Models and Analysis of Vocal Emissions for Biomedical Applications
(MAVEBA 1999)
Firenze, Italy, September 1-3, 1999

CHARACTERISING VOWEL PHONATION BY FUNDAMENTAL SPECTRAL NORMALISATION OF LX-WAVEFORMS

C J Moore[1], N Slevin[2] and S Winstanley[3]

[1]North Western Medical Physics, Christie Hospital, Manchester, UK, (infcjm@dalpha2.cr.man/ac/uk)
[2] Department of Radiation Oncology, Christie Hospital, Wilmslow Road, Manchester, UK
[3]South Manchester University Hospitals Trust, Withington Hospital, Manchester, UK

Abstract

Objective spectral reference standards for normal vowel phonation are described. Application in larynx cancer monitoring is envisaged. 120 individuals contributed to a database of vowels /æ/ and /i/ in the form of trans-larynx impedance time series captured using an electrolaryngograph. The impedance signals are used for iterated power spectral estimation followed by spectral intra-pooling for each individual. Pooling of spectra across many individuals is complicated by significant variations in the precise frequencies and powers of the fundamental, harmonics and any other characteristic peaks that may be present. Fundamental-harmonic normalisation (FHN) of individual spectra circumvents these obstacles by normalising all powers relative to that of the fundamental and by transforming the entire frequency range into floating point multiples of the fundamental-frequency f_0. Population pooling then results in stable FHN-spectral patterns and associated characteristic distributions of f_0 values. For females it is FHN-spectral pattern that matters most. For males the f_0 distribution is also highly characteristic.

Introduction

The choice of medical treatment, driven by economic considerations and issues of accepted practice, is increasingly based on objectively assessed outcome. A particular example driving this study is the competition between radiotherapy and surgery as treatment of choice for larynx cancer. Since both techniques are equally successful for local control of the disease, quality of life after treatment has become an influencing factor. In this context radiotherapy has obvious advantages for the maintenance of vocal fold function compared to the surgical removal of tissues in laryngectomy. However, conservative and reconstructive surgical techniques now claim to offer similar advantages. The result is a heightened interest in the relative degree of conservation offered by the two therapeutic modalities. To the patient the most direct evidence of normal vocal fold functionality is 'voice quality'. To the expert such evidence is to be found in measures of glottal waveform that are free from the complicating factors introduced by tract resonance. In turn this has spotlighted the lack of concise but characteristically detailed objective reference standards for normal glottal waveforms, or for that matter voice quality, against which a patient can be compared. Pilot studies for assessing patient vocal fold function before and at intervals after radiotherapy have already shown promising results, even though they are based on relatively unrefined standards [1,2]. This report describes what is believed to be significant progress in the generation of improved standards that will underpin further clinical investigations

Theoretical Background

Speech and language therapists, SALTs, subjectively assess and score voice quality using a range of parameters [3]. Many parameters are simply descriptive, e.g. whisper, although some have been adopted from the physical sciences to produce a hybrid terminology, e.g. fundamental frequency and shimmer (rather than variance). All are capable of interpretation in terms of spectral content in the frequency domain [4]. The periodic glottal waveform produced by the vibrating vocal folds of the larynx is the driving force behind the production of the complex human acoustic waveform. As such the glottal waveform is an indication of the functional integrity of the vocal folds which, as already explained, is of particular interest in the management of larynx cancer. In this and the wider medical context SALTs increasingly supplement their assessments with objective measurements that are closely correlated with glottal waveform, in particular those from electrolaryngography [5].

