Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection

Max A Little*[1,2,3], Patrick E McSharry[1] , Stephen J Roberts[2] , Declan AE Costello[4] and Irene M Moroz[3]

[1]Systems Analysis, Modelling and Prediction Group,Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK
[2]Pattern Analysis Research Group,Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK
[3]Applied Dynamical Systems Research Group,Oxford Centre for Industrial and Applied Mathematics, Mathematics Institute,University of Oxford, Oxford OX1 3JP, UK
[4]Milton Keynes General Hospital, Standing Way, Eaglestone, Milton Keynes, Bucks MK6 5LD, UK

Email: Max A Little*- littlem@ieee.org; Patrick E McSharry - mcsharry@robots.ox.ac.uk; Stephen J Roberts - sjrob@robots.ox.ac.uk; Declan AE Costello - declancostello@doctors.org.uk; Irene M Moroz - moroz@maths.ox.ac.uk;

*Corresponding author

Abstract

Background: Voice disorders affect patients profoundly, and acoustic tools can potentially measure voice function objectively. Disordered sustained vowels exhibit wide-ranging phenomena, from nearly periodic to highly complex, aperiodic vibrations, and increased "breathiness". Modelling and surrogate data studies have shown significant nonlinear and non-Gaussian random properties in these sounds. Nonetheless, existing tools are limited to analysing voices displaying near periodicity, and do not account for this inherent biophysical nonlinearity and non-Gaussian randomness, often using linear signal processing methods insensitive to these properties. They do not directly measure the two main biophysical symptoms of disorder: complex nonlinear aperiodicity, and turbulent, aeroacoustic, non-Gaussian randomness. Often these tools cannot be applied to more severe disordered voices, limiting their clinical usefulness.

Methods: This paper introduces two new tools to speech analysis: recurrence and fractal scaling, which overcome the range limitations of existing tools by addressing directly these two symptoms of disorder, together reproducing a "hoarseness" diagram. A simple bootstrapped classifier then uses these two features to distinguish normal from disordered voices.

Results: On a large database of subjects with a wide variety of voice disorders, these new techniques can distinguish normal from disordered cases, using quadratic discriminant analysis, to overall correct classification performance of $91.8 \pm 2.0\%$. The true positive classification performance is $95.4 \pm 3.2\%$, and the true negative performance is $91.5 \pm 2.3\%$ (95% confidence). This is shown to outperform all combinations of the most popular classical tools.

Conclusions: Given the very large number of arbitrary parameters and computational complexity of existing techniques, these new techniques are far simpler and yet achieve clinically useful classification performance using only a basic classification technique. They do so by exploiting the inherent nonlinearity and turbulent randomness in disordered voice signals. They are widely applicable to the whole range of disordered voice phenomena by design. These new measures could therefore be used for a variety of practical clinical purposes.

Background

Voice disorders arise due to physiological disease or psychological disorder, accident, misuse of the voice, or surgery affecting the vocal folds and have a profound impact on the lives of patients. This effect is even more extreme when the patients are professional voice users, such as singers, actors, radio and television presenters, for example. Commonly used by speech clinicians, such as surgeons and speech therapists, are acoustic tools, recording changes in acoustic pressure at the lips or inside the vocal tract. These tools [1], amongst others, can provide potentially objective measures of voice function. Although acoustic examination is only one tool in the complete assessment of voice function, such objective measurement has many practical uses in clinical settings, augmenting the subjective judgement of voice function by clinicians. These measures find uses, for example, in the evaluation of surgical procedures, therapy, differential diagnosis and screening [1, 2], and often augment subjective voice quality measurements, for example the GRB (Grade, Roughness and Breathiness) scale. [3] These objective measures can be used to portray a "hoarseness" diagram for clinical applications [4], and there also exists a variety of techniques for automatically screening for voice disorders using these measures [5–7].

Phenomenologically, normal and disordered sustained vowel speech sounds exhibit a large range of behaviour. This includes *nearly periodic* or regular vibration, *aperiodic* or irregular vibration and sounds

with no apparent vibration at all. All can be accompanied by varying degrees of noise which can be described as "breathiness". Voice disorders therefore commonly exhibit two characteristic phenomena: increased vibrational aperiodicity and increased breathiness compared to normal voices [4].

In order to better characterise the vibrational aperiodicity aspects of voice disorders, Titze [8] introduced a typology (extended with subtypes [1]). Type I sounds are those that are nearly periodic: coming close to perfect periodicity. Type II are those that show qualitative dynamical changes and/or modulating frequencies or subharmonics. The third class, Type III are those sounds that appear to have no periodic pattern at all. They have no single, obvious or dominant period and can described as *aperiodic*. Normal voices can usually be classed as Type I and sometimes Type II, whereas voice disorders commonly lead to all three types of sounds. This is because voice disorders often cause the breakdown of stable periodicity in voice production. The breathiness aspect of disordered voices is often described as dominating high-frequency noise. Although this original typology covered sounds of only apparently deterministic origin, a very large proportion of voice signals seen in the clinic are so noisy as to be better considered stochastic rather than deterministic [2]. Methods that are based upon theories of purely deterministic nonlinear dynamical systems, although they can be appropriate for sounds of deterministic origin covered by the original typology, cannot in principle be applied to such noise-dominated sounds, that is, to sounds that would be better modelled as stochastic processes rather than deterministic. This makes it impossible to characterise the full range of signals encountered in the clinic. For these reasons, in this paper, when we refer to Type III sounds we include random, noise-like sounds (which, in keeping with original typology, have no apparent periodic structure, by virtue of their randomness).

There exists a very large number of approaches to the acoustic measurement of voice function. The most popular of these are the *perturbation* measures *jitter* and *shimmer* and variants, and *harmonics-to-noise ratios (HNR)* [1, 4]. The effect of a variety of voice disorders on these measures has been tested under both experimental and clinical conditions [4, 9], showing that different measures respond to different disorders in different ways [10]. For example, in some disorders, jitter will increase with severity of the disorder, and for other disorders jitter is unaffected. Thus, although these measures can have value in measuring certain limited aspects of voice disorders such as speech pitch variability, there is no simple relationship between the extent or severity of voice disorder and these measures [4, 10]. Therefore, they cannot be used to directly quantify the two main biophysical symptoms of voice disorders: increasingly severe aperiodicity and breath noise, a quantification required to differentiate normal from disordered voices.

Another limitation of existing measures is that they are only properly applicable when near periodicity

holds: in Titze's typology only Type I sounds satisfy this property [1]. The behaviour of the algorithms for other sound types is not known theoretically and limited only to experimental results and informal arguments [4]. The source of this limitation is that they make extensive use of extraction of the *pitch period*, or *fundamental frequency* (defined as the inverse of the pitch period) from the acoustic signal [1]. Popular pitch period extraction techniques include zero-crossing detection, peak-picking and waveform matching [1]. The concept of pitch period is only valid for Type I sounds, therefore, application of these methods based upon periodicity analysis, to any other type of sound is problematic [6]. Type II and III have therefore received much less attention in the literature [8], such that there exist few methods for characterising these types, despite the fact that they exist in great abundance in clinical settings. This precludes the proper use of these tools on a large number of disordered voice cases, limiting the reliability of the analysis, since in fact some algorithms will not produce any results at all for Type II and III sounds, even though they contain valuable clinical information [2]. Another reason for the limitations of these methods is that they are based upon classical linear signal processing methods that are insensitive to the inherent biophysical nonlinearity and non-Gaussianity in speech [1]. These linear methods include autocorrelation, the discrete Fourier transform, linear prediction analysis and cepstral processing [11]. Since standardised, reliable and reproducible results from acoustic measures of voice function are required for clinical applications, these limitations of perturbation methods are problematic in clinical practice [2]. It is clear that there is a clinical need for reliable tools that can characterise all types of disordered voice sounds for a variety of clinical applications, regardless of whether they satisfy the requirements of near periodicity, or contain significant nonlinearity, randomness or non-Gaussianity [8].

Furthermore, analysis techniques are complicated by the use of any arbitrary algorithmic parameters whose choice affects the analysis method – changing these parameters can change the analysis results. The values of such parameters must be chosen in order to apply the algorithms, and to be principled, it is better, when making that choice, to have a theoretical framework to apply which offers specific guidance on that choice. Often however, no such general theory exists, and therefore these values must be chosen by experimental and empirical evaluation alone [12]. There exists the danger then that these parameters are highly "tuned" to the particular data set used in any one study, limiting the reproducibility of the analysis on different data sets or clinical settings. It is necessary therefore to reduce the number arbitrary parameters to improve the reproducibility of these measures.

