Voice Pathology Assessment based on a Dialogue System and Speech Analysis

Richard B. Reilly[φ], Rosalyn Moran[φ], Peter Lacy*

[φ]Department of Electronic and Electrical Engineering, University College Dublin, Ireland
*St James's Hospital, Dublin 8, Ireland
Richard.Reilly@ucd.ie

Abstract

A system of remotely detecting vocal fold pathologies using telephone quality speech recorded during a telephone dialogue is presented. This study aims at developing a dialogue system using VoiceXML for remote diagnosis of voice pathology. To assess the accuracy of the system, a database of 631 clean speech files of the sustained phonation of the vowel sound /a/ (58 normal subjects, 573 pathologic) from the Disordered Voice Database Model 4337 was transmitted over telephone channels to produce a test corpus. Pitch perturbation features, amplitude perturbation features and a set of measures of the harmonic-to-noise ratio are extracted from the clean and transmitted speech files. These feature sets are used to test and train automatic classifiers, employing the method of Linear Discriminant Analysis. Cross-fold validation was employed to measure classifier performances. While a sustained phonation can be classified as normal or pathologic with accuracy greater than 90%, results indicate that a telephone quality speech can be classified as normal or pathologic with an accuracy of 74.15%. Amplitude perturbation features proving most robust in channel transmission. This study highlights the real possibility for remote diagnosis of voice pathology.

Introduction

Patients with voice disorders can possess a variety of vocal fold pathologies. Voice pathologies are relatively common affecting almost 5% of the population [1]. These pathologies can be found in varying degrees of severity and development. They can be classed as physical, neuromuscular, traumatic and psychogenic and all directly affect the quality of the voice.

Developments in non-invasive methods for voice pathology diagnosis have been motivated by a need to conduct both objective and efficient analysis of a patient's vocal function. At present a number of diagnostic tools are available to the otolaryngologists and speech pathologists such as videostroboscopy [2] and videokymography. However these current methods are time and personnel intensive and lack objectivity.

Research has been reported on the development of reliable and simple methods to aid in early detection, diagnosis, assessment and treatment of laryngeal disorders. This research has lead to the development of feature extraction from acoustic signals to aid diagnosis. Much focus has been centred on perturbation analysis measures such as jitter and shimmer and on signal-to-noise ratios of voiced speech, which reflect the internal functioning of the voice. Through this research it has been shown that these features can discriminate between normal and pathologic speakers [3],[4],[5],[6]. Voice pathology detection systems using high quality voice recordings have achieved classification accuracies of over 90% in being able to discriminate between normal and pathologic speakers [7], [8].

One of the major difficulties in voice pathology assessment is the availability of trained clinicians and centres of excellence. With the limited number of centres available, long queues at voice outpatient clinics are a common occurrence. This tends to dissuade patients from attending, in particular those who must travel a distance to the clinic. Assessing progress of patients who have undergone surgery is also of concern for clinicians. Having a method of remote assessment patients for voice pathologies, for both new and continuing patients, would be of enormous benefit to both the clinician and patient.

Aim

The aim of this research was to investigate the performance of a telephone based dialogue to capture a voice recording of the patient producing a sustained phonation of the vowel sound /a/ and subsequently carry out voice pathology classification to categorise the voice as either normal or pathologic. The goal of this project was to produce a voice pathology classifier providing remote diagnosis in a non-intrusive and objective manner.

Methodology

The steps involved in a voice pathology classification system, as shown in Figure 1, are discussed below.
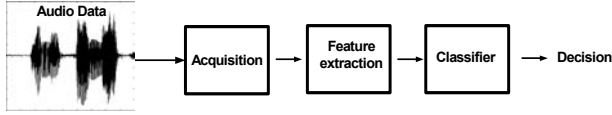
Figure 1. Processes involved in Voice Pathology Classification

Audio Data

The labelled voice pathology database "Disordered Voice Database Model 4337" [9] acquired at the Massachusetts Eye and Ear Infirmary Voice and Speech Laboratory and distributed by Kay Elemetrics is often cited in the literature for voice pathology assessment. A detailed description of the database can be found at [9]. The mixed gender database contains 631 voice recordings (Pathological: 573 and Normal: 58) with an associated clinical diagnosis. The types of pathologies are diverse, ranging from Vocal Fold Paralysis to Vocal Fold Carcinoma. Attention in this study was focused on the sustained phonation of the vowel sound /a/ (as in the English word "c<u>a</u>p"). This database was originally recorded at a sampling rate of 25kHz. For this study the database was downsampled to 10kHz.

