Speech Signal Processing Based on Wavelets and SVM for Vocal Tract Pathology Detection^{*}

P. Kukharchik, I. Kheidorov, E. Bovbel, and D. Ladeev

Belarusian State University, 220050 Nezaleshnasty av, 4, Minsk, Belarus

Abstract. This paper investigates the adaptation of modified wavelet-based features and support vector machines for vocal folds pathology detection. A new type of feature vector, based on continuous wavelet transform of input audio data is proposed for this task. Support vector machine was used as a classifier for testing the feature extraction procedure. The results of the experimental study are shown.

1 Introduction

Information achieved from speech analysis plays a great role in vocal tract pathology detection. In some cases such analysis is the only way to find pathologies. In medicine voice quality estimation is a very important task that caused a lot of researches in different spheres. Nowadays there are a lot of methods for direct observation and diagnostics of vocal pathologies, but they have a series of drawbacks. Human vocal tract during the sounds pronouncing is hardly observed and this is a problem for pathology detection. In addition, such examination causes discomfort to patient and influence the result reliability [1]-[2]. In this comparison the acoustical signal analysis does not have such drawbacks as pathology detection method. Except this such method has serious advantages. Firstly, acoustical signal analysis is a noncontact method, and thanks to this it lets to explore more patients in a small period of time. Secondly, it lets to detect diseases on early stages. There are done several researches in this direction based on analysis of some long vowels [3]-[4]. Last time accent in this sphere was shifted to the idea of usage of automatic speaker recognition methods for voice pathology detection [5]-[6]. The achieved accuracy is an encouraged one even for a small amount of training data. In this paper we propose the speech signals classification scheme specially developed for the vocal tract pathology detection. Base principles of this scheme are very close to those like physician analyses patient speech. As a basis for feature vector forming the continuous wavelet transform is used, and support vector machine was selected as a classifier. The main aim of this paper is to propose method for convenient continuous control of pathology evolution.

2 Methodology

Vocal pathology presence leads to changes in sounds pronunciation by a human. Depending on the pathology the changes can be more or less expressed.

^{*} The paper is supported by ISTC grant, project B-1375.

Fig. 1. Wavelet transformation of [e] sound, from the voice of speaker with normal voice

Fig. 2. Wavelet transformation of [e] sound, from the voice of speaker with polypus of vocal cord

Among sounds the most interesting are long vowels and some resonant sounds, the pathology is more evident for these sounds. On the first stage during the initial analysis the stressed vowels are to be manually selected from continuous speech and then processed by wavelet-analysis. Wavelet analysis is chosen as an optimal tool due to its effectiveness for analysis of short and non-stationary signals like phonemes. At fig.1. there is a wavelet transform of stressed sound [e] spoken by a healthy person.

If there is the pathology in a signal the picture is changed. At fig.2 there is a wavelet transform for the same vowel for patient with polypus of vocal cord. It is obvious the non-stability of fundamental frequency due to the flexibility loss by cords. It was analyzed more than 140 recordings of healthy voices and voices with pathologies, and the similar results were achieved. This fact makes us sure that wavelet transform will provide the good resolution performance for long speech fragments in order to find distortions caused by pathologies. Not any spectrum estimation method can produce the required accuracy in time-frequency domain, suitable for pathology detection.

2.1 Improved Algorithm for Wavelet Transformation

The continuous wavelet transform (CWT) of $f(t)$ can be presented as:

$$Wf(u, s) = \int_{-\infty}^{+\infty} f(t) \Psi_{u,s}(t) dt \quad (1)$$

Where wavelet Ψ function with zero mean and stretch parameter s and shift parameter u :

$$\Psi_{u,s}(t) = \frac{1}{\sqrt{s}} \Psi \left(\frac{t-u}{s} \right) \quad (2)$$

In our work we have used for CWT calculation algorithm from [7], which implement Morlet wavelet as time-frequency functions. Firstly, we used binary version of this algorithm based on powers of 2, to achieve the highest rate. The scale parameter s was changed as $s = 2^a 2^{frac{jJ}}$, where a - current octave, J - number of voices in a octave. We used $J = 8$. Secondly, the pseudo-wavelet was realized, which combines the averaging power of Fourier transform and accuracy of classical wavelet-transform. We used exponential change of base frequency and linear change of window size. This leads to the full correspondence of frequency scales of wavelet and pseudo-wavelet transform. In this case (1) transforms into:

$$W_{pseudo}f(u, s) = \int_{-\infty}^{+\infty} f(t) \rho_s(t-u)(t) dt \quad (3)$$

where $\rho_s(t)$ is a complex pseudo-wavelet with base frequency coordinated with wavelet frequency in scale s . The usage of pseudo-wavelets lets to average non-informative signal deviations during feature vector forming. In such a way we achieve higher accuracy for frequency analysis then it can achieved using FFT.

