultaneous masking. While highly sophisticated models have been developed to mimic the signal processing of the human auditory system (e.g., [1, 2, 3, 4]), they are very complex. Thus, only a few of these models have been put into the context of a distortion measure as needed in coding or speech/audio quality assessment (e.g., [4, 5]), while other models even lack the proof of predicting human auditory masking behavior properly. So far, objective tools have been missing to evaluate the reliability of a distortion value defined on the ouput of an auditory model. Especially in coding we have the problem that we have to use analysis-by-synthesis techniques to find the perceptually optimal code. Thus, a strong need exists for simple and yet accurate approximations of perceptual distortion measures that facilitate practical coding schemes.

Several authors have promoted the idea that the mathematical framework to solve these problems can be found in high-rate theory [6, 7, 8, 9, 10]. Under the assumption of small errors, it is often possible to describe systems analytically, and the results obtained often seem to hold even for low and thus practical rates. In this paper, we follow the work in [7] to derive a high-rate solution to the problems above. We show that for very low distortions a perceptual distortion measure based on an auditory model can be approximated by means of a single matrix multiplication. We can therefore analyze the signal dependent model behavior with standard tools from linear algebra.

Section 2 provides a short overview over the theory used to obtain the sensitivity matrix, a new method to obtain a masking threshold from the sensitivity matrix and how to analyze the perceptual distortion measure by means of eigenvalue decomposition. In section 3 we show how to apply this theory to an example auditory model. We evaluate the validity of the high-rate approximations in section 4 and show to which extent it is possible to linearize a non-linear perceptual distortion measure.

This work was partially funded by the E.U. grant IST-2001-34095 (ARDOR) and the Swedish Research Council (fund 271-2000-784).

2. THEORY

In this section we provide a brief introduction to the theory needed to derive a sensitivity matrix for a non-difference distortion measure. Let $\mathbf{y} = [y_0, \ldots, y_n, \ldots, y_{N-1}]^H$ be an N-dimensional vector of source samples, $\hat{\mathbf{y}} = Q(\mathbf{y})$ a vector-quantized version of this vector and let D be the expected value of some distortion measure $d(\mathbf{y}, \hat{\mathbf{y}})$,

$$D = E[d(\mathbf{y}, \hat{\mathbf{y}})], \tag{1}$$

under the conditions $d(\mathbf{y}, \hat{\mathbf{y}})$ is continuous and has continuous differentiables and $d(\mathbf{y}, \hat{\mathbf{y}}) \geq 0$ with equality iff $\hat{\mathbf{y}} = \mathbf{y}$. Then a Taylor series expansion of $d(\mathbf{y}, \hat{\mathbf{y}})$ around $\hat{\mathbf{y}} = \mathbf{y}$, leaving out terms vanishing for $D \to 0$ [7], results in

$$d(\mathbf{y}, \hat{\mathbf{y}}) \approx \frac{1}{2}(\mathbf{y} - \hat{\mathbf{y}})^H \mathbf{D}_y(\hat{\mathbf{y}})(\mathbf{y} - \hat{\mathbf{y}}), \tag{2}$$

where $\mathbf{D}_y(\hat{\mathbf{y}})$ is an N by N dimensional matrix, the so-called sensitivity matrix, with the i, jth element defined by

$$\left[D_y(\hat{\mathbf{y}})\right]_{i,j} = \left.\frac{\partial^2 d(\mathbf{y}, \hat{\mathbf{y}})}{\partial \hat{y}_i \partial \hat{y}_j}\right|_{\hat{\mathbf{y}} = \mathbf{y}}. \tag{3}$$

It can be shown that a high-rate vector quantizer minimizing the right-hand side of (2) will have the same centroid density, Voronoi region shapes and performance as the quantizer minimizing the original distortion measure.

