Proceedings of the 6th Nordic Signal Processing Symposium - NORSIG 2004
June 9 - 11, 2004
Espoo, Finland

Multiscale Detection of Transiently Evoked Otoacoustic Emissions

Vaidotas Marozas[1], Leif Sörnmo[2], Arturas Janušauskas[1], and Arunas Lukoševičius[1]
[1]Inst. of Biomed. Eng., Kaunas University of Technology, Studentu str 50, Kaunas, Lithuania
[2]Signal Processing Group, Dept. of Electroscience, Lund University, Lund, Sweden.
E-mail: vaidotas.marozas@ktu.lt

Abstract— **This study presents a novel approach to multiscale detection of transiently evoked otoacoustic emissions. Using statistical hypothesis testing, it is shown that the optimal detector involves a time windowing operation where the window is estimated from ensemble correlation information. The detector performs adaptive splitting of the signal into different frequency bands using wavelet or wavelet packet decomposition. The results show that the presented detectors have performances similar to that of the traditional TEOAE detector, however, the performance improves considerably when the signal energy is disregarded in the threshold test.**

I. Introduction

The detection of transiently evoked otoacoustic emissions (TEOAE) is important in applications where quick and objective hearing assessment is required, e.g., in screening of newborns, ototoxic drugs effect monitoring, and screening of large populations exposed to noisy environments [1].

"Wave reproducibility" is the commonly used parameter for detecting if TEOAEs are present or not, and is implemented in many commercial screening devices. Wave reproducibility is given by the crosscorrelation coefficient (as computed between two subaverages) and compared to a fixed detection threshold. This detection parameter is appealing when taking into account the early observation made by Kemp that TEOAEs exhibit a high degree of reproducibility [2]. However, when a less good signal-to-noise ratio (SNR) of the averaged signals is achieved, wave reproducibility is known to perform poorly.

In a recent comparative study, Stürzebecher and coworkers presented the performance of several detection parameters by using a data set which consisted of 420 TEOAEs recorded from adult subjects [3]. In addition to wave reproducibility, the comparison involved parameters such as the variance ratio [4], the modified variance ratio [5], the binominal test [6], and different SNR definitions. The examined frequency interval was divided into two subintervals (1.5–2.5 and 2.5–4.0 kHz) which were processed separately: an otoacoustic emission was considered as present when detections were declared in both subintervals. The best detection performance was obtained by the modified variance ratio, surpassing that of wave reproducibility which, in fact, was associated with the worst performance.

Transiently evoked OAEs are characterized by nonstationary signals whose frequency content changes as a function of time such that it becomes progressively more lowpass. This property has been subjected to several investigations with the common aim to better understand the underlying mechanism of TEOAE generation and the activity of outer hair cells along the basilar membrane. Time-frequency analysis represents a powerful tool for analyzing such nonstationary behavior, and may be based on the short-term Fourier transform, the Wigner-Ville distribution, and the wavelet transform [7]–[11]. However, the problem of developing detection schemes, which make use of the nonstationary TEOAE signal properties, remains and is therefore addressed in the present study.

In a recent paper, we introduced signal-dependent time windows, determined from the ensemble correlation properties [12]; these windows were found to considerably improve detection performance. In the present paper, we develop an optimal detector based on a statistical signal model, and show that the optimal detector implies the use of time windowing. The performance is evaluated in terms of ROCs using a large database with TEOAE signals.

II. Methods

The development of the present TEOAE detector involves three essential parts. First, the binary detection problem, defined by a random signal in Gaussian noise, is revisited. Second, the availability of an ensemble with several signals is made good use of for the purpose of estimating the unknown statistical parameters of the detector. Finally, the detector is modified to account for the nonstationary characteristics of TEOAEs using a time-scale decomposition.