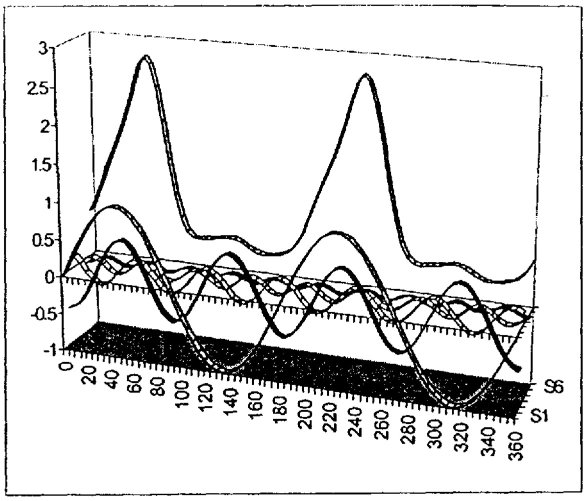

Figure-1
LX waveform (upper trace) synthesised from sine components (lower traces)

In electrolaryngography two transducers are placed across the thyroid cartilage of the larynx to measure electrical impedance changes, typically during phonation of vowels. The transducer signal, called an LX waveform, has a simple four phase periodic structure with superposed fine detail. Figure-1 shows a synthesised LX waveform made from simple sine components. Unusually for medical applications [6], the abundance and high signal to noise ratio of the impedance time series data is ideally suited to power spectral analysis [7].

The well known stability of LX waveforms for normal vowel phonation suggests that characteristic power spectral estimates can be generated for individuals. However, the varied positions and powers of peaks in the individual spectra demand that suitable normalisation be performed prior to any population pooling. Here we observe that SALTs tend to describe waveform phenomena intuitively, i.e. relative to fundamental frequency and harmonic pattern. Consequently we chose to normalise individual spectra with respect to the power of the fundamental and then to transform frequency from Hertz into a continuous floating point harmonic scale. True harmonic peaks always appear at integer multiples of f_0. Population pooling can then act to dramatically reinforce common harmonically related features. In summary this procedure separates spectral *pattern* from fundamental *frequency distribution.*

Methods

One hundred and twenty healthy volunteers, with equal male/female representation, were recruited through institutional advertising and the local news media. Under the expert guidance of speech and language therapists each volunteer was connected to a PC controlled electrolaryngograph, asked to phonate the vowels /æ/ and /i/, and the resulting impedance signals digitised at 20kHz for up to 4 seconds. The resultant binary LX data-files were transmitted by network to a DEC Unix Alpha-server-2000 dual 4/275 processor system for storage, visualisation and analysis using software written in the 4th generation VNI-WAVE language.

Impedance data streams are stationarised by differencing to remove any slow trends such as transducer drift, background sound, mains contamination etc. that could otherwise obstruct subsequent analysis. The processed data stream is then split into consecutive data frames, each 1000 samples long. The auto-covariance of each frame is computed, variance reduced by a Hanning window and then Fourier transformed to provide a power spectral estimate.

The consecutive power-spectral estimates, each corresponding to 0.05 second increments, provide a time developing view of impedance frequency characteristics that are highly correlated with underlying glottal waveform. Excluding the chaotic interval at the onset of phonation, the consecutive power spectral estimates are pooled to yield a reliable estimate of the underlying spectrum for an individual volunteer. This is the 'intra-pooled spectrum' where each peak now has an associated standard error analogous to the ubiquitous 'shimmer' for the fundamental frequency, f_0.

Pooling of spectra across the male or female populations is preceded by fundamental-harmonic normalisation (FHN) scheme. Taking each intra-pooled spectrum in turn, the power and frequency of the fundamental, $P(f_0)$ and f_0, are identified by sequential peak thresholding

