Telephony-Based Voice Pathology Assessment Using Automated Speech Analysis

Rosalyn J. Moran*, Richard B. Reilly, *Senior Member, IEEE*, Philip de Chazal, *Member, IEEE*, and Peter D. Lacy

***Abstract*—A system for remotely detecting vocal fold pathologies using telephone-quality speech is presented. The system uses a linear classifier, processing measurements of pitch perturbation, amplitude perturbation and harmonic-to-noise ratio derived from digitized speech recordings. Voice recordings from the Disordered Voice Database Model 4337 system were used to develop and validate the system.**

Results show that while a sustained phonation, recorded in a controlled environment, can be classified as normal or pathologic with accuracy of 89.1%, telephone-quality speech can be classified as normal or pathologic with an accuracy of 74.2%, using the same scheme. Amplitude perturbation features prove most robust for telephone-quality speech.

The pathologic recordings were then subcategorized into four groups, comprising normal, neuromuscular pathologic, physical pathologic and mixed (neuromuscular with physical) pathologic. A separate classifier was developed for classifying the normal group from each pathologic subcategory. Results show that neuromuscular disorders could be detected remotely with an accuracy of 87%, physical abnormalities with an accuracy of 78% and mixed pathology voice with an accuracy of 61%.

This study highlights the real possibility for remote detection and diagnosis of voice pathology.

***Index Terms*—Speech analysis, voice pathology, voiceXML.**

I. Introduction

VOICE pathologies are relatively common affecting almost five percent of the population [1] and are found in varying degrees of progression and severity. All pathologies directly affect the quality of the voice and can be classed as physical, neuromuscular, traumatic and/or psychogenic.

Developments in noninvasive 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 including videostroboscopy [2] and videokymography [34]. These approaches employ invasive endoscopy, are time and personnel intensive and lack objectivity. Schemes for detecting voice pathologies, based on acoustic analysis that have been reported with accuracies of over 90% [3]–[6], offer savings in both time and cost and also remove practitioner subjectivity. Development of these automatic detection systems has focused largely on feature extraction algorithms. Perturbation measures such as jitter (changes in pitch with time) and shimmer (changes in amplitude with time) and energy measures of voiced speech have been shown to reflect the internal functioning of the voice [7]–[12], and provide a basis for normal/pathologic discrimination.

In a study by Rosa *et al.* [13], a measure of pitch perturbation was shown to be most useful in pathological discrimination. Using a set of 48 dysphonic speakers a classification system employing pitch perturbation measures achieved a 55% accuracy in discriminating the normal speakers from the pathologic speakers. Godino-Llorente *et al.* [6] describe a neural network-based detector processing Mel frequency cepstral coefficients (MFCCs) and their derivatives for normal/abnormal discrimination. MFCCs are discrete Fourier transform (DFT)-based parameters originating from studies of the human auditory system and have been used extensively in speech analysis [14]–[16], yet the physical relevance of MFCCs to vocal fold pathologies has not been widely examined. The cepstral domain of the MFCCs has been employed in measures of energy present in pathologic voice. Noise energy, broadly understood as that outside of harmonic components during sustained phonation, represents air escaping during closed off periods of glottal flow, indicative of an interruption of the morphology of the larynx [17]. Several measures of noise energy have been defined in the literature [18], the prevailing measure is the harmonic-to-noise ratio (HNR) [19], computed in the cepstral domain. All of these studies have used high-quality speech samples acquired in low-noise recording environments.

To date, no studies have investigated the detection of voice pathologies by processing speech from variable or lower quality input channels although studies in the closely related field of speech recognition suggest a probable decrease in performance when processing telephone speech. In [14]–[16], the performance of speech-based systems that utilize telephone-transmitted-quality acoustic signals was determined. The TIMIT database (produced by researchers at Texas Instruments, TI and The Massachusetts Institute of Technology, MIT), of high-quality speech recordings, the NTIMIT (Network TIMIT speech transmitted via long distance channels) [20] and the CTIMIT (Cellular TIMIT) [21] have been used as comparative test sets for the study of both speech [14] and speaker [15], [16] recognition systems. These studies showed that the performance of speech and speaker recognition systems decreased when processing telephone-quality signals relative to systems processing high-quality recordings. The primary sources of information loss were band limiting, channel filtering

Manuscript received August 10, 2004; revised April 29 2005. This work was supported in part by Enterprise Ireland under ATRP/2002/231. *Asterisk indicates corresponding author.*

*R. J. Moran is with the Department of Electronic and Electrical Engineering, University College Dublin, Dublin 4, Ireland (e-mail: rosalyn.moran@ee.ucd.ie).

R. B. Reilly is with the Department of Electronic and Electrical Engineering, University College Dublin, Dublin 4, Ireland, and The Cognitive Neurophysiology Laboratory, St Vincent's Hospital, Fairview, Dublin, Ireland (e-mail: richard.reilly@ucd.ie).

P. de Chazal is CTO of BiacaMed Ltd., NOVA, University College Dublin, Dublin 4, Ireland (e-mail: philip.dechazal@biancamed.com).

P. D. Lacy is Consultant ENT Surgeon with Beaumont Hospital, Dublin 9, Ireland (e-mail: entdoc@iol.ie).

Digital Object Identifier 10.1109/TBME.2005.869776

and additive noise [15], [16]. These studies further illustrated the dependence for successful system implementation on the availability of a large amount of training data matched to the acoustic requirements of the test environment.

Successful systems for the remote treatment of laryngeal disorders have been reported. Mashima *et al.* [22], describe a proof-of-concept study, whereby successful voice rehabilitation and therapy was conducted by group video-teleconferencing. This study, while hardware intensive and requiring observation by a specialist clinician, demonstrated that the delivery of remote diagnostics ensured patient participation by those who may otherwise have discontinued treatment prematurely. The service also provided therapy for populations in remote regions.

This paper focuses on the development of automatic voice pathology detection schemes for use in nonideal recording environments. Specifically the paper focused on the use of telephone-quality speech for voice pathology assessment. By transmitting the voice over long distances, using standard methods (the Public Switched Telephone Network,) savings in time and costs for pathology screening, post-operative follow ups, and vocal-quality monitoring may be realized.

II. Aim

The aim of this study was to investigate the performance of a telephone-based voice pathology classifier to categorise a sustained phonation of the vowel sound /a/ into either a normal or pathologic class, focusing on which particular feature sets prove most robust to noisy channel transmission. A further aim was to determine classifier performance on specific pathology classes. The long-term goal of this research is to produce a voice pathology classifier providing remote diagnosis in a nonintrusive and objective manner.

