Voice Disorder Detection on the Basis of Continuous Speech

Conference Paper · January 2011

DOI: 10.1007/978-3-642-23508-5_24

3 authors, including:

Klara Vicsi
Budapest University of Technology and Economics
115 PUBLICATIONS **695** CITATIONS

SEE PROFILE

Voice Disorder Detection on the Basis of Continuous Speech

Vicsi, Klára[1], Imre, Viktor[1] and Dr. Mészáros, Krisztina[2]

[1] Laboratory of Speech Acoustics, Budapest University of Technology and Economics,
Department of Telecommunications and Media Informatics
[2] National Institute of Oncology, Department of Head and Neck Surgery

Abstract— **During vocal diagnostical analyses there were made more examinations in connection with that question, whether sustained voice, or continuous speech is more effective in distinguishing healthy voice from pathological voice. Now, it is presented by a series of classification experiments, that how it is possible to separate the healthy speech from the pathological one automatically, on the base of continuous speech. It is cleared out, that by a multi-step processing methodology, most of those examples at which uncertainties occurred at the measurement of the acoustical parameters can be grouped separately, and in this way, the automatic classification results of the healthy and pathological voices are improved in a big extent.**

Keywords— **pathological voices, voice disorder detection, automatic speech recognition, support vector machine**

I. Introduction

Generally in voice production, there is a close connection between the mutations of voice generation organs (differences in size, in tissue flexibility, ect.), and the measurable acoustical parameters (fundamental frequency, sound pressure, spectrum, ect.) of the generated speech product.

During vocal diagnostical analyses there were made more examinations in connection with that question, whether sustained voice, or continuous speech is more effective in distinguishing healthy voice from pathological voice[1], [2], [3] ,[4], [5],[6];) .

In phoniatrical practice mainly continuous speech is applied by the phoniatric specialists for classification of voice quality based on hearing. It is not accidental, because for generating speech there is a need for the cooperation of other important utterance functions besides vibration of vocal chords, and thus in case of any disorders clearness of the voice can easily disrupt. On the other hand, there is a possibility in continuous speech for the observation of the so called supra-segmental characteristics: emphasis, intonation, duration of sonants.

Let us examine the possibilities of analysis of continuous speech and sustained voice from the viewpoint of acoustic measurements. According to the claim of Rabinov et al [7] the most reliable „tool" for the evaluation of voice quality is the human ear, after all. It can be explained with the fact, that during measurement of oscillation of amplitude and frequency of vibration of vocal chords these parameters did not take into consideration the shape of generated voice waves, and that vibration of vocal chords comes together with a frictional noise. These issues may contain relevant information, mainly in case of a pathological voice.

Titze [8] suggests in his study, that acoustic measurements (jitter, shimmer) can provide authentic information only in case of the of sustained voice type, because those voices provide the most authentic jitter and shimmer results, because the characterization of periodicity can be determined easily because of quasi-periodicity of the signal, and sustained voice contains enough period numbers for the authentic calculation of oscillations. On the contrary, in case of analysis of continuous speech, where the length of examined sections is very short because of the quick voice transitions, it is less applicable.

Zhang & Jiang [1] examined the acoustics characteristics of sustained voice and continuous speech for distinguishing the healthy and pathological voice. Acoustic parameters: jitter, shimmer and HNR values were taken into consideration. They demonstrated that continuous speech is less suitable for distinction healthy and pathological voice.

While in phoniatrical practice mainly continuous speech is applied by doctors, we also want to examine which kind of sound material, the continuous speech or the sustained vowels are the better basic material for the acoustical parameters and for the automatic separation of normal and pathological speech. First a detailed statistical analyses of acoustical parameters of vowels in continuous speech and sustained voice databases were examined, and compared the results in case of healthy speech and pathological ones (Imre 2009),(Vicsi-Imre 2010). Now, in this lecture we present our classification experiments, how it is possible to separate the healthy speech from the pathological one automatically on the base of continuous speech.

Á. Jobbágy (Ed.): 5th European IFMBE Conference, IFMBE Proceedings 37, pp. 86–89, 2011.
www.springerlink.com

II. Pathological and healthy speech databases

A well processed pathological and healthy speech databases construction was necessary for the examination and for the automatic separation of the healthy and pathological voices.

