ISSN 1392-2114 ULTRAGARSAS, Nr.1(34). 2000.

Otoacoustic emissions and pass/fail separation using artificial neural network

V. Marozas[1), A. Lukoševičius[1), B. Engdahl[2), O. Svensson[3) and L. Sörnmo[3)

1) Dept. of Biomedical Engineering Kaunas University of Technology, Kaunas, Lithuania

2) Dept. of Environmental Medicine, National Institute of Public Health, Oslo, Norway

3) Signal Processing Group, Dept. of Applied Electronics, Lund University, Lund, Sweden

E-mail: vaidotas.marozas@tef.ktu.lt

Introduction

Transient evoked otoacoustic emissions (TEOAE) are low-level sounds produced by the healthy inner ear in response to a transient stimulus such as a short acoustic click 1. It is believed that outer hair cells within the cochlea are involved in complex active amplification mechanisms of weak incoming sound vibrations. The TEOAE generation is assumed to be a partial product of this amplification [1]. Several studies have evidenced that the presence of TEOAE correlates with the hearing level [1,2,3,4,5]. While this relationship is reported to be frequency specific, the relationship is too weak to predict the pure tone hearing levels. Nevertheless, TEOAE appeared to be useful in screening tests in neonates [7], in population exposed by noise [8] and to monitor the influence of drugs [9].

The detection of TEOAE is difficult because of noise in the recorded signal. The performance can, however, be improved by looking into separate frequency bands and time intervals and combining extracted parameters [10,11,12,13]. An average of cross correlation coefficients, as calculated between components of successive subaverages after splitting them into three frequency bands and windowing in time has been used as a criterion for TEOAE detection [12,13]. The TEOAE generation is well known to be closely related with nonlinearities which are present in the cochlea and which are responsible for the high dynamic range of the hearing system. Thus, it would be natural to think about more complex relationship than linear in associating hearing level and time-frequency features extracted from TEOAE. So far, the most published attempts to model the mentioned association have relied on linear relationship and they have used linear multivariate models [4,6] or simple averaging of features [12,13]. In one study by Buller and Lutman [15] an artificial neural network (ANN) was used in TEOAE classification. The task for the neural network was in this case to mimic a human expert in classifying the shapes of TEAOE waveforms into four beforehand-described classes. In our study we have used a neural network, which used a set of TEOAE features as inputs and audiometric data as the targets in the training procedure.

The purpose of this study is to compare two approaches in separation of normal and hearing impaired subjects: a linear and a more complex, which could account not only for linear, but also for possible nonlinear association of features extracted from TEOAE to hearing level as obtained from pure tone audiometry. Artificial neural networks are known to be capable to realize any complex relationship, when sufficient amount of training data is available [14]. Thus a nonlinear classifier was established by training of an ANN and for comparison a linear classifier implemented as a simple average of the features.

Material and methods

A. Database

A database consisting of 5213 TEOAE records was collected during the health screen of 65604 subjects in the Norwegian county of Nord–Trøndelag (HUNT) [5,6]. The ILO92 Otodynamics analyzer was used for recording of TEOAE data and air conduction pure tone audiograms were recorded using Interacoustics AD25 automatic audiometers. The audiometric criterion used to separate normal hearing from hearing impaired subjects was chosen as 30dB of mean hearing level as obtained at the frequencies 0.5, 1, 2 and 4kHz (MHL). Based on that separation level a total of 4404 subjects were classified as having normal hearing while the remaining 809 were classified as a having impaired hearing.

B. Feature extraction procedure

The initial problem in detecting TEOAE is to establish a criterion according to which TEOAE can be recognized. A TEOAE response exhibit, however, high intersubject variability in shape, which makes it difficult to define general criteria for detection. In addition, TEOAE responses are very weak signals buried in environmental, subject generated and noise of recording hardware. Averaging is, therefore, used to increase the signal- to-noise ratio. An example of TEOAE record consisting of two subaverages is shown in fig.1. This response has been achieved by averaging 1200 sweeps after which a close similarity between the subaverages was obtained evidencing stimulus synchronous activity that is interpreted as TEOAE.