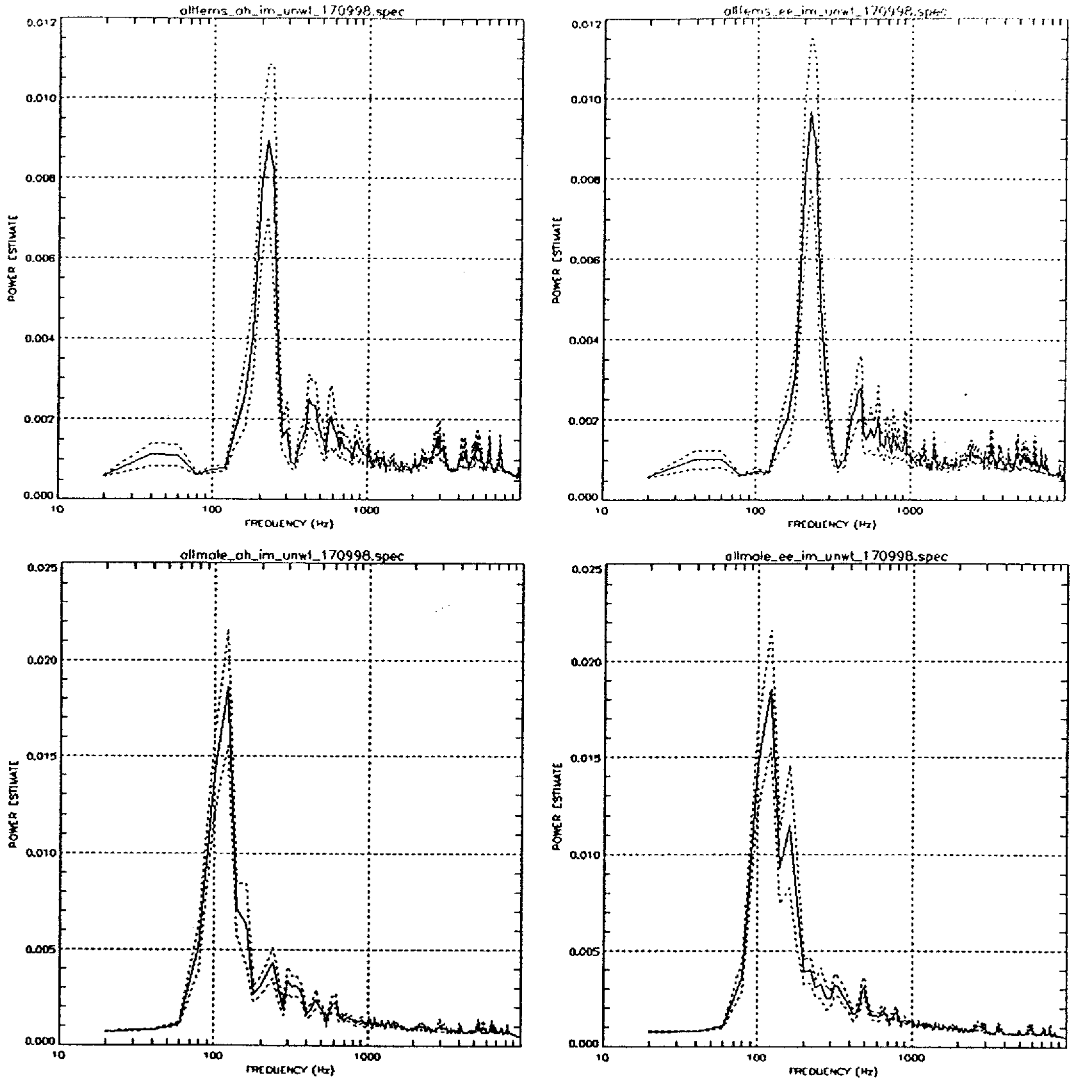

Figure-2
Non-normalised, pooled impedance power spectra.
Upper: Female /æ/ (left), /i/ (right) Lower: Male /æ/ (left), /i/ (right)

within the 50-300 Hz band. All spectral powers are then normalised relative to $P(f_0)$. The value of f_0 is then used to transform the discrete frequency scale, initially dimensioned in Hertz, into a discrete harmonic scale, where each bin is a floating point multiple of the particular f_0. True harmonics appear at integer multiples of f_0 and other spectral contributions appear at fractional locations on the scale.