2.2 Feature Vector

The classification scheme is shown at fig.3. The result is a time-frequency signal representation. The image of wavelet transform for each segment is the source for future feature vector extraction procedure. There are a lot of methods to construct feature vector from CWT image, but it was proposed to use the simplest one for vocal fold pathology detection task. In order to do this we use the averaging of neighbor wavelet-coefficients on time-frequency scale. The whole time-frequency range divided on sub-ranges along time and frequency scales. Then coefficients inside each mosaic element are averaged and used as feature vector parameters(fig. 4).

2.3 Support Vector Machines (SVM)

SVM is a separating classifier, simple in its structure but effective. We use SVM for the voice pathology detection and classification as an optimal classifier. Distinction in kind of SVM to commonly used classifiers as hidden Markov models (HMM), Gaussian mixture models (GMM), is that SVM directly approximates between-class borders, not modeling probability distributions of training sets. SVM classifier is defined by the elements of the training set. But not all the elements are used for the classifier creation. Usually support vector's share is not big and classifier will be thinned. Training set defines the complexity of the classifier. Classification using SVM model is a simply calculation of vector relation to the border between classes, which was built during the training procedure.

Using the SVM as classifier for the task of vocal tract pathology detection is righteous due to following reasons:

- Speech signals classification for the task of voice pathologies detection can be described as a set of two-class classifications. Classifier structure in this case is a tree, where the first class contains of the most similar in structure pathologies and second class contains all others. Then classification in every of the classes is performed. It is also to perform classification of more than two classes optimizing SVM so, that all classes are processed simultaneously [8].
- Training sequence determines complexity and accuracy of the classifier. In our experiment we use feature vectors as training elements. Bigger differences between each element of two class's vectors make easier to build classes boundaries with the SVM classifier. Space dimension is equal to the dimension of the feature vectors.
- Recognition quality is sensible to the samples topology: compact distribution of the same class samples can help the recognition task. However, wider distribution of the samples leads to the recognition difficulties. Euclidian distance cannot help solving this problem. Training sequence should be well balanced. First, number of the records of both classes should be comparable. If one class is represented with much more records than another, classifier cannot build class boundaries correctly, and misclassification rate will be high. Each record contribution in the training sequence also has to be controlled to be equal to others, and all pathologies are represented adequately.

3 Experiment

For the common case, experiment of pathology recognition task consist of:

- Database creation. Database for pathology detection and recognition must contain records of many people with different types of pathologies and without any pathology. It is better if database contains records made on different languages, so classifier effectiveness and robustness can proved.

Fig. 3. Classification scheme using continuous wavelet transformation and SVM

Fig. 4. Feature vector creation

- Choosing speech signal parameters for feature vector creation. Former we must specify acoustic signal type and classifier structure.
- Creation of the model for good and pathology voices using database. Former we choose learning and parameters optimization procedures.
- Model evaluation. Data is separated into two parts: learning sequence and testing sequence. Learning part we use for model creation, testing sequence we use for evaluation.
- Using real voice signals for system evaluation. It can be speech of anybody in appropriate format.

3.1 Database Description

We use database which was created in Republic Center of Hearing, Voice and Speech Pathologies (Minsk, Belarus). All records represented in audio format PCM WAVE with 44 kHz sample rate and 16 bit sample size, mono. Patients were asked to read some text during several minutes. There were no any requirements about pronunciation, clearness articulation. Patients also didn't need to pronounce long vowels. Each record was specified a diagnose made by a phoniatriest after a patient check up using special equipment. Thus was created database of around 70 hours for good voices and around 20 hours of voices with pathologies. What distinguishes this database from others (for example free available database from Massachusetts hospital lab of voice and hearing) is that our database contains patterns for natural spontaneous voice records without pre-processing. Using this database guaranties good resembling of the experiment conditions to the situation of natural voice in noisy environment. Database was created of 90 speakers: 30 speakers with the normal voices, 30 speakers with the vocal cords neps and 30 speakers with the functional pathologies. All phrases have been processed with the speech-detector and contain just numbers (from "2" to "9").

3.2 Experimental Protocol

During the experiment speech signal was divided into separate words. Each word was parameterized and represented with 8×8 and 16×4 feature vectors of

Table 1. Classification of the normal voices and voices with vocal cord neps

WORD	INPUT SIGNAL	OUTPUT SVM 8×8		OUTPUT SVM 16×4	
		correct classification	wrong classification	correct classification	wrong classification
"2"	normal (20)	16	4	19	1
	pathology(20)	17	3	20	0
"3"	normal (20)	14	6	19	1
	pathology(20)	17	3	20	0
"4"	normal (20)	19	1	19	1
	pathology(20)	17	3	20	0
"5"	normal (20)	19	1	19	1
	pathology(20)	20	0	20	0
"6"	normal (20)	19	1	19	1
	pathology(20)	20	0	20	0
"7"	normal (20)	19	1	19	1
	pathology(20)	20	0	20	0
"8"	normal (20)	19	1	19	1
	pathology(20)	20	0	20	0
"9"	normal (20)	19	1	19	1
	pathology(20)	20	0	20	0
ALL	normal (160)	144(90.0%)	16(10.0%)	152(97.5%)	8(2.5%)
	pathology(160)	151(94.3%)	9(5.7%)	160(100%)	0(0.0%)

continuous wavelet transformation: in time-frequency domain each word is divided into 8 segments along time axis and 8 along frequency axis, and averaging is performed for each of 64 2D segments. In case of 16×4 feature vector the word is divided into 16 segments along frequency axis and 4 segments along time axis. Two SVM models were trained for the classification of the records belonging to speakers with the normal voices and speakers with the pathologies: model for the classification of the normal voices and voices with the vocal cords neps, model for the classification of the normal voices and voices with the functional pathology. Testing sequence went through the classifiers and according to the output segment belonging is decided.

