

An Iterative Inverse Filter Model linking the Acoustic and EGG Spectral Domains

Kathiresan Manickam, Terry Willard and Christopher Moore
Developing Technologies-Radiotherapy, N W Medical Physics
Christie Hospital, Manchester United Kingdom

Abstract

A non-parametric iterative model is developed to predict the EGG from the acoustic normalised harmonic spectrum. This study is a follow up from the previous work in quantifying EGG harmonic complexity pattern using single figure computed Approximate Entropy. This single figure computation aids in understanding the voice quality changes and vocal functionality. EGG measurements are not widely available so consequently a need to estimate EGG characteristics from the acoustic spectrum is inevitable. 30% of the healthy male population recovered with the same pattern complexity and 65% have shown little changes in complexity and the remaining 5% showed substantial changes in complexity.

1. Introduction

Vocal fold functionality simultaneously characterises the quality of voice, its impairment and the dysfunction of laryngeal anatomy. The sub glottis air pressure variation is amplified at resonance frequencies in the vocal tract depending on the type of vowel being phonated. These resonated effects, termed formants, make the acoustic signal more complicated. An electro-laryngograph, or electro-glottograph (EGG) can be used to measure vocal fold vibration via trans-larynx electrical impedance variation or 'LX-waveform'. Recently it has been shown that the modified EGG spectrum can be used to simply but objectively define voicing normality against which to measure pathological cases using a single figure of merit based on approximate entropy [1]. However, the laryngograph is not yet available in many institutions.

Inverse filtering techniques are widely used to recover the glottal air pressure variation from the acoustic signals and this has undoubtedly aided the recognition of pathological voicing [2], both in the time and frequency domains. Underlying this technique is

the development of a model to recover the glottal air pressure variation. Parameters used in the model rely on anatomical measurements and in some circumstances sophisticated devices are used to estimate these parameters [3].

The main focus of this study is to eliminate the need for prior knowledge, measurements and sophisticated equipment in the objective assessment of voicing quality from the vowel acoustic signal. Instead of estimating the glottal air pressure variation during vowel phonation, a model has been developed to recover the essential spectral features that would have appeared in trans-larynx impedance variations of the EGG, and can be used to characterise voice quality using approximate entropy, but using only a simple record of the acoustic signal as the initial input to the recovery process.

2. Method

Eighty-one healthy male volunteers were asked to phonate vowel, /i:/ under expert guidance. Both impedance (EGG) and acoustic (SP) signals were recorded using laryngograph. The signals were digitised at a sampling rate of 20 KHz and split into frames of 1000 points. Prior to discrete Fourier transformation, each frame was de-trended, autocovarianced and a HANNING variance reduction window was then applied. Using the maximum in the autocovarianced frame to estimate the fundamental frequency (F_0), the frames were normalised with respect to the computed F_0 frequency and power to form fundamental harmonic normalised (FHN) spectrum [1]. To recover the EGG features the following procedure was adopted:

a. The low frequency noise present in the SP-FHN, characterised by non-uniform troughs, was reduced by an iterative, multiple Gaussian fitting that was constrained to remain positive but minimise trough filling shown in Figure-1.

b. After noise reduction, the modified SP-FHN

spectrum was truncated from the 2nd to the 8th harmonic based on previously reported work that characterises normality in voicing using approximate entropy [1].

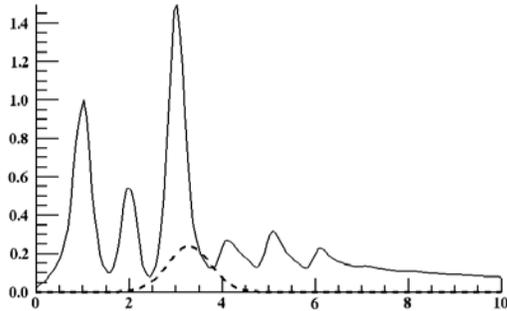


Figure 1

**Low frequency error reduction
FHN acoustic (SP) spectrum (up to 10th harmonic) and iterative Gaussian (dashed line)**

c. A multiple log-normal iterative weighting model, to be applied to the modified SP-FHN, was developed based on the first formant position and its height, Figure-2.