Acquisition

A voice recording is typically acquired using a microphone and digital storage in a clean audio environment. As the aim of this study is for a remote diagnosis classification system, each of the 631 voice recordings were played over a long distance telephone channel using a specifically written VoiceXML script and recorded under the control of another VoiceXML script. The VoiceXML development was carried out using the online development system *VoxBuilder* [10]. This process created a telephone quality voice pathology database for all 631 voice recordings in the "Disordered Voice Database Model 4337" database. A VoiceXML application currently exists to allow new telephone quality audio samples to be gathered.

Feature Extraction

Features typically extracted from the audio data for voice pathology analysis include the fundamental frequency (F0), jitter (short-term, cycle to cycle, perturbation in the fundamental frequency of the voice), shimmer (short-term, cycle to cycle, perturbation in the amplitude of the voice), signal-to-noise ratios and harmonic-to-noise ratios [7]. The features used in this study include Pitch Perturbation Features, Amplitude Perturbation Features and Harmonic to Noise Ratio (HNR).

Pitch and Amplitude Measures. Pitch and Amplitude Perturbation measures were calculated by segmenting the speech waveform (3-5 seconds in length) into overlapping 'epochs'. Each epoch is 20msecond with an overlap of 75% between adjacent epochs. A 20msecond epoch is necessary to give an accurate representation of pitch. Table 1 and Table 2 provides a list of the twelve pitch and amplitude features employed in this study.

No.	Description	Formula
1	Mean F0	$\frac{1}{n}\sum_{i=1}^{n} F_i$
2	Maximum F0 Detected	$\max(F_i)$
3	Minimum F0 Detected	$\min(F_i)$
4	Standard Deviation of F0 contour	$\frac{1}{n-1}\sum_{i=1}^{n}\left(F_i - \overline{F}\right)^2$
5	Phonatory Frequency Range	$12 \times \frac{\log\left(\frac{F0\_hi}{F0\_lo}\right)}{\log 2}$
6	Mean Absolute Jitter *(MAJ)*	$\frac{1}{n-1}\sum_{i=n-1}^{1}\mid F_{i+1} - F_i \mid$
7	Jitter (%)	$\frac{MAJ}{F0\_av}$
8	Relative Average Perturbation smoothed over 3 pitch periods	$\frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\mid \frac{F_{i+1}+F_i+F_{i-1}}{3} - F_i \mid}{F0\_av} \times 100$
9	Pitch Perturbation Quotient smoothed over 5 pitch periods	$\frac{\frac{1}{n-4}\sum_{i=3}^{n-2}\mid \frac{\sum_{k=i-2}^{i+2}F(k)}{5} - F_i \mid}{F0\_av} \times 100$
10	Pitch Perturbation Quotient smoothed over 55 pitch periods	$\frac{\frac{1}{n-54}\sum_{i=28}^{n-27}\mid \frac{\sum_{k=i-27}^{i+27}F(k)}{55} - F_i \mid}{F0\_av} \times 100$
11	Pitch Perturbation Factor	$\frac{N_{p\geq threshold}}{N_{voice}} \times 100$
12	Directional Perturbation Factor	$\frac{N_{\Delta\pm}}{N_{voice}} \times 100$

Table 1   Pitch Perturbation Features

No.	Description	Formula
1	Mean Amp	$\frac{1}{n}\sum_{i=1}^{n} A_i$
2	Maximum Amp Detected	$\max(A_i)$
3	Minimum Amp Detected	$\min(A_i)$
4	Standard Deviation of Amp contour	$\frac{1}{n-1}\sum_{i=n-1}^{1}\left(A_i - \overline{A}\right)^2$
5	Mean Absolute Shimmer *(MAS)*	$\frac{1}{n-1}\sum_{i=n-1}^{1}\mid A_{i+1} = A_i \mid$