To address these limitations, empirical investigations and theoretical modelling studies have been conducted which have lead to the suggestion that *nonlinear dynamical systems theory* is a candidate for a

unified mathematical framework modelling the dynamics seen in all types of disordered vowel speech sounds [1, 8, 13]. The motivation for the introduction of a more general model than the classical linear model is the principle of parsimony: the more general model explains more phenomena (more types of speech) with a smaller number of assumptions than the classical linear model [12, 14].

These suggestions have lead to growing interest in applying tools from nonlinear time series analysis to speech signals in order to attempt to characterise and exploit these nonlinear phenomena [1, 13]. For normal Type I speech sounds, fractal dimensions, Lyapunov exponents and bispectral methods have all been applied, also finding evidence to support the existence of nonlinearity [15, 16]. Extracting dynamical structure using local linear predictors, neural networks and regularised radial basis functions have all been used, with varying reported degrees of success. Algorithms for calculating the correlation dimension have been applied, which was successful in separating normal from disordered subjects [17]. Correlation dimension and second-order dynamical entropy measures showed statistically significant changes before and after surgical intervention for vocal fold polyps [18], and Lyapunov exponents for disordered voices were found to consistently higher than those for healthy voices [19]. It was also found that jitter and shimmer measurements were less reliable than correlation dimension analysis on Type I and unable to characterise Type II and (non-random) Type III sounds [20]. Mixed results were found for fractal dimension analysis of sounds from patients with neurological disorders, for both acoustic and electroglottographic signals [21]. Instantaneous nonlinear AM and FM formant modulations were shown effective at detecting muscle tension dysphonias [22]. For the automated acoustic screening of voice disorders, higher-order statistics lead to improved normal/disordered classification performance when combined with several standard perturbation methods [7].

In order to categorise individual voice signals into classes from the original typology (excluding severely turbulent, noise-like sounds), recent studies have found that by applying correlation dimension measurements to signals of these types, it was possible, over a sample of 122 stable vowel phonations, to detect a statistically significant difference between the three different classes [23, 24]. This provides further evidence in favour of the appropriateness of nonlinear signal analysis techniques for the analysis of disordered voices.

These studies show that nonlinear time series methods can be valuable tools for the analysis of voice disorders, in that they can analyse a much broader range of speech sounds than perturbation measures, and in some cases are found to be more reliable under conditions of high noise. However, very noisy, breathy signals have so far received little attention from nonlinear time series analysis approaches, despite

these promising successes. Common approaches such as correlation dimension, Lyapunov exponent calculation and predictor construction require that the scaling properties of the embedded attractor are not destroyed by noise, and for thus very noisy, breathy signals, there is the possibility that such nonlinear approaches may fail. There are also numerical, theoretical and algorithmic problems associated with the calculation of nonlinear measures such as Lyapunov exponents or correlation dimensions for real speech signals, casting doubt over the reliability of such tools [21, 25–27]. For example, correlation dimension analysis shows high sensitivity to the variance of signals in general, and it is therefore necessary to check that changes in correlation dimension are not due simply to changes in variance [28]. Therefore, as with classical perturbation methods, current nonlinear approaches cannot yet directly measure the two most important biophysical aspects of voice disorder.

A limitation of deterministic nonlinear time series analysis techniques for random, very noisy signals is that the implicit, deterministic, nonlinear dynamical model, which is ordinarily assumed to represent the nonlinear oscillations of the vocal folds [29] is no longer appropriate. This is because randomness is an inherent part of the biophysics of speech production [30, 31]. Thus there is a need to expand the nonlinear dynamical systems framework to include a random component, such that random voice signals and breath noise can also be fully characterised within the same, unified framework.

This paper therefore introduces a new, framework model of speech production that splits the dynamics into both deterministic nonlinear *and* stochastic components. The output of this model can then be analysed using methods that are able to characterise both nonlinearity and randomness. The deterministic component of the model can exhibit both periodic and aperiodic dynamics. It is proposed to characterise this component using *recurrence analysis*. The stochastic components can exhibit statistical time dependence or autocorrelation, which can be analysed effectively using *fractal scaling analysis*. This paper reports the replication of the "hoarseness" diagram [4] illustrating the extent of voice disorder, and demonstrates, using a simple quadratic classifier, how these new measures may be used to screen normal from disordered voices from a large, widely-used database of patients. It also demonstrates that these new measures achieve superior classification performance overall when compared to existing, classical perturbation measures, and the derived irregularity and noise measures of Michaelis [4].

Methods

In this section we first discuss in detail the evidence that supports the development of a new stochastic/deterministic model of speech production.

The classical linear theory of speech production brings together the well-developed subjects of classical linear signal processing and linear acoustics (where any nonlinearities in the constitutive relationships between dynamic variables of the fluid are considered small enough to be negligible) to process and analyse speech time series [11]. The biophysical, acoustic assumption is that the vocal tract can be modelled as a linear resonator driven by the vibrating vocal folds that are the source of excitation energy [11]. However, extensive modelling studies, experimental investigations and analysis of voice signals have shown that dynamical nonlinearity and randomness are factors that should not be ignored in modelling speech production.

For the vocal folds, there are two basic, relevant model components to consider. The first is the vocal fold tissue, and the second is the air flowing through that structure. The vocal folds, during phonation, act as a vibrating valve, disrupting the constant airflow coming from the lungs and forming it into regular puffs of air. In general the governing equations are those of fluid dynamics coupled with the elastodynamics of a deformable solid. In one approach to solving the problem, the airflow is modelled as a modified quasi-one-dimensional Euler system which is coupled to the vocal fold flow rate, and the vocal folds are modelled by a lumped two mass system [32]. A widely used and simplified model is the two-mass model in Ishizaka [33], further simplified in asymmetric [29, 34] and symmetric versions [35]. These models demonstrate a plausible physical mechanism for phonation as nonlinear oscillation: dynamical forcing from the lungs supplies the energy needed to overcome dissipation in the vocal fold tissue and vocal tract air. The vocal folds themselves are modelled as elastic tissue with nonlinear stress-strain relationship.

Classical nonlinear vocal fold models coupled with linear acoustic vocal tract resonators appear to account for the major part of the mechanisms of audible speech, but from these mechanisms an important component is missing: that of *aspiration*, or "breath" noise [36]. Such noise is produced when the air is forced through a narrow constriction at sufficiently high speeds that "turbulent" airflow is generated, which in turn produces noise-like pressure fluctuations [37]. Aspiration noise is an unavoidable, involuntary consequence of airflow from the lungs being forced through the vocal organs, and can be clearly heard in vowel phonation [38]. Certain voice pathologies are accompanied by a significant increase in such aspiration noise, which is perceived as increased "breathiness" in speech [4]. This noise is therefore an important part of sound generation in speech.

One significant deficiency in the classical linear acoustic models is due to the assumptions about fluid flow upon which their construction is based [39]. These classical linear models make very many simplifying assumptions about the airflow in the vocal organs, for example, that the *acoustic limit* [40] holds in which

the fluid is nearly in a state of uniform motion. Similarly, the *Bernoulli's equation* assuming energy conservation along streamlines, upon which many classical vocal fold models rely, applies only if the fluid is assumed inviscid and irrotational [41, 42]. The important point for this study is that these classical assumptions forbid the development of complicated, "turbulent" fluid flow motion, in which the flow follows convoluted paths of rapidly varying velocity, with eddies and other irregularities at all spatial scales [43]. This breakdown of laminar flow occurs at high *Reynolds number*, and for the relevant length scales of a few centimetres in the vocal tract and for subsonic air flow speeds typical of speech [38], this number is very large (of order 10^5), indicating that airflow in the vocal tract can be expected to be turbulent. Under certain assumptions, turbulent structures, and vortices in particular (fluid particles that have rotational motion), can be shown to be a source of aeroacoustic sound [37]. Turbulence is a highly complex dynamical phenomenon and any point measurement such as acoustic pressure taken in the fluid will lose most of the information about the dynamics of the fluid. Consequently, even if turbulence has a purely deterministic origin, it is reasonable to model any one single dynamical variable measured at a point in space as a random process [43].