The sound records were made at National Institute of Oncology. The following illness occurred in the recorded database: functional dysphonia, recurrent paresis, a tumor at different places of vocal tract, gastro-oesophagealis reflux disease, chronical inflammation of larynx, bulbar paresis, its symptoms(paralysis of lips, tongue, soft palate, pharnyx and muscles of larynx), amyotrophic lateral sclerosis, leukoplakia, spasmodic dysphonia and glossectomia**.** Records, for comparison, were prepared with absolutely healthy patients too, who had gone to the consulting hours only for control.

Speech were recorded by nearfield microphone (Monacor ECM-100), with Creative Soundblaster Audigy 2 NX: an outer USB sound card with 44100Hz, a 16-bit sampling rate.

The following tasks were recorded from each patients: 3 „o" vowel sustained for a long time, with a large breathing before the utterance of each of them, and reading of a folk tale, frequently used in the phoniatric practice, the „ Northern wind, and the sun".

The recorded sound examples were classified by a leading phoniatric specialist by a sound perception evaluation scale RBH*, which is a popular scale in the practice of phoniatry. The scale classifies the voice samples to 4 classes on the basis of a subjectively felt parameters provided by the RBH code. This scale was used, to differentiate the degree of the voice generation disorders in the database. Speech examples of patients were labeled on the base of this numeric scale.

While predetermined voiced sequences of the continuous speech were planned to process, a phoneme-leveled segmentation of voice files was necessary. It was made on a semi-automatic way, using our automatic speech recognizer.

The continuous speech part (folk tail) of 59 speakers were used now for the classification experiment (33 pathological and 26 healthy speakers).

**RBH ((Rauhigkeit)* (roughness) *(Behauchtkeit)* (breathing) *(Heiserkeit)* (hoarseness)*): 0 = normal voice quality, 3 = heavy huskiness.*

III. Measured acouastical parameters

Earlier it was examined (Vicsi-Imre 2010), which acoustical analyzing methods reflect the degree of the voice generation disorders better (or which reflect the sound perception evaluation RBH scale). Statistical distributions of the acoustical parameters were examined by measuring these parameters at sustained vowels and at the middle of the vowels in continuous speech. In this experiment it was found, that the selected acoustical parameters in the quasi-stationer part of the vowels in continuous speech can represent the perceptual classification of experts much better than in the traditionally used steady state sounds. The measured and analyzed parameters were used for the classification experiments too:

jitter: This is the average absolute difference between consecutive time of periods (T) in speech, divided by the average time period. Generally two form of jitter are in the practice:

$$\text{jitter}_{\text{local}} = \frac{\sum_{i=1}^{N-1} |T_i - T_{i+1}|}{\sum_{i=1}^{N-1} T_i} \cdot 100\ [\%] \quad (1)$$

$$\text{jitter}_{ddp} = \frac{\sum_{i=2}^{N-1} |2T_i - T_{i-1} - T_{i+1}|}{\sum_{i=2}^{N-1} T_i} \cdot 100[\%] \quad (2)$$

where N is the number of periods, and T is the time of the periods.

shimmer: This is the average absolute difference between consecutive differences between the amplitudes of consecutive periods.

$$\text{shimmer}_{\text{local}} = \frac{\sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\sum_{i=1}^{N-1} A_i} \cdot 100\ [\%] \quad (3)$$

$$\text{shimmer}_{dda} = \frac{\sum_{i=2}^{N-1} |2A_i - A_{i-1} - A_{i+1}|}{\sum_{i=2}^{N-1} A_i} \cdot 100[\%] \quad (4)$$

where A is the amplitude of the period.

HNR: Harmonics-to-Noise Ratio. Represents the degree of acoustic periodicity.

$$\text{HNR} = 10 * \log \frac{E_H}{E_Z}\ [\text{dB}] \quad (5)$$

where E_H and E_Z is the energy of the harmonic and noise component respectively.

IV. Classification of healthy and pathological voices in continuous speech

All of the sounds „e" in the reading test were used for the classification. At the middle of the „e" sound the following acoustical parameters were measured: the local jitter, ddp jitter, local shimmer, dda shimmer and HNR values. The average, the spread, min, max, and median of the measured values were calculated. These vectors were the input of the classifier. A special neural net, the Support Vector Machine (LibSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm/)) were used for the classification, and the Leave-One-Out Cross-

Validation (LOOCV) technics was used for the training and testing. For the selection of the most important parameters a series of test experiments were prepared. Different groups of the acoustical parameters were used and the recognition (classification) results were examined.