In Section III, the audio database is presented and the methods used in the development of the classifier discussed. This section also presents the infrastructure developed for the acquisition of the telephone transmitted speech corpus. The classification performance of the system processing distorted audio recordings and the tuned performance when trained and tested using a subset of specific pathologies are given in Section IV. Following in Section V is a discussion of our findings, comparing results with previously published work. Section VI draws conclusions from this study.

III. Methodology

The steps involved in the presented voice pathology classification system, as shown in Fig. 1, are discussed below.

A. Audio Data

The labeled voice pathology database "Disordered Voice Database Model 4337" [23] acquired at the Massachusetts Eye and Ear Infirmary Voice and Speech Laboratory was used throughout this study (herein referred to as "the database"). A detailed description of the database can be found in [23]. The mixed gender database contains 631 voice recordings (573 Pathological, 58 Normal), of the sustained phonation of the vowel sound /a/ (as in the English word "cap") with an associated clinical diagnosis. The types of pathologies include Vocal Fold Paralysis and Vocal Fold Carcinoma. The vocalizations last 1 to 3 s. All test files were recorded in a controlled environment with low ambient noise, direct digital 16-bit sampling, robust signal conditioning, and at a sampling rate of 25 kHz. Although the aim was maintain a constant microphone distance during acquisition, the developers reported that this was not the case during initial trials [23] and this may add some variability to signal fidelity across the recordings. It has been shown that steady-state phonations [24] provide higher detection accuracy over nonsteady-state, sentence-based normal/abnormal classification. Within steady-state phonations, the phoneme /a/ outperforms /i/ and /e/ [13].

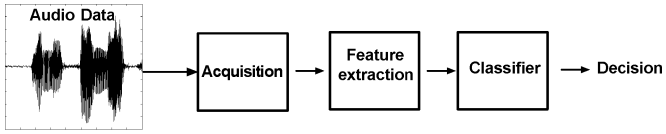

Fig. 1. Processes involved in voice pathology classification.

B. Acquisition

1) Telephone Simulation (Corpora 1–4): Four speech corpora were created from the original database simulating different types of distortions introduced by the telephone transmission and used to examine the robustness of different feature sets to discriminate normal from pathological voice. For Corpus 1, the 631 voice recordings were down sampled to 10 kHz (files were low pass filtered to prevent aliasing before down sampling). This distortion provides for the prevailing model of the telephone transmission path used in speech processing [14], [16]. The distortions applied to the four corpora were as follows:

- Corpus 1) predistortion, 10 kHz;
- Corpus 2) $1 +$ Downsampled to 8 kHz: (effective bandwidth 4 kHz);
- Corpus 3) $1 + 2 +$ spectrally shaped linear filter 200 Hz–3400 Hz;
- Corpus 4) $1 + 2 + 3 +$ noise additive white Gaussian noise at 30-dB signal-to-noise ratio (SNR).

2) Telephone Transmission:

a) Corpus 5: The fifth corpus was developed by transmitting all voice recordings over a long distance telephone channel and recording each at the receiver end. This was achieved using VoiceXML scripts developed with VoxBuilder.[1]

b) Categorized pathologic data sets: Diagnoses accompanying the database (Table I) were subcategorized into three independent classes; normal, neuromuscular pathologic and physical pathologic by specialist ENT (ear, nose and throat) surgeon (P. D. Lacy). Using this categorization, three smaller, homogenous train/test datasets comprising 1) neuromuscular pathologic and normal voice; 2) physical pathologic and normal voice; and 3) mixed pathologic and normal voice were produced using samples from the telephone-quality corpus (Corpus 5). The three datasets were used to measure the performance of the system on the different pathology types. For direct comparison of results obtained from the three datasets, each included a similar number of voice recordings (Fig. 2).

C. Feature Extraction

Features typically extracted from the audio data for voice pathology analysis include the fundamental frequency (F0),

[1]Voxpilot Ltd., Dublin, Ireland (www.voxpilot.com).

TABLE I
PATHOLOGY SUBSETS

Subset A *: Neuromuscular, Normal:* Altered nerve supply (Too much, too little, irregular), Interrupted Nerve Supply.	***Subset B****: Physical, Normal :* Vocal cord anomalies, masses adjacent to vocal cords, discrete laryngeal anomalies, impaired movement	***Subset C****: Mixed, Normal:* A&B
No. of pathologic voice samples: 56	*No. of pathologic voice samples: 39*	*No. of pathologic voice samples: 56*
No. of normal voice samples: 54	*No. of normal voice samples: 54*	*No. of normal voice samples: 54*

Disorder	*No. of Cases*	*Disorder*	*No. of Cases*	*Disorder*	*No. of Cases*
Abductor Spasmodic dysphonia	1	Cordectomy	1	*Neuromuscular)*	
Adductor Spasmodic dysphonia	8	Cyst	3	Adductor Spasmodic dysphonia	3
A-P Squeezing	10	Erythema	1	A-P Squeezing	13
Bowing	5	Erythoplasia	1	Bowing	2
Conversion aphonia	1	Fusiform mass/swelling	1	Conversion dysphonia	1
Conversion dysphonia	6	Gastric Reflux	6	Hyperfunction	48
Dyskinesia	1	Granulation Tissue	1	Presbyphonia	1
Episodic functional dysphonia	1	Hemmorhagic Reinke's	1	Ventricular Compression	4
Essential Tremor	1	Hemmorhagic Polyp	2	Vocal Tremor	1
Hyperfunction	42	Malignant Tumor	1	Ventricular Phonation	1
Idiopathic Laryngeal Discoordination	1	Normal	54		
Idiopathic neurological Disorder	1	Pachydermia	1	*(Physical)*	
Muscular Dystrophy	1	Papilloma	1	Arytenoid dislocation	1
Normal	54	Polpoid degeneration		Blunt Trauma	1
Paradoxical VC movement	2	(Reinke's)	6	Contact granuloma	1
Paralysis	17	Post microflap resection	2	Cyst	2
Paresis	8	Post radiation	1	Erythema	2
Presbyphonia	3	Subglottis stenosis	1	Fusiform mass/swelling	4
Ventricular Compression	6	Sulcus vocalis	1	Gastric Reflux	8
Vocal Tremor	5	Teflon granuloma	2	Generalised Edema	1
		Tissue change	1	Inflamed Arytenoid	1
		Varix	3	Keratosis/Leukoplasia	4
		Vocal fold edema	5	Nodular swelling	1
		Vocal fold lesion	1	Normal	53
		Vocal fold polyp	4	Pachydermia	1
		Vocal fold nodules	8	Papilloma	1
				Polpoid degeneration (Reinke's)	2
				Post laryngoplasty	2
				Post radiation	2
				Post Thyroplasty	3
				Prenodular swellings	1
				Varix	2
				Vocal fold edema	13
				Vocal fold lesion	1
				Vocal fold polyp	7
				Vocal fold nodules	3

Diagnoses of samples included from the database "Disordered Voice Database Model 4337" [9] acquired at the Massachusetts Eye and Ear Infirmary (MEEI). At the highest categorization level (i.e., either normal or pathologic) the categories are mutually exclusive. At lower levels of categorization the patient's diagnoses are no longer mutually exclusive and a subject may be diagnosed into more than one category, i.e., They may have a pathology that is both physical and neuromuscular.