There are two broad classes of physically plausible mathematical models of the effects of aeroacoustic noise generation in speech. The first involves solving numerically the full partial differential equations of gas dynamics (e.g. the Navier-Stokes equations), and the second uses the theory of *vortex sound* [37]. For example, the study of Zhao [44] focused on the production of aspiration noise generated by vortex shedding at the top of the vocal folds, simulated over a full vocal fold cycle. This study demonstrates that the computed sound radiation due to vorticity contains significant high frequency fluctuations when the folds are fully open and beginning to close. On the basis of these results, it can be expected that if the folds do not close completely during a cycle (which is observed in cases of more "breathy" speech), the amplitude of high frequency noise will increase.

The second class of models, which makes use of *Lighthill's acoustic analogy* [37], are based around the theory of vortex sound generated in a cylindrical duct, [37] where, essentially, each vortex shed at an upstream constriction acts as a source term for the acoustic wave equation in the duct, as the vortex is convected along with the steady part of the airflow. The resulting source term depends upon not only the attributes of the vortex itself, but also upon the motion of the vortex through the streamlines of the flow [31, 37]. One model that uses this approach involves the numerical simulation of two overall components: the mean steady flow field and the acoustic wave propagation in the vocal tract [38]. Vortices are assumed to be shed at random intervals at constrictions at particular locations in the vocal tract, for

example, at the vocal folds or between the roof of the mouth and the tongue. Each vortex contributes to the acoustic source term at each spatial grid point. Here, an important observation is that simulated pressure signals from this model are stochastic processes [45], i.e. a sequence of random variables. It is also noticeable from the spectra of simulated pressure signals that although the signals are stochastic, they exhibit significant non-zero autocorrelation since the spectral magnitudes are not entirely constant, leading to statistical self-similarity in these signals [15, 43].

Other potential sources of biophysical fluctuation include pulsatile blood flow, muscle twitch and other variability and randomness in the neurological and biomechanical systems of the larynx [30].

The typical nonlinear dynamical behaviour of models of the vocal folds, such as period-doubling (subharmonics), bifurcations [29], and transitions to irregular vibration [35] have been observed in experiments with excised larynxes, a finding that helps to support the modelling hypothesis that speech is an inherently nonlinear phenomena [29, 46]. Similarly, models of turbulent, random aeroacoustic aspiration noise have been validated against real turbulent airflow induced sound generated in acoustic duct experiments [38]. Such studies show that the models of vortex shedding at random intervals are plausible accounts for the dynamical origins of breath noise in phonation.

Complementing modelling and experimental studies, the final source of evidence for nonlinearity and randomness in speech signals comes from studies of voice pressure signals. Using surrogate data analysis, it has been shown that nonlinearity and/or non-Gaussianity is an important feature of Type I sounds [47–49]. Nonlinear bifurcations have been observed in excised larynx experiments [46], and period-doubling bifurcations and chaos-like features have been observed in signals from patients with organic and functional dysphonias [13]. Aspiration noise has been observed and measured as a source of randomness in voiced phonation, in both normal [50] and disordered speech sounds [1, 4]. The fractal, self-similarity properties of aspiration noise as a turbulent sound source have also been observed and quantified in normal [15] and disordered speech [27].

Taken as a whole, these modelling, simulation, validation and signal analysis studies suggest that during voiced phonation there will be a combination of both deterministic and stochastic elements, the deterministic component attributable to the nonlinear movements of the vocal fold tissue and bulk of the air in the vocal tract, and the stochastic component the high frequency aeroacoustic pressure fluctuations caused by vortex shedding at the top of the vocal folds, whose frequency and intensity is modulated by the bulk air movement (and other sources of biophysical fluctuation). During voiced phonation, the deterministic oscillations will dominate in amplitude over the noise component which will show high

frequency fluctuations around this oscillation. During unvoiced or breathy pathological phonation, the turbulent noise component will dominate over any deterministic motion.

In order to capture all these effects in one unified model, we introduce a continuous time, two component dynamical model that is taken to generate the measured speech signal. The state of the system at time $t \in \mathbb{R}$ is represented by the vector $\mathbf{u}(t)$ of size d. Then the equation of motion that generates the speech signal is the following vector stochastic differential equation, commonly known as a *Langevin equation* [25].

$$\dot{\mathbf{u}}(t) = \mathbf{f}(\mathbf{u}(t)) + \varepsilon(t), \tag{1}$$

where $\varepsilon(t)$ is a vector of stochastic forcing terms. It is not necessary to assume that these fluctuations are independent and identically distributed (i.i.d.). The function $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$ is unknown and represents the deterministic part of the dynamics. Given an initial condition vector $\mathbf{u}(0)$ then a solution that satisfies equation (1) is called a *trajectory*. Ensembles of trajectories can be shown to satisfy the properties of a stochastic process with finite memory (a higher-order Markov chain) [25].

Of importance to both deterministic and stochastic systems is the notion of recurrence in state space. Recurrent trajectories are those that return to a given region of state space [51]. Recurrence time statistics provide useful information about the properties of both purely deterministic dynamical systems and stochastic systems [51]. Recurrence analysis forms the basis of the method of recurrence plots in nonlinear time series analysis [25].

In the context of the model (1), state-space recurrence is defined as:

$$\mathbf{u}(t) \subset B(\mathbf{u}(t + \delta t), r), \tag{2}$$

where $B(\mathbf{u}, r)$ is a closed ball of small radius $r > 0$ around the point $\mathbf{u}$ in state-space, and $\mathbf{u}(t) \not\subset B(\mathbf{u}(t + s), r)$ for $0 < s < \delta t$. Each different $t \in \mathbb{R}$ may be associated with a different δt, called the *recurrence time*. We will define *aperiodic* as recurrent in the sense of (2) but not *periodic*. Periodicity is recovered from the definition of recurrence in the special case when $r = 0$ and δt is the same for all t, so that the system vector $\mathbf{u}(t)$ assumes the same value after a time interval of δt:

$$\mathbf{u}(t) = \mathbf{u}(t + \delta t), \tag{3}$$

for all $t \in \mathbb{R}$. Then δt is the *period* of the system. Therefore, although these concepts of periodicity and aperiodicity are mutually exclusive, both are special cases of the more general concept of recurrence. The requirement of periodicity is central to many linear signal processing methods (the basis of the Fourier

transform, for example), but aperiodicity is a common feature of many voice disorders. It can be seen therefore that in order to characterise the inherent irregularity of disordered speech sounds, we require more general processing techniques that can directly account for such departures from strict periodicity. Using this analysis, nearly periodic speech sounds of Type I can be described as recurrent for some small $r > 0$, with δt nearly the same for each t. Type II sounds are more irregular than Type I, and for the same r, the δt will assume a wider range of values than for Type I. Similarly, Type III sounds which are highly irregular and aperiodic, will have a large range of values of δt again for the same r.

Similarly, of importance in the analysis of stochastic processes is scaling analysis [25]. Stochastic time series in which the individual measurements in time are not entirely independent of earlier time instants are often *self-similar*, that is, when a sub-section of the time series is scaled up by a certain factor, it has geometrical, approximate or statistical similarity to the whole time series [25]. Such self-similarity is a property of *fractals*. [43] The tools of dimension measurement and scaling analysis may be used to characterise the self-similarity in signals such as speech. As discussed above, theoretical models of aeroacoustic turbulent sound generation in speech predict randomly-timed impulses convolved with an individual impulse response for each vortex that induces *autocorrelated* random noise sequences [31], so that turbulent speech signals at one time instant are not independent of earlier time instants.

It has also been found experimentally that changes in the statistical time dependence properties of turbulent noise in speech, as measured by a particular fractal dimension of the graph of the speech signal, are capable of distinguishing classes of phonemes from each other [15]. As introduced above, disordered voices are often accompanied by increased "breathiness", due in part to the inability of the vocal folds to close properly, so that air escapes through the partial constriction of the vocal folds creating increased turbulence in the airflow [31]. Thus scaling analysis (and more general graph dimension measures) could be useful for characterising vocal fold disorders.

Recent experiments have shown that the use of recurrence analysis coupled with scaling analysis can distinguish healthy from disordered voices on a large database of recordings with high accuracy [27]. These techniques are computationally simple and furthermore substantially reduce the number of arbitrary algorithmic parameters required, compared to existing classical measures, thus leading to increased reproducibility and reliability.