A. Gaussian Detection of Random Signals

The binary detection problem is based on hypothesis testing for Gaussian, random signals in noise [13],

$$\mathcal{H}_1: \quad \mathbf{x} = \mathbf{s} + \mathbf{v} \tag{1}$$

$$\mathcal{H}_0: \quad \mathbf{x} = \mathbf{v}, \tag{2}$$

where

$$\mathbf{x} = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} s(0) \\ s(1) \\ \vdots \\ s(N-1) \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v(0) \\ v(1) \\ \vdots \\ v(N-1) \end{bmatrix}. \tag{3}$$

The signal $\mathbf{s}$ is assumed to have a zero-mean, Gaussian probability density function which is completely characterized by its correlation matrix $\mathbf{R}_s$. The noise $\mathbf{v}$ is assumed to be zero-mean, Gaussian, white, and with variance σ_v^2.

Using the Neyman-Pearson (NP) approach to signal detection, a hypothesis test is defined in which hypothesis $\mathcal{H}_1$ is selected whenever the likelihood ratio $\Lambda(\mathbf{x})$ exceeds a certain threshold η, i.e.,

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > \eta. \qquad (4)$$

Otherwise, hypothesis $\mathcal{H}_0$ is selected indicating that no signal is present. By inserting the previous statistical assumptions, the optimal NP detector is given by the following estimator-correlator structure,

$$T(\mathbf{x}) = \mathbf{x}^T \hat{\mathbf{s}} > \eta, \qquad (5)$$

where $\hat{\mathbf{s}}$ denotes the Wiener estimate of $\mathbf{s}$, i.e., the minimum mean square estimate of $\mathbf{s}$. This estimate is obtained by

$$\hat{\mathbf{s}} = \mathbf{R}_s(\mathbf{R}_s + \sigma_v^2 \mathbf{I})^{-1}\mathbf{x}. \qquad (6)$$

Using the eigendecomposition of the correlation matrix into its eigenvalues λ_i and eigenvectors $\mathbf{v}_i$,

$$\mathbf{R}_s = \sum_{k=0}^{N-1} \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \qquad (7)$$

the detector in (5) can be expressed in its canonical form

$$T(\mathbf{x}) = \sum_{k=0}^{N-1} \frac{\lambda_k}{\lambda_k + \sigma_v^2} y_k^2, \qquad (8)$$

where y_k results from correlating the observed signal with the eigenvector $\mathbf{v}_k$, i.e., $y_k = \mathbf{v}_k^T \mathbf{x}$.

B. Detection and Ensemble Correlation

In general, non-stationary signals are difficult to handle within the above detection context. However, when $\mathbf{s}$ is assumed to be white and nonstationary with a time-varying variance, its correlation matrix is given by

$$\mathbf{R}_s = \mathrm{diag}(\sigma_s^2(0), \sigma_s^2(1), \ldots, \sigma_s^2(N-1)). \qquad (9)$$

The corresponding optimal detector is given by

$$T(\mathbf{x}) = \mathbf{w}^T \mathbf{x}^2 > \eta, \qquad (10)$$

where $\mathbf{w}$ is a vector whose entries are defined by

$$w(n) = \frac{\sigma_s^2(n)}{\sigma_s^2(n) + \sigma_v^2}, \quad n = 0, \ldots, N-1. \qquad (11)$$

Hence, the detector weights the squared observations $x(n)$ in relation to their signal-to-noise ratio (SNR).

In practice, the statistical parameters $\sigma_s^2(n)$ and σ_v^2 are not *a priori* known and therefore need to be estimated. For situations when an ensemble of observed signals is available,

$$\{\mathbf{x}_i = \mathbf{s} + \mathbf{v}_i\}, \quad i = 1, 2, \ldots, M, \qquad (12)$$

where each $\mathbf{x}_i$ is assumed to derive from the signal model defined by (1) and (9), it has previously been shown by us that the Wiener weight $w(n)$ in (11) can be directly estimated from the ensemble correlation [12]. This quantity is defined as the cross-correlation coefficient between $x_i(n)$ and $x_j(n)$, i.e.,

$$\rho_{ij}(n) = \frac{E\left[x_i(n)x_j(n)\right]}{\sqrt{E[x_i^2(n)]E[x_j^2(n)]}}, \qquad (13)$$

and is identical to the weight in (11) since

$$\rho_{ij}(n) = \begin{cases} 1 & i = j \\ \dfrac{\sigma_s^2(n)}{\sigma_s^2(n) + \sigma_v^2} \ (= \rho(n)) & i \neq j. \end{cases} \qquad (14)$$