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), SNRs and HNRs [3]. A design philosophy was to ensure that such a remote automatic system was both assessable and acceptable to specialized clinicians and speech therapists. To this end, feature groups already understood and employed as a measure of vocal quality by these professionals, were used in this study.

Three independent sets of features were used in this study: pitch perturbation features, amplitude perturbation features and HNR. Combining all of the above features formed a fourth feature set. These form the input vector processed by the classifier and are discussed below.

1) Pitch and Amplitude Measures: Pitch and amplitude perturbation measures [10], [24], [25] were calculated by segmenting the speech waveform (3–5 s in length) into overlapping "epochs." These measures include values of perturbation estimated on a per epoch basis (Tables II and III). The overlap of adjacent epochs (see below) provides for assessment of the changes in glottal cycles. In [8], a peak-detection algorithm was used to measure pitch periods to determine changes across glottal cycles for pathologic voice. This study utilized the autocorrelation in the time domain [26] to track the pitch and amplitude contour to prevent extraneous peaks introduced by channel transmission influencing the calculation of pitch periods.

Each epoch was 20-ms and overlapped neighboring epochs by 75% [24]. A 20-ms epoch was necessary to give an accurate representation of pitch. For each epoch, the value of the fundamental frequency, or pitch F_i, was calculated and returned with its corresponding amplitude measure A_i. These epoch values

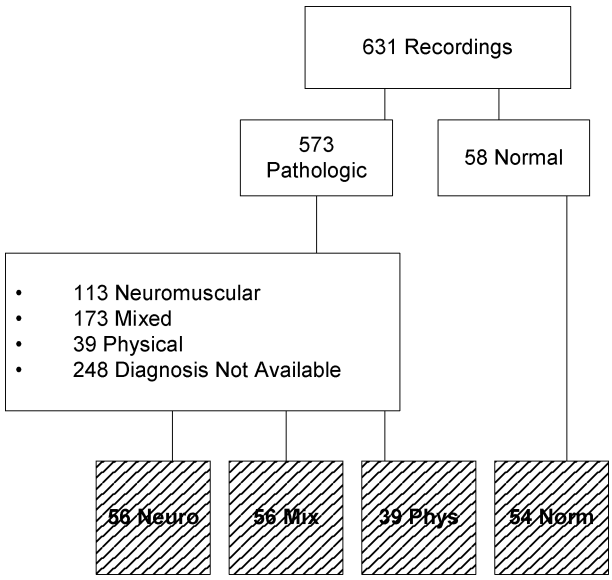

Fig. 2. Database catagorization.

were used to create two one-dimensional vectors, which define the "pitch contour" and "amplitude contour" of a particular voice. N_{voice} is a counting measure of the difference in pitch/amplitude between epoch value i and epoch value $i+1$ and n is the number of epochs extracted.

2) Harmonic to Noise Ratio: The Cepstral domain is employed in speech processing as the lower valued cepstral "quefrencies" model the vocal tract spectral dynamics, while the higher valued quefrencies contain pitch information, seen as equidistant peaks in the spectra.

The HNR [19] employed here indicates the stability of voiced speech. Eleven HNR measures (Table IV) at different frequency bands were calculated in the Cepstral domain, transformation to this domain is illustrated in Fig. 3.

1) Speech signal is normalized to have zero mean and unit variance.
2) One 100-ms epoch is extracted from the middle of the phonation.
3) A hamming window is applied to the segment and the inverse Fourier transform of the log spectrum is calculated.
4) A peak-detection algorithm locates the peaks at multiples of the fundamental frequency.
5) 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.
6) 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, $O(f)$.
7) The HNR for a given frequency band B is then calculated as per

$$\mathrm{HNR}_{\beta}(f) = \mathrm{mean}\left(O(f)\right)_{\beta} - \mathrm{mean}\left(N(f)\right)_{\beta}$$

3) Aggregate Feature Set: Combining all of the above features formed a fourth feature set. It contained 35 features.

TABLE II
PITCH PERTURBATION FEATURES

Feature Identifier	Description	Calculation Method
1	Mean F0 *(F0_av)*	$\frac{1}{n}\sum_{i=1}^{n} F_i$
2	Maximum F0 Detected *(F0_hi)*	$\max(F_i)$
3	Minimum F0 Detected *(F0_lo)*	$\min(F_i)$
4	Standard Deviation of F0 contour *(F0_sd)*	$\frac{1}{n-1}\sum_{i=1}^{n}\left(F_i - \overline{F}\right)^2$
5	Phonatory Frequency Range (PFR)	$\frac{\log\left(\frac{F0\_hi}{F0\_lo}\right)}{\log 2} x12$
6	Mean Absolute Jitter *(MAJ)*	$\frac{1}{n-1}\sum_{i=n-1}^{1} \lvert F_{i+1} - F_i \rvert$
7	Jitter (%) (JITT)	$\frac{MAJ}{F0\_av}$
8	Relative Average Perturbation; smoothed over 3 pitch periods (RAP)	$\frac{\frac{1}{n-2}\sum_{i=2}^{n-1} \lvert \frac{F_{i+1}+F_i+F_{i-1}}{3} - F_i \rvert}{F0\_av}\times 100$
9	Pitch Perturbation Quotient; smoothed over 5 pitch periods (PPQ_5)	$\frac{\frac{1}{n-4}\sum_{i=3}^{n-2} \lvert \frac{\sum_{k=i-2}^{i+2} F(k)}{5} - F_i \rvert}{F0\_av}\times 100$
10	Pitch Perturbation Quotient; smoothed over 55 pitch periods (PPQ_55)	$\frac{\frac{1}{n-54}\sum_{i=28}^{n-27} \lvert \frac{\sum_{k=i-27}^{i+27} F(k)}{55} - F_i \rvert}{F0\_av}\times 100$
11	Pitch Perturbation Factor (PPF)	$\frac{N_{p \geq threshold}}{N_{voice}}\times 100$ *where,* $*N_p$ *: epoch perturbation across time greater than 0.5msec in magnitude*
12	Directional Perturbation Factor (DPF)	$\frac{N_{\Delta\pm}}{N_{voice}}\times 100$ *where,* $*\ N_{\Delta\pm}$ *: epoch perturbation across time for which there is a change in algebraic sign.*