Suppose we want to quantize other dimension-N representations $\mathbf{z}$ of the vectors $\mathbf{y}$, where $\mathbf{z} = \mathbf{z}(\mathbf{y})$ is a continuous one-to-one vector function, such that the inverse transform $\mathbf{y} = \mathbf{y}(\mathbf{z})$ exists. The i, jth element of the sensitivity matrix in the domain of $\mathbf{z}$ is then given by (compare (3))

$$[\mathbf{D}_z(\mathbf{z})]_{i,j} = \left.\frac{\partial^2 d(\mathbf{y}(\mathbf{z}), \mathbf{y}(\hat{\mathbf{z}}))}{\partial \hat{z}_i \partial \hat{z}_j}\right|_{\hat{\mathbf{z}} = \mathbf{z}} \tag{4}$$

and straightforward application of the chain rule yields

$$\mathbf{D}_z(\mathbf{z}) = \mathbf{J}_z^H(\mathbf{z}) \mathbf{D}_y(\mathbf{y}(\mathbf{z})) \mathbf{J}_z(\mathbf{z}), \tag{5}$$

where $\mathbf{J}_z(\mathbf{z})$ is the N by N Jacobian matrix of the transform $\mathbf{y}(\mathbf{z})$ with the i, jth element defined by

$$[\mathbf{J}_z(\mathbf{z})]_{i,j} = \left.\frac{\partial y_i(\mathbf{z})}{\partial \hat{z}_j}\right|_{\hat{\mathbf{z}} = \mathbf{z}}. \tag{6}$$

Minimizing

$$\hat{d}(\mathbf{y}(\mathbf{z}), \mathbf{y}(\hat{\mathbf{z}})) = \frac{1}{2}(\mathbf{z} - \hat{\mathbf{z}})^H \mathbf{D}_z(\mathbf{z})(\mathbf{z} - \hat{\mathbf{z}}) \tag{7}$$

yields the same performance in high-rate as minimizing (2). In the same manner we obtain the sensitivity matrix for the vector $\mathbf{z}$ in the frequency domain $\mathbf{Z} = \mathbf{F}\mathbf{z}$ with the Fourier transform $\mathbf{F}$ in matrix notation,

$$\mathbf{D}_Z(\mathbf{Z}) \quad = \quad \mathbf{F}\mathbf{D}_z(\mathbf{z})\mathbf{F}^H. \tag{8}$$

Let us denote an error vector in perceptual domain as $\mathbf{r}_a$. Our high-rate distortion measure assumes locally quadratic distortions in the perceptual domain, that is $D = \mathbf{r}_a^H \mathbf{r}_a$. All perceptual error vectors resulting in the same distortion are located on a sphere around the original value. Using this assumption and (8), we can derive the shape of a masking curve in frequency domain from the sensitivity matrix, keeping the perceptual distortion at a constant level

$$\begin{aligned} 1 &= \mathbf{r}_a^H \mathbf{r}_a && (9) \\ &= \mathbf{r}_Z^H \mathbf{D}_Z(\mathbf{Z})\mathbf{r}_Z, && (10) \end{aligned}$$

where $\mathbf{r}_Z$ is an error vector in frequency domain.

Except for a scaling factor, the masking curve at frequency i is the gain g_i that results in

$$g_i |\mathbf{u}_i^H \mathbf{D}_Z(\mathbf{Z})\mathbf{u}_i| = 1, \tag{11}$$

with the unit vector $\mathbf{u}_i$. Solving (11) we get

$$g_i = \frac{1}{|[\mathbf{D}_Z(\mathbf{Z})]_{i,i}|}. \tag{12}$$

This clearly shows the weakness of a two-tone masking curve: it assumes a diagonal weighting matrix and does therefore not take into consideration all available perceptual information.

In case the sensitivity matrix $\mathbf{D}$ (in any domain) is symmetric, an eigenvalue decomposition yields $\mathbf{D} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$, where $\mathbf{\Lambda} = \mathrm{diag}[\lambda_0, \ldots, \lambda_{N-1}]$ contains the eigenvalues and $\mathbf{Q} = [\mathbf{q}_0, \ldots, \mathbf{q}_{N-1}]$ the corresponding orthonormal eigenvectors as columns. We will see later that $\mathbf{D}$ for our measure is indeed symmetric. Note, that for an input error vector $\mathbf{r} = \mathbf{q}_i$ the distortion value is $\frac{1}{2}\mathbf{q}_i^H \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H \mathbf{q}_i = \frac{\lambda_i}{2}$ (from (7) and the orthonormality of $\mathbf{Q}$), i.e. the eigenvalue decomposition tells us in a sorted manner, in which direction $\mathbf{q}_i$ of the input the high-rate distortion measure is most sensitive.