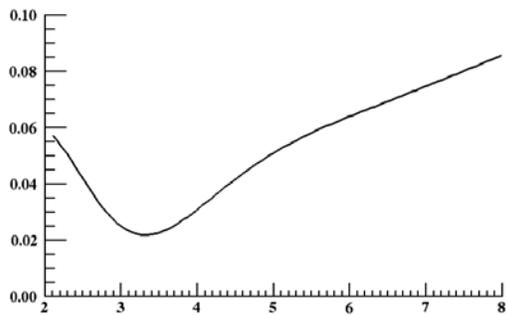


Figure 2

**Iterative Weighting Model
Abscissa: Harmonics (2nd to 8th)
Ordinate: Normalised Power**

d. The weighted modified SP-FHN was optimised by minimisation relative to the known, training EGG-FHN. Then, using reduced chi-squared optimisation technique, the error is minimised further using the recovered and the original impedance harmonics.

e. Finally, the complexity based on approximate entropy (ApEn) was computed and compared to the original and recovered EGG harmonics for future analysis.

f. The generic acoustic weighting model was used to provide recovered estimates for EGG approximate entropy and these were compared to known values for a healthy male population.

3. Results

An individual example of an iteratively recovered EGG harmonic spectrum, using the weighted acoustic model above, is shown in Figure-3.

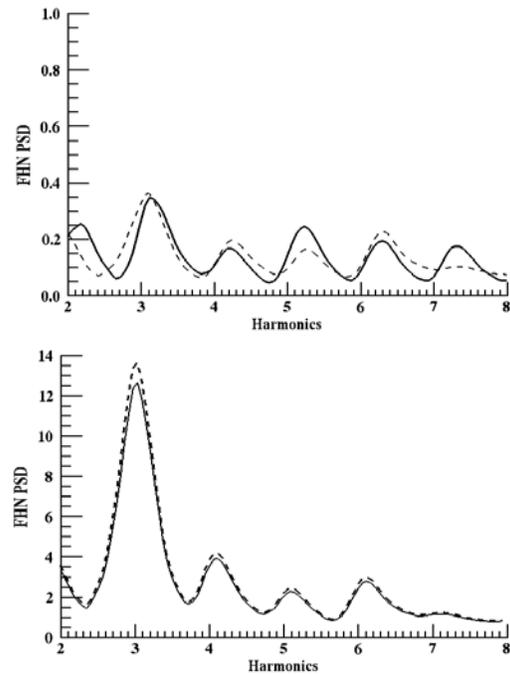


Figure 3

**Healthy Male Individual
Top: EGG Normalised Power Spectral Density (LX PSD)**

Solid line: Original EGG spectrum

Dashed line: Recovered EGG spectrum

Below: Acoustic Normalised Power Spectral Density (SP PSD)

Solid line: Original SP spectrum

Dashed line: Low frequency noise reduced SP spectrum

The aim of modelling is to recover specific harmonics known to be important in speech analysis, however the shape of the acoustic spectrum will inevitably determine what is propagated through the inverse iterative processing. Hence, peaks in the EGG domain can only be recovered if there are indicative peaks in

the acoustic spectrum itself. After 27 iterations the impedance harmonics were recovered with a minimised error of 3.1% between the original and recovered spectrum. The peak formant height has increased from 12.67 to 13.67 after minimising the low frequency noise. The most important parameter, namely approximate entropy of the original spectrum was 0.350, whilst the recovered spectrum yielded a closely matching value of 0.352.

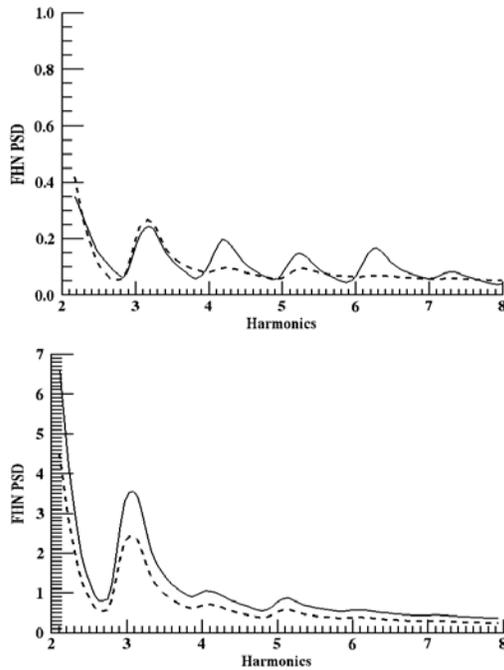


Figure 4

Male Larynx Cancer Patient

Top: EGG Normalised Power Spectral Density (LX PSD)