Time-Delay State-Space Recurrence Analysis

In order to devise a practical method for applying the concept of recurrence defined earlier and measuring the extent of aperiodicity in a speech signal, we can make use of *time-delay embedding* [25]. Measurements of the output of the system (1) are assumed to constitute the acoustic signal, s_n:

$$s_n = h\left(\mathbf{u}\left(n\Delta t\right)\right), \tag{4}$$

where the measurement function $h : \mathbb{R}^d \to \mathbb{R}$ projects the state vector $\mathbf{u}(t)$ onto the discrete-time signal at time instances $n\Delta t$ where Δt is the sampling time, and $n \in \mathbb{Z}$ is the time index.

From these sampled measurements, we then construct of an m-dimensional *time delay embedding vector*:

$$\mathbf{s}_n = \left[s_n, s_{n-\tau}, \ldots s_{n-(m-1)\tau}\right]^T. \tag{5}$$

Here τ is the *embedding time delay* and m is the *embedding dimension*.

We will make use of the approach introduced in Ragwitz [52] to *optimise* the embedding parameters m and τ such that the recurrence analysis produces results that are close as possible to known, analytically derived results upon calibration with known signals. (We use this as an alternative to common techniques for finding embedding parameters, such as false-nearest neighbours, which explicitly require purely deterministic dynamics [25, 52]). Note that, under this approach, for very noisy signals, we will not always resolve all signals without self-intersections. However, in the context of this study, achieving a completely non-intersecting embedding is not necessary. For very high-dimensional deterministic or stochastic systems, any reconstruction with self-intersections due to insufficiently high embedding dimension can be considered as a different stochastic system in the reconstructed state-space. We can then analyze the stochastic recurrence of this reconstructed system. This recurrence, albeit different from the recurrence properties in the original system, is often sufficient to characterize the noisy end of the scale of periodicity and aperiodicity.

Figure 1 shows the signals s_n for one normal and one disordered voice example (Kay Elemetrics Disordered Voice Database). Figure 2 shows the result of applying the above embedding procedure for the same speech signals.

In order to investigate practically the recurrence time statistics of the speech signal, we can make use of the *method of close returns* [53], originally designed for studying recurrence in deterministic, chaotic dynamics. Here we adopt this method to study stochastic dynamics as well as deterministic dynamics. In this method, a small, closed ball $B(\mathbf{s}_{n_0}, r)$ of radius $r > 0$ is placed around the embedded data point $\mathbf{s}_{n_0}$.

Then the trajectory is followed forward in time $\mathbf{s}_{n_0+1}, \mathbf{s}_{n_0+2} \ldots$ until it has left this ball, i.e. until $\|\mathbf{s}_{n_0} - \mathbf{s}_{n_0+j}\| > r$ for some $j > 0$, where $\|.\|$ is the Euclidean distance. Subsequently, the time n_1 at which the trajectory first returns to this same ball is recorded (i.e. the first time when $\|\mathbf{s}_{n_0} - \mathbf{s}_{n_1}\| \leq r$), and the difference of these two times is the (discrete) *recurrence time* $T = n_1 - n_0$. This procedure is repeated for all the embedded data points $\mathbf{s}_n$, forming a histogram of recurrence times $R(T)$. This histogram is normalised to give the *recurrence time probability density*:

$$P(T) = \frac{R(T)}{\sum_{i=1}^{T_{\max}} R(i)}, \tag{6}$$

where $T_{\max}$ is the maximum recurrence time found in the embedded state space. The choice of r is critical to capture the properties of interest to this study. For example, if the trajectory is a nearly periodic (Type I), we require that r is large enough to capture all the recurrences, but not too large to find recurrences that are due to spurious intersections of $B(\mathbf{u}, r)$ with other parts of the trajectory, violating the conditions for proper recurrence. The appropriate choice of embedding delay τ has a role to play: selecting τ too small means that any trajectory lies close to a diagonal in the reconstructed state-space, potentially causing spurious recurrences. Thus τ must be chosen large enough to avoid spurious recurrences. Similarly, if τ is too large then the points in the embedded state-space fill a large cloud where recurrences will be difficult to find without using an inappropriately large value of r. There will be an optimum value of τ which traditionally is set with reference to autocorrelation or time-delayed mutual information estimates, for more details see [25].

In order to understand the behaviour of this algorithm (partly for optimising the embedding parameters), we consider two extreme forms that the density (6) may assume. The first is the ideal limiting case in which the recurrence distance r tends to zero for a periodic trajectory. The recurrence time probability density is:

$$P(T) = \begin{cases} 1 & \text{if } T = k \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where k is the period of the trajectory. In the second extreme case we consider a purely random, uniform i.i.d. signal which is normalised to the range $[-1, 1]$. The recurrence probability density is approximately uniform:

$$P(T) \approx \frac{1}{T_{\max}}. \tag{8}$$

Proofs for the results in equations (7) and (8) are given in the Appendix.

We can then optimise m, τ and r such that for a synthetic signal of perfect periodicity, $P(T)$ is determined using the close returns method such that it is as close as possible to the theoretical expression (7). This

optimisation is carried out by a straightforward systematic search of values of these parameters $m = 2, 3 \ldots 10$, $\tau = 2, 3 \ldots 50$, and for $r = 0.02, 0.04, \ldots 0.5$, on a perfectly periodic test signal. This search can be considered as a scan for the optimum parameter values through all points on a three-dimensional parameter cube with m, τ and r as the co-ordinates of that cube.

All disordered voice speech signals will lie somewhere in between the extremes of perfect periodicity and complete randomness. Thus it will be useful to create, from the recurrence time probability density, a straightforward sliding scale so that normal and disordered voice signals can be ranked alongside each other. This depends upon a simple characterisation of the recurrence probability density $P(T)$. One simple measure of any probability density is the *entropy* which measures the average uncertainty in the value of the discrete-valued density $p(i)$, $i = 1, 2 \ldots M$ [25]:

$$H = -\sum_{i=1}^{M} p(i) \ln p(i), \tag{9}$$

with units of *nats* (by convention, $0 \ln 0$ is taken to be zero). This measure can then rank disordered voice signals by representing the *uncertainty in the period* of the disordered voice signal in the following way. For perfectly periodic signals the recurrence probability density entropy (RPDE) is:

$$H_{\text{per}} = -\sum_{i=1}^{T_{\max}} P(i) \ln P(i) = 0. \tag{10}$$

since $P(k) = 1$ and the rest are zero. Conversely, for the purely stochastic, uniform i.i.d. case (derived in the appendix), the uniform density can be taken as a good approximation, so that the RPDE is:

$$H_{\text{iid}} = -\sum_{i=1}^{T_{\max}} P(i) \ln P(i) = \ln T_{\max}, \tag{11}$$

in units of nats. The entropy scale H therefore ranges from H_{per}, representing perfectly periodic examples of Type I sounds, to H_{iid} as the most extreme cases of noise-like Type III sounds. In practice, all sounds will lie somewhere in between these extremes.

However, the entropy of a probability density is maximum for the uniform density, so that H_{iid} is the maximum value that H can assume. Thus, in addition to ranking signals on a scale of aperiodicity, we can know precisely the two extremes of that scale. For different sampling times Δt the value $T_{\max}$ will change. Therefore, we can normalised the RPDE scale for subsequent usage:

$$H_{\text{norm}} = \frac{-\sum_{i=1}^{T_{\max}} P(i) \ln P(i)}{H_{\text{iid}}}, \tag{12}$$

a unit less quantity that assumes real values in the range $[0, 1]$.

The method of close returns, upon which this technique is based, was originally designed to characterise deterministic, chaotic systems [53]. In this case, if the chaotic system is ergodic and has evolved past any transients, then the recurrence properties of the system are independent of the initial conditions and initial time, i.e. they are invariants of the system. Similarly, if the system is purely stochastic and ergodic, then it has a stationary distribution. Again, after any transient phase, the recurrence properties will be invariant in the above sense. Thus the derived measure H will also be an invariant of the system. We note that traditional jitter and shimmer measurements do not share this invariance property, in this sense they do not give a reliable description for chaotic or Type III time series. Often, when stable phonation is initiated in speech, the vocal folds will take some time to settle into a pattern of stable periodic or chaotic oscillation. The behaviour of speech signals during this "settling time" is similar to the transient behaviour typically observed in nonlinear dynamical systems. In this study, we make use of voice signals which are stable phonations, and we discard any of these transient phases. Thus, to a reasonable approximation H can be considered as an invariant of the dynamics of the speech organs.