The maximum likelihood estimate of the ensemble correlation $\rho(n)$ is given by [14]

$$\hat{\rho}(n) = \frac{\sum_{i=1}^{M}\sum_{\substack{j=1 \\ i\neq j}}^{M} x_i(n)x_j(n)}{(M-1)\sum_{i=1}^{M} x_i^2(n)}. \qquad (15)$$

Thanks to the availability of the ensemble $\{\mathbf{x}_i\}$, the SNR of the observed signal is improved by ensemble averaging in which $\mathbf{x}$ is replaced by the ensemble average $\overline{\mathbf{x}}$ in (5) and (6),

$$\mathbf{x} \to \overline{\mathbf{x}} = \frac{1}{M}\begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_M\end{bmatrix}\mathbf{1}, \qquad (16)$$

where the vector $\mathbf{1}$ denotes a column vector with ones in all entries. Since the averaging operation reduces the noise variance σ_v^2 by a factor of M, the ensemble correlation in (14) cannot be directly used but is modified by the following nonlinear transformation,

$$w(n) = \frac{\rho(n)}{\rho(n)\left(1 - \frac{1}{M}\right) + \frac{1}{M}} \left(= \frac{\sigma_s^2(n)}{\sigma_s^2(n) + \frac{\sigma_v^2}{M}} \right). \qquad (17)$$

Finally, by inserting $\hat{\rho}(n)$ in (17) the weight estimates are obtained which are needed for computing the detection statistic

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} \hat{w}(n)\overline{x}^2(n). \qquad (18)$$

C. Multiscale Detection

Transiently evoked OAEs exhibit a frequency-decreasing behavior as a function of latency. In order to arrive at a detector structure that is well-suited to such behavior, the detector of the preceding subsection is extended to handle splitting of the observed signal into different frequency components. This is done by first considering the case when $s(n)$ is a wide sense stationary process and accordingly characterized by a symmetric, Toeplitz correlation matrix. The corresponding eigenvalues λ_k and eigenvectors $\mathbf{v}_k$ have a particular structure for large data records (i.e., $N \to \infty$), namely

$$\lambda_k = P_s(f_k) \qquad (19)$$

$$\mathbf{v}_k = \begin{bmatrix}1 & \exp(\jmath 2\pi f_k) & \cdots & \exp(\jmath 2\pi (N-1) f_k)\end{bmatrix}^T \qquad (20)$$

for $k = 0, 1, \ldots, N-1$; the function $P_s(f)$ denotes the power spectrum of $s(n)$, and $f_k = k/N$. In this case, each weight $\lambda_k/(\lambda_k + \sigma_v^2)$ reflects the SNR at a certain frequency f_k. Hence, $T(\mathbf{x})$ results from first correlating the observed signal with a narrowband filter $\mathbf{v}_k$ whose center frequency is located at f_k, weighting it in relation to its SNR, and then summing the terms that result from the N different frequencies.

Putting the preceding results together, the following detector structure is proposed in which the filtered ensemble averages are weighted with their corresponding time-varying weight $\hat{w}_k(n)$,

$$T(\mathbf{x}) \approx \sum_{k=0}^{N-1} \sum_{n=0}^{N-1} \hat{w}_k(n) \overline{y}_k^2(n). \tag{21}$$

The signal $\overline{y}_k(n)$ denotes the output of the eigenvector filter $\mathbf{v}_k$ with the ensemble average $\overline{x}(n)$ as input. The estimate of the weight $\hat{w}_k(n)$ is obtained as before but with the difference that the ensemble consists of signals which have been filtered with $\mathbf{v}_k$.