6	Shimmer (%)	$\dfrac{MAS}{Amp\_av}$
7	Shimmer :Decibels	$\dfrac{1}{n-1}\sum\limits_{i=1}^{n-1}20\times\log\left(\dfrac{A_i}{A_{i+1}}\right)$
8	Amplitude Relative Average Perturbation smoothed over 3 pitch periods	$\dfrac{\frac{1}{n-2}\sum\limits_{i=2}^{n-1}\lvert\frac{A_{i+1}+A_i+A_{i-1}}{3}-A_i\rvert}{Amp\_av}\times100$
9	Amplitude Perturbation Quotient smoothed over 5 pitch periods	$\dfrac{\frac{1}{n-4}\sum\limits_{i=3}^{n-2}\lvert\frac{\sum_{k=i-2}^{i+2}A(k)}{5}-A_i\rvert}{Amp\_av}\times100$
10	Amplitude Perturbation Quotient smoothed over 55 pitch periods	$\dfrac{\frac{1}{n-54}\sum\limits_{i=28}^{n-27}\lvert\frac{\sum_{k=i-27}^{i+27}A(k)}{55}-A_i\rvert}{Amp\_av}\times100$
11	Amplitude Perturbation Factor	$\dfrac{N_{p\geq threshold}}{N_{voice}}\times100$
12	Amplitude Directional Perturbation Factor	$\dfrac{N_{\Delta\pm}}{N_{voice}}\times100$

Table 2   Amplitude Perturbation Features

Harmonic to Noise Ratio. Mel Frequency Cepstral Coefficients (MFCC) features are commonly used in Automatic Speech Recognition (ASR) and also Automatic Speaker Recognition systems [11]. The Cepstral domain is employed in speech processing as the lower valued cepstral "quefrequencies" model the vocal tract spectral dynamics, while the higher valued quefrequencies contain pitch information, seen as equidistant peaks in the spectra.

The Harmonic to Noise Ratio is calculated in the Cepstral domain, as follows:

1. Speech signal is processed to have zero mean and unit variance.
2. A 100msecond epoch is extracted.
3. A peak-picking algorithm locates the peaks at multiples of the fundamental frequency.
4. A bandstop filter in the Cepstral domain is applied to the signal. The stopband of the filter is limited to the width of each peak. The remaining signal is known as the rahmonics (harmonics in the cepstral domain) comb-liftered signal and contains the noise information.
5. The Fourier transform of this comb-liftered signal is taken, generating an estimate of the noise energy present N(f). Similarly, the Fourier Transform of the original cepstral-domain signal, including rahmonics is taken, O(f).
6. The HNR for a given frequency band B is then calculated as per

$$HNR_\beta(f) = mean(O(f))_\beta - mean(N(f))_\beta$$

Eleven HNR measures were calculated as in Table 3.

Band Number	Incorporating Frequencies (Hz)
1	0 - 500
2	0 - 1000
3	0 - 2000
4	0 - 3000
5	0 - 4000
6	0 - 5000
7	500 – 1000
8	1000 -2000
9	2000 – 3000
10	3000 – 4000
11	4000 - 5000

Table 3   HNR bands.

Pattern Classifier

Linear discriminants (LD) [12] partition the feature space into the different classes using a set of hyper-planes. The parameters of this classifier model were fitted to the available training data by using the method of maximum likelihood. Using this method the processing required for training is achieved by direct calculation and is extremely fast relative to other classifier building methods such as neural networks. This model assumes that the feature data has a Gaussian distribution for each class. In response to input features, linear discriminants provide a probability estimate of each class. The final classification is obtained by choosing the class with the highest probability estimate.

The cross-validation scheme [13] was used for estimating the classifier performance. The variance of the performance estimates was decreased by averaging results from multiple runs of cross validation where a different random split of the training data into folds is used for each run. In this study ten repetitions of ten-fold cross-validation were used to estimate classifier performance figures. For each run of cross fold validation the total normal population and a randomly selected group of abnormals equal in size to the normal population was utilised. This results in a more realistic reflection of the predictive ability of the system.

In this study the performance of the classifier is quoted using class sensitivities, predictivities and overall accuracy.

		Classification Pathology	Classification Normal
Diagnosed Pathology	**P**	*True Positive* TP	*False Negative* FN
Diagnosed Normal	**N**	*False Positive* FP	*True Negative* TN

Table 4   Classification Matrix.

Definitions of sensitivity, specificity, predictivities and the overall accuracy are given in Table 4b.

Sensitivity	$TP/TP + FN$	Fraction of speech files from the set of all pathologic files correctly classified.
Specificity	$TN/TN + FP$	Fraction of speech files from the set of all normal voices correctly classified.
Positive Predictivity	$TP/TP + FP$	Fraction of speech files detected as pathologic that are correctly classified.
Negative Predicitivity	$TN/TN + FN$	Fraction of speech files detected as normal that are correctly classified.
Overall Accuracy	$\dfrac{TP + TN}{TP + TN + FP + FN}$	Fraction of the total number of subjects' voices that are classified correctly.