Let us define the ε*-space* as the space spanned by the set $\mathbf{Q}_\varepsilon$ of all eigenvectors $\mathbf{q}_i$ corresponding to very small eigenvalues λ_i

$$\mathbf{Q}_\varepsilon = \{\mathbf{q}_i : 0 < \lambda_i \le \varepsilon, \forall \varepsilon \in \mathbb{R}^+\}. \tag{13}$$

Then, quantization error vectors should preferably lie in the ε*-space*, as all vectors in this space correspond to a low sensitivity of the measure and are therefore most likely to be masked. An elegant way to analyse the ε*-space* is to form a matrix $\mathbf{H}_\varepsilon$ containing all vectors from the set $\mathbf{Q}_\varepsilon$ as columns and use it to project a vector $\mathbf{n}$ of white noise onto the ε*-space* according to

$$\mathbf{n}_\varepsilon \quad = \quad \mathbf{H}_\varepsilon(\mathbf{H}_\varepsilon^H \mathbf{H}_\varepsilon)^{-1}\mathbf{H}_\varepsilon^H \mathbf{n}. \tag{14}$$

where $\mathbf{n}_\varepsilon$ contains all noise components that lie in the ε*-space*.

3. APPLICATION TO AN EXAMPLE AUDITORY MODEL

In this section we explain how to apply the theory above to a distortion measure based on an example spectro-temporal perceptual model. The model we selected was proposed by Dau et al. in [2] and was exhaustively tuned and tested against experimental data in [3]. In this paper it is referred to as the Dau auditory model and the basic system including a definition of a distortion measure on its output is described in 3.1. In 3.2 we show how to linearize the non-linear stages of the Dau model for small errors, using general techniques that can be applied to similar auditory models.

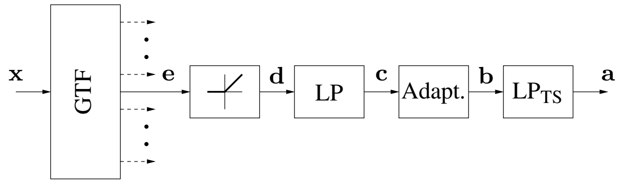

Figure 1: The single stages of a channel of the Dau auditory model.

3.1 The Dau auditory model

One channel of the Dau auditory model is shown in Fig. 1. The input signal $\mathbf{x}$ is filtered by a gammatone filterbank (GTF) that mimics the behavior of the basilar membrane. Our implementation has one gammatone filter per ERB, resulting in $M_c = 29$ channels for 8 kHz sampling frequency. The gammatone filter is followed by a half-wave rectifier and a low-pass filter with cut-off frequency 1 kHz. Then, the signal is processed by five so-called adaptation loops, which, after applying a positive threshold on the signal $\mathbf{c}$ related to the threshold of hearing, perform a signal-dependent gain compression that models the non-simultaneous masking behavior of the human ear. These adaptation loops converge to a logarithmic transform for stationary input signals $\mathbf{x}$, while rapid changes in the input signal are transformed linearly. The last stage $\mathrm{LP_{TS}}$ is a first-order low-pass filter for temporal smoothing with a very low cut-off frequency of 8 Hz, yielding the internal representation $\mathbf{a}(\mathbf{x})$ per channel. If we index the channels by $m < M_c$, $m \in \mathbb{N}$ and denote the internal representation per channel by $\mathbf{a}_m$, the complete internal representation is a matrix $\mathbf{A}(\mathbf{x}) = [\mathbf{a}_0, \ldots, \mathbf{a}_m, \ldots, \mathbf{a}_{M_c-1}]^H$.

To facilitate block-by-block distortion measurement of N-dimensional signal vectors $\mathbf{x}$ we have to pay special attention to the ringing of the non-linear internal processing. The model input has to be constructed by M-dimensional, $M > N$, signal vectors $\mathbf{x}'$ and $\hat{\mathbf{x}}'$, where $\hat{\mathbf{x}}'$ contains distorted signal values only in the first N elements. This way the model is provided with the (clean) signal future and the signal future dependent perceptual error can ring out correctly.

We can define a simple distortion measure on the matrices $\mathbf{A}(\mathbf{x}')$ and $\hat{\mathbf{A}}(\hat{\mathbf{x}}')$

$$d_{\mathrm{Dau}}(\mathbf{x}, \hat{\mathbf{x}}) \quad = \quad \sum_{m=0}^{M_c-1}\sum_{n=0}^{M-1} |A_{m,n}(\mathbf{x}') - \hat{A}_{m,n}(\hat{\mathbf{x}}')|^2, \tag{15}$$

assuming synchrony in $\mathbf{x}'$ and $\hat{\mathbf{x}}'$. For stationary input signals, where the adaptation loops correspond to a logarithmic transform, (15) will have a meaning similar to a log spectral distortion measure on critical bands. We assume that measuring the Euclidean distance in the perceptual domain per channel is a good measure for very low distortions even for non-stationary input signals. Unlike the work in [5], we assume in this study that the model is accurate and do not apply any additional weighting on the distortion measure per channel.