The detector in (21) involves equidistant splitting of the frequency axis. However, it is often advantageous to tailor the frequency splitting to individual TEOAE signals, and therefore wavelet packets are considered to allow for an efficient, adjustable resolution of frequencies [15]. Based on considerations of an earlier study [11], Coiflets were chosen as the mother wavelet. Using wavelet packets for detection, it becomes natural to compute the weights $\hat{w}_k(n)$ in the wavelet packet domain, and to apply them to the corresponding frequency components $\overline{y}_k^2(n)$ in the same domain. The detector structure in (21) is then a double sum which involves the selected set of scales, denoted by Ω_k, and the set of samples in the wavelet packet domain, denoted by $\Omega_n(k)$, whose number depends on the selected scale:

$$T(\mathbf{x}) = \sum_{k \in \Omega_k} \sum_{n \in \Omega_n(k)} \hat{w}_k(n) \overline{y}_k^2(n). \tag{22}$$

This structure may also allow us to constrain the detection information to those frequencies which are judged to be relevant for TEOAE signals and to those time intervals where certain frequencies are expected to be present.

Yet another variation on detector structure of particular interest is the one where the signal energy is simply disregarded in $T(\mathbf{x})$ such that

$$T'(\mathbf{x}) = \sum_{k \in \Omega_k} \sum_{n \in \Omega_n(k)} \hat{w}_k(n). \tag{23}$$

This structure is of interest since it avoids the need for a detection threshold which accounts for signal energy; the weights $\hat{w}_k(n)$ are normalized and are therefore more easily thresholded.

III. TEOAE DATABASE

A database of 4900 TEOAE records from the study "Otoacoustic emissions in the general adult population of Nord-Trøndelag, Norway" was used [16]. The database contained the TEOAE signals from men and women with an average age of of 49 years (the standard deviation was 16 years) ranging from 20 to 96 years. The average mean hearing level (MHL) of the subjects was 16 dB HL (the standard deviation was 15 dB HL) ranging from -10 to 114 dB HL. Transiently evoked OAE's were recorded in an attenuation booth or in a relatively quiet, but not sound-proof, room using the ILO92 OAE recording device (Otodynamics Ltd). The signals were acquired during 20 milliseconds (ms) at a sampling rate of 25600 Hz and stored as raw data. The acquisition was terminated when the quality of the response met one of the criteria involving TEOAE level, SNR, reproducibility or noise level [16]. The health screen also included air conduction, pure-tone audiograms recorded by an Interacoustics AD25 audiometer. An audiogram was related to each TEOAE record and used for classifying the record according to hearing threshold. Since it may be assumed that TEOAE responses are absent in subjects with a hearing loss of 30 dB MHL or worse, the database of TEOAE responses was divided into two groups: one group with mean hearing threshold better than 30 dB MHL, classified as normal hearing subjects (NH), and another with hearing threshold at 30 dB MHL or worse, classified as hearing impaired subjects (HI). The NH group consisted of 4231 subjects which all were assumed to have TEOAEs, while the remaining 669 HI subjects no responses.

IV. RESULTS

The performance of different TEOAE detectors are compared in terms of receiver operating characteristics (ROCs). The detectors are based on 1. wave reproducibility, 2. the modified variance ratio, 3. wavelet decomposition and weighted signal energy (cf. (22)), and 4. wavelet packets and weighted signal energy (cf. (22)). The latter two detectors were also studied when signal energy was disregarded (cf. (23)). The ROC provides a quantitative description of diagnostic accuracy in terms of sensitivity and specificity and is for different detection thresholds.

The performance results are presented in Figs. 1 and 2. Figure 1 compares ROCs of detectors based on wave reproducibility and modified variance ratio to those of multiscale detectors which make use of weighted signal energy (calculated according to (22)). It can be seen that the performance of all four detectors are almost the same since their ROCs almost coincide.

Figure 2 presents ROCs for the case when detection criterions for multiscale methods were calculated according to (23), i.e. only weights were considered in construction of the detectors. The ROCs for wave reproducibility and modified variance ratio have been repeated from Figure 1. It is evident from these results that information on signal energy does not offer any performance improvement. Rather, information on ensemble correlation, in combination with how it is distributed in different frequcny bands, is the more important.

