TREMOR IN SPEAKERS WITH SPASMODIC DYSPHONIA

Maria Koutsogiannaki, Yannis Pantazis, Yannis Stylianou [1] and Philippe Dejonckere [2]
[1]Institute of Computer Science, FORTH, and Multimedia Informatics Lab, CSD, UoC, Greece
[2] Utrecht University, Utrecht, The Netherlands
email: {mkoutsog, pantazis, yannis}@csd.uoc.gr, Philippe.Dejonckere@fmp-fbz.fgov.be

***Abstract*—The objective of this work is the estimation of vocal tremor in patients with spasmodic dysphonia before and after treatment, and the comparison of their tremor characteristics with those estimated from healthy speakers. As an outcome, a new tremor attribute is introduced, the deviation of the modulation level and a novel method is proposed for classifying speakers according to the prevalence of tremor in their voice. Results are consistent with subjective evaluations on patients who suffer from spasmodic dysphonia and confirm that the proposed method can be used for accurate estimation and objective ranking of the severity of tremor.**

***Index Terms*—Voice quality, vocal tremor, dysphonia, deviation of modulation level**

I. Introduction

Vocal tremor, a rhythmic change in pitch and loudness, appears both in healthy speakers and in speakers with voice disorders. In normal speaking voice, no tremor is audible, but it can be elicited by emotions, either spontaneous or volitional (actors). Central (mostly degenerative) neurological diseases, particularly those involving cerebellum and basal ganglia, frequently elicit voice tremor. In spasmodic dysphonia (or laryngeal dystonia), task-related tremor ("spasms") may considerably hamper fluency and intelligibility [1]. This work focuses on estimating tremor in speakers with spasmodic dysphonia before and after treatment, and compares their tremor characteristics (level and frequency) with those estimated from healthy speakers.

Acoustic analysis of tremor is usually based on the accurate estimation of fundamental frequency and then the characterization of the fundamental frequency's variations [2], [3]. Modulation frequency and modulation level are prominent attributes that are extracted from the instantaneous fundamental frequency [2], [3]. Previous studies in tremor analysis assume modulation frequency and modulation level being as time-invariant characteristics of tremor, by considering short-time analysis windows of speech. Then, stationary frequency estimation approaches are used for the estimation of these tremor attributes, like the classical Fourier transform. However, tremor characteristics and in general modulations in speech are time-varying. Actually, analysis of large segments of speech showed interesting time-varying characteristics on vocal tremor [4], [5].

The detection of tremor attributes in a speech signal involves the accurate extraction of the signal that modulates the time-varying fundamental frequency. We employ a recently proposed method to extract time-varying tremor attributes; the level and the frequency of the modulating signal [6]. This method is applied to sustained vowels and decomposes the speech signal into its time-varying quasi-harmonics. Quasi-harmonics are components with frequencies which are near to be harmonics of a fundamental frequency. It has been shown that speech is better modeled as a sum of quasi-harmonics rather than a sum of harmonics [7]. Next, we will refer to the components rather to harmonics. After the decomposition of speech into components, one component is chosen for further analysis; the desired signal that modulates the component is extracted and its time-varying amplitude and frequency are estimated.

This method is applied in speech vowels uttered by normophonic speakers and speakers who suffer from spasmodic dysphonia before and after imposed on medical treatment [8]. Our analysis shows that the mean modulation level in dysphonic speakers is distinguishably greater than that in normophonic speakers. However, the modulation level is not the only criterion for classifying speakers as normophonic or dysphonic. This study introduces a novel attribute of tremor which derives from the time-varying characteristic of the modulation level, namely the deviation of the modulation level. The mean modulation level and its deviation are combined in a quality indicator trying to classify speakers according to the amount of tremor in their voice. It is shown that this objective classification of speakers matches subjective evaluations by experts in the case of spasmodic dysphonia patients.

The organization of the paper is as follows. Section II describes briefly the tremor estimation method. Section III presents the analysis on normophonic and dysphonic speakers, introduces the proposed tremor classification method and compares the results with the subjective evaluations. Finally, Section IV concludes the paper.

II. Estimation of Vocal Tremor

The method used for tremor features estimation assumes speech as a sum of time-varying sinusoids [7], [9]. The extraction of vocal tremor characteristics is carried out in three steps, following the procedure in [6]. The first step estimates the instantaneous amplitude and instantaneous frequency of every sinusoid component of the speech signal using a recently proposed AM-FM decomposition algorithm, the so-called Adaptive Quasi-Harmonic Model (AQHM) [7], [9].