D. Classifier

1) Linear Discriminant (LD) Analysis: Histograms of the features across the normal and pathological classes revealed that each had an approximate Gaussian distribution suggesting that a LD classifier was an appropriate classifier model. LD analysis has been employed by Umapathy *et al.* [35] for discriminating normal from pathologic voice. Cross-fold validation was used to estimate classification accuracies. Their scheme, which used

TABLE III
AMPLITUDE PERTURBATION FEATURES

Feature Identifier	Description	Calculation method
1	Mean Amp *(Amp_av)*	$\frac{1}{n}\sum_{i=1}^{n} A_i$
2	Maximum Amp Detected *(A_hi)*	$\max(A_i)$
3	Minimum Amp Detected *(A_lo)*	$\min(A_i)$
4	Standard Deviation of Amp contour *(A_sd)*	$\frac{1}{n-1}\sum_{i=n-1}^{1}\left(A_i-\bar{A}\right)^2$
5	Mean Absolute Shimmer *(MAS)*	$\frac{1}{n-1}\sum_{i=n-1}^{1}\lvert A_{i+1}=A_i\rvert$
6	Shimmer % (SHIM%)	$\frac{MAS}{Amp\_av}$
7	Shimmer :Decibels (SHIM)	$\frac{1}{n-1}\sum_{i=1}^{n-1}20\times\log\left(\frac{A_i}{A_{i+1}}\right)$
8	Amplitude Relative Average Perturbation smoothed over 3 pitch periods (ARP)	$\frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\lvert\frac{A_{i+1}+A_i+A_{i-1}}{3}-A_i\rvert}{Amp\_av}\times 100$
9	Amplitude Perturbation Quotient smoothed over 5 pitch periods (APQ_5)	$\frac{\frac{1}{n-4}\sum_{i=3}^{n-2}\lvert\frac{\sum_{k=i-2}^{i+2}A(k)}{5}-A_i\rvert}{Amp\_av}\times 100$
10	Amplitude Perturbation Quotient smoothed over 55 pitch periods (APQ_55)	$\frac{\frac{1}{n-54}\sum_{i=28}^{n-27}\lvert\frac{\sum_{k=i-27}^{i+27}A(k)}{55}-A_i\rvert}{Amp\_av}\times 100$
11	Amplitude Perturbation Factor (APF)	$\frac{N_{p\geq threshold}}{N_{voice}}\times 100$
12	Amplitude Directional Perturbation Factor (ADPF)	$\frac{N_{\Delta\pm}}{N_{voice}}\times 100$

high-quality recordings of continuous speech, achieved a correct classification rate of 96.1% for normal voice samples and 92.5% for abnormal voice samples. LDs 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 "plug-in" maximum likelihood estimates [27], [28].

Training of the LD classifiers proceeded as follows. Let $\mathbf{x}$ be a column vector containing d feature values which is to be assigned to one of two classes. Assume there are N_1 feature vectors available for training the classifier from class 1 and N_2 feature vectors from class 2. The nth feature vector for training in class k is designated as $\mathbf{x}_n^{(k)}$. Training involves determining the class-conditional mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_1$ using

$$\boldsymbol{\mu}_1 = \frac{1}{N_1}\sum_{n=1}^{N_1}\mathbf{x}_n^{(1)},\ \boldsymbol{\mu}_2 = \frac{1}{N_2}\sum_{n=1}^{N_2}\mathbf{x}_n^{(2)} \quad (1)$$

TABLE IV
HNR BANDS

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

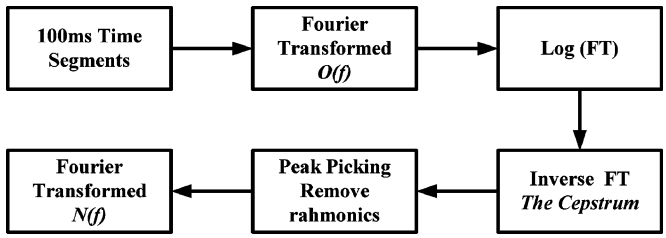

Fig. 3. HNR domain construction.

and the common covariance matrix $\boldsymbol{\Sigma}$ using

$$\boldsymbol{\Sigma} = \frac{1}{N_1+N_2-2}\sum_{k=1}^{2}\sum_{n=1}^{N_k}\left(\mathbf{x}_n^{(k)}-\boldsymbol{\mu}_k\right)\left(\mathbf{x}_n^{(k)}-\boldsymbol{\mu}_k\right)^T. \quad (2)$$

The common covariance matrix is employed since within class covariance is similar and offers better representation for the limited normal training set. After determining the $\boldsymbol{\mu}_k$'s and $\boldsymbol{\Sigma}$ from the training data, a feature vector $\mathbf{x}$ is classified by assuming values for the prior probability of class 1, π_1 (note that $\pi_2 = 1-\pi_1$), and calculating the discriminant value, y using

$$y = (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2) + \log\left(\frac{\pi_1}{(1-\pi_1)}\right). \quad (3)$$

The posterior probabilities for classes 1 and 2 are then calculated using

$$P(1|\mathbf{x}) = \frac{\exp(y)}{\exp(y)+1},\ P(2|\mathbf{x}) = 1-P(1|\mathbf{x}). \quad (4)$$

The final classification is obtained by choosing the class with the highest posterior probability estimate from (4). For all experiments performed in this study, the prior probability of class 1 was set to $\pi = 0.5$.

2) Assessing Classifier Performance: The cross-validation scheme [29] 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

TABLE V
DEFINITIONS OF TRUE POSITIVES, TRUE NEGATIVES, FALSE POSITIVES AND FALSE NEGATIVES

		True Classification	
		Pathology	Normal
Predicted Classification	Pathology	*True Positive* **TP**	*False Negative* **FN**
	Normal	*False Positive* **FP**	*True Negative* **TN**

each run. In this study, ten repetitions of tenfold cross-validation were used to estimate classifier performance. For each run of cross fold validation the number of normal and abnormal cases were equal.

