BIO-MECHANICAL CHARACTERIZATION OF VOICE FOR SMOKING DETECTION

S. Ben Jebara

COSIM Laboratory
Ecole Supérieure des Communications de Tunis
Carthage University
Route de Raoued 3.5 Km, Cité El Ghazala, Ariana, 2088, TUNISIA
sofia.benjebara@supcom.rnu.tn

ABSTRACT

The purpose of this work is to discriminate between smoker and non-smoker speakers by analyzing their voice. In fact, the vocal folds, the main organ responsible of producing voice, is damaged by smoke so that its structure and its vibration are altered. Some bio-mechanical features, describing vocals folds behavior and status are used. They are based on the two-mass model which characterizes vocal folds by the mass, the stiffness and the losses of their cover and body parts. Bio-mechanical features of smokers and non-smokers are analyzed and compared to select relevant features permitting to discriminate between the two categories of speakers. The Quadratic Discriminant Analysis is used as a tool of classification and shows a relatively good rate of detection of smokers.

***Index Terms*—** Smoking detection, voice analysis, bio-mechanical features

1. INTRODUCTION

Smoking and exposure to smoke has a negative impact on health (see for example [1][2]): it damages the mucous-producing glands and delicate cell lining of the mouth, the throat and the vocal folds. Gastric ulcer and gastric reflux are frequent for persistent smokers. The Reinke's space, the upper layer of the covering of the vocal fold, is enlarged by the accumulation of gelatinous fluid. Pre-cancerous growths or plaques appear as irregular thickening of the vocal fold cover. It can become extensive and go beyond to intermediate and deep layers, signaling early carcinoma.

The consequence of smoking on voice is easily noticeable: it may sound weak, breathy, scratchy, husky, strained... Moreover, the pitch which is the frequency of vocal folds vibration is lowered. For that reason, women more frequently notice the symptoms than men, who already have a low-pitched voice.
The idea to detect the negative effects of smoking from voice analysis is inspired from theses observations: instead of using invasive techniques such as videoscopy to observe vocal folds, one can analyse voice by extracting some features which can indicate the vocal folds status.

Voice analysis for smoking detection purposes begins to be timidly developed in the literature. Some previous works mainly examined and compared some acoustic voice parameters. In [3], it was shown that the fundamental frequency is lower and average jitter and shimmer are higher for smokers than non smokers. In [4], the effect of smoking habit at a relatively early stage is also studied. It was noticed that main differences are observed in frequency perturbation parameters for both genders, in fundamental frequency parameters for women and in tremor parameters for men. Another study [5] considered the phenomena of reversibility, which means how vocal parameters change following a period of abstinence from cigarette smoking. It was observed that jitter and shimmer decreased significantly, whereas Harmonic to Noise Ratio (HNR) increased. In [6], it seemed that the increased time of silence during connected speech of smokers is related to the defective quality of the closed phase of vocal cords movement.

In this work, another kind of parametric characterization of vocal folds is used to study the effect of smoke on vocal folds. It is based on a mechanical modeling of vocal folds which enables to define some bio-mechanical features. Physically, they describe the mass, the stiffness and the losses in the cover part and in the body part of vocal folds. Note that this kind of parametrization is used for other purposes such as forensic applications, vocal folds pathologies characterization, neuronal disease detection,...

The paper is organized as follows. Section 2 describes the origin of the bio-mechanical characterization. Section 3 gives the features, the material and the data used in this study. Section 4 is devoted to the features statistical analysis of smoker and non smoker speakers. Section 5 shows the interest of the bio-mechanical features for smoker/non smoker speakers discrimination. Finally, some conclusion remarks are given in section 6.

2. FROM BIOMECHANICS TO ELECTRO-MECHANICS

2.1. Biomechanical model

The vocal folds consist of a set of tissue layers able to vibrate thanks to the aerodynamic interaction between the vocal folds system and the airflow from the trachea. This airflow coming from the lung is described by Bernoulli's principle and leads to a bio-mechanical model that depicts the main vocal folds motions and reproduces their dynamics. Based on the study of Ishizaka and Flanagan [7], a well-known model, termed as two-mass model, assumes that each vocal fold side (left and right) is described by a pair of two coupled oscillators composed of two masses, three springs and two dampers (see Fig.1.a). One mass is called the cover mass and concerns the Reinke's space (upper part m_{1l} and m_{1r} in Fig. 1) while the second one, called the body mass, concerns the body and visco-elastic ligaments (down part m_{2l} and m_{2r} in Fig. 1). Each mass is attached to a linear spring (characterized by its stiffness k_{ij}, $i \in \{1, 2\}$ and $j \in \{l, r\}$) and to a linear viscous damper (characterized by its damping ratio ζ_{ij} or its equivalent viscous resistance $r_{ij} = 2\zeta_{ij}\sqrt{m_{ij}k_{ij}}$). Masses are coupled together by linear spring of stiffness k_{3j}.