V. DISCUSSION AND CONCLUSIONS

Good detection reliability of TEOAE is crucial for the outcome of hearing screening programs. The purpose of this study was to analyze the task of TEOAE signals detection from the point of view of statistical signal detection theory and to investigate the possibility to further increase the detection performance of a recently developed algorithm [12].

Many authors investigated the possibility to exploit *a priori* information about the nonstationary properties of the TEOAE signals for the purpose of increasing the detection performance. There were proposed methods that used simple time windowing of the acquired signal as to account additional

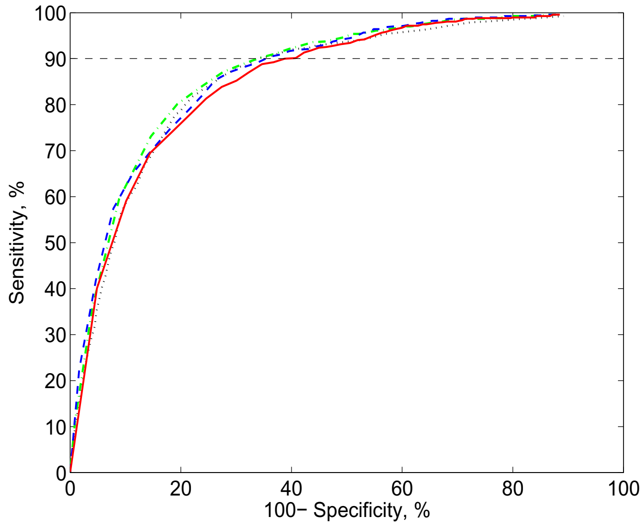

Fig. 1. ROC curves for database of 4900 TEAOE signals. Detection parameters: wave reproducibility (black dots), modified variance ratio (green dash-dot), wavelet and energy (blue dashes), wavelet packets and energy (red line))

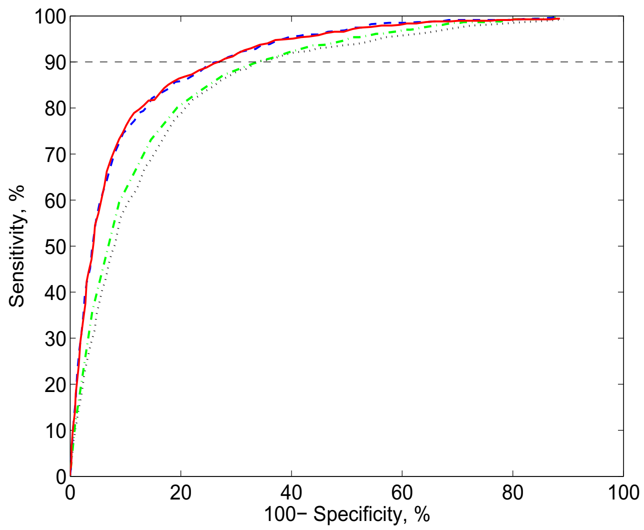

Fig. 2. ROC curves for database of 4900 TEAOE signals. Detection parameters: wave reproducibility (black dots), modified variance ratio (green dash-dot), wavelet and no energy (blue dashes), wavelet packets and no energy (red line))

information about the most probable location of TEOAE activity in time [17]. Then methods were proposed which account for information about the most probable location of the activity of different TEOAE components in time and also in frequency [11], [12]. These methods were shown to have significantly better performance over methods that do not account for TEOAE signal nonstationarity.

The proposed multiscale TEOAE detectors, which include information on signal energy and whose design are based on optimal detection theory, do not appear to offer any advantage over the reference methods "wave reproducibility" and "modified variance ratio". However, by simply skipping the energy in multiscale detectors resulted in a significant improvement of detection performance.