3.2 The sensitivity matrix for the Dau auditory model

Finding a sensitivity matrix for the Dau model is not practical from a straightforward application of (3), because the function $d(\mathbf{x}, \hat{\mathbf{x}})$ cannot be described in a simple fashion. However, we can use the fact that, for each channel of the model, the distortion measure on the internal representation $\mathbf{a}_m$ is already of quadratic form (15),

$$d_{\mathrm{Dau}}^{(m)}(\mathbf{a}_m, \hat{\mathbf{a}}_m) \; = \; \frac{1}{2}(\mathbf{a}_m - \hat{\mathbf{a}}_m)^H \mathbf{D}_a^{(m)}(\mathbf{a}_m)(\mathbf{a}_m - \hat{\mathbf{a}}_m), \tag{16}$$

with $\mathbf{D}_a^{(m)}(\mathbf{a}_m) = 2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and $d_{\mathrm{Dau}}(\mathbf{x}, \hat{\mathbf{x}}) = \sum_{m=0}^{M_c-1} d_{\mathrm{Dau}}^{(m)}(\mathbf{a}_m, \hat{\mathbf{a}}_m)$. Using (16) and by recursive application of the chain rule (5), we can express the sensitivity matrix per channel in the signal domain as

$$\mathbf{D}_x^{(m)}(\mathbf{x}) \quad = \quad 2\left(\mathbf{W}_x^{(m)}\right)^H \mathbf{W}_x^{(m)}. \qquad (17)$$

$\mathbf{W}_x^{(m)}$ is an M by N matrix, $M > N$, containing the first N columns of the product of M by M Jacobian matrices for the single stages $\mathbf{J}_b^{(m)}\mathbf{J}_c^{(m)}\mathbf{J}_d^{(m)}\mathbf{J}_e^{(m)}\mathbf{J}_x^{(m)}$, and

$$\mathbf{J}_c^{(m)} \quad = \quad \mathbf{J}_{c^{(5)}}^{(m)}\mathbf{J}_{c^{(4)}}^{(m)}\mathbf{J}_{c^{(3)}}^{(m)}\mathbf{J}_{c^{(2)}}^{(m)}\mathbf{J}_{c^{(1)}}^{(m)}\mathbf{J}_{c^{(0)}}^{(m)}, \qquad (18)$$

where $\mathbf{J}_{c^{(0)}}^{(m)}$ denotes the Jacobian for the signal threshold on signal $\mathbf{c}^{(0)} = \mathbf{c}$ and $\mathbf{J}_{c^{(l)}}^{(m)}$ the Jacobian for adaptation loop l with input signal $\mathbf{c}^{(l)}$, $l \in \{1,2,3,4,5\}$. M needs to be much larger than N to ensure that every sample of the error vector $(\mathbf{x}-\hat{\mathbf{x}})$ is transformed into perceptual domain without boundary effects. Indeed, the matrix $\mathbf{W}_x^{(m)}$ then represents a convolution matrix for a time-varying linear filter and for small errors $(\mathbf{x}-\hat{\mathbf{x}})$ the product $\mathbf{W}_x^{(m)}(\mathbf{x}-\hat{\mathbf{x}})$ resembles the changes $(\mathbf{a}_m - \hat{\mathbf{a}}_m)$ in the output of the channel m. Finally, we obtain a sensitivity matrix for the entire model from (15) and (17)

$$\mathbf{D}_x(\mathbf{x}) = 2\sum_{m=0}^{M_c-1}\left(\mathbf{W}_x^{(m)}\right)^H \mathbf{W}_x^{(m)}. \qquad (19)$$

All Jacobians for the linear filters of the model ($\mathbf{J}_x$, $\mathbf{J}_d$ and $\mathbf{J}_b$) are directly given by the lower triangular Toeplitz matrix $\mathbf{H}$ with as first column the filter impulse response h_n, $n = 0\ldots M-1$. Note that for the low-frequency basilar membrane filters, as well as the temporal smoothing filters, a very long block length M is needed to minimize aliasing.