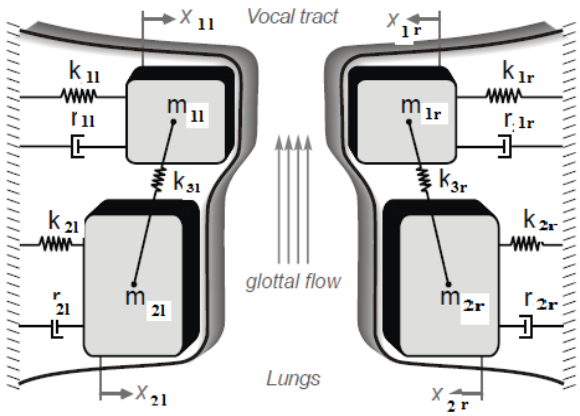

Fig. 1. Two-mass model of vocal folds [8].

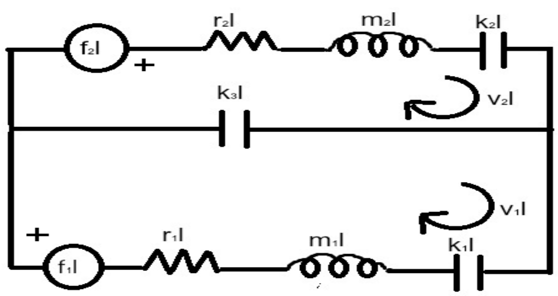

Fig. 2. Electromechanical equivalent circuit of vocal folds.

2.2. Equivalent electromechanical circuit

The vocal folds motion estimation is associated to an electromechanical equivalent circuit shown in Fig. 2, where electromotive forces play the role of mechanical forces and currents play the role of velocities.
The analysis of the electrical circuit using system theory tools such as Laplace transform permits to determine the response of vocal folds to an excitation. The mathematical equation of the transfer function gives the main characteristics (minima, maxima,...) which are directly related to the two-mass model characteristics (mass, stiffness, loss). Hence, the characterization of vocal folds can be observed from a signal processing point of view.

2.3. Bio-mechanical parameters estimation

To estimate the parameters of the two-mass model, one efficient approach is developed (see for example [9] for details). From the acquired voice signal, the vocal tract efect is removed to generate the glottal source, which corresponds to the vocal folds transfer function excited by air pressure from lungs. The glottal source is decomposed in two parts. The first one, called Average Acoustic Wave (AAW), provides the low-order vibration of the vocal folds (average main movement) whereas the second part, called Mucosal Wave Correlate (MWC) gives the higher-order vibration regime of the vocal folds. The signature obtained from MWC is more specifically related to the biomechanics of the vocal fold cover, while that from the glottal source includes the biomechanics of both the body and the cover of the vocal fold. The power spectral density of AAW is calculated cycle by cycle, relevant maxima and minima are labeled and their corresponding frequencies are pointed. Body parameters (mass, loss, stiffness) are calculated using mathematical formula [9]. The same procedure is applied for the cover part using the MWC signal.

3. BIOMECHANICAL FEATURES AND MATERIALS

3.1. Selected features

The selected features describing vocal folds behavior are derived from the biomechanical parameters as follows.
• For each cycle of vibration, the body and cover masses, losses and stiffness are calculated. Their mean values over all cycles are considered as good indicators of overall vocal folds mechanics, they hence constitute the first batch of features.
• The smoking, yielding to pathologies and vocal fold dysfunctions modify the vocal folds behavior over time. That's why the analysis should consider the biomechanical parameters changes from one cycle to another. Thus, for each mean feature, we associate an unbalance one, it is defined as the mean over time of the difference divided by the sum of the considered parameters during two successive cycles. Any perturbation, modification or fatigue on smoker's voice over time will be detected on this category of features.
• Another interesting indicator of smoking effect on vocal folds is the difference of the features values range in comparison with those of healthy vocal folds. Hence, the same features are estimated for a large population and overall mean values are calculated to describe healthy vocal folds. The normalized difference between the studied subject and the overall healthy average is called the deviation feature and is estimated for each category (mean and unblalance) of already defined features.