Classifier performance was measured using sensitivity, specificity, positive predictivity, negative predictivity, and the overall accuracy. These measures were calculated as per the definition of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) presented in Table V:

1) $\text{sensitivity} = \text{TP}/(\text{TP}+\text{FN})$;
2) $\text{specificity} = \text{TN}/(\text{TN}+\text{FP})$;
3) $\text{positive predictivity} = \text{TP}/(\text{TP}+\text{FP})$;
4) $\text{negative predictivity} = \text{TN}/(\text{TN}+\text{FN})$;
5) $\text{overall\_accuracy} = (\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN})$

3) Receiver Operator Characteristics(ROC): ROC curves provide valuable information on a test's ability to discriminate between two classes over the complete spectrum of decision thresholds. The ROC curve is a graph of sensitivity versus (100%—specificity) as the *a priori* probabilities of the two classes are swept between zero and one. It provides information on clinical usefulness since it presents a trade-off in costs between false positives and false negatives and can be used to decide the threshold for different clinical requirements e.g., screening versus presurgical diagnosis. The area enclosed by the ROC plot is a metric against which other similar tests can be compared [30]. The area was computed using the trapezoidal rule.

IV. RESULTS

A first assessment of the classifier was made to determine its performance in separating normal recordings from nonhomogeneous pathology recordings. Following this a second assessment of the performance of the classifier was tested in order to examine: 1) if performance varied for different types of voice pathology; 2) if system performance improved using more homogenous training and testing data. Key results are described as follows.

A. Normal and Nonhomogeneous Pathology Discrimination

Table VI shows the performance of the classifier processing the pitch perturbation, amplitude perturbation, HNR and the combined feature sets for corpora one to four. For the telephone database (Corpus 5), the classifier performance (including 95% confidence intervals) processing the three features sets is presented in Table VII. For this corpus, bands 6 and 11 were omitted from the HNR feature vector. However, due to channel variability bands 1–5 were measured, as for Corpora 1–4, from 0 Hz. Performance figures for the best and worst contributing features for the transmitted database are presented in Table VIII. Fig. 4 shows the ROC curve of the classifier processing the aggregate feature set. The area under the curve was 73.7%.

B. Normal and Homogenous Pathology Discrimination

Using the jitter and shimmer feature sets, the classifier was trained and tested on the telephone-quality data, using the three pathology groupings (neuromuscular disorders, physical disorders and mixed disorders) independently. The results using this subset of Corpus 5 are presented in Table IX

V. DISCUSSION

This paper reports on the investigation of a remote system for the detection of laryngeal pathologies. The discussion below includes:

- an examination of classification system degradation when performing autodiagnosis remotely versus autodiagnosis using controlled recordings;
- an examination of telephone simulation reliability, whereby sources of information loss were tracked from the control set to the transmitted set using simulation corpora;
- a comparison of our system methodology with previously published research;
- pathologic class subcategorization into three homogenous groups and the resulting improved system performance;
- practical implications for remote pathology assessment.

A. Examination of Discriminating Feature Vectors Across the Corpora

Twelve pitch and twelve amplitude perturbation measures were extracted. The overall accuracies attained, presented in Table VI and Table VII, indicate these features performed steadily across corpora. Comparing the high-quality test set and transmitted test set, jitter feature discrimination fell by approximately 4% from 68.9% to 64.7%. Similarly, performance loss of 4.1% was observed for the shimmer feature set. Using just one measure of pitch perturbation, the Pitch Perturbation Factor (PPF) and amplitude perturbation, the Amplitude Relative average Perturbation smoothed over three pitch periods (ARP), accuracies of 63.3% and 63.7%, respectively, were achieved on *Corpus 5* (Table VIII).

HNR measures were not as suitable for remote pathological/normal identification. Table VI shows that the largest source of information loss appears when frequencies from 0–200 Hz and 3400–4000 Hz are removed after spectral shaping. HNR measures produce 79.8% overall accuracy after down sampling from 10 kHz to 8 kHz (*Corpus 2*) but just 63.7% after simulated channel filtering (*Corpus 3*). This indicates that key discriminating information is present in the cepstrum's low frequency components. HNR measures, which proved most useful in differentiating normal and abnormal for the high-quality samples (Table VI: 77.8%), resulted in the lowest classification performance of the three feature sets for the telephone data (Table VII: 57.9%). For the band comprising the greatest breadth of frequencies (Table VIII) the sensitivity measure 4.8% indicates a strong bias of HNRs in favor of the normal class. This contributed most significantly to final performance loss for the transmitted system tests. Based on these results, the HNR measures were not included in the feature

TABLE VI
CLASSIFICATION PERFORMANCE OF THE CANDIDATE CONFIGURATIONS ON CONTROL AND SIMULATED TELEPHONE DATASETS. THE FEATURE SET RESULTING IN THE BEST CLASSIFIER PERFORMANCE FOR EACH CORPORA ARE SHOWN IN BOLD

Corpus	Description	Feature Set(s)	Acc (%)	Spec (%)	Sens (%)	Pos Pred (%)	Neg Pred (%)
1.	Control 10kHz	Combined: Jitt, Shim, HNR	89.10	93.26	85.14	87.63	86.25
		Jitter Accuracy (%) : 68.93					
		Shimmer Accuracy (%) : 77.10					
		HNR Accuracy (%) : 77.80					
2.	Downsampled to 8kHz	Combined: Jitt, Shim, HNR	83.35	85.52	81.15	82.12	84.70
		Jitter Accuracy (%) : 66.09					
		Shimmer Accuracy (%) : 77.97					
		HNR Accuracy (%) : 79.79					
3.	Spectrally Shaped	Combined: Jitt, Shim, HNR	82.49	84.15	80.68	82.06	83.61
		Jitter Accuracy (%) : 66.00					
		Shimmer Accuracy (%) : 75.63					
		HNR Accuracy (%) : 63.66					
4.	Noisy	Combined: Jitt, Shim, HNR	79.14	79.86	78.25	77.03	80.16
		Jitter Accuracy (%) : 75.70					
		Shimmer Accuracy (%) : 74.85					
		HNR Accuracy (%) : 57.16					

TABLE VII
THE CLASSIFICATION PERFORMANCE (WITH 95% CONFIDENCE INTERVALS) OF THE SYSTEM PROCESSING THE TELEPHONE DATASET

Feature Set(s)	Acc (%)	Spec (%)	Sens(%)	Pos Pred (%)	Neg Pred(%)
Combined: Jitt, Shim, HNR	74.15	75.69	72.60	73.66	74.69
	+/- 7.2	+/- 8.3	+/- 9.9	+/- 11.4	+/- 5.6

Jitter Accuracy (%): 64.7 +/- 5.7
Shimmer Accuracy (%): 73.03 +/- 6.7
HNR Accuracy (%): 57.85 +/- 10.6

vector used in the second study utilising the homogeneous pathologic sets (Table IX).