AQHM is an adaptive algorithm which is able to represent accurately multi-component AM-FM signals like speech. In the second step, the very slow modulations ($< 2Hz$), derived mainly from the pulsation of the heart, are subtracted from the instantaneous component. This is achieved by filtering the instantaneous component using a Savitzky-Golay smoothing filter [10]. In the final step, the time-varying modulation frequency and the time-varying modulation amplitude of the analyzed instantaneous component are estimated by employing again the AQHM algorithm for just one component. The time-varying modulation amplitude with an appropriate scaling corresponds to the modulation level. The scaling is necessary because the modulation amplitude is relative to the mean value of the instantaneous component and involves the normalization of the amplitude by this mean value. More details of the estimation algorithm are provided in [6].

III. Results

A. Data Analysis

The suggested tremor estimation method, as described in Section II, is applied to two different databases of sustained vowels to extract the time-varying modulation level and the time-varying modulation frequency. The first database consists of sixteen healthy subjects. Sustained vowels $/a/$, $/e/$, $/i/$, $/o/$ and $/u/$ of varying duration ($2s-8s$) have been recorded. The second database was provided by the last coauthor (Prof. P. Dejonckere). Speakers in this database suffer from spasmodic dysphonia and are subjected to treatment (botulinum toxin injections). Recordings and subjective evaluations by experts have been made before and after the treatment. For every patient, the sustained vowels of $/a/$ are extracted to create the signals for our analysis. In the current study, five untreated speakers could not be analyzed because they could only provide phonemes with very limited duration (less than a second).

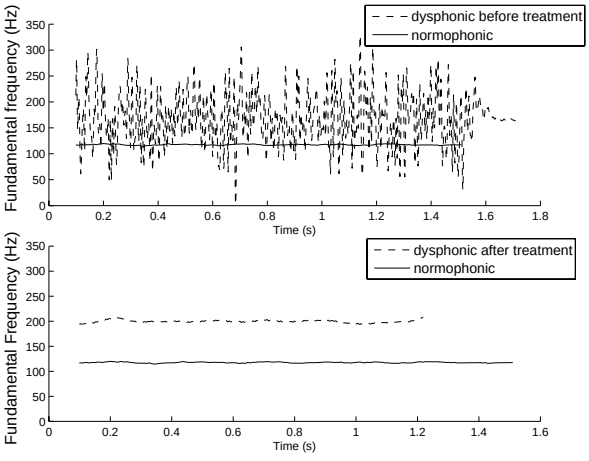

Fig. 1. The time-varying instantaneous component of a normophonic speaker and of a speaker with spasmodic dysphonia before and after treatment.

The upper panel of Fig.1 shows a typical example for the time-varying frequency characteristics of the first component (nearly the fundamental frequency) for a dysphonic and a normophonic male speaker. It is worth noticing the high fluctuations of the component for the case of the dysphonic speaker in contradiction to that of the healthy speaker who keeps his voice almost steady in time. After treatment the dysphonic speaker achieves to stabilize his voice (lower panel of Fig.1). The tremor attributes of these signals, the modulation level and the modulation frequency, are depicted in Fig.2. The upper panel of Fig.2 shows the time-varying modulation levels of a normophonic and that of a dysphonic speaker before and after his treatment. The lower panel of Fig.2 depicts the corresponding modulation frequencies. As it can been seen, the normophonic speaker appears to have much lower mean modulation level than the dysphonic speaker before treatment. Moreover, the modulation level of the dysphonic speaker before treatment presents high fluctuations over time. After treatment, both speakers have similar modulation levels; the modulation level of the treated dysphonic speaker has decreased significantly, meaning that the tremor is no longer audible after treatment. In all cases, modulation frequency values are quite comparable (lower panel in Fig.2).

