Abstract: Larynx cancer patients receive radiotherapy as a non-invasive alternative to surgery and cure rates are high. Inevitably this impacts vocal fold functionality. Hence, voice recovery as a pre-requisite for resuming normal life is of special interest. Voicing recovery following radiotherapy is studied in this paper.

Complexity analysis, using approximate entropy to concisely quantify the collective spectral pattern derived from the electro-glottogram, has revealed a double banded male normal voicing reference standard. Forty-eight male larynx cancer patients have been studied by applying this technique in parallel with an unrestricted perceptual analysis before and one year after radiotherapy.

Two thirds of radiotherapy patients had improved voice quality one year after treatment. Approximate entropy increased to reach normal population reference levels. These patients were predominantly in the less aberrant perceptual categories. However, a quarter of patients showed reduced approximate entropy and were predominantly in the most aberrant perceptual categories.

Complexity analysis has the potential to be a reliable, single parameter measure of voicing quality for use in monitoring radiotherapy patient recovery.

I. INTRODUCTION

United Kingdom cancer statistics for 2001 show that the larynx is the site for nearly a third of all 7820 new head and neck cancers and that well over 4 times as many men than women suffered from the disease [1]. Hence, it is as prevalent as cervix cancer in women, though it attracts far less public attention. The five year survival of larynx cancer patients following treatment is good, at approximately two-thirds. Hence, quality of life in terms of voice preservation is important for a large number of individuals wishing to resume normal life.

Radiotherapy arguably has fewer side effects than surgery, which is self evidently more invasive. However, the measure of recovery of voice quality after radiotherapy has not been concisely and objectively quantified. Irradiation effects may leave the targeted tissues intact but they do impact the tissue mechanics and perturb vocal fold functionality for months after treatment [2], which in turn directly influence voice quality [3-6].

Speech and Language Therapists (SALTs) working at the Christie Hospital have been engaged to assess the impact of radiotherapy one year after treatment. A patient’s normal voice, prior to the appearance of cancer, is rarely recorded. Hence, SALT subjective assessment reflects experience and the audibility of aberrant voicing. Inevitably this is complicated by differences in clinical technique and opinion [7]. Assessment at the Christie, requires patients to phonate vowels and provide a sample of connected speech. An electro-glottogram (EGG) and acoustic digital recording form the core record of such an examination [8]. The VPAS scheme guides assessment with voice quality eventually binned into a multi-category scale, which, at its most challenging, ranges from 0 (normal) to 6 (severely aberrant). Throughout assessment a SALT knows the phase of treatment reached by the patient, which, along with full knowledge of the stage of the cancer itself and examinations such as endoscopy, inevitably heights expectations and introduces bias.

Mathematical analysis of the entire range of spectral features in voicing is rarely deployed in the routine cancer clinic. Most disconcertingly there is no definition of what constitutes normal voicing and therefore no scientific or physical reference standard. As a result, it has not been possible to explain how cancer patients subjected to intense vocal fold irradiation during radiotherapy can recover vocal fold functionality to a level that could be considered to be “normal”.

This paper shows that vocal fold vibration as evidenced through the EGG impedance time series of vowel phonation can be deployed to differentiate the healthy normal population, quantify cancer patient voicing and to track the pattern of cancer patient recovery following radiotherapy. The approach reported is based on the regularity statistic ‘approximate entropy’ (ApEn) [9] applied for the first time to detect collective changes across the entire EGG spectral pattern.

II. THEORY

Vowel phonation is predominantly driven by vocal fold vibration. Vocal folds function is impaired by physical damage arising from malignant disease and associated therapy. Fold vibration is accompanied by impedance variations across the thyroid area. These trans-larynx impedance changes can be detected during phonation using a laryngograph. Successive measurements form a time series that usually has a
distinctive waveform structure that is known as the EGG [1,2]. This correlates well with vocal fold vibration and is virtually free from tract resonance. The EGG has not found widespread use amongst SALTs, at least in the UK. Sustained vowel phonation produces a more or less regular EGG waveform, which is ideally suited to characterisation in the frequency domain via the changes seen in the corresponding power spectral pattern [10,11].