B. Telephone Degradations

By comparing the results from Corpora *1–4* (Table VI) with the results of the actual transmitted corpus (*Corpus 5*, Table VII, one can observe each major source of information loss contributing independently to the final classification accuracy measure. This is consistent with other studies of intelligent telephone speech systems. In [16], it was reported that speaker recognition decreased from 99.1% using high-quality recordings (TIMIT) to 69.0% using a telephone-quality corpus (NTIMIT). Here, 14% of performance loss was reported as due to band limiting, filtering and noise addition, with the remaining 16.1% being accounted for by channel and microphone nonlinearity's. Using our system of acquisition the call/answer VoiceXML program does not require any microphone for

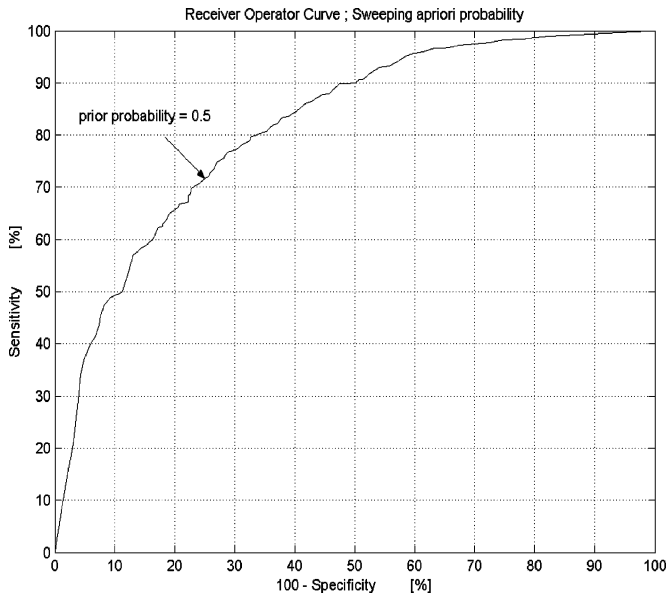

Fig. 4. ROC plot for Transmitted Corpus 5. The classifier performance at the *a priori* probability $\pi_k = 0.5$ is indicated. The area under the curve is 73.7%.

TABLE VIII
SUMMARY TABLE OF CLASSIFICATION PERFORMANCE OBTAINED WITH INDIVIDUAL FEATURES ON THE TELEPHONE DATASET INLCUDING 95% CONFIDENCE INTERVALS

Feature Set	Pathology /Normal	*Feature Identifier* (as per tables 3&4)	ACCURACY(%)	SPEC (%)	SENS(%)	POS PRED(%)	NEG PRED(%)
Jitter	*Best Discrimination:*	10. (PPF)	63.31 +/- 10.3	94.24 +/- 12.2	32.76 +/- 16.1	85.20 +/- 6.2	58.60 +/- 12.2
	Worst Discrimination:	12. (DPF)	49.35 +/- 9.5	62.48 +/- 13.7	36.38 +/- 12.3	49.53 +/-6.6	49.24 +/- 18.0
Shimmer	*Best Discrimination:*	8. (ARP)	63.66 +/- 8.2	44.15 +/- 7.6	82.93 +/- 11.2	60.05 +/-9.3	71.88 +/- 10.0
	Worst Discrimination:	3. (*A_lo*)	47.70 +/- 9.1	45.72. +/- 13.1	49.66 +/- 13.0	48.08 +/- 9.3	47.29 +/-11.4
HNR	*Best Discrimination:*	10. (3000-4000 Hz)	57.15 +/- 10.2	52.41 +/- 10.3	74.00 +/- 12.2	67.11 +/- 14.0	60.51 +/- 7.5
	Worst Discrimination:	4. (0-3000 Hz)	49.78 +/- 6.6	95.29 +/- 12.5	4.83 +/- 7.2	50.91 +/- 5.9	49.73 +/- 10.3

TABLE IX
SUMMARY TABLE OF HOMOGENOUS PATHOLOGY CLASSIFIER PERFORMANCE USING JITTER FEATURES 1–12 AND SHIMMER FEATURES 1–12

		Performance Measure				
		Acc (%)	Spec (%)	Sens (%)	Pos Pred (%)	Neg Pred (%)
Pathology Group	Neuromuscular	87.27	90.74	83.93	84.48	90.38
	Physical	77.97	80.19	74.87	81.54	73.18
	Mixed	61.08	62.43	60.18	50.88	70.80

voice sample playback, however, a transducer was used for the original recordings (unlike NTIMIT) and so transducer nonlinearity's are present in *Corpus 5*.

C. Comparisons With Other Automatic Classification Schemes

A number of research groups [5], [6], [31], [32] have reported results for detection rates for voice pathologies of 95.1%, 96%, 76%, and 91.3%, respectively.

Godino-Llorente *et al.* [6] used samples from the database employed in this study. After trialing 5000 different classifier configurations their best accuracy was 96.0%. The input feature vectors for the artificial neural networks (ANNs) comprised of Mel-Frequency Cepstral coefficients and features based on first and second derivatives of the MFCCs. Using learning vector quantization, sensitivity was reported as 94.90%, specificity of 95.89%, false positives of 2.05% and false negatives of 2.54%. On initial inspection, this appears to be a more successful detection scheme than that developed here. However, the authors used a subset containing 135 files, 53 normal and 82 pathologic voices (chosen randomly from the 573—but not published) which, in turn was divided into a fixed training set (70%) and a fixed test set (30%). Without the specific knowledge of the files used by the authors for testing their system we have been unable to do a side-by-side comparison of the performance of the two systems.

In studies [5], [31], and [32], different databases (with high-quality acoustic data) were used and a direct comparison of results cannot be made.

Nonlinear classification methods used for discriminating pathologic and normal voice have been employed in previous detection schemes. These include hidden Markov models (HMM) [36], and Neural Networks [6]. Dibazar *et al.* [36] have compared LDs with a three-state, three-mixture HMM method using the same database as employed here [23]. They report minor improvements of 2.4% when using HMMs over LD analysis. In [5], the authors compared a classification method based on self-organizing maps (SOMs) to the LD classifiers used in this study and found that the SOM outperformed the LDs by approximately 5%. The focus of our investigations was on the features themselves and we did not undertake a comparison of classifier models. An appropriate model for HMM pattern recognition of pathologic voice may lead to differences in classification accuracy, but require dynamic rather than static measures of pitch and amplitude perturbations for sustained phonations. Short-term window measures that track the pitch and amplitude contours would act as an appropriate input, provided window lengths were sufficiently long to not incur extraneous measures from transmission interference.