Solid line: Original EGG spectrum

Dashed line: Recovered EGG spectrum

Below: Acoustic Normalised Power Spectral Density (SP PSD)

Solid line: Original SP spectrum

Dashed line: Low frequency noise reduced SP spectrum

For example, in Figure-4, the recovered pattern has the standard deviation of the error between the recovered and the original impedance spectrum as low as 2.5%. However, the complexity is 0.148 of the recovered pattern and 0.319 for the original impedance pattern. Following 191 iterations, the complexity of the recovered pattern is low since there are barely

any identifiable peaks above the fourth harmonic in the acoustic spectrum.

Previous work identifies two, separated but well defined EGG approximate entropy bands, $G1_M$ & $G2_M$ as shown in Figure-5, which define the normal male population voicing [1]. GT_M is the transition band with seven individuals and the GP_M is the band below normality with four individuals. In this study 30% of the population had recovered values that placed them in the same category; 32% of the population changed a category e.g. GP_M to $G2_M$ or $G1_M$ to GT_M ; 33% of the population had two categories changes i.e. GT_M to GP_M or $G2_M$ to $G1_M$. Only 5% showed recovered approximate entropy that was not consistent with known values. Five healthy males were grouped in GP_M in the original complexity analysis and in the recovered complexity 13 individuals were grouped in GP_M phase. Majority of them had recovered complexity of a “normal” phonation range.

The ratio of $G1_M$ & GT_M and $G2_M$ & GP_M is 1.5:1 in the original complexity and in the recovered groups it is that four times as many patients are seen in the weaker categories than the stronger ones, shown in figure 6. 11 patients have no band difference; and a large proportion of 16 patients have a band difference; again 11 patients have indicated two band differences and only three have three band differences. More subjects have increased in complexity as the band increases while the reverse is seen for those who have decreased in complexity.

4. Discussion and Conclusion

Locating formants in the normalised acoustic spectrum is problematic. Some individuals have extremely weak formants compared to the fundamental peak [4]. This could be caused by individuals not keeping the pharyngeal wall stiff or the nasal port being closed [5]. Vowels like [ε] and [æ] are chosen because of their relatively high F1 values assuring a better separation of F_0 and F_1 compared to vowels e.g. [i] or [u] [6]. Another obstacle observed is suboptimal recording of signals. Incorrect placement of the sensors and fatty tissues around the neck has an impact on EGG impedance recordings whilst defects at the microphone level have similar consequences in the acoustic recordings. Despite these obstacles, the initial implementation of the inverse filter

model has been extremely encouraging and represents a significant breakthrough in our ability to measure and quantify vocal fold functionality. This will greatly extend the single figure characterisation of voice quality based upon approximate entropy.

5. Reference

[1] Moore C, Manickam K, Willard T, Jones S, Slevin N, Shalet S, 'Spectral pattern complexity analysis and the quantification of voice normality in healthy and radiotherapy patient groups', *Med Eng Phys*, 26:291-301,2004.

[2] Rothenberg M, 'A new inverse filtering technique for deriving the glottal air flow waveform during voicing', *Journal of the Acoustical Society of America*, 53:6:1632-1645, 1973.

[3] Titze I R, *Principles of Voice Production*, Prentice Hall, 1994.

[4] Simpson A P, 'Dynamic Consequences of Differences in Male and Female Vocal Tract Dimensions', *J. Acoust. Soc. Am*, 109:5:2153-2164, 2001.

[5] Rothenberg M, 'Cosi Fan Tutte and What it means – or – Nonlinear Source Tract Interaction in the Soprano Voice and some implications for the definition of vocal efficiency', *Vocal fold Physiology: Laryngeal Function in Phonation and respiration*, College Hill Press, San Diego , 254-263,1986.