In summary, a set of 24 features are estimated, 12 for the body part and 12 for the cover part. For each part, features are categorized into mean and deviation. For each category (of 6 features), subcategories of mean features and mean unbalance features are defined. Each subcategory, deals with mass, stiffness and loss characteristics. This tree and features names are illustrated in the first lines of Tab. 1.

3.2. Materials and data

All voice samples analyzed in the study were extracted from the Massachusetts Eye and Ear Infirmary (MEEI) voice disorders database [10]. A population of 292 speakers, producing the sustained vowel /a/, is considered. For each gender, the subjects are sub-divided into two groups: smokers and non smokers. The apportionment of the whole database is the following: 65 female smoker, 109 female non smoker, 62 male smoker and 56 male non smoker.
The software Glottex [11] is used for biomechanical features extraction.

4. STATISTICAL FEATURES ANALYSIS

4.1. Correlation of features with smoking/non smoking status

As a first step, we tried to know the degree of correlation of each feature with smoking. For such reason, a binary variable is associated to the classification into the two classes of smoking and non-smoking. The correlation coefficient with smoking/non smoking status is cal-

Table 1. Bio-mechanical features and their correlation with smoking. The capital letters in feature names are B for body, C for cover, M for mass, S for stiffness, L for loss, U for Unbalance, A for average (or mean) and D for deviation.

	Cover											
	Mean						**Standard deviation**					
	Mass		**Stiffness**		**Loss**		**Mass**		**Stiffness**		**Loss**	
	CMA	CMUA	CSA	CSUA	CLA	CLUA	CMD	CMUD	CSD	CSUD	CLD	CLUD
Male	-0.11	0.0052	-0.02	-0.01	0.05	-0.002	-0.06	0.028	0.02	0.05	0.12	0.04
Female	0.12	0.05	0.06	0.004	0.23	0.035	0.11	0.08	0.06	0.03	0.18	0.16
Both	0.039	0.08	0.05	-0.03	0.19	0.005	0.035	0.022	0.04	0	0.16	0.07

	Body											
	BMA	BMUA	BSA	BSUA	BLA	BLUA	BMD	BMUD	BSD	BSUD	BLD	BLUD
Male	-0.02	-0.11	-0.01	-0.13	0.12	-0.15	0.01	-0.01	0.02	-0.15	-0.22	-0.14
Female	-0.12	-0.08	0.07	-0.08	0.12	-0.03	-0.05	-0.1	0.01	-0.13	-0.01	-0.12
Both	-0.03	-0.12	-0.02	-0.14	0.16	-0.1	-0.01	-0.08	-0.006	-0.17	-0.13	-0.16

culated for each feature:

$$c = \frac{\sum_{i=1}^{N} (x_i - m_x)(y_i - m_y)}{\sqrt{\sum_{i=1}^{N} (x_i - m_x)^2 (y_i - m_y)^2}}, \qquad (1)$$

where x_i is the considered biomechanical feature of each subject i, y_i binary indicator of the class to which it belongs and N is the number of subjects. c belongs to the range $[-1, 1]$. If $|c|$ is close to one, the considered feature is a good indicator of smoking/non smoking state. If $|c|$ is close to zero, the feature is not correlated with the smoking/non smoking information.

For male and female speakers considered separately and together, Tab. 1 gives the correlation values of each feature. One can notice the overall weak values of correlation coefficients. It means, that when considering them separately, they are bad indicators of smoking state of the speaker. That's why, some of them should be discarded and some others should be retained in a different way to improve the analysis. Hence, when considering, the highest absolute values of correlation coefficients, the following features retain attention simultaneously for male, female and genders together: the cover mass average, the cover and body losses averages and deviations (CLA, CLD, BLA, BLD), body average unbalance features of mass, stiffness and loss (BMUA, BSUA, BLUA) and body unbalance deviations (BMUD, BSUD, BLUD).