Figure 3 shows the result of this recurrence analysis, applied to a synthetic, perfectly periodic signal created by taking a single cycle from a speech signal and repeating it end-to-end many times. It also shows the analysis applied to a synthesised, uniform, i.i.d. random signal (on the signal range $[-1, 1]$) after optimising m, τ and r by gridded search. Even though exact results are impossible to obtain due to the approximation inherent to the algorithm and only finite-length signals, the figure shows that a close match is obtainable between the theoretical, predicted results and the simulated results.

Detrended Fluctuation Analysis

In order to investigate the second aspect of disordered voices, that of increased breath noise, we require a practical approach to applying the scaling ideas introduced above. Detrended fluctuation analysis is one straightforward technique for characterising the self-similarity of the graph of a signal from a stochastic process [54].

It is designed to calculate the *scaling exponent* α in nonstationary time series (where the statistics such as mean, variance and autocorrelation properties might change with time), and has been shown to produce robust results when there are slowly moving trends in the data. These will naturally include low frequency vibrations [54], which are similar in nature to the nonlinear vibrations of the vocal folds described by the function $\mathbf{f}$ in the model (1). Thus this technique can be used to characterise the properties of only the stochastic part $\varepsilon(t)$ of the model (1).

In this technique, the scaling exponent α is a measure of the ratio of the logarithm of the *fluctuations* or vertical height of the graph to the logarithm of the horizontal width of a chosen time window over which that vertical height is measured. The scaling exponent is calculated as the slope of a log-log graph of a range of different horizontal time window sizes against, the vertical height of the signal in those time windows. Mathematically, for self-similar signals with positive scaling exponent α the self-similarity property of the graph of the signal s_n should hold on all time scales, but we are limited by the finite amplitude range of physical measurements to a certain maximum scale. Thus the signal is first integrated in order to create a new stochastic process that exhibits self-similarity over a large range of time scales (then, for example, a purely independent, stochastic process will result in a self-similar random walk). First, the time series s_n is integrated:

$$y_n = \sum_{j=1}^{n} s_j, \tag{13}$$

for $n = 1, 2 \ldots N$, where N is the number of samples in the signal s_n. Then, y_n is divided into windows of length L samples. A least-squares straight line *local trend* is calculated by analytically minimising the squared error E^2 over the slope and intercept parameters a and b:

$$\arg\min_{a,b} E^2 = \sum_{n=1}^{L} (y_n - an - b)^2 . \tag{14}$$

Next, the root-mean-square deviation from the trend, the fluctuation, is calculated over every window at every time scale:

$$F(L) = \left[\frac{1}{L} \sum_{n=1}^{L} (y_n - an - b)^2 \right]^{\frac{1}{2}} . \tag{15}$$

This process is repeated over the whole signal at a range of different window sizes L, and a log-log graph of L against $F(L)$ is constructed. A straight line on this graph indicates self-similarity expressed as $F(L) \propto L^{\alpha}$. The scaling exponent α is calculated as the slope of a straight line fit to the log- log graph of L against $F(L)$ using least-squares as above. For a more in-depth presentation and discussion of self-similarity in signals in general, and further information about DFA, please see Kantz, Hu [25, 54].

We are assuming that the signal, as the measured output of the new model, represents a combination of deterministic and stochastic dynamics. The deterministic part of the dynamics, dictated by the function $\mathbf{f}$ in equation (1) will result in slower changes in the signal s_n taking place over a relatively long time scale. Similarly, the stochastic fluctuations in the signal indicated changes taking place over a much shorter time scale. Since the goal of DFA is to analyse the faster changing, stochastic properties of the signal, only a limited range of window sizes is investigated, over which the stochastic component of the signal exhibits

self-similarity as indicated by a straight-line on the log-log graph of window size against fluctuation. As an example, Type III would include some speech signals that are actually chaotic, where the chaos is due to slow, nonlinear dynamics in the vocal fold tissue and airflow, the characteristic time of this nonlinear oscillation will be much longer than the window sizes over which the scaling exponent is estimated. Thus, the nature of the chaotic oscillation will not affect the scaling exponent, which will respond only to any random fluctuations occurring on a much shorter time scale.

The resulting scaling exponent can assume any number on the real line. However, it would be more convenient to represent this scaling exponent on a finite sliding scale from zero to one, as we have done for the RPDE measure. Thus we need a mapping function $g : \mathbb{R} \to [0, 1]$. One such function finding common use in statistical and pattern recognition applications is the *logistic function* $g(x) = (1 + \exp(-x))^{-1}$ [55], so that the *normalised scaling exponent* is:

$$\alpha_{\mathrm{norm}} = \frac{1}{1 + \exp(-\alpha)}. \tag{16}$$

Therefore, each sound will lie somewhere between the extremes of zero and one on this scale, according to the self-similarity properties of the stochastic part of the dynamics. As will be shown later, speech sounds for which α_{norm} is closer to one are characteristic of general voice disorder.

Application to Examples of Normal and Disordered Voices

In this section we will apply these two new measures to some examples of normal and disordered voices, as a limited demonstration of characterising the extent of aperiodicity and breathiness in these signals.

Figure 4 shows the RPDE value H_{norm} calculated on the same two speech signals from the Kay database as shown in figure 1. Note that the second, disordered example is of Type III and shows significantly irregular vibration, which is detected by a large increase in H_{norm}.

Similarly, figure 5 shows two more speech examples, one normal and one disordered from the same database and the corresponding values of the scaling exponent α and α_{norm}. In these cases, the disordered example is extremely "breathy", and this turbulent noise is detected by an increase in the scaling exponent.

Quadratic Discriminant Analysis

In order to test the effectiveness of these two measures in practice, one approach is to set up a *classification task* to separate normal control subjects from disordered examples using these measures alone. Here we choose one of the simplest approaches, *quadratic discriminant analysis (QDA)*, which allows separation by

modelling the data conditional upon each class, here the normal (class C_1) and disordered (class C_2) cases, using joint Gaussian probability density functions [55]. For a $J \times K$ data matrix $\mathbf{v} = v_{jk}$ of observations consisting of the measures $j = 1, 2$ for RPDE and DFA respectively, and all subjects k, these likelihood densities are parameterised by the mean and covariance matrices of the data set:

$$\boldsymbol{\mu} = E\left[\mathbf{v}\right], \;\; \mathbf{C} = E\left[\left(\mathbf{v} - \boldsymbol{\mu}\right)\left(\mathbf{v} - \boldsymbol{\mu}\right)^T\right], \tag{17}$$

where E is the expectation operator, and $\boldsymbol{\mu}$ is the mean vector formed from the means of each row of $\mathbf{v}$. The class likelihoods are:

$$f_C(\mathbf{w} \mid C_i) = (2\pi)^{-J/2} \left|\mathbf{C}_i\right|^{-1/2} \exp\left[-\frac{1}{2}\left(\mathbf{w} - \boldsymbol{\mu}_i\right)^T \mathbf{C}_i^{-1}\left(\mathbf{w} - \boldsymbol{\mu}_i\right)\right], \tag{18}$$

for classes $i = 1, 2$ and an arbitrary observation vector $\mathbf{w}$. It can be shown that, given these Gaussian class models, the maximum likelihood regions of the observation space $\mathbb{R}^J$ are separated by a *decision boundary* which is a (hyper-)conic section calculated from the difference of log-likelihoods for each class, which is the unique set of points where each class is equally likely [55]. The maximum likelihood classification problem is then solved using the decision rule that $l(\mathbf{w}) \geq 0$ assigns $\mathbf{w}$ to class C_1, and $l(\mathbf{w}) < 0$ assigns it to class C_2, where:

$$l\left(\mathbf{w}\right) = -\frac{1}{2}\mathbf{w}^T\mathbf{A}_2\mathbf{w} + \mathbf{A}_1\mathbf{w} + A_0, \tag{19}$$

$$\mathbf{A}_2 = \mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}, \mathbf{A}_1 = \boldsymbol{\mu}_1^T\mathbf{C}_1^{-1} - \boldsymbol{\mu}_2^T\mathbf{C}_2^{-1}, \tag{20}$$

$$A_0 = -\frac{1}{2}\ln\left|\mathbf{C}_1\right| + \frac{1}{2}\ln\left|\mathbf{C}_2\right| - \frac{1}{2}\boldsymbol{\mu}_1^T\mathbf{C}_1^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T\mathbf{C}_2^{-1}\boldsymbol{\mu}_2. \tag{21}$$

In order to avoid the problem of overfitting, where the particular chosen separation model shows good performance on the training data but fails to *generalise* well to new, unseen data, the classifier results require validation.