To generate an EGG spectrum, the EGG time series are segmented into short frames, stationarised by finite differencing, variance reduced with a suitable function such as the Hanning window, autocorrelated and then fast Fourier transformed. This produces a sequence of frame power spectral density estimate (FPSD) [3]. However, the dynamic effects of fundamental frequency variation from frame to frame need to be removed in order to maximally reveal spectral shape. Therefore, the FPSD are individually normalised relative to the frequency and power of the frame fundamental (F0) itself. This fundamental harmonic normalisation approach produces what the authors call the FHN-PSD for each frame in which all features are on a common normalised harmonic scale rather than frequency scale [12]. The frame FHN-PSD can then be averaged to reinforce any shared spectral pattern ready for characterisation. A normal individual would be expected to have vocal folds that vibrate most regularly, producing rich harmonic patterns within an envelope showing lengthy decay. If these patterns have common characteristics, and the literature is full of examples, then a suitable form of regularity statistic sensitive to the collective pattern will resolve this into a useful normal population reference standard.

In a single value ApEn quantifies the repeatability of the pattern sampled from a time series itself. No assumptions need to be made about the shape or functional basis of the patterns being sought. Given N data points \( \{u(i)\} = u(1), u(2), \ldots, u(N) \) and commencing with the first point, vector sequences \( x(1), x(N-m+1) \) are formed from m values \( x(i) = [u(i), u(i+1), \ldots, u(i+m-1)] \). The Pincus ApEn [9] is interpreted heuristically as a measure of the average logarithmic likelihood, over all sequences \( x(1), x(N-m+1) \), such that any sequence in the data series \( \{u(i)\} \), which is within a tolerance \( r \) of the given sequence \( x(j) \) of length m, remains within the same tolerance when the length of both sequences is increased by one data point. Tolerance \( r \) is proportional to the measured series standard deviation \( \sigma \), i.e. \( r = k \sigma \) where \( k \) is a constant. It is necessary to determine \( k \) empirically so that the widest range of complexity values is achieved. ApEn had been used to study complexity changes in cardiac ECG time series, which show the presence or absence of vital, highly individual feedback mechanisms placing demands on the heart.

ApEn was primarily developed for use in time series analysis and was not used for characterising changes in the spectral pattern, being reserved for comparing fluctuations in a small number of pre-selected peaks. In the realms of speech analysis Moore et al reported the development of a vowel phonation reference standard in for normal males using the ApEn of the truncated FHN-PSD spectral pattern considered collectively [13].

Cancer patients with malignant lesions, possibly infiltrating the vocal folds, would be expected to have abnormal voicing characteristics. However, patients present with cancer in different stages of development and their treatment planned accordingly. Consequently, their vowel FHN-PSDs and corresponding ApEn values might reasonably be expected to vary from nearly normal to completely aberrant. Moore et al [13] have shown that this is indeed the case. Clinical opinion [2] suggests that the most obvious side effects of curative radiotherapy are likely to resolve, leaving stabilised voicing, after one year. In this paper pre-therapy and one year post-therapy ApEn complexities, derived from vowel phonation, are compared. The aim is to identify recovery patterns in male larynx cancer patients, relative to the ApEn reference standard already established for normal males.

### III. METHODOLOGY

Eighty-nine male volunteers provided the reference standard for this study. Each subject was connected to an electro-laryngograph and asked to phonate sustained vowel /i/ for up to 4 seconds. The output impedance signal was digitised at a sampling rate of at 20kHz. The digital EGG data files, excluding 4 compromised files, were subjected to ApEn complexity analysis using software written in IDL from Research Systems International (UK). This software first stationarised the time series to remove background noise and mains contamination. The resultant time series were then split into consecutive data frames, each 1000 samples long. The auto-covariance of each frame was computed and the maximum used to determine F0. A multiplicative Hanning window was applied to each frame to reduce the variance at high lags before estimating the PSD by fast Fourier transformation. Each frame PSD was then normalised using the FHN approach described by Moore et al [12]. This left the harmonics as integer multiples of the frame F0 with all other spectral components at non-integer multiples. The frame FHN-PSD were then averaged for each subject. Since, spectral shape variation in and around the normalised F0 peak is minimal, by design, the averaged FHN PSD was removed below the maximum of the first true harmonic peak, H2, and above the maximum of the seventh harmonic peak, H7. The logarithm of the truncated FHN PSD was then taken in order to minimise any trend in the spectral pattern. ApEn values were then calculated as described by Moore et al. [13].
Forty-eight male larynx cancer patients attending the Christie Hospital for radiotherapy, volunteered and were consented for approved study. EGG data was collected prior to and one year after radiotherapy and ApEn analysed as described for the normal voicing volunteers. On both occasions each patient was also perceptually assessed by an experienced SALT. No restriction was imposed on the data used for the perceptual assessment, which included acoustic data and access to patient hospital records. Guided by VPAS, the SALT categorised patient voice quality onto a seven point scale ranging from normal (category-0, CAT0) to completely aberrant (category-7, CAT7).