The most popular criterion for TEOAE detection is the cross correlation coefficient between the two subaverages. If the calculated cross correlation coefficient between the two subaverages exceeds some predetermined threshold, it is considered as an evidence of presence of deterministic activity in the recorded signal and a conclusion about detected TEOAE is made. The TEOAE response shown in fig.1 can be clearly distinguished with a high cross correlation value, but this is not always the case. The calculated cross correlation value depends on both initial signal-to-noise ratio and available averaging time (i.e. number of sweeps in the average). Long averaging time is often difficult to maintain in the clinical practice, especially in child investigations. In order to reduce further the influence of the remaining noise in the averaged signal, additional measures can be considered, which use a priori information on TEOAE. Many investigations [13,10,20] have shown that TEOAE exhibit frequency dependant latency: where higher frequencies have shorter post-stimulus time while lower frequencies have longer. This particular feature of TEOAE can be also observed in the example of TEOAE subaverages shown in fig.1, where oscillations of higher frequency, starting 3ms post-stimulus, precede oscillations with lower frequency.

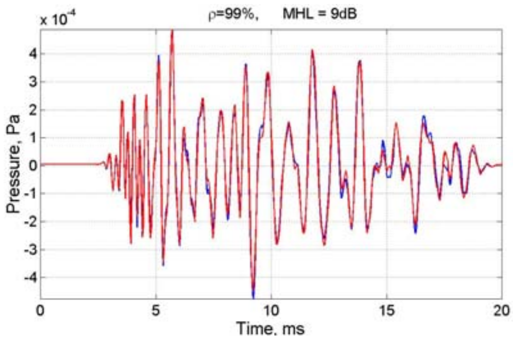

Fig.1. **Example of TEOAE subaverages as recorded from a normal hearing subject having mean hearing threshold 9dB**

In an earlier investigation we have shown that decomposition of the time-frequency plane into regions of interest improves the detection substantially [17]. One possibility to form regions of interest in time frequency plane is to split the signal into frequency bands and to apply specially designed time windows to each bandpass filtered component of the signal. This procedure can be very efficiently accomplished in the wavelet domain due to the existence of fast wavelet decomposition algorithms [21].

The discrete wavelet transform maps the signal into the time-frequency plane according to:

$$s(n) = \sum_{j=1}^{K} \sum_{k=-\infty}^{\infty} w_j(k)\psi(2^j n - k) \qquad (1)$$

where $\psi(n)$ denotes the analysis wavelet and $w_j(k)$ are wavelet coefficients representing the signal at the decomposition level j and time index k.

Choosing wavelet coefficients from decomposition level j is equivalent to bandpass filtering of the signal, while choosing wavelet coefficients with indexes from k to l from given decomposition level is equivalent to time windowing. The fast orthogonal discrete wavelet transform decomposes the signal by definition into octave frequency bands, which can not be chosen independently. The time windowing in wavelet domain can, however, be accomplished with no restriction in the choice of the indexes k and l.

The choice of time indexes k and l in our TEOAE specific feature extraction problem was based on the data from a study of Janušauskas et. al [13], where the average time locations of TEOAE in a database of normal hearing subjects was studied by use of an ensemble correlation technique. The time windowing was thus carried out directly in wavelet domain by selecting or, equivalently, by applying the rectangular windows to the wavelet coefficients from the given level of decomposition (as seen in fig. 2). The features, cross correlation coefficients between two windowed frequency components of two TEOAE subaverages A and B in the wavelet domain, are then obtained as:

$$\rho_j = \frac{\sum_{n=k}^{l} w_{A,j}(n)\cdot w_{B,j}(n)}{\sqrt{\sum_{n=k}^{l} w_{A,j}^2(n)}\cdot\sqrt{\sum_{n=k}^{l} w_{B,j}^2(n)}} \qquad (2)$$

where $w_{A,j}$ and $w_{B,j}$ are wavelet coefficients of subaverages A and B from level j and where k and l are indexes of the first and last coefficient in the respective window.