Since all the FHN-spectra are derived from 1000 sample data frames with the same Nyquist limit, those associated with low values of f_0 span more harmonics than those associated with high values of f_0. Consequently, to facilitate pooling it is necessary to interpolatively map every FHN-spectrum into a common harmonic scale with bin size determined by the highest value of f_0 found in the population. All the FHN-spectra are then pooled for a given population and vowel sound, i.e. male and female, /æ/ and /i/.

Results

For a sample of volunteers in this study individual spectra were found to be highly stable, distinctly featured and characteristic of each individual, even when generated from data gathered on four different occasions. This encouraged attempts to directly pool the data, particularly in view of the remarkably similar peak structure found in the female spectra. Figure-2 shows the result of direct pooling for

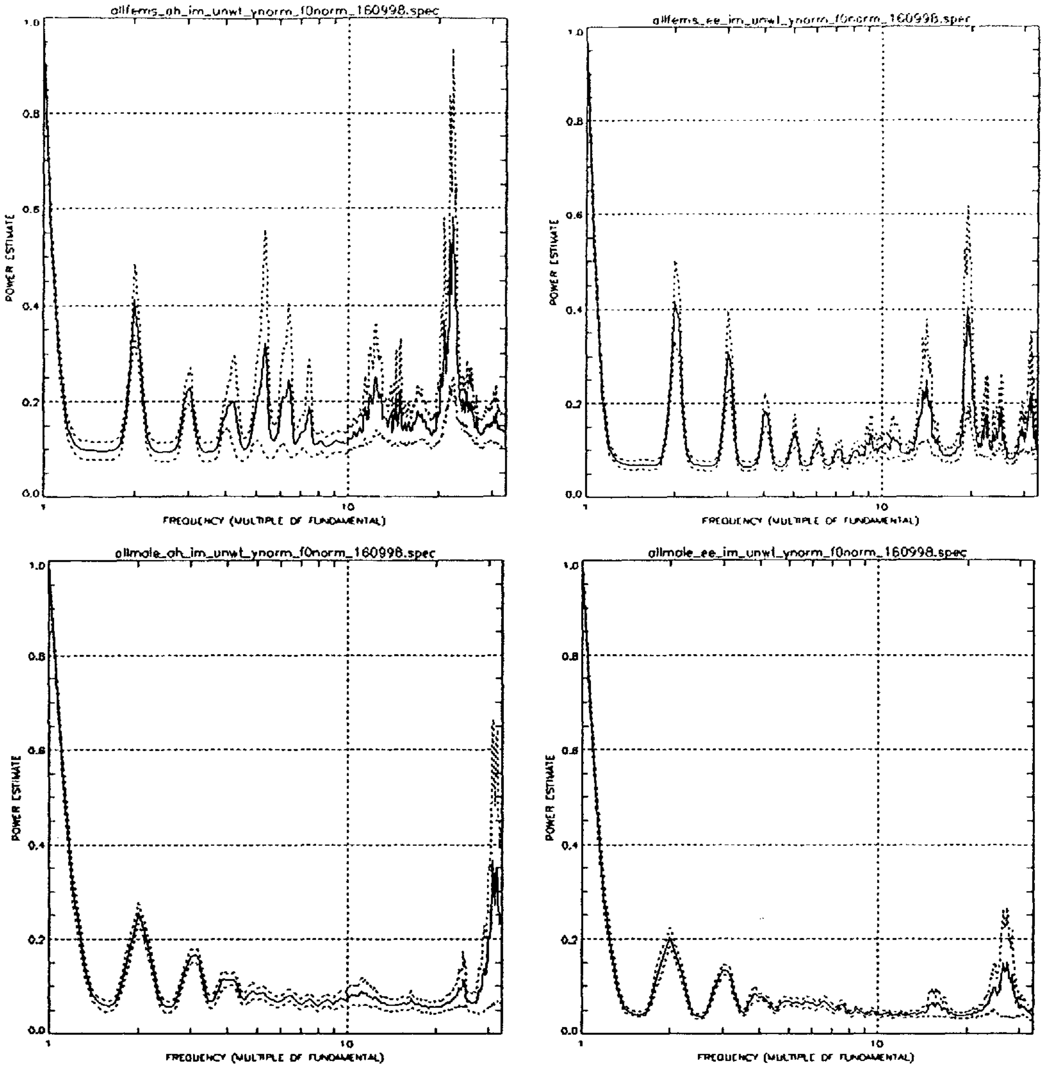

Figure-4

Pooled normal impedance power spectra where each contributing spectrum is normalised to fundamental-power and the frequency rescaled to multiples of fundamental frequency.
Upper: Female /æ/ (left), /i/ (right) Lower: Male /æ/ (left), /i/ (right)

Discussion and Conclusion

Convergence to the particular FHN spectral patterns shown in Figure-4 has been tested against the number of volunteers from whom data has been gathered. Stability is achieved after approximately 30 individuals have been pooled. However, the present 60 male, 60 female pooling is to be extended to 100 in each population.

The distributions of f_0 in Figure-5 for males suggest there may be two normal sub-groupings, as indicated in an earlier study [1]. The majority group has a low frequency fundamental peak between 100-120 Hz whilst the minority group is at 160 Hz. Clarifying the existence of the latter has important consequences for the practical assessment of voice quality in patients. Male patients with larynx cancer have been observed to switch between low and high pitch phonation at particular stages in their therapy. Voice quality as it stabilises a year or so after therapy is an emotive issue for patients. They alone are aware of their own ‘normal sub-group’ voice characteristics i.e. prior to the onset of disease.

Armed with objective reference standards, in the form of the FHN patterns, clinical studies of the progress of larynx cancer patients undergoing radiotherapy for cancer of the larynx are underway at the Christie Hospital.