Table 4b   Definition of Measures of Performance.

Results

Each feature was tested for class sensitivities, predictivities and overall accuracy. The contribution provided by each pitch perturbation measures, amplitude perturbation measures and HNR, on the subsampled (10kHz) and telephone quality databases is given in Tables 5-10.

Feature Number	Acc	Sens	Spec	Pos. Pred	Neg. Pred
1	52.82	50.17	55.50	53.30	52.35
2	65.75	53.10	78.53	71.46	62.33
3	51.78	48.28	55.32	52.24	51.38
4	60.02	35.52	84.82	70.31	56.51
5	60.97	39.66	82.55	69.70	57.47
6	58.28	30.00	86.91	69.88	55.09
7	57.42	27.59	87.61	69.26	54.45
8	57.33	27.59	87.43	68.97	57.40
9	57.41	48.66	43.56	60.49	58.13
10	60.10	25.52	95.11	84.09	55.78
11	62.45	37.93	87.26	75.09	58.14
12	49.52	49.14	49.91	49.83	49.23

Table 5   Pitch Perturbation Measures: Clean 10kHz database

Feature Number	Acc	Sens	Spec	Pos. Pred	Neg. Pred
1	47.79	43.28	52.36	47.90	47.69
2	63.66	57.93	69.46	65.75	61.99
3	50.91	43.10	58.81	51.44	50.52
4	62.36	41.03	83.94	72.12	58.49
5	61.01	42.24	82.02	70.40	58.39
6	61.75	37.07	86.74	73.88	58.49
7	60.54	34.48	86.91	72.73	57.72
8	60.36	34.32	86.56	72.20	56.62
9	58.37	35.00	82.02	66.34	55.49
10	63.31	32.76	94.24	85.20	58.6
11	58.11	36.90	79.58	64.65	53.47
12	49.35	36.38	62.48	49.53	49.24

Table 6   Pitch Perturbation: Transmitted database

Feature Numbe	Acc	Sens	Spec	Pos. Pred	Neg. Pred
1	49.87	52.59	47.12	50.16	49.54
2	51.17	53.28	49.04	51.41	50.91
3	50.04	52.24	47.82	50.33	49.73
4	56.63	70.17	42.93	55.45	58.71
5	72.94	92.76	52.88	66.58	87.83
6	67.82	88.97	46.42	62.70	80.61
7	71.21	94.66	47.47	64.59	89.77
8	67.65	88.62	46.42	62.61	80.12
9	66.61	87.76	45.20	61.85	78.48
10	62.36	76.21	48.34	59.89	66.75
11	56.55	71.72	41.19	55.25	59.00
12	67.04	75.52	58.46	64.79	70.23

Table 7   Amplitude Perturbation: Clean 10kHz database

Feature Number	Acc	Sens	Spec	Pos. Pred	Neg. Pred
1	51.17	53.10	49.21	51.42	50.90
2	56.63	57.24	56.02	56.85	56.41
3	47.70	49.66	45.72	48.08	47.29
4	63.14	73.63	52.53	61.09	66.30
5	72.68	89.72	52.36	66.34	87.72
6	63.57	82.76	44.15	60.00	71.67
7	51.08	11.72	90.92	56.67	50.44
8	63.66	82.93	44.15	60.05	71.88
9	62.97	81.55	44.15	59.65	70.28
10	58.81	68.45	45.03	55.76	58.50
11	56.90	72.41	41.19	55.48	59.60
12	61.23	68.62	53.75	60.03	62.86

Table 8   Amplitude Perturbation: Transmitted database

Band Number	Acc	Sens	Spec	Pos. Pred	Neg. Pred
1	76.25	79.83	78.18	78.74	75.00
2	75.55	83.79	78.58	80.07	82.78
3	79.88	85.00	74.69	77.27	83.11
4	77.62	83.28	71.90	75.00	80.94
5	65.74	77.59	53.75	62.94	70.32
6	47.61	52.59	42.58	48.11	47.01
7	70.86	83.79	57.77	66.76	77.88
8	50.22	34.66	65.97	50.76	49.93
9	75.20	73.28	77.14	76.44	74.04
10	79.63	78.79	82.14	82.19	79.40
11	75.46	84.14	66.67	71.87	80.59