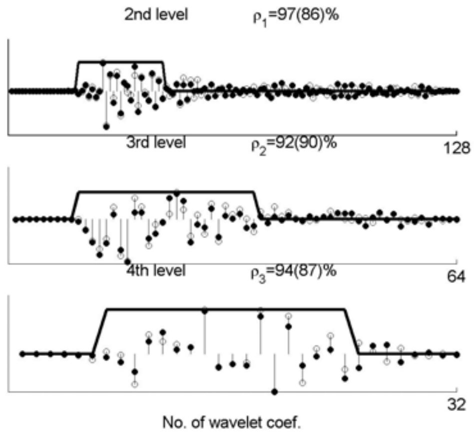

Fig.2. **Three levels of wavelet decomposition of two subaverages and corresponding time windows. The two subaverages A and B are shown as open and filled circles, respectively. Solid lines indicate corresponding time windows. The cross correlation values ρ_j with and without windows (in parentheses) are indicated.**

For the evaluation we have chosen the three levels of decomposition, which contain most of the TEOAE response energy. These levels represent the octave frequency bands whose central frequencies are 1.15kHz, 2.2kHz and 4.4kHz, respectively. The high frequency TEOAE components are represented by the time region 2.5-6.5ms from the 2nd level, the middle frequency components by the time region 2.5-10ms from the 3rd level and low frequency components by the time region 5-14.5ms from the 4th level. An example of TEOAE subaverages transformed into wavelet domain is shown in fig. 2. It can be seen that the highest similarity between the wavelet coefficients and equivalently the highest cross correlation appears in rectangular windows as defined by Janušauskas et. al. The calculated cross correlation values between the TEAOE subaverages with and without windows, as indicated in the fig. 2, exemplify the improvement achieved by windowing.

The three cross correlation coefficients ρ_j, which represent each recorded signal consisting of two subaverages with 512 time samples each are in the following used as TEOAE features. By this procedure we reduced the dimensionality of our problem in the TEAOE detection algorithm.

C. Artificial neural network for classification

Artificial neural networks are used in applications, where predetermined analytical relationships are difficult to establish because of lack of knowledge about the phenomenological background, but where rich empirical datasets are available for teaching of the network of the desired relationship. The artificial neural network is represented by a structure, consisting of units called neurons and connections called weights as seen in fig. 3. Each neuron is a unit that computes the weighted inputs from neighboring neurons. The output of a neuron depends on the input values and an activation function. This output can in turn serve as one of the input values for other neurons. The weights are multiplicative coefficients that can change the influence of one neuron's output to another neuron's input. By changing the connection weights during the training procedure a very complex, possibly nonlinear, relationship between the network inputs and the output can be established.

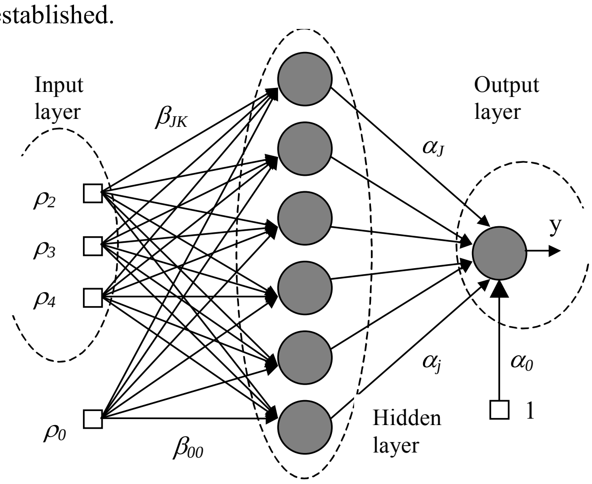

Fig.3. **Schematic representation of a three layer ANN**

The multilayer perceptron was chosen due to its ability to model both simple and very complex functional relationships [14]. We restricted, however, ourselves to consider artificial neural network having only one hidden layer and only hyperbolic tangent activation functions.