[6] Price P J, 'Male and Female Voice Source Characteristics: Inverse Filtering Results', *Journal of Speech Communication*, 8:3:261-277, 1989.

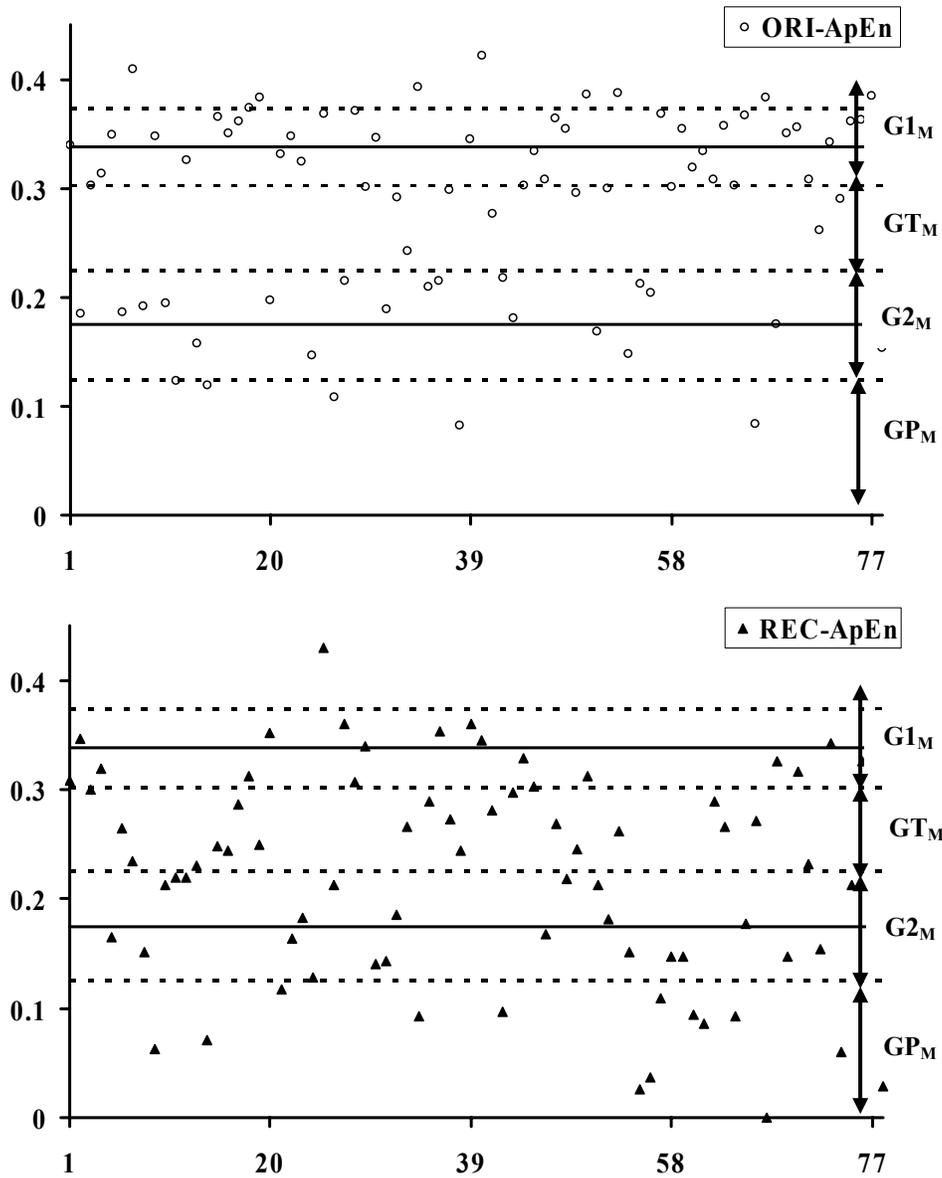


Figure 5

Healthy male complexity (ordinate) plotted against original impedance complexity (discs) and recovered impedance complexity (triangles). Normal mean complexity is shown as a continuous horizontal line with standard deviation indicated by parallel, adjacent dashed lines dashed lines ($G1_M$ ideal normal upper set, $G2_M$ normal lower set).