Table 1. Testing of the classifier with different acoustical incoming parameters.(jitter = jitter (local) and jitter (ddp) together; shimmer = shimmer(local) and shimmer(dda) together)

acoustical parameters	statistics for sound „e”	
	average	average, spread, min, max, median
jitter, shimmer, hnr, mfcc	73%	63%
jitter, shimmer, mfcc	73%	63%
jitter, shimmer	79%	79%
jitter(local), shimmer(local)	79%	79%
jitter(ddp), shimmer(dda)	**84%**	79%
jitter, shimmer, hnr	73%	73%
hnr, mfcc	68%	63%
mfcc	73%	63%

The best classification results of the healthy and pathological speech were obtained when the average of jitter(ddp), shimmer(dda) were the incoming acoustical parameter. See Table 1. We wanted to analyse this result further, how the healthy examples were separated from the pathological cases on the base of these two parameters. Thus we plotted the spread values into the function of average values in the case of jitter(ddp) and shimmer(dda). See it in Fig.1. and Fig. 2.

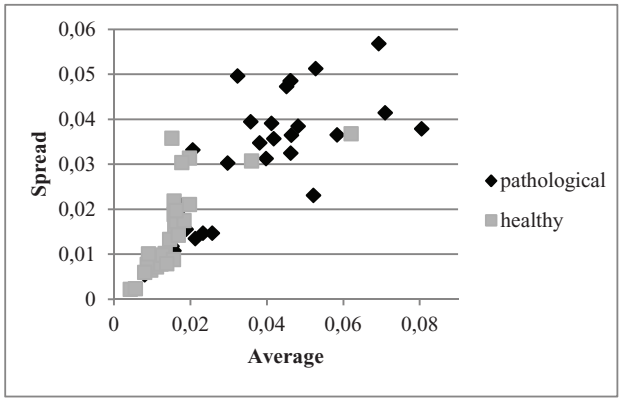

Fig. 1. Spread of jitter (ddp) in the function of the average in case of sound „e”

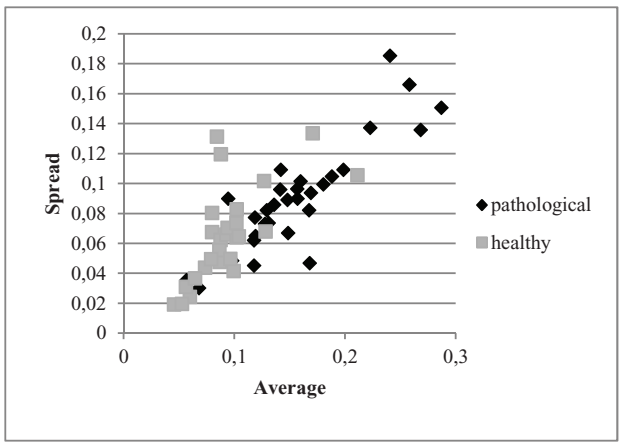

Fig. 2. Spread of shimmer (dda) in the function of the average, in case of sound „e”

There are a few salient examples in case of the spread of the jitter and the shimmer too. Analyzing these examples one by one, it cleared out that the origin of all of those salient examples came from measuring problem. Either the fundamental frequency was too low or too high, or the voiced period in speech was very small. In that case when the fundamental frequency was too high, the fluctuation in % was smaller than expected. In case of too low frequencies, the measurement of the fundamental frequencies became ambiguous, making mistakes. Both the jitter and shimmer measurements are based on the measurement of fundamental frequency, thus in the before mentioned cases we get senseless results, giving salient results. In the case when the voiced period in speech was very small, much smaller that the unvoiced one, inadequate examples cause mistakes. At the first moment it seems that Zhang and Jiang [1] are right, when they told that there are small examples in continuous pathological speech for the authentic calculation of the jitter and shimmer parameter, but the calculation difficulty occurs only at some examples, and those examples can be selected and evaluated in a different way.

V. Separation of the uncertain examples

In the first step, it was necessary to decide the threshold of the voiced/unvoiced frame rate, under which the examples are selected. Thus voiced/unvoiced frame rate were calculated in continuous part of the speech of the patients. 75 ms window was used with 18.75 ms frame steps. The voiced/unvoiced frame rate was calculated as follows:

$$Fr_{v/u} = \frac{\sum_{i=1}^{N} Fr(v)_i}{\sum_{i=1}^{N} Fr(u)_i} \quad (6)$$

where $Fr(v)_i$, is the number of voiced frames, and $Fr(u)_i$ is the number of unvoiced frames of the i-th sound. The distribution of these voiced/unvoiced frame rate in case of healthy and pathological voices is presented in Figure 3.