The output of this network y can be written as:

$$y = \tanh\left(\sum_{k=0}^{K} \alpha_k \cdot \tanh\left(\sum_{j=0}^{J} \beta_{jk} \cdot \rho_j\right)\right) \qquad (3)$$

where ρ_j denotes the feature vector parameters with $\rho_0 = 1$, J is the number of inputs, K is the number of hidden neurons, β_{jk} are the weights between inputs and hidden layer and α_k are the weights between hidden and output layers.

The neural network training procedure is based on adjusting the weight parameters α_k and β_{jk}. The neural network is considered to be trained when it gives small errors when applied on the training set of data but also responds properly to a new testing set not used in the training procedure. When a network is able to perform as well on both the testing and the training sets of data, we say that the network generalizes well. In order to improve the generalization, training with regularization was used. The regularization method constrains the size of the network weights, causing the network response to be smooth. The Bayesian technique of regularization proposed by D. Foresee and F. Hagan [16] to improve generalization was used in our case. In addition, this regularization procedure gives the number of weights in the neural network that are effectively used in reducing the error function. This feature can be employed to choose optimal number of network neurons. We can simply add more neurons and retrain. If the larger network has the same final effectively used number of parameters, then the smaller network was large enough. In our case, the final network had 3 inputs, 6 neurons in the hidden layer and one neuron in the output layer. The weights were adjusted using Levenberg- Marquardt algorithm during the training procedure. This algorithm has the most rapid convergence properties for networks with moderate complexity [14].

The network was trained using supervised learning with a training set of inputs and targets. This procedure is described by:

$$f(\rho_1, \rho_2, \rho_2) = \begin{cases} 1, & if \quad MHL < 30dB \\ -1, & if \quad MHL \geq 30dB \end{cases} \qquad (4)$$

where $f(\rho_1, \rho_2, \rho_3)$ is the discriminant function which have to be determined by the ANN to minimize the mapping error of the features ρ_1, ρ_2, ρ_3 to the targets, the binary values 1 and –1. These binary values represent the subjects having audiometric MHL in frequency range 500-4000Hz less than 30dB and more than 30dB, respectively. Thus, normal hearing subjects are coded by "1" and hearing impaired subjects by "-1".

The preliminary attempts of the neural network training showed that the network generalized well if the training set consisted of approximately equal number of hearing impaired and normal hearing cases. The training set was therefore made of a database representing 385 hearing impaired and 400 normal hearing subjects. The testing set contained all the subjects: 809 hearing impaired and 4404 normal hearing.

In the testing stage, separation of subjects belonging to one of the groups is made according to this rule:

$$y \underset{IH}{\overset{NH}{\gtrless}} \gamma_0 \qquad (5)$$

where NH represents the normal hearing subjects, IH the impaired hearing subjects, y is the ANN output value and γ_0 is the threshold value at which we obtained 90% of sensitivity.

D. Procedure to evaluate the subject separation results

The principles of statistical decision theory [18] were used for the comparison of linear and nonlinear classifiers, where the linear classifier used the average of the features and the nonlinear classifier used the features combined by the trained ANN.

According to the statistical decision theory the subjects that were identified correctly as normal hearing are named as True Negatives (TN) and those, that are identified correctly as hearing impaired, are named True Positives (TP). Some subjects were however identified incorrectly and contribute to errors. Subjects that posses normal hearing, but were identified by the separation algorithm as hearing impaired are called False Negatives (FN) and subjects that have impaired hearing but were identified as normals are called False Positives (FP).