4.2. Correlation between features

The features are extracted from cover and body parts of vocal folds to describe the mass, the stiffness and losses in different manners (mean, deviation, unbalance). We think that there should be a similarity between some of them so that there is a redundancy when considering the whole set. The degree correlation is an indicator of the redundancy and can be calculated for all distinct pairs of features. The formula of Eq. 1 is applicable, where x and y are now the couple of features. One proposal is to find features whose pairwise correlation is small. Fig. 3 shows the pairs of features (one feature in the x-axis and the other one in the y-axis) whose correlation is less than a threshold (in absolute value sense) which is fixed empirically. In this case, the value of 0.2 is chosen as an example of illustration. Note that an entry is not paired with itself, and changes in abscissa and ordinate do not constitute a new pairing. For the three cases of separate genders and both genders, one can notice that lower correlation is observed in the upper left part. In fact, the body features are not correlated with cover features. Moreover, some horizontal lines appear and correspond to BMA, BSA, BMD and BSD. It means that these features are not correlated with the major part of other features. The same reasoning is applied for vertical lines where we can retain mainly CMA, CLA, CSA, CMD, CSD.

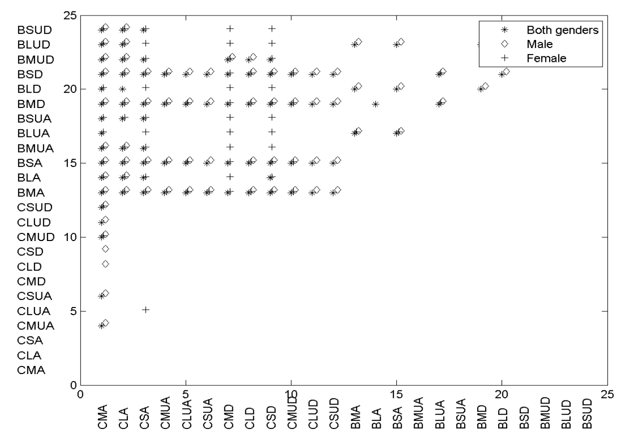

Fig. 3. Features whose pairwise absolute correlation is less than 0.2.

As a conclusion and according to the analysis of Fig. 3, the most important features which are not so much pairwise correlated are: CMA, CLA, CSA, CMD, CSD, BMA, BSA, BMD and BSD. It means that features based on unbalance property are correlated with others, the average features of mass and stiffness are relevant, their deviation relatively to healthy features is also significant. However, the loss characteristic is relatively correlated with others, except in the case of cover loss average.

When adding the information about the correlation of these features with smoking/non smoking decision see Tab. 1), the most relevant features are the cover mass and loss means, the cover mass deviation and the body mass average.

4.3. Features box-plots

In statistical analysis, the box-plot is a useful tool to display median, quartiles, range and possibly extreme values of a set of data. The central box represents the values from the lower to upper quartile (25 to 75 percentile). The middle line represents the median. The vertical line extends from the minimum to the maximum value, excluding outside and far out values which are displayed as separate points.

Due to lack of space, some box-plots of some features are presented in Fig. 4. The label FS (resp. MS) is for female smoker (resp. male smoker) and FNS (resp. MNS) represent female non smoker (resp. male non somker). Note that, sometimes, the upper part of box-plots is not drawn, because features variation range is very large so that the box-plot of smoking subjects, which are located in the lower part, can not be clearly displayed.

We have selected two kinds of features: those whose box-plots are quite different and can help to discriminate between smoker and non smoker subjects (BMUA, BSUA and BMUD) and those whose correlation with smoking/non smoking is relatively high and their intercorrelation is low (CMA, CLA, BMA).

A first look to box-plots of features with interesting correlation properties shows that they share the main range values and most important statistics. It hence seems difficult to discriminate between smokers and non smokers using these features. However, those with bad correlation properties could be used to describe the smoking effect on vocal folds. We propose in the following section to overcome this difficulty by combining features in vectorial form and applying appropriate tools of classification.

5. CLASSIFICATION

5.1. Classification methodology

The experiments are conducted with the database described in subsection 3.2. Among the 65 female smokers, 109 female non smokers, 62 male smokers and 56 male non smokers, two out of three of them (in each category) is used for training and the remaining ones are used for test.