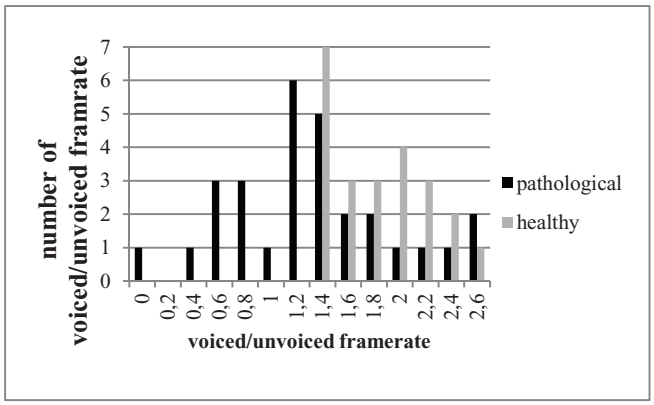

Fig. 3. Distribution of the voiced/unvoiced frame rate of the healthy and pathological voices

It is quite clear from the measurement that healthy speech does not exist under 1,4 frame rate. The quality of these sounds is the worst. The error of the measurement of the fundamental frequency is high. Those examples, where the voiced/unvoiced frame rate was less than 0,5 was filtered out, those are surely pathological voices. Then the classification was repeated, and now the classification exactness increased to 86 %.

In the second step, on the base of the distribution of fundamental frequencies, those examples were separated, where the fundamental frequency was higher than 160 Hz in case of men, and 270 Hz of women. These examples are also pathological voices, because in these cases the patients wanted to make their hoarse voices better by increasing the fundamental frequency. The classification was repeated again, and now its exactness increased to 88 %.

VI. Conclusion

It was impressive for us how easily we could increase the exactness of the classification, by avoiding some measuring problems. In spite of the opinion of Rabinov et al [7] that the most reliable „tool" for the evaluation of voice quality is the human ear, and the results of Zhang and Jiang [1], it is clear for us, that it is worth to go further in our way and use continuous speech for the detection and classification of pathological voices. Of course much more data collections have to be prepared for obtaining better results, and further investigation is necessary to decide which acoustical parameters are the best for the classification.

References

1. Yu Zhang, Jack J. Jiang: Acoustic Analyses of Sustained and Running Voices From Patients With Laryngeal Pathologies, accepted for publication: 2006, Journal of Voice, Vol. 22, No. 1, pp. 1-9, 0892-1997/$
2. Ce Peng, Wenxi Chen, Xin Zhu, Baikun Wan, Daming Wei: Pathological voice classification based on a single vowel's acoustic features, 0-7695-2983-6/07 $25.00, 2007 IEEE
3. Parsa, V.and Jamieson, D. G. : Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech, Journal of Speech, Language, and Hearing Research 44, 2001, p. 327–339.
4. Anders G. Askenfelt, Britta Hammarberg: Speech Waveform Perturbation Analysis, A Perceptual-Acoustical Comparison of Seven Measures , Journal of Speech and Hearing Research Vol.29 50-64 March 1986.
5. Ce Peng, Wenxi Chen, Xin Zhu, Baikun Wan, Daming Wei: Pathological Voice Classification Based on a Single Vowels's Acoustic Features, 7th.International Conf. on Computer and Information Techn, IEEE, 2007. 1106-1110.
6. R. T. Ritchings, M. McGillion, C. J. Moore: Pathological voice quality assessment using artificial neural networks, 2002, Medical Engineering & Physics 24 (2002) 561-564
7. Rabinov et al (1995) Virginia Wolfe és David Martin: Acoustic correlates of dysphonia: type and severity, J. COMMUN. DISORD. 30 (1997), 403-416, Elsevier Science Inc. 1997
8. Titze, I., Wong, D., Milder, M., Hensley, S., & Ramig, L. (1995). Comparison between clinician-assisted and fully automated procedures for obtaining a voice range profile. J. Speech Hear. Res., 35,526-535.
9. Imre, Viktor: Acoustical examination of pathological voices, Diploma work. Budapest University of Technology and Economics 2009.
10. Vicsi, Klára, Imre, Viktor: Voice disorder detection on the base of continuous speech, 4th Advanced Voice Function Assessment Workshop, COST Action 2103. York 2010, pp.42.

Author: Vicsi, Klára

Institute: Laboratory of Speech Acoustics, Budapest University of Technology and EconomicsDepartment of Telecommunications and Media Informatics
Street: Magyar tudósok körútja 2.
City: Budapest
Country: Hungary
Email: vicsi@tmit.bme.hu