In this paper, we make use of *bootstrap resampling* for validation [55]. In the bootstrap approach, the classifier is trained on K cases selected at random with replacement from the original data set of K cases. This trial resampling processes is repeated many times and the mean classification parameters $E\left[\mathbf{A}_2\right], E\left[\mathbf{A}_1\right], E\left[A_0\right]$ are selected as the parameters that would achieve the best performance on entirely novel data sets.

Bootstrap training of the classifier involves calculating H^k_{norm} and α^k_{norm} (the observations) for each speech sample k in the database, (where the superscript k denotes the measure for the k-th subject). Then, K random selections of these values with replacement H'^k_{norm} and α'^k_{norm} form the entries of the vector

$v_{1k} = H'^{k}_{\text{norm}}$ and $v_{2k} = \alpha'^{k}_{\text{norm}}$ Then the mean vectors for each class $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and covariance matrices $\mathbf{C}_1, \mathbf{C}_2$ for the whole selection are calculated. Next, for each subject, the decision function is evaluated:

$$l(\mathbf{w}_k) = l([H^k_{\text{norm}}, \alpha^k_{\text{norm}}]^T). \tag{22}$$

Subsequently, applying the decision rule assigns the subject k into either normal or disordered classes. Then the performance of the classifier can be evaluated in terms of percentage of true positives (when a disordered subject is correctly assigned to the disordered class C_1) and true negatives (when a normal subject is correctly assigned to the normal class C_2). The overall performance is the percentage of correctly classified subjects, in both classes. This bootstrap trial process of creating random selections of the measures, calculating the class mean vectors and covariance matrices, and then evaluating the decision function on all the measures to obtain the classification performance is repeated many times. Assuming that the performance percentages are normally distributed, then the 95% confidence interval of the classification performance percentages can be calculated. The best classification boundary can then be taken as the mean boundary overall all the trials.

Efficient implementations of the algorithms described in this paper written in C with Matlab MEX interface accompany this paper: close returns [see Additional files 1 and 2] and detrended fluctuation analysis [see Additional files 3, 4 and 5].

Algorithms for Classical Techniques

For the purposes of comparison, we calculate the classical measures of jitter, shimmer and HNR (Noise-to-Harmonics Ratio) [1]. There are many available algorithms for calculating this quantity, in this study we make use of the algorithms supplied in the software package Praat [56]. These measures are based on an autocorrelation method for determining the pitch period (see Boersma [57] for a detailed description of the method).

We also use the methods described in Michaelis [4]. This first requires calculating the measures EPQ (Energy Perturbation Quotient), PPQ (Pitch Perturbation Quotient), GNE (Glottal to Noise Excitation Ratio) and the mean correlation coefficient between successive cycles, measures which require the estimation of the pitch period using the waveform matching algorithm (see Titze [58] for a detailed description of this algorithm). The EPQ, PPQ, GNE and correlation coefficients are calculated over successive overlapping "frames" of the speech signal. Each frame starts at a multiple of 0.26 seconds, and is 0.5 seconds long. For each frame, the EPQ, PPQ, GNE and correlation coefficients are combined into a

pair of component measures, called Irregularity and Noise. We use the average of the Irregularity and Noise components over all these frames [4].

Classification Test Data

This study makes use of the Kay Elemetrics disordered voice database (KayPENTAX Model 4337, New Jersey, USA), which contains 707 examples of disordered and normal voices from a wide variety of organic, neurological and traumatic voice disorders. This represents examples of all three types of disordered voice speech signals (Types I, II and III). There are 53 control samples from normal subjects. Each speech sample in the database was recorded under controlled, quiet acoustic conditions, and is on average around two seconds long, 16 bit uncompressed PCM audio. Some of the speech samples were recorded at 50kHz and then downsampled with anti-aliasing to 25kHz. Used in this study are sustained vowel phonations, since this controls for any significant nonstationarity due changes in the position of the articulators such as the tongue and lips in running speech, which would have an adverse effect upon the analysis methods. For calculating the Irregularity and Noise components, the signals are resampled with anti-aliasing to 44.1kHz.

Results

Figure 6 shows hoarseness diagrams after Michaelis [4] constructed using the speech data and RPDE and DFA measures, the derived irregularity and noise components of Michaelis, along with the same diagrams using two other combinations of the three classical perturbation measures for direct comparison. The three classical measures are jitter, shimmer and NHR (Noise-to-Harmonics Ratio) [1]. The normalised RPDE, DFA scaling exponents and derived irregularity and noise components are calculated for each of the $K = 707$ speech signals. Where the traditional perturbation algorithms did not fail to produce a result, the traditional perturbation values were calculated for a smaller subset of the subjects, see [1] for details of these traditional algorithms. Also shown in figure 6 is the result of the classification task applied to the dataset; the best classification boundary is calculated using bootstrap resampling over 1000 trials. Table 1 summarises all the classification performance results for the classification tasks on the hoarseness diagrams of figure 6. The RPDE parameters were the same as for figure 3, and the DFA parameters were the same as for figure 5.

Discussion

As shown in table 1, of all the combinations of the new and traditional measures, and derived irregularity and noise components, the highest overall correct classification performance of $91.8 \pm 2.0\%$ is achieved by the RPDE/DFA pair. The combination of jitter and shimmer leads to the next highest performance. These results confirm that, compared under the same, simple classifier approach, the new nonlinear measures are more accurate on average than traditional measures or the derived irregularity and noise components. We will now discuss particular aspects of these results in comparison with traditional perturbation measures.

Feature Dimensionality

The *curse of dimensionality* afflicts all challenging data analysis problems [55]. In pattern analysis tasks such as automated normal/disordered separation, increasing the size of the feature vector (in this case, the number of measures J in the classifier vector $\mathbf{v}$) does not necessarily increase the performance of the classifier in general. This is because the volume of the *feature space* (the space spanned by the possible values of the measures) grows exponentially with the number of features. Therefore, the limited number of examples available to train the classifier occupy an increasingly small volume in the feature space, providing a poor representation of the mapping from features to classes that the classifier must learn [55]. Therefore the new measures help to mitigate this problem of dimensionality, since only these two new measures are required to obtain good separation performance. By comparison, we need to calculate four different measures in order obtain the irregularity and noise components [4].

Feature Redundancy – Information Content

It is also important to use as few features as possible because in practice, increasing the number of features causes excessive data to be generated that may well contain redundant (overlapping) information. The actual, useful information contained in these vectors has a much smaller dimensionality. For clinical purposes, it is important that only this useful data is presented. This effect of redundant information for the traditional measures can be clearly seen in figure 6, where combinations of pairs of (the logarithms of) these measures are seen to cluster around a line or curve in the feature space, indicating high positive correlation between these measures. Traditional measures occupy an effectively one-dimensional object in this two-dimensional space. The irregularity and noise components occupy more of the area of the feature space than traditional measures, and the new measures are spread evenly over the same space.

Arbitrary Parameters – Reproducibility

Minimising the number of arbitrary parameters used to calculate these measures is necessary to avoid selecting an excessively specialised set of parameters that leads, for example, to good normal/disordered separation on a particular data set but does not generalise well to new data.

Many parameters are required for the algorithms used in calculating traditional perturbation measures [4, 5, 7]. For example, the waveform matching algorithm [1] requires the definition of rough markers, upper and lower pitch period limits, low-pass filter cutoff frequencies, bandwidth and order selection parameters, and the number of pitch periods for averaging should these pitch period limits be exceeded [58]. Similarly, in just one of the noise measures (glottal-to-noise excitation ratio) used in Michaelis [4], we can determine explicitly at least four parameters relating to linear prediction order, bandpass filter number, order, cutoff selection, and time lag range parameters. There are two additional parameters for the length and starting sample of the part of the signal selected for analysis.

Our new measures require only five arbitrary parameters that must be chosen in advance: the length of the speech signal N, the maximum recurrence time $T_{\max}$, and the lower value, upper value and increment of the DFA interval lengths L. We have also shown, using analytical results, that we can calibrate out the dependence upon the state space close recurrence radius r, the time-delay reconstruction dimension d and the reconstruction delay τ.

Interpretation of Results

We have found, in agreement with Titze [8] and Carding [2], that perturbation measures cannot be obtained for all the speech sounds produced by subjects (see table 1). This limits the clinical usefulness of these traditional measures. By contrast, the new measures presented in this chapter do not suffer from this limitation and are capable of measuring, by design, all types of speech signals.