The Quadratic Discriminant Analysis (QDA) is used as classification tool. It is a parametric classification approach which uses a decision function that tries to maximize the distance between the centroids of each class of the training data and at the same time minimizes the distance of the data from the centroid of the class to which it belongs. To find the most relevant features among the 24 ones described above, the rates of correct and false classification of speakers to smoker class and no smoker class are computed. The rate of false decision R_e is defined as the ratio of incorrectly identified speakers to the total number of speakers. The rate of correct smoker (resp. non smoker) speaker classification R_s (resp. R_{ns}) s defined as the ratio of correctly identified smoker (resp. non smoker) speaker to the total number of smoker (resp. non smoker) speaker.

The approach to identify the most relevant features is the following:

• Features are sorted in descending order according to their correlation coefficients with smoking/non smoking decision.
• Features are grouped together in vectors.
• The vector size is varied from one to the total number of features.
• For each vector size, all possible combinations of features are determined.
• For each combination, the classification technique based on QDA is applied and the error rate is calculated. The features combination leading to lowest error rate is retained.

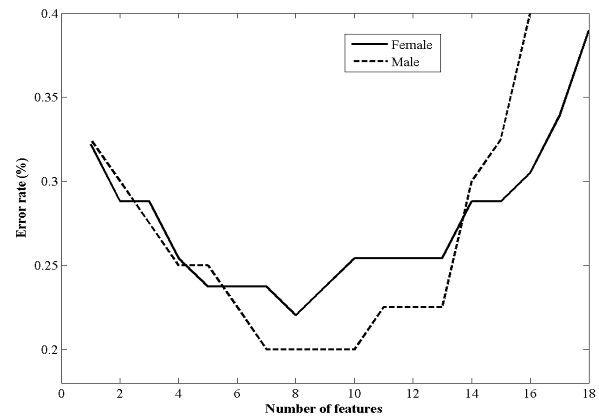

Fig. 5. Error rate of classification versus features number.

Table 2. The eight relevant features and rates of correct classification.

	Male		Female	
Relevant features	CMA, CMS BLS, CLS CMUS, BSUS CSUS, CLA		CMA, CMS BMA CMUS, BSUS CSS, BMS	
Correct	$R_s(\%)$	$R_{ns}(\%)$	$R_s(\%)$	$R_{ns}(\%)$
rate	81	79	54	89

• After considering all possible sizes of vectors, the global minimal error rate is obtained as the minimum of resulting error rates. The corresponding combination of features is the optimal one.

Fig. 5 shows the error rate of classification between smokers and non smokers for both male and female speakers. One can notice that better results are obtained with eight features. The rate of error is $22\,\%$ for female and $20\,\%$ for male. It means that one speakers out of 5 is badly identified. This error rate can be justified by the fact that no noticeable modification is observed on vocal folds of some smokers. It is for example the case of recent smokers.

Tab. 2 gives the eight most relevant features giving better results in terms of error rate. Both genders share four features which are the cover mass average, the cover mass deviation, the cover mass unbalance deviation and the body stiffness unbalance deviation. But they differ in the four other features, as it is indicated in the table.

Tab. 2 gives also the rate of correct classification of smokers and non smokers using the eight features giving better results in terms of error rate. It shows that the rate of good classification is almost the same for smoker and non smoker males. Oddly, female speakers are less identified than female non smoker. Perhaps, when adding acoustic features, such as pitch which changes considerably for female smokers, the rate of good classification will be improved.

6. CONCLUSION

In this paper, a bio-mechanical characterization of speech is used in order to identify smoker speakers. The used features describe the cover and the body parts of vocal folds in terms of mass, loss and stiffness. Among 24 features, the analysis showed that some