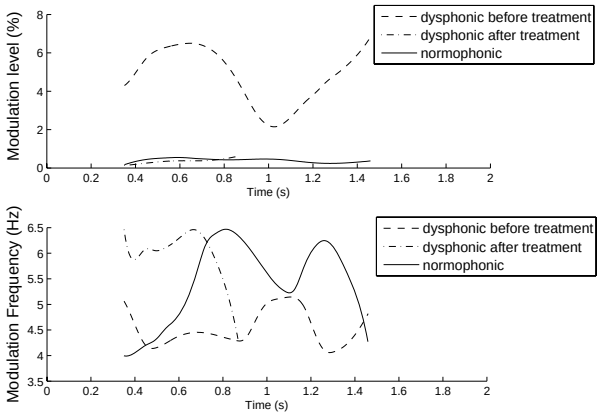

Fig. 2. Modulation level and modulation frequency of a normophonic speaker and of a speaker with spasmodic dysphonia before and after treatment.

Significant results derive from the analysis of the two databases. Fig.3 shows the mean values of the two time-varying tremor attributes for every normophonic speaker in the first database; the mean modulation level (upper panel of Fig.3) and the mean modulation frequency (lower panel of Fig.3). Frequencies vary from $2-7Hz$, while the mean modulation levels are all but one below 1% of the mean value of the instantaneous component for the corresponding normophonic speakers. In a similar way, the upper panel of Fig. 4 shows the mean modulation levels and the lower panel of Fig.4 the mean modulation frequencies for dysphonic speakers before and after their treatment. Comparing Fig.4 and Fig.3 it can be seen that the modulation frequencies are quite comparable for the normophonic and dysphonic speakers. However, this is not true for the modulation level. Indeed, five out of six untreated dysphonic speakers have modulation level above 1% and seven out of nine treated dysphonic speakers have modulation level below 1%. This is more evident in Fig.5, where the modulation level for each dysphonic speaker before

and after treatment is illustrated. For the speakers coded as Lul, Roo and Stu the modulation level has decreased after treatment, while for Bru and Vro there is a slight increase in the modulation level after the treatment. The general trend, however, is that the treated patients have modulation level values below 1% of the mean value of their component and this is comparable with that of the normophonic speakers.

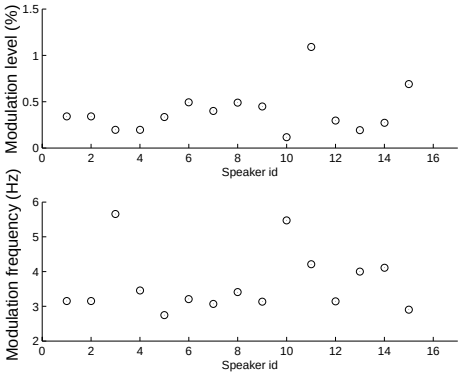

Fig. 3. Modulation levels and modulation frequencies of normophonic speakers.

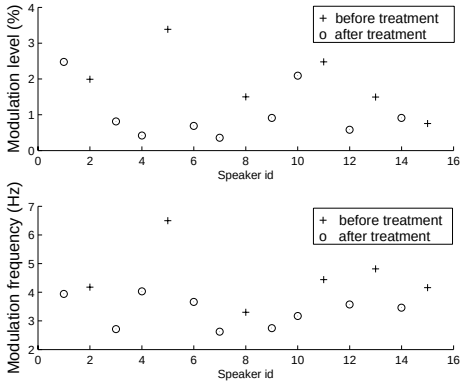

Fig. 4. Modulation levels and modulation frequencies of treated and untreated dysphonic speakers.

As illustrated by the evolution of the modulation level in the upper panel of Fig.2, the deviation of the modulation level from its mean value is quite high in the case of the dysphonic speaker before treatment. This was also observed in other dysphonic speakers from the same database. Based on this observation a new characteristic of tremor is introduced, which will be referred to as deviation of modulation level, or DML. It is worth noticing that this new tremor attribute is based on the capability of the suggested tremor-estimator to produce time-varying modulation frequency and modulation level, overcoming the limitations of short signal duration.