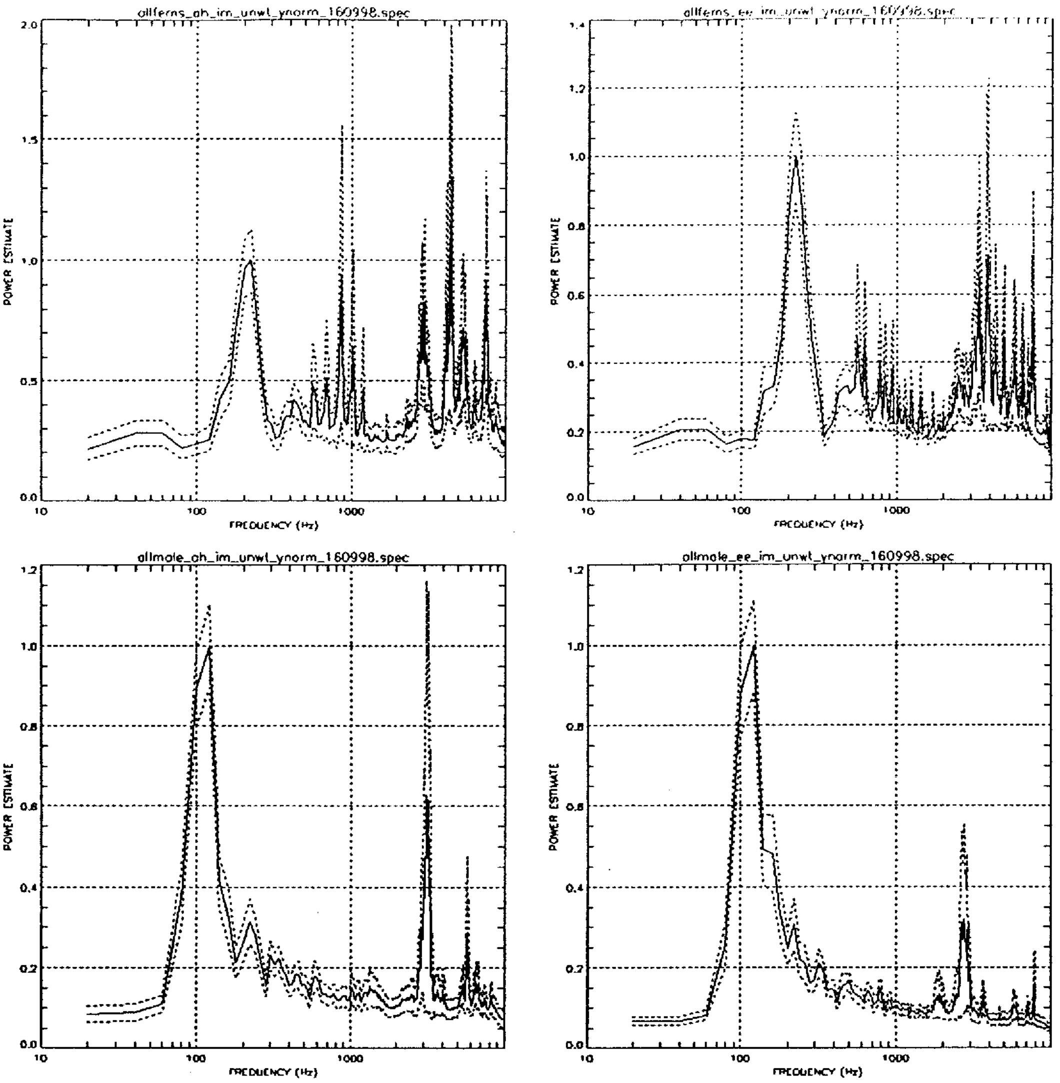

Figure-3

Pooled impedance power spectra, each contributing spectrum has been normalised to fundamental-power.
Upper: Female /æ/ (left), /i/ (right) Lower: Male /æ/ (left), /i/ (right)

both male or female populations. The smeared spectra are promising but not ideal as 'normal' reference standards. Normalisation of the individual spectra relative to fundamental power produces the results shown in Figure-3. Whilst spectral peaks become sharply defined (solid lines) the standard errors (dashed lines) are large, due to misalignment of corresponding features.

In addition to fundamental power normalisation, harmonic rescaling of the frequency axes of the individual spectra produces the results shown in Figure-4. This FHN scheme clearly aligns corresponding harmonics from each individual spectrum so that pooling produces characteristic peaks with low variance. For /æ/ and /i/ the female population FHN-spectra each exhibit 6 narrow, relatively high power harmonic peaks. However, the /æ/ spectrum has enhanced 4th-6th harmonics, increased variance and a higher frequency f_0 distribution compared to /i/. The two male population FHN-spectra have 3 wide, relatively low power harmonics. Both show remarkably similar patterns of harmonic decline and low variance. Unlike the females, where FHN-spectral pattern matters most, for males it is the f_0 distribution that is the primary distinguishing feature.

Figure-5 shows the f_0 distributions that correspond to the FHN-spectra of Figure-4. Male /æ/ and /i/ distributions peak at 120 Hz, well below the peak of 220 Hz for the equivalent female groupings.