The detection results suggest that splitting the TEOAEs into different frequency components has no significant impact on performance: similar performance is obtained whether wavelet style fixed, octave-wide frequency subbands or adaptive frequency axis division by wavelet packets is used. Wavelet packet transform includes the search for the best basis (best tree) or equivalently best frequency axis division into subbands, resulting in the signal representation, which minimizes some suitable criterion. We investigated several known criteria for best basis selection based on entropy or l^p norm cost with different power p. However, no significant differences in the resulting frequency axis divisions were observed. Thus, we stayed with the popular entropy criterion for search of the best basis.

In conclusion, taking into account any a priory known information about the signal, such as inherent nonstationarity of TEOAE signals, can help to design improved signal detector.

REFERENCES

[1] J. W. Hall III, *Handbook of Otoacoustic Emissions*. San Diego: Singular, Thomson Learning, 2000.

[2] D. T. Kemp, "Stimulated acoustic emissions from within the human auditory system," *J. Acoust. Soc. Am.*, vol. 64, pp. 1386–1391, 1978.

[3] E. Stürzebecher, M. Cebulla, and K. D. Wernecke, "Objective detection of transiently evoked otoacoustic emissions," *Scand. Audiol.*, vol. 29, pp. 78–88, 2001.

[4] C. Elberling and M. Don, "Quality estimation of averaged auditory brainstem responses," *Scand. Audiol.*, vol. 13, pp. 187–197, 1984.

[5] M. Cebulla, E. Stürzebecher, and K. D. Wernecke, "Untersuchung verschiedener SNR-schätzer für den nachweis von biologischen reizantworten im rauschen," *Z. Audiol.*, vol. 39, pp. 14–22, 2000.

[6] P. Zoth, "Automatisiertes hörscreening basierend auf otoakustischen emissionen mit matematisch exakt definierten wahrscheinlichkeitskritieren." Patent application DE 4441127 A1, 1994.

[7] H. P. Wit, P. van Dijk, and P. Avan, "Wavelet analysis of real and synthesized click-evoked otoacoustic emissions," *Hearing Res.*, vol. 73, pp. 141–147, 1994.

[8] J. Cheng, "Time-frequency analysis of transient-evoked otoacoustic emissions via smoothed Wigner distribution," *Scand. Audiol.*, vol. 24, pp. 91–96, 1995.

[9] G. Tognola, P. Ravazzani, and F. Grandori, "Wavelet analysis of click-evoked otoacoustic emissions," *IEEE Trans. Biomed. Eng.*, vol. 45, pp. 686–697, 1998.

[10] L. Zheng, Y. Zhang, F. Sheng, and D. Ye, "Synthesis and decomposition of transient-evoked otoacoustic emissions based on an active auditory model," *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 1098–1106, 1999.

[11] A. Janušauskas, V. Marozas, B. Engdahl, H. Hoffmann, O. Svensson, and L. Sörnmo, "Otoacoustic emissions and improved pass/fail separation using wavelet based denoising," *Med. Biol. Eng. & Comp.*, vol. 39, pp. 134–139, 2001.

[12] A. Janusauskas, L. Sörnmo, O. Svensson, and B. Engdahl, "Detection of transient-evoked otoacoustic emissions and the design of time windows," *IEEE Trans. Biomed. Eng.*, vol. 49, pp. 132–139, 2002.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing - Detection Theory*. New Jersey: Prentice-Hall, 1998.

[14] R. Atarius and L. Sörnmo, "Signal-to-noise ratio enhancement of cardiac late potentials using ensemble correlation," *IEEE Trans. Biomed. Eng.*, vol. 42, pp. 1132–1137, 1995.

[15] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inf. Theory*, vol. 38, pp. 713–718, 1992.

[16] B. Engdahl, "Otoacoustic emissions as a function of age, gender and hearing loss in a large general adult population," *Proc. AHEAD Eur. Meeting: Psychoacoust. and OAEs*, 1998.

[17] M. L. Whitehead, A. M. Jimenez, B. B. Stagner, M. J. McCoy, B. L. Lonsbury-Martin, and G. K. Martin, "Time windowing of click-evoked otoacoustic emissions to increase signal-to-noise ratio," *Ear and Hearing*, vol. 16, pp. 599–611, 1995.