Taking into account the classification performance achievable using a simple classifier, the number of these measures that need to be combined to achieve effective normal/disordered separation, the number of arbitrary parameters used to calculate the measures, and the independence of these measures, traditional approaches and derived irregularity and noise components are seen to be considerably more complex than the new measures developed in this paper. The results of the classification comparison with traditional measures and the irregularity and noise components suggest that, in order to reach the classification performance of the new measures, we will either need much more complex classifiers, or need to combine many more classical features together [5–7]. It is therefore not clear that traditional approaches capture

the *essential biomechanical differences* between normal and disordered voices in the most parsimonious way, and an excessively complicated relationship exists therefore between the values of these measures and extent of the voice disorder. As a final comment, we note that the classical perturbation measures were, for the majority of signals, able to produce a result regardless of the type of the signal. This is consistent with the findings of other studies [4], where for Type II/III and random noise signals, the correct interpretation of these measures breaks down. Therefore, although it is no longer possible in these cases to assign a coherent meaning to the results produced by these measures, this does not *per se* mean that there is not some as yet unknown connection between disorder and the these measures. For this reason, we do not discard the results of these measures for Type II/III and random cases.

Limitations of the New Measures

There are certain limitations to the new measures in clinical practice. These measures rely upon sustained vowel phonation, and sometimes subjects experience difficulty in producing such sounds, which limits the applicability. Also, at the beginning of a sustained vowel phonation, the voice of many subjects may require some time to settle into a more stable vibration. As such, discarding the beginning of the phonation is usually a prerequisite (but this does not adversely affect the applicability of these methods). Nonetheless, the extent of breathiness in speech is not usually affected in this way. In practice we require that the subject maintains a constant distance from the microphone when producing speech sounds; this can be achieved, for example, with the use of head-mounted microphones. We note that these limitations also apply to existing measures.

Possible Improvements and Extensions

There are several improvements that could be made to these new measures. Firstly, every arbitrary parameter introduces extra variability that affects the reliability of the results. Much as it has been possible to calibrate out the dependence upon the RPDE parameters using analytical results, a theoretical study of the DFA interval lengths based upon typical sustained phonation recurrence periods could reveal values that would be found for all possible speech signals. These would be related to the sampling time Δt. The particular choice of normalisation function g for the scaling exponent might affect the classification performance, and better knowledge of the possible range of α values using theoretical studies of the DFA algorithm would be useful for this. It should also be possible to increase the recurrence time precision of the RPDE analysis by interpolating the state space orbits around the times of close recurrence n_0, n_1. It

should then be possible to achieve the same high-resolution as waveform matching techniques which would make RPDE competitive for the detailed analysis of Type I periodic sounds.

Conclusions

In this study, in order to directly characterise the two main biophysical factors of disordered voices: increased nonlinear, complex aperiodicity and non-Gaussian, aeroacoustic breath noise, we have introduced recurrence and scaling analysis methods. We introduced a new, combined nonlinear/stochastic signal model of speech production that is capable, in principle, of producing the wide variation in behaviour of normal and disordered voice examples. To exploit the output of this model in practice, and hence all types of normal and disordered voices, we explored the use of two nonlinear measures: the recurrence period density entropy and detrended fluctuation analysis.

Our results show that, when the assumptions of the model hold under experimental conditions (in that the speech examples are sustained vowels recorded under quiet acoustic conditions), we can directly characterise the two main factors of aperiodicity and breath noise in disordered voices and thus construct a "hoarseness" diagram showing the extent of normality/disorder from a speech signal. The results also show that on average, over all bootstrap resampling trials, these two measures alone are capable of distinguishing normal subjects from subjects with all types of voice disorder, with better classification performance than existing measures.

Furthermore, taking into account the number of arbitrary parameters in algorithms for calculating existing perturbation measures, and the number of these existing measures that need to be combined to perform normal/disordered separation, we have shown that existing approaches are considerably more complex.

We conclude that the nonlinearity and non-Gaussianity of the biophysics of speech production can be exploited in the design of signal analysis methods and screening systems that are better able characterise the wide variety of biophysical changes arising from voice disease and disorder. This is because ultimately the biophysics of speech production generate the widely varying phenomenology.

Appendix – Mathematical Proofs

Periodic Recurrence Probability Density

We consider the purely deterministic case, i.e. when the model of equation (1) has no forcing term $\varepsilon(t)$. Thus the measured time series is purely deterministic and points in the time series follow each other in an exactly prescribed sequence. When the measured, trajectory $\mathbf{s}_n$ is a purely periodic orbit of finite period k

steps, there is an infinite sequence of points $\{\mathbf{r}_n\}, n \in \mathbf{Z}$ in the reconstructed state space with $\mathbf{r}_n = \mathbf{r}_{n+k}$, and $\mathbf{r}_n \neq \mathbf{r}_{n+j}$ for $0 < j < k$.

Picking any point $\mathbf{s}$ in the reconstructed state-space, there are two cases to consider. In the first case, if $\mathbf{s} = \mathbf{r}_n$ for some n, then $\mathbf{s}$ is not the same as any other points in the periodic orbit except for $\mathbf{r}_{n+k}$, so that the trajectory returns with certainty for the first time to this point after k time steps. This certainty, with the requirement the that probability of first recurrence is normalised for $T = 1, 2 \ldots$ implies that:

$$P_{\mathbf{s}}(T = r) = \begin{cases} 1 & \text{if } r = k \\ 0 & \text{otherwise} \end{cases} . \tag{23}$$

In the second case when $\mathbf{s} \neq \mathbf{r}_n$ for any n, the trajectory never intersects the point so that there are also never any first returns to this point. All the points in the reconstructed space form a disjoint partition of the whole space. Thus the probability of recurrence to the whole space is the sum of the probability of recurrence to each point in the space separately, appropriately weighted to satisfy the requirement that the probability of first recurrence to the whole space is normalised However, only the k distinct points of the periodic orbit contribute to the total probability of first recurrence to the whole space. Therefore, the probability of first recurrence is:

$$P(T) = \frac{1}{k} \sum_{i=0}^{k-1} P_{\mathbf{r}_i}(T = r) = \begin{cases} 1 & \text{if } r = k \\ 0 & \text{otherwise} \end{cases} . \tag{24}$$

Uniform i.i.d. Stochastic Recurrence Probability Density

Consider the purely stochastic case when the nonlinear term $\mathbf{f}$ is in equation (1) is zero and the stochastic forcing term is an i.i.d. random vector. Then the measured trajectory $\mathbf{s}_n$ is also a stochastic, i.i.d. random vector. Since all the time series are normalised to the range $[-1, 1]$ then each member of the measurement takes on a value from this range. Then the trajectories $\mathbf{s}_n$ occupy the reconstructed state-space which is the region $[-1, 1]^m$, and each co-ordinate s_n is i.i.d. uniform. We form a equal-sized partition of this space into N^m (hyper)-cubes, denoting each cubical region R. The length of the side of each cube R is $\Delta s = 2/N$. Then the probability of finding the trajectory in this cube is $P_R = \Delta s^m / 2^m$. Since the co-ordinates s_n are uniform i.i.d., then the probability of first recurrence of time T to this region R is geometric [51]:

$$P_R(T) = P_R [1 - P_R]^{T-1} = \frac{\Delta s^m}{2^m} \left[1 - \frac{\Delta s^m}{2^m}\right]^{T-1} . \tag{25}$$

This is properly normalised for $T = 1, 2 \ldots$. However, we require the probability of first recurrence to all possible cubes. The cubes are a disjoint partition of the total reconstruction space $[-1, 1]^m$. Thus the

probability of recurrence to the whole space is the sum of the probability of recurrence to each cube separately, appropriately weighted to satisfy the requirement that the probability of recurrence to the whole space is normalised. Since the probability of first recurrence to each cube R, $P_R(T)$ is the same, the probability of recurrence to all cubes is:

$$P(T) = \sum_{i=1}^{N^m} \frac{\Delta s^m}{2^m} P_R(T) = N^m \frac{\Delta s^m}{2^m} P_R(T) \tag{26}$$

$$= \frac{2^m}{\Delta s^m} \frac{\Delta s^m}{2^m} P_R \left[1 - P_R\right]^{T-1} = \frac{\Delta s^m}{2^m} \left[1 - \frac{\Delta s^m}{2^m}\right]^{T-1}. \tag{27}$$

For small cube side lengths Δs and close returns algorithm radius r, the first recurrence probability determined by the close returns algorithm is then:

$$P(T) = \frac{\Delta s^m}{2^m} \left[1 - \frac{\Delta s^m}{2^m}\right]^{T-1} \approx \frac{r^m}{2^m} \left[1 - \frac{r^m}{2^m}\right]^{T-1}. \tag{28}$$

Similarly, for small close returns radius r and/or for large embedding dimensions m, $1 - r^m/2^m \approx 1$ so that:

$$P(T) \approx \frac{r^m}{2^m}. \tag{29}$$

Note that for fixed m and r this expression is constant. Since the close returns algorithm can only measure recurrence periods over a limited range $1 \leq T \leq T_{\max}$, and we normalise the recurrence histogram $R(T)$ over this range of T, then the probability of first recurrence is the uniform density:

$$P(T) \approx \frac{1}{T_{\max}}, \tag{30}$$

which is proportional to the expression $r^m/2^m$ above. Thus, up to a scale factor, for a uniform i.i.d. stochastic signal, the recurrence probability density is uniform.