The performance parameters for the separation algorithms can be defined as:

- Specificity, the probability to correctly detect the normal hearing subject,
- Sensitivity, the probability to correctly detect hearing impaired subjects.

These probabilities can be varied by choosing different decision thresholds in the outputs of the classifiers. However, when one of the probabilities is increased, the other is decreased. A function of sensitivity as a function of the variable *1-specificity*, when the decision threshold is varied over the range of the classifier output values, is called receiver operating characteristics (ROC). In TEOAE detection applications it is very important to have high sensitivity of the classifier to identify the main part of the hearing impaired subjects. The comparison of the performance of the different classifiers was therefore made by keeping the sensitivity at a fixed level of 90% and comparing the resulting specificity.

Results

Both, the linear and the nonlinear, classifiers transformed a vector of three TEOAE features, representing one subject, into one single output. The linear classifier output was the mean cross correlation between the two subaverages for each scale in the range of 0 to 100 while the ANN gave the output in the range from -1 to 1. The results from both classifiers in separating the NH group from HI group are shown in fig. 4 and fig. 5.

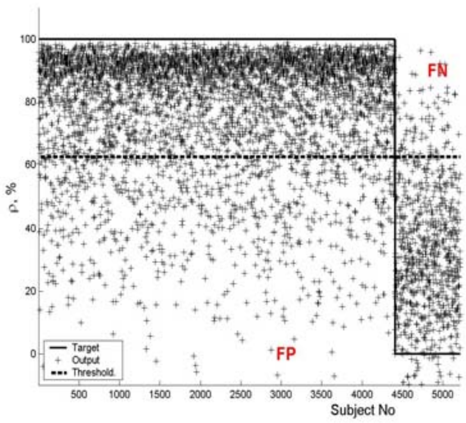

Fig.6. **Separation of normal hearing and hearing impaired subjects using a linear classifier. Here ρ is the average of the features- ρ_1, ρ_2 and ρ_3**

Normal hearing subjects are grouped in the left part of the figures (4404 subjects), while hearing impaired are grouped in the right part of the figures (809 subjects). The

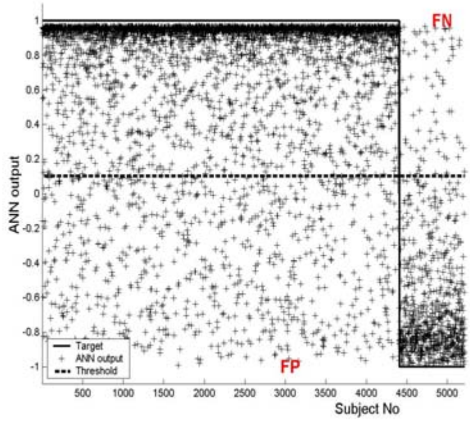

Fig.4. **Separation of normal hearing and hearing impaired subjects using ANN separation**

decision threshold (dashed horizontal line) for separation of hearing impaired from normal hearings is selected such that a sensitivity of 90% is obtained. It can be observed that most of normal hearing subjects are above the decision threshold in the left-hand side of the figures. They constitute the true negatives. The normal hearing subjects below the decision threshold are the false positives. Similarly, hearing impaired subjects (right side of the graphs) below the threshold are the true positives and above the threshold are the false negatives. It can obviously be seen that the linear approach distributes subjects more evenly in comparison with ANN, which seems to separate most of subjects in two distant regions. The improvement expressed as specificity at 90% of sensitivity is, however, showing a very small difference between the methods, (82.7±0.57) % for the linear versus (84.1±0.55) % for the nonlinear (the specificities are shown together with one standard deviation).

It might be assumed that bigger differences would appear at other levels of sensitivity. This is, however, is not the case, as can be seen in the ROC curves in fig.6. The curves indicate that the separation methods are very similar, though, nonlinear approach exceeds linear at some regions. This means a small advantage of nonlinear method to linear at some particular decision threshold values in terms of correctly identified subjects.