Table 9   HNR Bands: Clean 10kHz database

Band Number	Acc	Sens	Spec	Pos. Pred	Neg. Pred
1	53.77	72.14	28.10	52.70	57.09
2	44.41	42.76	46.07	44.52	44.30
3	45.19	48.45	41.88	45.77	44.53
4	49.78	4.83	95.29	50.91	49.73
5	53.86	67.24	40.31	53.51	58.86
6	NA	NA	NA	NA	NA
7	51.08	16.55	86.04	54.55	50.46
8	50.65	31.03	70.51	51.58	50.25
9	58.28	24.48	92.50	76.76	54.75
10	57.15	52.41	74.00	67.11	60.51
11	NA	NA	NA	NA	NA

Table 10   HNR Bands: Transmitted database

For the telephone database bands 6 and 11 were omitted. However due to channel variability bands 1-5 were measured, as above, from 0Hz.

The various feature groups provide independent and complimentary classification information. By combining the feature groups it was anticipated that the overall classification performance would be improved. Classification results were obtained for the combination of these features, as shown in Table 11.

Test Corpus	Acc	Sens	Spec	Pos. Pred	Neg. Pred
Clean 10kHz	89.10	93.26	85.14	87.63	86.25
Telephone	74.15	75.69	72.60	73.66	74.69

Table 11   Classification results based on the combination of feature sets.

Discussion

Twelve pitch and twelve amplitude perturbation measures were extracted from the pitch and amplitude contours respectively. The pitch and amplitude perturbation measures detect short-term changes in the pitch contour. It was hypothesized that such measures, when extracted from normal and pathologic subjects, would also be statistically different and therefore allow a classifier to distinguish between the two groups. Normal speech is known to have certain levels of jitter and shimmer. However pathologic speech should exhibit larger perturbations in both the pitch and the amplitude contours. However, the difficulty in accurately tracking the pitch contour especially in speech could severely limit the ability of the perturbation measures to separate between normal and pathologic voice.

The classification performances of the pitch perturbation measures to differentiate between normal and pathologic voice are presented in Table 5 and 6. Using just one feature, best test set accuracies of 65.75% and 63.66% were achieved for individual features on the 10kHz and telephone databases. The classification performances of the amplitude perturbation measures to differentiate between normal and pathologic voice are presented in Table 7 and 8. Best test set accuracies of 72.94% and 63.66% were achieved for just one feature (shimmer) for individual features on the 10kHz and telephone databases.

The Harmonics-to-Noise ratio (HNR) measures were extracted for eleven frequency bands described in Table 3. The HNR method transforms the speech signal to the cepstral domain, removes the *rahmonic* information from the cepstrum and applies the DFT to this signal, which is defined as the spectrum of the estimated noise signal. The classification performances of the harmonic to noise ratio measures to differentiate between normal and pathologic

voice are presented in Table 9 and 10. Using just one feature, best test set accuracies of 79.88% and 58.28% were achieved for individual features on the 10kHz and telephone databases.

Pathologic subjects should theoretically have increased levels of jitter, shimmer and additive noise. Much research exists that demonstrates that the signal-to-noise measure does successfully distinguish between the normal and pathologic groups. From the results in Table10 the telephone transmission of the audio files does not allow preservation of HNR information to identify the pathologic and normal voice, compared to the 10kHz database allows.

By taking combinations of features, the ability of the features to separate normal and pathologic subjects on the two databases, 89.10% and 74.15%, is a definite improvement over using the individual features groups. More research is needed on the specific combinations of features for these databases.

A number of research groups [14], [15], [16] have reported results for detection rates for voice pathologies of 94.87%, 76% and 96.30% respectively. In [14] the "Disorder Voice Database Model 4337" sampled at 25kHz was employed and their results may be compared with the results obtained in this study, although they have used the higher quality audio data. In study [15] different databases were used and a direct comparison of results cannot be made.

The database used in the present study provides a large number of pathologic subjects that might not fairly represent the pathologies present in other studies conducted in this area or those encountered by the medical profession on a day-to-day basis. The predictive ability of this model could be confirmed through external validity.

As mentioned in Section II the "Disorder Voice Database Model 4337" is accompanied by a diagnostic description for each subject. The diagnostic description provided with the database is very detailed. These detailed diagnostic descriptions were grouped into several diagnostic categories by our medical consultant. It was observed from the distributions of the diagnoses for each subject that only at the highest level (i.e. either normal or pathologic) that a patient's diagnosis is mutually exclusive. However, as the level of categorisation proceeds to sublevels the patient's diagnoses are no longer mutually exclusive. Thus each subject may be diagnosed into more than one category i.e. they may have a pathology that is both physical and neuromuscular. This has a significant effect on the potential for an automatic classification system to differentiate between the categorisation types. One could categorise the database to allow for a vocal quality classification scheme. In this way, a speech recording may be categorised based on the vocal quality of the speech recording; breathy, strained or noisy. Future work in this

area, based on the methods developed in this study, would allow investigation into the differentiation, for example, between normal subjects and subjects with nodules.