For an input signal sample z_n at time n, the half-wave rectifier as well as all other thresholds contained in the model can be defined as a threshold function $T_n(z_n,\Delta) = r(z_n-\Delta)+\Delta$ with the ramp function $r(t)$. $T_n(z_n,\Delta)$ is continuous and differentiable almost everywhere. Taking the partial derivative with respect to z_m, we get:

$$\frac{\partial T_n(z_n,\Delta)}{\partial z_m} = \begin{cases} u(z_n-\Delta) & \text{for } m = n \\ 0 & \text{for } m \neq n, \end{cases} \qquad (20)$$

with the unit step function $u(t)$. Thus, the Jacobian of the half-wave rectifier is a diagonal matrix with elements

$$[\mathbf{J}_e]_{n,n} = u(z_n). \qquad (21)$$

The derivation of the Jacobian $\mathbf{J}_{c^{(l)}}$ for the adaptation loop l is more complicated. Fig. 2 shows the single adaptation loop l in discrete-time. It implements a division of the positive input signal z_n by the first order low-pass filtered version f_n of the output signal y_n, with coefficients $a^{(l)} = e^{-1/(f_s\tau^{(l)})}$ and $b^{(l)} = 1-a^{(l)}$. The denominator f_n is limited by an absolute threshold $f_{\min}^{(l)}$ to avoid a division by zero. The single adaptation loops only differ in the time constants $\tau^{(l)}$ of the low-pass filters and the thresholds $f_{\min}^{(l)}$, $l \in \{1,2,3,4,5\}$ [2]. From Fig. 2, the difference equation for this system is

$$z_n y_{n-1} \quad = \quad T_n\left(\beta_n^{(l)}, f_{\min}^{(l)} y_n y_{n-1}\right), \qquad (22)$$

with

$$\beta_n^{(l)} = (1-a^{(l)})y_n y_{n-1}^2 + a^{(l)} y_n z_{n-1}. \qquad (23)$$

Taking the derivatives $\frac{\partial}{\partial z_{n-k}}$, $k < M-1$, $k \in \mathbb{N}$ on both sides of (22) and sorting the single terms yields the diagonals of the Jacobian for the adaptation loop. Note that the loop is causal and therefore

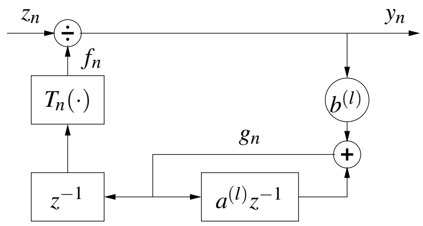

Figure 2: One adaptation loop in discrete-time implementation.

$\frac{\partial y_n}{\partial z_{n+k}} = 0$ for all n, resulting in a lower triangular Jacobian. After some algebra, the elements of the Jacobian for adaptation loop l are given by

$$\left[\mathbf{J}_{c^{(l)}}\right]_{n,n} = \xi_n^{(l)}\frac{y_n}{z_n} \qquad (24)$$

$$\left[\mathbf{J}_{c^{(l)}}\right]_{n,n-1} = \xi_n^{(l)}\left[\mathbf{J}_{c^{(l)}}\right]_{n,n}\left(\left[\mathbf{J}_{c^{(l)}}\right]_{n-1,n-1}\gamma_n^{(l)} - \frac{a^{(l)}y_n}{y_{n-1}}\right) \qquad (25)$$

$$\left[\mathbf{J}_{c^{(l)}}\right]_{n,n-k} = \xi_n^{(l)}\left[\mathbf{J}_{c^{(l)}}\right]_{n,n}\left[\mathbf{J}_{c^{(l)}}\right]_{n-1,n-k}\gamma_n^{(l)}, \qquad (26)$$

for $1 < k < M$ with

$$\xi_n^{(l)} = u\left(\beta_n^{(l)} - f_{\min}^{(l)} y_n y_{n-1}\right), \qquad (27)$$

and

$$\gamma_n^{(l)} = \frac{z_n}{y_{n-1}} - 2(1-a^{(l)})y_n. \qquad (28)$$

If (24) is calculated first, (25) second and then (26) with increasing k, the right hand sides of (25) and (26) only contain previously calculated values. Note that evaluating this Jacobian at the point $\hat{\mathbf{z}} = \mathbf{z}$ corresponds to reading $\mathbf{z}$ and $\mathbf{y}$ from the model when run on the original signal $\mathbf{x}'$ and then applying (24)-(28). With the same arguments as in section 3.1, the model needs to be run once on an M-dimensional signal vector to obtain an N by N sensitivity matrix.