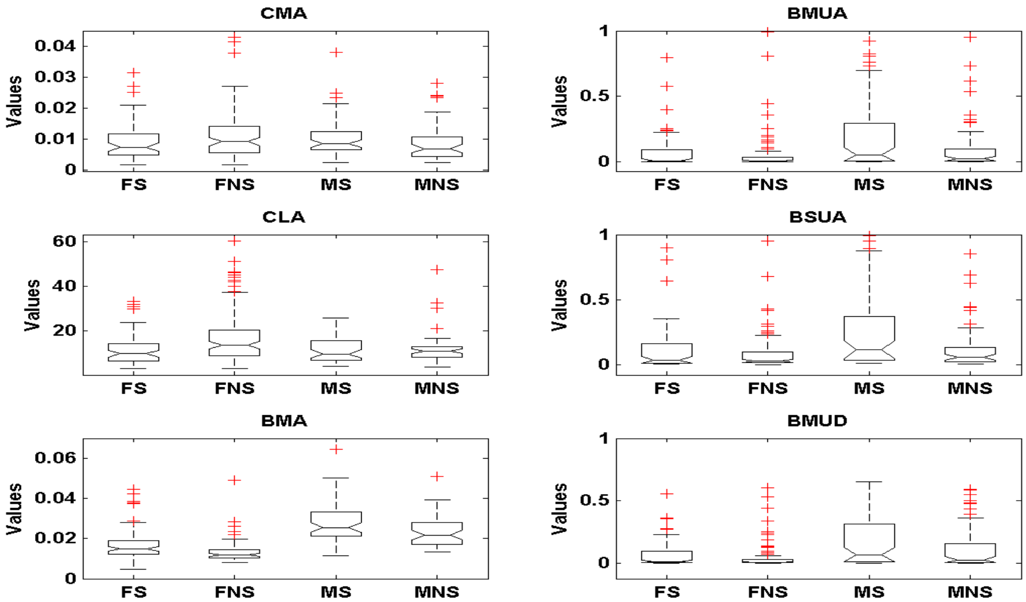

Fig. 4. Box-plots of some features.

of them are relatively correlated with the smoking state since vocal folds are damaged by tobacco. A statistical analysis is carried to identify relevant features. A classification procedure is carried to show that 4 smokers over 5 are correctly identified.

Acknowledgements

The author is thankful to Professor Pedro Gomez Vilda, from Universidad Politechnica de Madrid (in Spain), for providing the Glottex software to extract bio-mechanical features.

REFERENCES

[1] Voice problem, retrieved December 2013, from http://www.voiceproblem.org.

[2] Swallowing, airway, and voice problems treated at NYU voice center, retrieved December 2014, from http://voicecenter.med.nyu.edu

[3] J. Damborenea Tajada, R. Fernandez Liesa R, E. Llorente Arenas, M. J. Naya Glvez, C. Marn Garrido, P. Rueda Gormedino and A. Ortiz Garca, "The effect of tobacco consumption on acoustic voice analysis (Spanish)," *Acta Otorrinolaringol Esp*, vol. 50, pp. 448-52, 1999.

[4] J. Gonzalez and A. Carpi, "Early effects of smoking on the voice: a multidimensional study," *Medical Science Monitor*, vol. 10, no. 12, pp. CR649–656, 2004.

[5] L. Dirk and A. Braun, "Voice parameter changes in smokers during abstinence from cigarette smoking," *Proc. of International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, China, August 2011.

[6] O. Mohamadi, A. Sadolahi and R. Ghorbani, "Effects of smoking cigarette on some acoustic voice parameters and elements of speech pattern of male subjects," *Journal of Semnan University of Medical Science*, vol. 12, no. 3, pp 253–258, 2011.

[7] K. Ishizaka and J.L. Flanagan , "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell System Technical Journal*, vol. 51, pp. 1233-1268, 1972.

[8] A. P. Pinheiro, T. Raitio, D. S. Gomes and P. Alku, "Voice source analysis using biomechanical modeling and glottal inverse filtering," *Proc. of Interspeech*, Oregon-Portland, Sept. 9–13, 2012.

[9] P. Gomez, R. Fernandez-Baillo, V. Rodellar-Biarge, V. Nieto-Lluiz, A. Alvarez-Marquina, L. M. Mazaira-Fernnadez, R. Martinez-Olalla, J. I. Godino-Llorente, "Glottal source biomedical signature for voice pathology detection," *Speech Communication Journal*, vol. 51 no. 9, pp 759–781, Sept. 2009.

[10] Massachusetts Eye and Ear Infirmary, *Voice Disorders Database*, Version 1.03. Kay Elemetrics Corp., Lincoln Park, NJ, 1994, CD-ROM. [41] V. Parsa and D. G. Jamieson, Identification of pathological voices

[11] Glottex sofware, retrieved Decemeber 2012, from http://www.glottex.com, .