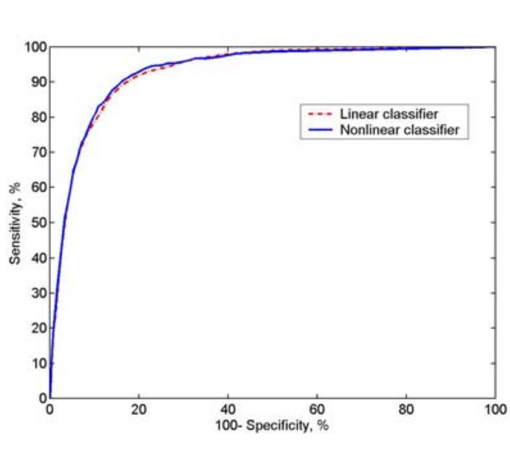

Fig.5. **ROC curves for both methods**

Although the resulting difference between the two methods was small we could prove that the difference is statistically significant by applying a hypothesis test with the null hypothesis defined, as "no differences among the results from the classifiers". After making the assumption about normal distribution of the results, the hypothesis test showed that we could reject the null hypothesis, as the evaluated P value was 0.*0002*.

Discussion

We compared in this study two classifiers to separate hearing impaired and normal hearing subjects, using TEOAE based features. The first classifier was constituted by the linear average of the set of features, while the second, based on neural network, which could account for a more complex relationship, possibly nonlinear, among the set of features as extracted from TEOAE and the mean hearing level in the frequency range 0.5-4kHz. We expected, based on the facts about nonlinear TEOAE generation mechanism in the cochlea, a substantial improvement in subject separation using more complex nonlinear classifier as compared to a linear. The results were, however, very similar with a small but still statistically significant advantage of the neural network based classifier.

A possible reason why the neural network did not decrease the number of errors more compared to the linear classifier might be due to the fact that the database includes a certain amount of errors caused by deficient measurements of TEOAE or audiograms. The outlyers may prevent the neural network to establish the right separation function during the training procedure. We have manually inspected some cases with erroneous behavior: hearing impaired subjects with a TEOAE like response (false negatives) and normal hearing subjects with a response in which TEOAE can not be detected (false positives). There are several reasons that may contribute to false positives: a) poor fitting of the probe into the ear canal (loose seal to the ear canal reduces stimulus pressure and TEOAE amplitudes), b) a blockage of the microphone or speaker ports against the ear canal wall or by ear wax, which prevents the recording of the TEOAE response, c) strong ambient noise during the session of recording, d) conductive hearing loss in middle ear of 10 to 20 dB can make emission undetectable. One example of a false positive case is shown in fig. 7, where the subject has MHL=4dB indicating the potential to generate TEAOE and where the ρ value equal to 40% indicates mainly the random activity. The possible reasons for low ρ value might be a) somewhat low stimulus pressure -72dB (scaled stimulus is shown in the left-hand side of the figures), while the average pressure is 80dB in the database and b) strong ambient noise.

False negatives, may also appear due to technical failures: a) bad fitting of the probe into the ear canal may cause prolonged stimulus artifact, which will give increased cross correlation values, b) noise from instrumentation which is synchronized with stimulus may be detected as the TEOAE.