4. RESULTS

We tested the above described distortion measures with blocks of narrow band music signals and artificial test signals. The dimension of the Jacobians for all experiments was $M = 800$. The first artificial signal consisted of a single sinusoid at 458 Hz, the center frequency of the eleventh gammatone filter of our model, with an additive white Gaussian noise floor at either -20 dB or -35 dB. This signal was used to test the simultaneous masking properties of our distortion measure. The second artificial signal was used for non-simultaneous masking experiments and consists of a white noise blip of 6.25 ms duration, with an additive white Gaussian noise floor at -35 dB. First we comment on the validity of the high-rate approximations and then show a few results obtained with the analysis techniques described in section 2.

4.1 Range of the Linearized Model

Fig. 3 a) shows the true distortion from the Dau auditory model versus the estimated value obtained with the sensitivity matrix, averaged over 200-sample blocks of 4 seconds of narrow-band music with i.i.d. Gaussian noise at different SNRs. The difference is below 3 dB above an SNR of 30 dB, and the approximation becomes exact at higher SNRs. Fig. 3 b) shows a very high accuracy in the same experiment for a simplified model without the adaptation loops and the temporal smoothing filter. The results of this figure indicate that the discrepancies in Fig. 3 a) are mainly due to the non-linearities in the adaptation loops, since for large block lengths the linear temporal smoothing filter is described exactly by a Jacobian matrix.

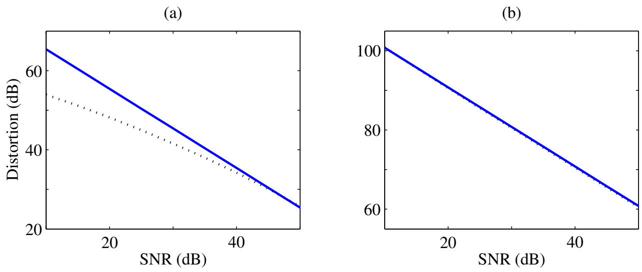

Figure 3: True (dotted line) and estimated (solid line) distortion values for narrow-band music signals with additive white gaussian noise for different input SNR; (a) full Dau model, (b) Dau model without adaptation loops and temporal smoothing.

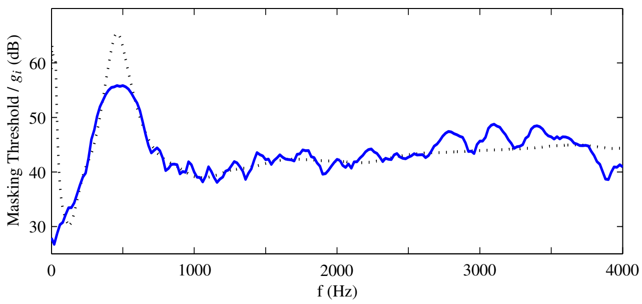

Figure 4: A masking curve shape obtained from (12) (solid line) compared to a spectral masking model [11] (dotted line) for a sinusoid at 458 Hz with additive white Gaussian noise.

4.2 Analysis

Fig. 4 shows the shape of the masking curve obtained from (12) for a 400-sample block of the sinusoidal test signal with -20 dB noise floor compared to the masking curve obtained from a spectral psychoacoustical model described in [11]. In general, a good correspondence is found, although our masking curve shape predicts a peak at 458 Hz that is around 9 dB lower than the predicted threshold from the spectral model. It is also clearly visible that the hearing threshold in quiet is not modeled by our implementation of the Dau model. Note that (12) yields frequencies of equal perceptual distortion for small distortions, which should be more valuable to most coding applications than an absolute masking threshold without any information about the perceptual sensitivity.

An analysis based on the eigenvalue decomposition allows a more detailed view on the model characteristics both in time and frequency. Fig. 5 (a) shows the sinusoidal test signal at 35 dB SNR (first row) and the lowest-sensitivity eigenvectors $\mathbf{q}_1$,$\mathbf{q}_2$ and $\mathbf{q}_3$ from $\mathbf{Q}_\varepsilon$. As expected, these eigenvectors are sums of sinusoids that have frequencies close to the sinusoidal test signal, and this is confirmed by spectral analysis.