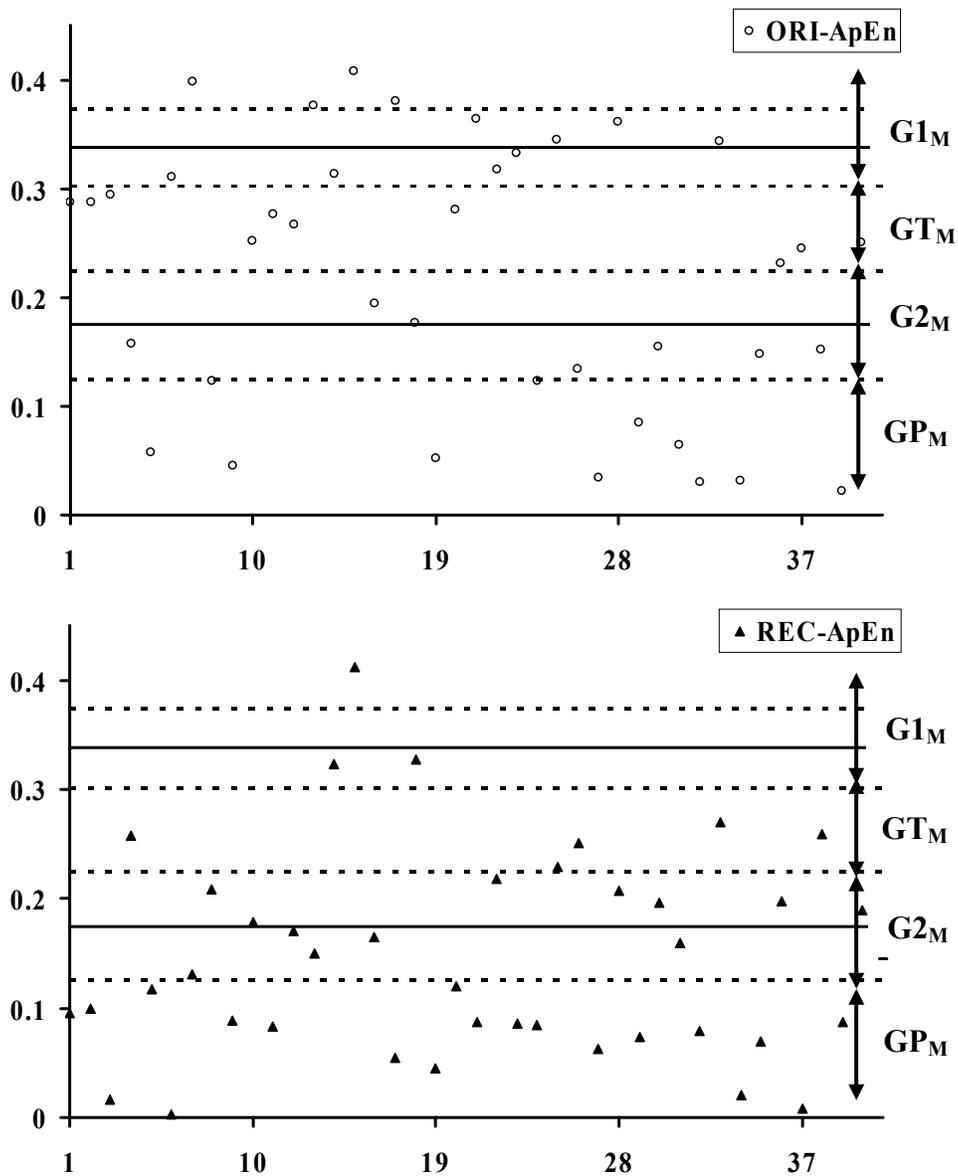


Figure 6

Male larynx cancer patients (ordinate) plotted against original impedance complexity (discs) and recovered impedance complexity (triangles). Normal mean complexity is shown as a continuous horizontal line with standard deviation indicated by parallel, adjacent dashed lines dashed lines (G1_M ideal normal upper set, G2_M normal lower set).