D. Normal and Homogenous Pathology Discrimination

The performance of the system in discriminating normal from homogeneous pathologic groups was higher than the re-

sults from the nonhomogeneous pathological groups (Table IX). Overall accuracy for detecting neuromuscular voice disease remotely was 87.3% with a positive predictivity of 84.5% and negative predictivity of 90.4%. Physical pathologies were also classified more successfully using homogenous training, with system accuracy rate of 77.9%. Samples from those patients presenting with a mixture of pathologies were least likely to be correctly classified, with a classifier input feature vector containing aggregate jitter and shimmer features providing 61.1% overall accuracy. These results confirm the overall findings composed using *Corpus 5*. The average classification accuracy across these three configurations was 75.4%, which compares well with overall accuracy of 74.2%. (Table VII) obtained from *Corpus 5*.

E. Implementation Issues for Remote Pathology Assessment

VoiceXML is the emerging standard for intelligent web-based Interactive Voice Response (IVR) systems. Based on the Worldwide Web Consortium's (W3C) Extensible Markup Language (XML) [33], it leverages the web infrastructure for the design and deployment of voice enabled applications. Developers of these systems utilize VoiceXML "gateways," which provide IVR capabilities and call management services. By using this infrastructure, a three-way communication between the patient under test, signal processing decision software, and physician's database can be initialized. This provides for a fully automated assessment system, whereby a telephone call to the application can result in a voice pathology assessment being automatically posted to a website for consideration along with other patient information by the specialist or physician. This system may prove advantageous as a prescreening tool for clinics located far from specialist voice centres and without local access to videostroboscopy [2] and videokymography.

VI. Conclusion

In this study, a system for remote screening of voice pathologies over telephone lines was developed. Tests of the system demonstrated that it successfully separated 87.2% of voices with neuromuscular disorders from normal voices. When physical and mixed disorders were included in the abnormal voice set the accuracy was 74.2%. The significance of this study is that by combining a telephony interface and server side speech processing, the possibility of a practical system for remote diagnosis of voice pathology is elucidated. Such a system may reduce the costs currently associated with voice analysis and prove more convenient for patients.

Acknowledgment

The authors would like to thank the Informatics Research Initiative of Enterprise Ireland and Voxpilot Ltd. They would also like to thank the Speech and Language Therapy Department at Tallaght Hospital, Dublin for assistance in data collection.

References

[1] W. Becker, H. H. Naumann, and C. R. Faltz, *Ear, Nose and Throat Diseases*, 2nd ed. New York: Thieme Medical, 1994.

[2] B. Schneider, J. Wendler, and W. Seidner, "The relevance of stroboscopy in functional dysphonias," *Folia Phon.*, vol. 54, no. 1, pp. 44–54, 2002.

[3] C. Maguire, P. de Chazal, R. B. Reilly, and P. Lacy, "Automatic classification of voice pathology using speech analysis," presented at the *World Congr. Biomedical Engineering and Medical Physics*, Sydney, Australia, Aug. 2003.

[4] ——, "Identification of voice pathology using automated speech analysis," presented at the *3rd Int. Workshop Models and Analysis of Vocal Emission for Biomedical Applications*, Florence, Italy, Dec. 2003.

[5] S. Hadjitodorov, B. Boyanov, and G. Baudoin, "Laryngeal pathology detection by means of class specific neural maps," *IEEE Trans. Inf. Technol. Biomed.*, vol. 4, no. 1, pp. 68–73, Mar. 2000.

[6] J. I. Godino-Llorente and P. Gomez-Vilda, "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network-based detectors," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 2, pp. 380–384, Feb. 2004.

[7] R. J. Baken and R. Orlikoff, *Clinical Measurement of Speech and Voice*, 2nd ed. San Diego, CA: Singular, 2000.

[8] S. Feijoo and C. Hernandez, "Short-term stability measures for the evaluation of vocal quality," *J. Speech. Hear. Res.*, vol. 33, pp. 324–334, Jun. 1990.

[9] D. Michaelis, M. Frohlich, and H. W. Strube, "Selection and combination of acoustic features for the description of pathologic voices," *J. Acoust. Soc. Am.*, vol. 103, no. 3, pp. 1628–1639, 1998.

[10] P. Liebermann, "Perturbations in vocal pitch," *J. Acoust. Soc. Am*, vol. 33, no. 5, pp. 597–603, 1961.

[11] National Centre for Voice and Speech and I. R. Titze, "Summary Statement," presented at the Workshop Acoustic Voice Analysis, Denver, CO, 1994.

[12] G. de Krom, "Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments," *J. Speech. Hear. Res.*, vol. 38, pp. 794–811, 1995.

[13] M. de Oliveira Rosa, J. C. Pereira, and M. Grellet, "Adaptive estimation of residue signal for voice pathology diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 1, pp. 96–102, Jan. 2000.

[14] P. J. Moreno and R. M. Stern, "Sources of degradation of speech recognition in the telephone network," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing.*, vol. 1, Adelaide, Australia, Apr. 1994, pp. 109–112.

[15] D. A. Reynolds, "Large population speaker identification using clean and telephone speech," *IEEE Signal Process. lett.*, vol. 2, no. 3, pp. 48–48, Mar. 1995.

[16] D. A. Reynolds, M. A. Zissman, T. F. Quatieri, G. C. O'Leary, and B. A. Carrlson, "The effects of telephone transmission degradations on speaker recognition performance," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, vol. 1, Detroit, MI, May 1995, pp. 329–332.

[17] N. Yanagihara, "Significance of harmonic changes and noise components in hoarseness," *J. Speech Hear. Res.*, vol. 10, no. 3, pp. 531–541, 1967.

[18] H. Kasuya, S. Ogawa, K. Mashima, and S. Ebihara, "Normalized noise energy as an acoustic measure to evaluate pathologic voice," *J. Acoust. Soc. Am.*, vol. 80, no. 5, pp. 1329–1334, 1986.

[19] G. Krom, "A cepstrum-based technique for determining a harmonic-to-noise ratio in speech signals," *J. Speech. Hear. Res.*, vol. 36, pp. 254–266, 1993.