3.3 Experimental Results

Table 1 presents results of classification of the normal voices and voices with the vocal cord neps. Correct classification rate reached for this task using continuous wavelet transformation feature vector of size: $8 \times 8 - 92.2\%((144 + 151)/(160 + 160))$ $16 \times 4 - 97.5\%((152 + 160)/(160 + 160))$. It can be noticed from the results that vector size 16×4 is preferable for the task of pathology detection.

Table 2 presents results of classification of the normal voices and voices with functional pathology. Correct classification rate reached for this task using continuous wavelet transformation feature vectors of size: $8 \times 8 - 93.4\%((145 + 154)/(160 + 160))$ $16 \times 4 - 97.5\%((152 + 160)/(160 + 160))$

Table 2. Classification of normal voices and voices with functional pathologies

WORD	INPUT SIGNAL	OUTPUT SVM 8×8		OUTPUT SVM 16×4	
		correct classi- fication	wrong classifi- cation	correct classi- fication	wrong classifi- cation
"2"	normal (20)	15	5	19	1
	pathology(20)	18	2	20	0
"3"	normal (20)	16	4	19	1
	pathology(20)	18	2	20	0
"4"	normal (20)	19	1	19	1
	pathology(20)	18	2	20	0
"5"	normal (20)	19	1	19	1
	pathology(20)	20	0	20	0
"6"	normal (20)	19	1	19	1
	pathology(20)	20	0	20	0
"7"	normal (20)	19	1	19	1
	pathology(20)	20	0	20	0
"8"	normal (20)	19	1	19	1
	pathology(20)	20	0	20	0
"9"	normal (20)	19	1	19	1
	pathology(20)	20	0	20	0
ALL	normal (160)	145(90.6%)	15(9.4%)	152(97.5%)	8(2.5%)
	pathology(160)	154(96.2%)	6(3.8%)	160(100%)	0(0.0%)

Certain decreasing in classification rate takes place in case of the type of pathology: the neps of the vocal cords or the functional pathology. For the case of pathology presence detecting (normal voice or pathological voice) correct classification reaches 90%. Archived results can be considered as encouraging for reasons:

- They show that pathology information can be caught by continuous wavelet transformation and SVM classifier even though there is a few speech material is available.
- It is possible to caught not just pathology presence but also predict the type of the pathology.

4 Conclusion

This article investigates the task of pathology recognition in voice signals using wavelets and SVM. It has been shown that acoustic analysis of recorded voices is capable of making decision about pathology presence and type in the signal. Building feature vectors from wavelet transformations is a very promising approach for the task of voice pathology detection. Adjusting parameters of the classifier to the optimal levels provides acceptable precision of normal and pathology voices classification. Obtained results prove that the proposed approach is able to work in case of not sufficient amount of learning data as

well. Following work in the defined direction will be devoted to recognition rate increasing using different types of SVM classifiers and signal parameterizations.

References

1. Alonso, J.B., de Leon, J., Alonso, I., Ferrer, M.A.: Automatic Detection of Pathologies in the Voice by HOS Based PArameters. *EURASIP Journal on Applied Signal Processing* 4, 275–284 (2001)
2. Gavidia-Ceballos, L., Hansen, J., Kaiser, J.: A Non-Linear Based Speech Feature Analysis Method with Application to Vocal Fold Pathology Assessment. *IEEE Trans. Biomedical Engineering* 45(3), 300–313
3. Manfredi, C.: Adaptive Noise Energy Estimation in Pathological Speech Signals. *IEEE Trans. Biomedical Engineering* 47(11), 1538–1543 (2000)
4. Wallen, E.J., Hansen, J.H.: A Screening Test for Speech Pathology Assessment Using Objective Quality Measures. In: *ICSLP 1996*, vol. 2, pp. 776–779 (1996)
5. Fredouille, C.: Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia). In: *Proc. of Eurospeech* (2005)
6. Maguire, C.: Identification of voice pathology using automated speech analysis. In: *Third International Workshop on Models and Analysis of Vocal Emission for Biomedical Applications*, Florence, Italy (2003)
7. Mallat, S.: *A wavelet tour of signal processing*. Academic, San Diego (1998)
8. Cristianini, N., Shawe-taylor, J.: *Introduction to Support Vector Machines*, p. 139. Cambridge University Press, Cambridge (2001)