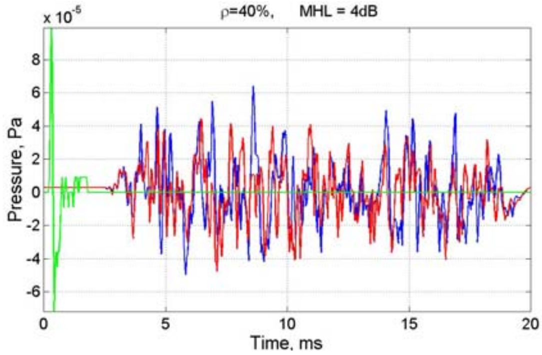

Fig.7. **The example of TEOAE response in case of false positive subject**

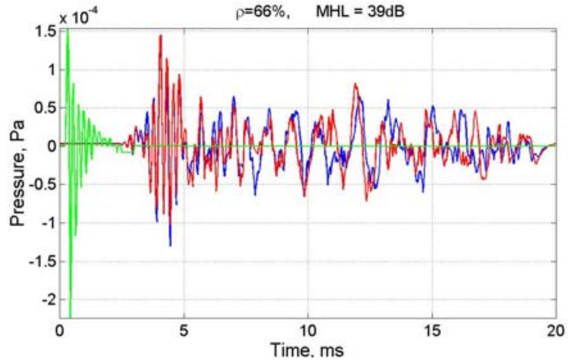

Fig.8**. An example of prolonged artifact**

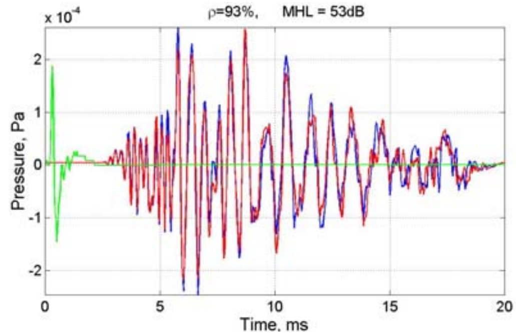

Fig.9. **The example of false negative case**

Fig.8 shows an example representing a false negative case, which might have increased cross correlation value due to prolonged stimulus artifact. This subject is classified as normal hearing although the MHL is 39dB.

It is, however, sometimes difficult to find a simple explanation for the achieved error. Figure 9 shows a TEOAE response from a subject, which is hearing impaired according to audiometric data with MHL 53dB. This response looks like a response from a completely normal hearing subject, where: high, middle and low frequencies easily can be distinguished in the response. One possible explanation for this example is that it may be a retrocochlear hearing loss, which means that the hearing loss is caused by dysfunction the auditory pathway after the cochlea. Such a condition cannot be detected by a TEOAE test, since these cases have a normal cochlea producing a normal TEOAE. Another possible explanation is error in the measurement of the audiogram.

Conclusions

The results of this investigation have shown a small difference between a method using average of TEOAE cross correlation based time-frequency features and a method, which could account more complex relationships, possibly nonlinear, in association of these TEOAE features to the mean hearing level. Although the difference between the methods was small, a statistical hypothesis testing confirmed this difference to be statistically significant.

Acknowledgment

This work was supported by Swedish Institute, Lithuanian Sciences, Studies Fund and in part by the U.S. National Institute on Deafness and Other Communication Disorders (NIDCD), National Institutes of Health (NIH).