[20] C. Janowski and A. Kalyanswamy *et al.*, "NTIMIT: a phonetically balanced, continuous speech, telephone bandwidth speech database," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol. 1, Albuquerque, NM, Apr. 1990, pp. 109–112.

[21] K. L. Brown and E. B. George, "CIMIT: a speech corpus for the cellular environment with application to automatic speech recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol. 1, Detroit, MI, May 1995, pp. 105–108.

[22] P. A. Mashima and D. P. Birkmire-Peters *et al.*, "Telehealth: voice therapy using telecommunications technology," *Am. J. Speech Lang. Pathol.*, vol. 12, no. 4, pp. 432–439, Nov. 2003.

[23] Massachusetts Eye and Ear Infirmary Voice and Speech Lab, Boston, MA, "Disorder Voice Database Model 4337 (Kay Elemetrics Corp.),", Jan. 1994.

[24] Y. Koike, "Application of some acoustic measures for the evaluation of laryngeal dysfunction," *Studia Phonologica*, vol. 7, pp. 17–23, 1973.

[25] E. J. Wallen and J. H. L. Hansen, "A screening test for speech pathology assessment using objective quality measures," in *Proc 4th Int. Conf. Spoken Language*, vol. 2.3-6, Philadelphia, PA, Oct. 1996, pp. 776–779.

[26] L. R. Rabiner, "On the use of autocorrelation analysis for pitch detection," *IEEE Trans. Acoust., SpeechSignal Process.*, vol. ASSP-25, no. 1, Feb. 1977.

[27] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge, U.K.: Cambridge Univ. Press, 1996.

[28] R. O. Duda, P. E. Hart, and H. G. Stork, *Pattern Classification*. New York: Wiley-Interscience, 2000.
[29] R. Kohavi, "A study of cross validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Conf. Art. Intel.*, 1995, pp. 1137–1143.
[30] M. H. Zweig and G. Campbell, "Receiver operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," *J. Clin. Chem.*, vol. 39, no. 4, 1993.
[31] D. G. Childers, "Detection of laryngeal function using speech and electrographic data," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 1, pp. 19–25, Jan. 1992.
[32] C. E. Martínez and H. L. Rufiner, "Acoustic analysis of speech for detection of laryngeal pathologies," presented at the Chicago 2000 World Congr. IEEE EMBS, Chicago, IL, Jul. 2000.
[33] Proc. W3C VoiceXML Forum. [Online]. Available: www.voicexml.org
[34] H. K. Schutte, J. G. Svec, and F. Sram, "First results of clinical application of videokymography," *Larygoscope*, pt. 1, vol. 108, no. 8, pp. 1206–1210, Aug. 1998.
[35] K. Umapathy, S. Krishnan, V. Parsa, and D. G. Jamieson, "Discrimination of pathological voices using a time-frequency approach," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 3, pp. 421–430, Mar. 2005.
[36] A. A. Dibazar, S. Narayanan, and T. W. Berger, "Feature analysis for automatic detection of pathological speech," in *Proc. IEEE Joint EMBS/BMES Conf.*, Houston, TX, Oct. 2002, pp. 182–184.

Rosalyn J. Moran received the B.E. degree in electronic engineering, from University College Dublin, Dublin, Ireland, in 2004 where she is currently working toward the Ph.D. degree in electronic engineering.

Ms. Moran was awarded a research grant by The Embark Initiative, the research funding programme run by the Irish Research Council for Science Engineering and Technology (IRCSET) in 2004 for work involving the application of information theory to biomedical signal processing.

Richard B. Reilly (M'92–SM'04) received the B.E., M.Eng.Sc., and Ph.D. degrees in 1987, 1989 and 1992, all in electronic engineering, from University College Dublin, Dublin, Ireland.

In 1988, he joined Space Technology Ireland and the Department de Recherche Spatiale (CNRS group), Paris, France, developing DSP-based on-board experimentation for the NASA satellite WIND. In 1990, he joined the National Rehabilitation Hospital and in 1992 became a Post-Doctoral Research Fellow at University College Dublin, focusing on signal processing for speech and gesture recognition. Since 1996, He has been on the academic staff in the Department of Electronic and Electrical Engineering at University College, Dublin. He is currently Senior Lecturer and researches into neurological signal processing and multimodal signal processing.

Dr. Reilly was the 1999/2001 Silvanus P. Thompson International Lecturer for the Institution of Electrical Engineers (IEE). In 2004, he was awarded a US Fulbright Award for research collaboration into multisensory integration with the Nathan Kline Institute for Psychiatric Research, New York. He is a member of the IEEE Engineering in Medicine and Biology Society and Signal Processing Society. He is the Republic of Ireland representative on the Executive Committee of the IEEE United Kingdom and Republic of Ireland Section. He is an Associate Editor for IEEE TRANSACTIONS ON MULTIMEDIA and also a reviewer for IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, the *Journal of Applied Signal Processing*, *Signal Processing*, and *IEE Proceedings Vision, Image & Signal Processing*.

Philip de Chazal (M'94) received the B.E. degree in electronic engineering and the M.Biomed.E. and Ph.D. degrees in biomedical engineering from University of New South Wales, Sydney, Australia, in 1989, 1995, and 1999, respectively.

He was a Research Scientist for the CSIRO, Sydney, from 1990 to 1992, a Biomedical Engineer for MedCare Systems, Sydney from 1993 to 1999, and a Research Fellow at University College Dublin, Dublin, Ireland from 1999 to 2003. He is cofounder and currently Chief Technical Officer of BiancaMed, Dublin Ireland, a company providing intelligent computer-based analysis of medical signals. His research interests include signal processing and pattern recognition for biomedical applications and image processing for multimodal applications.

Dr. de Chazal is a member of the IEEE Engineering in Medicine and Biology Society. He is a reviewer for IEEE TRANSACTIONS ON MULTIMEDIA and IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING.

Peter D. Lacy qualified from medical school at Dublin University, Dublin, Ireland, in 1989 and trained as an Otolaryngologist/Head and Neck Surgeon in Dublin. He spent a year training in Clinical Epidemiology and Outcomes research a t Washington University, St Louis, before undertaking clinical Fellowships in the Royal Victorian Eye and Ear Hospital, Melbourne, Australia., and Cincinnati Children's Hospital

Dr. Lacy is a member of the Irish Institute of Otorhinolaryngology and the Royal Academy of Medicine. His research interests include staging of head and neck cancer, pediatric airway surgery, and research on clinical speech and gesture recognition with the department of electronic and electrical engineering at University College Dublin.