Fig.6 combines the two characteristics, the modulation level and the DML in one graph for normophonic and dysphonic speakers; each data point has two tremor coordinates; the DML and the mean modulation level. The arrows show the

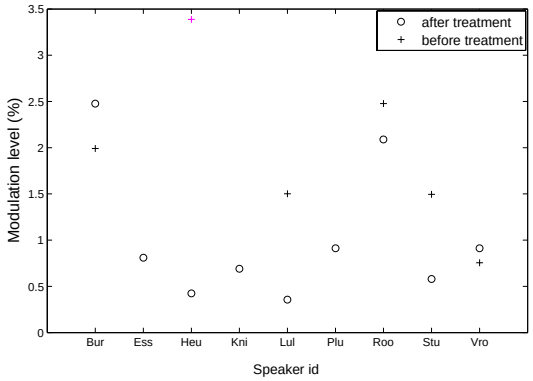

Fig. 5. Modulation level of each dysphonic speaker before and after treatment.

change of the tremor coordinates for dysphonic speakers after treatment. The beginning of the arrow corresponds to the tremor coordinates of the dysphonic speaker before treatment and the end of the arrow to the tremor coordinates after treatment. Each arrow is named after the speaker. The normophonic speakers occupy the low left part of the graph, where the modulation level and the DML take low values, defining therefore a "normophonic area" of these attributes. As it is shown in Fig.6, the untreated dysphonic speakers diverge from the normophonic area. The dysphonic speakers after treatment tend to reach the normophonic region as the arrows show. However, some patients (Roo, Bur) seem to have no improvement. Notice that for some treated speakers there is no estimation of their previous state (before treatment) since, due to the severity of their disease, their phonemes could not be analyzed (small signal duration).

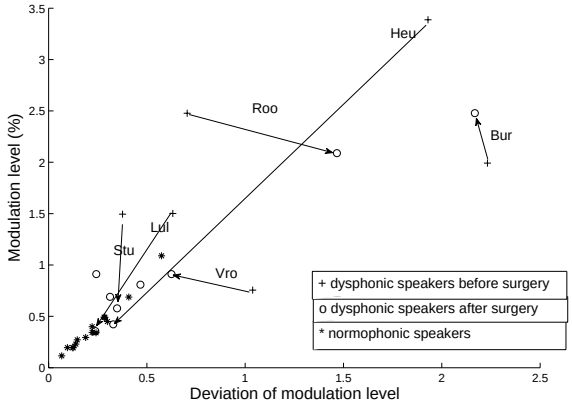

Fig. 6. Mean modulation level as a function of its deviation for normophonic speakers and for speakers with spasmodic dysphonia before and after treatment.

The above analysis suggests that the modulation level and the its corresponding deviation are significant values defining tremor and these attributes can be used either for classifying speakers as normophonic or dysphonic, or for classifying speakers according to the severity of dysphonia. Furthermore,

it may be used as an objective measure for the patient's progress evaluation before and after treatment.

B. Objective Tremor Classification Method

As we saw in the previous section, the tremor signal that modulates an instantaneous component differs significantly in a healthy and in a dysphonic speaker. The outcome of our analysis is that the instantaneous modulation level of the untreated patients with spasmodic dysphonia present high variations in time. A dysphonic speaker appears to have higher modulation level and significant DML than a normophonic speaker. Therefore, we suggest the introduction of a quality indicator that classifies speakers according to their tremor value in their voice. The quality indicator is called Weighted Mean Tremor Value (WMTV) and is defined as:

$$WMTV = w\overline{x} + (1-w)\sigma(x) \ , \qquad (1)$$

where $\overline{x}$ is the mean modulation level, $\sigma(x)$ the standard deviation of the modulation level of the tremor signal, and w is a weighting factor.

The severity of the spasmodicity of each speaker is ranked using the WMTV with a 40% weighting factor, deriving from the analysis. Our classification is compared with the subjective ranking of tremor for the same speakers and same speech files. Both classifications are presented in Table I. The subjective evaluation was conducted by specialized doctors. In Table I the "-pre" ending corresponds to dysphonic speakers before treatment and the "-pos" to the dysphonic speakers after treatment. For instance, speaker Bur, according to the subjective evaluations, had a slight enhancement after surgery (from 1.00-Burpre to 0.94-Burpos). Notice, that there are differences in the subjective and in the proposed objective classification. However, both evaluations "separate" the patients with severe tremor. For example, both evaluations agree that patients Bur and Roo have high tremor despite treatment and that patients Heu, Stu, Lul, Plu and Ess have low tremor values after treatment. It is found that the correlation between our ranking and the subjective ranking is significant; the correlation coefficient is 0.72 and the p-value is 0.0024.