References

1. **Kemp D.T.** Stimulated acoustic emissions from within the human auditory system.// J. Acoust. Soc. Am. 1978. Vol. 64. P. 1386-1391.
2. **Prieve B.A.** Click and tone-burst-evoked otoacoustic emissions in normal and hearing-impaired ears// J.Acoust.Soc.Am. 1996. Vol.99. No.5. P.3077-3086.
3. **Kapadia S., Lutman M.E.** Are normal hearing thresholds a sufficient condition for click- evoked otoacoustic emissions?// J.Acoust.Soc.Am. 1997. Vol.101. No.6. P.3566-3576.
4. **Hussain D.M., Gorga M.P., Neely S.T., Keefe D.H. and Peters J** Transient evoked otoacoustic emissions in patients with normal hearing and in patients with hearing loss// Ear Hear. 1998. 19. P. 434-449.
5. **Engdahl B.** Otoacoustic emissions in general adult population I: Age, gender and ear side.// Prepared for submittion.
6. **Engdahl B., Tambs K., Borchgrevnik H., Hoffman H**. Otoacoustic emissions in general adult population II: Pure tone hearing thresholds.// Prepared for submittion.
7. **Kemp D.T., Rayn S.** Otoacoustic emission tests in neonatal screening programs// Acta Otolaryngol. (Stockh.). 1991. Suppl.482. P.73-84.
8. **Engdahl B., Woxen O., Arnesen A. R., Mair I. W.** Transient evoked otoacoustic emissions as screening for hearing losses at the school for military training// Scand. Audiology. 1996. 25. P. 71-78.
9. **McFadden D. and Pasanen G. E.** Otoacoustic emissions and quinine sulphate// J. Acoust. Soc. Am. 1994. Vol.95(6). P.3460-3474.
10. **Tognola G., Grandori F. and Ravazzani P.** Latency distribution of click-evoked otoacoustic emissions// Proc. IEEE Eng. Med. Biol. Amsterdam. The Netherlands. 1996.
11. **Whitehead M. L., Jimenez A. M., Stagner B. B., McCoy M. J. Lonsbury-Martin B. L., and Martin G. K**. Time windowing of click evoked otoacoustic emissions to increase signal-to-noise ratio// Ear Hear. 1995. Vol.16. P.599-611.
12. **Janušauskas A., Marozas V., Engdahl B., Svensson O. and Sörnmo L.** Otoacoustic emissions and improved pass/fail separation using wavelet based denoising// Proc. of 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Hong Kong. 1998. P.3129- 3131.
13. **Janušauskas A., Engdahl B., Svensson O. and Sörnmo L.** Latency of otoacoustic emission frequency components// Proc. of 11th Nordic-Baltic Conference on Biomedical Engineering. Tallinn, Estonia. 1999. P.332-334.
14. **Haykin S.** Neural networks. A comprehensive foundation// Prentice-Hall. 1994.
15. **Buller G., Lutman M. E.** Automatic classification of transiently evoked otoacoustic emissions using an artificial neural network// British Journal of Audiology. 1998. 32. P.235-247.
16. **Foresee D., Hagan F.** Gauss-Newton approximation to Bayesian learning// International Conference on Neural Networks. 1997. Vol.3. P. 1930 –1935.
17. **Marozas V., Janušauskas A., Engdahl B., Svensson O. and Sörnmo L.** Detection of Otoacoustic emission in the wavelet domain// Proc. of 11th Nordic-Baltic Conference on Biomedical Engineering. Tallinn, Estonia. 1999. P.334-335.
18. **Nielsen W., Turner R. G.** Application of clinical decision analysis to audiological tests// Ear. Hear. 1984. P. 125-133.
19. **Altman D. K.** Practical Statistics for Medical Research// Chapman & Hall. Edinburgh, 1992.
20. **Ueda Hiromi.** Do click-evoked otoacoustic emissions have frequency specificity?// J. Acoust. Soc. Am. 1999. Vol. 105 (1). P.306-310.
21. **Mallat S.** Wavelet tour of signal processing// Academic Pres. 1998.

V. Marozas, A. Lukoševičius, B. Engdahl, O. Svensson, L. Sörnmo

Otoakustinė emisija ir rūšiavimas yra/nėra naudojant dirbtinį neuroninį tinklą

Reziumė

Straipsnyje nagrinėjama, ar, nustačius galimą netiesišką priklausomybę tarp otoakustinės emisijos signalų koreliacinių charakteristikų ir vidutinio klausos slenksčio, įmanoma tiksliau suskirstyti subjektus į girdinčius ir neprigirdinčius. Otoakustinės emisijos signalams skirstyti buvo panaudoti du klasifikatoriai: tiesinis su vienodais parametrų svoriais ir apmokytas dirbtinių neuronų tinklas. Neuronų tinklas apmokymo metu galėjo "išmokti" sudėtingesnę negu pirmojo klasifikatoriaus diskriminantinę funkciją, tačiau rezultatai parodė, kad specifiškumo padidėjimas, esant fiksuotam jautrumui, yra nedidelis, nors ir statistiškai patikimas.

Pateikta spaudai 2000 04 6