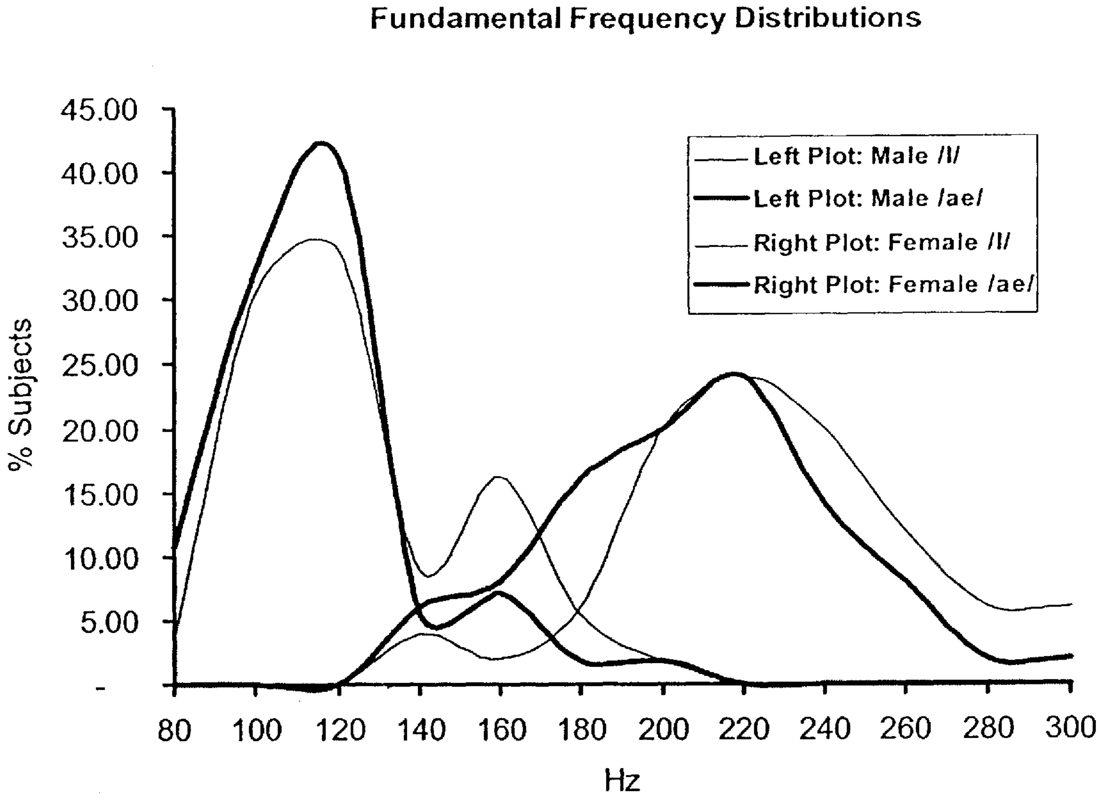

Figure-5
Fundamental frequency distributions corresponding to the FHN patterns in Figure-4.
The Left hand plots are for male /æ/ (upper, bold) and /i/ (lower, standard)
with the equivalent female plots shown to the right.

References

1. Moore C J, Slevin N , Winstanley S, Woods H, Allan E, Birzgalis A R, W T Farrington' "Computerised Quantification and 3D-Visualisation of Voice Quality Changes following Radiotherapy for Carcinoma of the Larynx", Brit Comp Soc, Procs HC-99 Current Perspectives in Healthcare Computing, Harrogate UK, 137-145, 1999.
2. McGillion M A, Moore C J, Ritchings R T, 'Objective Voice Quality Assessment Using Multi-Layer Perceptrons', Procs 4th Intl Workshop Neural Networks in Applications NN'99, Magdeburg Germany, 141-149, 1999.
3. Baken R J. *Clinical Measurement of Speech and Voice.* College Hill Press, 1987.
4. Titze I R, *Principles of Voice Production*, Prentice Hall, 1994.
5. Fourcin A J, 'Electrolaryngographic Assessment of Vocal Fold Function', Journal of Phonetics, 14, 435-442, 1986
6. Wallace W H B, Crowne E C, Shalet S M, Moore C, Gibson S, Littley M D and White A. 'Epidsodic ACTH and cortisol secretion in normal children', Clinical Endocrinology, 34, 215-222, 1991
7. Priestley M B. *Spectral Analysis and Time Series,* Academic Press 1981.