A) Subjective classification		B) Proposed classification	
	Normalized TR		WMTV
Burpre	1.00	Heupre	1.00
Burpos	0.94	Burpos	0.91
Roopre	0.82	Burpre	0.85
Stupre	0.71	Roopos	0.68
Roopos	0.59	Roopre	0.56
Vropre	0.53	Lulpre	0.39
Vropos	0.47	Vropre	0.37
Heupre	0.41	Stupre	0.33
Knipos	0.41	Vropos	0.30
Lulpre	0.24	Esspos	0.24
Plupos	0.12	Plupos	0.20
Esspos	0.06	Knipos	0.18
Heupos	0.06	Stupos	0.18
Lulpos	0.06	Heupos	0.15
Stupos	0.0	Lulpos	0.11

TABLE I
DYSPHONIC SPEAKERS CLASSIFICATION BASED ON: A) SUBJECTIVE EVALUATION, B) DESCENDING WMTV (WEIGHTING FACTOR = 40%)

Fig.7 compares the two evaluations. The ideal match between the two evaluations is the solid line. The closer the markers are to the line the more our method agrees with the subjective evaluations.

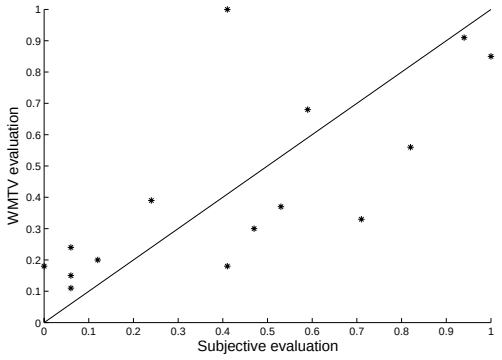

Fig. 7. Subjective evaluation to WMTV evaluation. The solid line corresponds to the ideal match between the two evaluations.

IV. CONCLUSION

Our proposed method aims at estimating tremor in speakers with spasmodic dysphonia. Evaluation results show that it achieves to estimate accurately the time-varying characteristics of tremor. From the analysis in normophonic and dysphonic speakers, a new tremor attribute is introduced, the deviation of the modulation level. This attribute derives from the time-varying characteristics of the modulation level and plays a prominent role in the objective classification of speakers according to their tremor. The two significant attributes, the modulation level and its deviation are combined in one value; the weighted mean tremor value, or WMTV. It was shown that WMTV is a quality indicator of tremor in voice and can be used as an objective measure for evaluating speakers with spasmodic dysphonia.

REFERENCES

[1] P. H. Dejonckere, K. J. Neumann, M.B.J. Moerman, and J.P. Martens. Perceptual and Acoustic Assesment of Adductor Spasmodic Dysphonia Pre-and PostTreatment with Botulinum Toxin. Proceedings Madrid, 2009.

[2] W. S. Winholtz and L. O. Ramig. Vocal tremor analysis with the vocal demodulator. *Journal of Speech Hearing Research*, 35:562–573, 1992.

[3] J. Schoentgen. Stohastic models of jitter. *Journal of Acoustic Society of America*, 109:1631–1650, 2001.

[4] J. Kreiman, B. Gabelman, and B.R. Gerratt. Perception of vocal tremor. *Journal of Speech, Language and Hearing Research*, 46:203–214, 2003.

[5] H. Ackermann and W. Zeigler. Acoustic analysis of vocal instability in cerebellar dysfunctions. *Annals of Otology, Rhinology and Laryngology*, 103:98–104, 1994.

[6] Y. Pantazis, M. Koutsogiannaki, and Y. Stylianou. A Novel Method for the Extraction of Vocal Tremor. In *MAVEBA*, Florence, 2009.

[7] Y. Pantazis, O. Rosec, and Y. Stylianou. Adaptive AM-FM Signal Decomposition with Application to Speech Analysis. *IEEE Trans. on Audio Speech and Language Processing*, 19(2):290–300, February 2011.

[8] D.I. S Luhring, M. Moerman, J.P. Martens, D. Deuster, F. Muller, and P. Dejonckere. Spasmodic Dysphonia, Perceptual and Acoustic Analysis: Presenting New Diagnostic Tools. *Eur Arch Otorhinolaryngol*, 2009.

[9] Y. Pantazis, O. Rosec, and Y. Stylianou. AM-FM Estimation for Speech based on a Time-varying Sinusoidal Model. In *Interspeech*, Brighton, 2009.

[10] A. Savitzky and M.J.E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36:1627–1639, 1964.