Fig. 5 (b) shows white Gaussian noise projected onto the ε-*space* of the sensitivity matrix according to (14) for the blip test signal. Here the ε-*space* is spanned by the first 118 out of 400 eigenvectors $\mathbf{q}_i$. In the first row we see the test signal, the second row contains the noise vector $\mathbf{n}$ used for projection, the third row shows the corresponding $\mathbf{n}_\varepsilon$, and in the last row the result of a projection on the orthogonal space is shown. Errors corresponding to a low sensitivity of the Dau auditory model apparently have an energy envelope in time with support from $1-3$ milliseconds before the blip to about 30 ms after the blip, while the orthogonal space exhibits opposite behavior. This clearly shows the non-simultaneous masking prediction capabilities of the model.

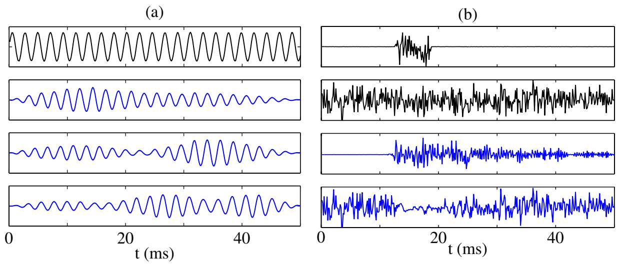

Figure 5: Characteristics of the eigenvectors in $\mathbf{Q}_\varepsilon$.

5. CONCLUSION

In this paper, we described a general method to derive a sensitivity matrix for distortion measures based on spectro-temporal auditory models under the assumption of small errors. The linearizations are shown to be accurate for high SNRs above 30 dB. Using methods from linear algebra, an analysis of the sensitivity matrix revealed insight into the characteristics of the auditory model for low distortions. Known properties of human auditory masking behavior were clearly visible in the studied model.

REFERENCES

[1] R. D. Patterson, M. Allerhand, and C. Giguère, "Time-Domain Modeling of Peripheral Auditory Processing: A Modular Architecture and a Software Platform," *J. Acoust. Soc. Am.*, vol. 98, no. 4, pp. 1890–1894, Oct. 1995.

[2] T. Dau, D. Püschel, and A. Kohlrausch, "A Quantitative Model of the Effective Signal Processing in the Auditory System. I. Model Structure," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3615–3622, June 1996.

[3] ——, "A Quantitative Model of the Effective Signal Processing in the Auditory System. II. Simulations and Measurements," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3623–3631, June 1996.

[4] F. Baumgarte, "Ein psychophysiologisches Gehörmodell zur Nachbildung von Wahrnehmungsschwellen für die Audiocodierung," Ph.D. dissertation, University of Hannover, 2000.

[5] M. Hansen and B. Kollmeier, "Objective Modeling of Speech Quality with a Psychoacoustically Validated Auditory Model," *J. Audio Eng. Soc.*, vol. 48, no. 5, pp. 395–409, May 2000.

[6] Y. Yamada, S. Tazaki, and R. M. Gray, "Asymptotic Performance of Block Quantizers with Difference Distortion Measures," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 1, pp. 6–14, Jan. 1980.

[7] W. R. Gardner and B. D. Rao, "Theoretical Analysis of the High-Rate Vector Quantization of LPC Parameters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 3, no. 5, pp. 367–381, Sept. 1995.

[8] T. Linder and R. Zamir, "High-Resolution Source Coding for Non-Difference Distortion Measures: The Rate-Distortion Function," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 533–547, Mar. 1999.

[9] T. Linder, R. Zamir, and K. Zeger, "High-Resolution Source Coding for Non-Difference Distortion Measures: Multidimensional Companding," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 548–561, Mar. 1999.

[10] J. Li, N. Chaddha, and R. M. Gray, "Asymptotic Performance of Vector Quantizers with a Perceptual Distortion Measure," *IEEE Trans. Inform. Theory*, vol. 45, no. 4, pp. 1082–1091, May 1999.

[11] S. van de Par, A. Kohlrausch, G. Charestan, and R. Heusdens, "A New Psychoacoustical Masking Model for Audio Coding Applications," in *Proc. IEEE Int. Conf. Acoust. Speech Sign. Process.*, vol. 2, Orlando, Florida, USA, 2002, pp. 1805–1808.