IV. RESULTS

Fig. 1 shows the ApEn complexity distribution for the healthy male normals reported by Moore et al. The bimodal nature of these data was tested by Gaussian mixtures model fitting using maximum likelihood [28]. They concluded (p<0.001) that two normal groups G1 and G2 existed, characterised by complexity values 0.340 (+/- 0.035) and 0.183 (+/- 0.057) with relative weights 62% and 38% respectively. Members of G1 exhibited strong EGG FHN-PSD features whilst those in G2 were weak, especially in the higher frequency harmonics.

Fig. 2 and Fig. 3 show the ApEn complexity results for larynx cancer patients measured before and, health permitting, one year after radiotherapy, arranged by pre-treatment SALT perceptual category, CATn (n=1,2,…7). Post treatment categorisation is indicated by single digit numbers placed side-on and above the CAT indication. Dashed boxes indicate the G1 and G2 standard deviation boundaries as a complexity reference standard for normal voicing. Patients showing increased ApEn after treatment appear in Fig. 2 whilst patients showing reduced ApEn appear in Fig. 3. ApEn values before treatment are indicated using circular symbols, whilst triangular symbols indicate those one year post treatment.

V. DISCUSSION

Of the 48 cancer cases considered, ApEn analysis indicated that one year after radiotherapy two-thirds would develop improved vocal fold functionality and
the G1 and G2 reference standards). Only one quarter of cases would be below normal voicing bounds and distinctly pathological.

Fig. 2 demonstrates that patients assigned a less aberrant pre-treatment category by SALT perceptual analysis have improved ApEn post treatment. This takes the individual into a normal voicing pattern, with spectral features enhanced at least to the lower level of normality seen in the G2 male population. Those individuals already in the G2 normal band prior to treatment predominantly improve after one year to become members of the ideal G1 population characterised by well developed harmonics in the vowel FHN-PSD.

Fig. 3 shows the converse is true for patients assigned by SALT perceptual analysis to the most aberrant pre-treatment categories. Whilst a handful show almost no change, many deteriorate and actually fall below the normal band defined by the G2 male population.

In 28 cases, the direction of complexity analysis changes agreed with SALTs perceptual assessment. Out of nine cases where the SALT indicated a large improvement, only four showed a corresponding improvement in complexity and five showed a reduction in complexity. The ApEn spectral evidence for these cases prompted SALT re-assessment of three individuals and a reduced categorisation more in line with that suggested by ApEn analysis.

What must not be forgotten, as mentioned in the introduction, is that the direction of change in SALT categorisation is undoubtedly biased since the SALTs must be aware of the patients’ details including their treatment stage and pre-treatment categorisation. They expect an improvement in patients’ voice quality one-year after radiotherapy. It is plausible that, in the context of SALT dealings with cancer patients, the perceptual definition of normal voicing equates to the lower ApEn, G2 reference standard. This would explain how SALTs could describe post-cancer, post-irradiation individuals as entirely normal in CAT0. These factors could be pursued further if unlabelled recordings were used for normal volunteers and patients taken pre and post treatment.

Most of the differences between SALT perception and ApEn complexity analysis occur in CAT 4-5. The authors believe that where patients present before radiotherapy with poor voicing then it is simply easier to perceptually detect, and as a result overrate, voice improvements. Furthermore, it should be remembered that the utility of perceptual categorisation depends on reliability. In this study there is some evidence, though not conclusive, that perceptual categorisation onto a 7 point scale has a standard deviation of at least 1 bin, i.e. a variation of up to 2 bins, is highly likely.

VI. CONCLUSION

Spectral ApEn complexity analysis of trans-larynx impedance measurements has allowed the recovery pattern of vocal fold functionality and voicing in male radiotherapy cancer cases to be examined. Using a single objective parameter to quantify the collective spectral pattern of vowel phonation, the majority of radiotherapy patients are seen to recover to levels of normality seen in the general, healthy population. Many patients recover to the normal G2 band with its characteristically weak harmonic structures. This probably reflects residual damage that SALTs find entirely acceptable.

REFERENCES