Ongoing research is examining feature configurations for further improvement of pathology diagnosis, with categorizations of the database into further sublevels of pathologies within the classes of neuromuscular, physical and mixed. Also further patient data is being acquired through collaboration with a local hospital Speech and Language department, specifically to ensure that the pathologies acquired are those encountered by the medical profession on a day-to-day basis. Audio samples are recorded in the clinic at 44kHz and simultaneously using a VoiceXML based telephone application. This allows the remote classification system to perform analysis on both telephony quality and high quality audio. The classification of the audio data is performed automatically on receipt of the audio data, with results posted to a web interface.

The results of the project suggest that by combining VoiceXML as a telephony interface and server side speech processing, an automatic classification system to differentiate between normal and pathologic voice can be achieved. Further investigations will continue into this aspect.

It is hoped that as the performance of the telephony system enhances with increased training samples, the system could provide the clinical staff with a useful pre-screening service for voice pathology.

Conclusion

The results of the study suggest that by combining VoiceXML as a telephony interface, having a dialogue to aid the patient in the production of a sustained phonation and having server side speech processing, an automatic classification system to differentiate between normal and pathologic voice can be achieved. This study highlights the real possibility for remote diagnosis of voice pathology.

Acknowledgements

References

[1] Becker, W., Naumann, H.H., and Faltz, C.R., 1994. *Ear, Nose and Throat Diseases*. Thieme Medical Publishers, 2nd Edition.

[2] Schneider, B., Wendler, J., and Seidner, W. 2002. The relevance of stroboscopy in functional dysphonias, *Folia Phon*., 54, 1: 44-54.

[3] Liebermann, P. 1961. Perturbations in vocal pitch. *J. Acoust. Soc. Am*, 33, 5.

[4] Titze, I.R. 1994. *Workshop on Acoustic Voice Analysis*, National Centre for Voice and Speech, America, Iowa City, IA:.

[5] de Krom, G. 1995. Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *J. Speech. Hear. Res*., Vol. 38: 794-811.

[6] Michaelis, D., Frohlich, M., Strube, H.W. 1998. Selection and combination of acoustic features for the description of pathologic voices", *J. Acoust. Soc. Am*., 103, 3:1628-1639.

[7] Maguire C., de Chazal P., Reilly R.B., and Lacy P. 2003. Automatic Classification of voice pathology using speech analysis. In *Proceedings of the World Congress on Biomedical Engineering and Medical Physics*, Sydney.

[8] Maguire C., de Chazal P., Reilly R.B., and Lacy P. 2003. Identification of Voice Pathology using Automated Speech Analysis. In *Proceedings of the 3rd International Workshop on Models and Analysis of Vocal Emission for Biomedical Applications*, Florence. 259-262.

[9] Kay Elemetrics Corporation. 1994. Disorder Voice Database Model 4337. Massachusetts Eye and Ear Infirmary Voice and Speech Lab, Boston, MA.

[10] Voxpilot Ltd., Dublin. www.voxpilot.com.

[11] Campbell, J.P. 1997. Speaker Identification: A tutorial. *Proc. of the IEEE*, 85, 9:1437-1462.

[12] Duda, R. O., Hart, P.E. and Stork H.G. 2000. *Pattern Classification*, Wiley-Interscience, New York, NY.

[13] Kohavi, R. 1995. A study of cross validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Conference on Artificial Intelligence*. 1137-1143.

[14] Llorente G. and Navarro S. 1997. On The Selection of Meaningful Speech Parameters Used Pathologic/Nonpathologic Voice Register Classifier. In *Proceedings of Eurospeech '99*, 1: 563-566.

[15] Childers, D. G. 1992. Detection of Laryngeal Function using Speech and Electrographic Data. *IEEE Transactions on Biomedical Engineering*, 39, 1:19-25.

[16] Cesar M. E. and Hugo, R. L. 2000. Acoustic Analysis of Speech for Detection of Laryngeal Pathologies. In *Proceedings of the 22nd Annual Engineering in Medicine and Biology Conference*. 2369-2372.