Authors' Contributions

MAL lead the conceptual design of the study, developed the mathematical methods, wrote the software and data analysis tools, and prepared and analysed the data. PEM participated in the conceptual design of the study and the mathematical methods. SJR participated in the discriminant analysis of the data. DAEC participated in the data preparation. IMM contributed to the development of the mathematical methods. All authors read and approved the manuscript.

Acknowledgements

Max Little acknowledges the financial support of the Engineering and Physical Sciences Research Council, UK, and wishes to thank Prof. Gesine Reinert (Department of Statistics, Oxford University, UK), Mr

Martin Burton (Radcliffe Infirmary, Oxford, UK) and Prof. Adrian Fourcin (University College London, London, UK) for valuable discussions and comments on early drafts of this paper.

References

1. Baken RJ, Orlikoff RF: *Clinical Measurement of Speech and Voice*. San Diego: Singular Thomson Learning, 2nd edition 2000.
2. Carding PN, Steen IN, Webb A, Mackenzie K, Deary IJ, Wilson JA: **The reliability and sensitivity to change of acoustic measures of voice quality**. *Clinical Otolaryngology* 2004, **29**(5):538–544.
3. Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, Van De Heyning P, Remacle M, Woisard V: **A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS)**. *Eur Arch Otorhinolaryngol* 2001, **258**(2):77–82.
4. Michaelis D, Frohlich M, Strube HW: **Selection and combination of acoustic features for the description of pathologic voices**. *Journal of the Acoustical Society of America* 1998, **103**(3):1628–1639.
5. Boyanov B, Hadjitodorov S: **Acoustic analysis of pathological voices**. *IEEE Eng Med Biol Mag* 1997, **16**(4):74–82.
6. Godino-Llorente JI, Gomez-Vilda P: **Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors**. *Ieee Transactions on Biomedical Engineering* 2004, **51**(2):380–384.
7. Alonso J, de Leon J, Alonso I, Ferrer M: **Automatic detection of pathologies in the voice by HOS based parameters**. *EURASIP Journal on Applied Signal Processing* 2001, **4**:275–284.
8. Titze IR: **Workshop on acoustic voice analysis: Summary statement** 1995.
9. Schoentgen J: **Jitter in sustained vowels and isolated sentences produced by dysphonic speakers**. *Speech Communication* 1989, **8**:61–79.
10. Hirano M, Hibi S, Yoshida T, Hirade Y, Kasuya H, Kikuchi Y: **Acoustic analysis of pathological voice. Some results of clinical application**. *Acta Otolaryngol* 1988, **105**(5-6):432–8.
11. Quatieri TF: *Discrete-Time Speech Signal Processing: Principles and Practice*. Upper Saddle River, NJ: Prentice Hall 2002.
12. Box GEP: **Science and Statistics**. *Journal of the American Statistical Association* 1976, **71**(356):791–799.
13. Herzel H, Berry D, Titze IR, Saleh M: **Analysis of vocal disorders with methods from nonlinear dynamics**. *Journal of Speech and Hearing Research* 1994, **37**(5):1008–1019.
14. Akaike H: **A new look at the statistical model identification**. *IEEE Trans. Automat. Contrl.* 1974, **AC-19**(6):716–723.
15. Maragos P, Potamianos A: **Fractal dimensions of speech sounds: computation and application to automatic speech recognition**. *J Acoust Soc Am* 1999, **105**(3):1925–32.
16. Banbrook M, McLaughlin S, Mann I: **Speech characterization and synthesis by nonlinear methods**. *IEEE Transactions on Speech and Audio Processing* 1999, **7**:1–17.
17. Zhang Y, Jiang JJ, Biazzo L, Jorgensen M: **Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis**. *Journal of Voice* 2005, **19**(4):519–528.
18. Zhang Y, McGilligan C, Zhou L, Vig M, Jiang JJ: **Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps**. *Journal of the Acoustical Society of America* 2004, **115**(5):2270–2277.
19. Giovanni A, Ouaknine M, Triglia JL: **Determination of largest Lyapunov exponents of vocal signal: Application to unilateral laryngeal paralysis**. *Journal of Voice* 1999, **13**(3):341–354.
20. Zhang Y, Jiang JJ, Wallace SM, Zhou L: **Comparison of nonlinear dynamic methods and perturbation methods for voice analysis**. *Journal of the Acoustical Society of America* 2005, **118**(4):2551–2560.

21. Hertrich I, Lutzenberger W, Spieker S, Ackermann H: **Fractal dimension of sustained vowel productions in neurological dysphonias: An acoustic and electroglottographic analysis**. *Journal of the Acoustical Society of America* 1997, **102**:652–654.

22. Hansen JHL, Gavidia-Ceballos L, Kaiser JF: **A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment**. *Ieee Transactions on Biomedical Engineering* 1998, **45**(3):300–313.

23. Zhang Y, Jiang JJ: **Nonlinear dynamic analysis in signal typing of pathological human voices**. *Electronics Letters* 2003, **39**(13):1021–1023.

24. Jiang JJ, Zhang Y, McGilligan C: **Chaos in voice, from modeling to measurement**. *Journal of Voice* 2006, **20**:2–17.

25. Kantz H, Schreiber T: *Nonlinear Time Series Analysis*. Cambridge; New York: Cambridge University Press, new edition 1999.

26. Behrman A, Baken RJ: **Correlation dimension of electroglottographic data from healthy and pathologic subjects**. *Journal of the Acoustical Society of America* 1997, **102**(4):2371–2379.

27. Little M, McSharry P, Moroz I, Roberts S: **Nonlinear, biophysically-informed speech pathology detection**. In *Proc ICASSP 2006*, New York: IEEE Publishers 2006.

28. McSharry PE, Smith LA, Tarassenko L: **Prediction of epileptic seizures: are nonlinear methods relevant?** *Nat Med* 2003, **9**(3):241–2.

29. Herzel H, Berry D, Titze I, Steinecke I: **Nonlinear dynamics of the voice - signal analysis and biomechanical modeling**. *Chaos* 1995, **5**:30–34.

30. Jiang JJ, Zhang Y: **Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds**. *Journal of the Acoustical Society of America* 2002, **112**(5):2127–2133.

31. Krane MH: **Aeroacoustic production of low-frequency unvoiced speech sounds**. *Journal of the Acoustical Society of America* 2005, **118**:410–427.

32. LaMar MD, Qi YY, Xin J: **Modeling vocal fold motion with a hydrodynamic semicontinuum model**. *Journal of the Acoustical Society of America* 2003, **114**:455–464.

33. Ishizaka K, Flanagan JL: **Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords**. *ATT Bell System Technical Journal* 1972, **51**(6):1233–1268.

34. Steinecke I, Herzel H: **Bifurcations in an Asymmetric Vocal-Fold Model**. *Journal of the Acoustical Society of America* 1995, **97**(3):1874–1884.

35. Jiang JJ, Zhang Y, Stern J: **Modeling of chaotic vibrations in symmetric vocal folds**. *Journal of the Acoustical Society of America* 2001, **110**(4):2120–2128.

36. Dixit R: **On defining aspiration**. In *Proceedings of the XIIIth International Conference of Linguistics*, Tokyo, Japan 1988:606–610.

37. Howe MS: *Theory of vortex sound*. New York: Cambridge University Press 2003.

38. Sinder D: **Synthesis of unvoiced speech sounds using an aeroacoustic source model**. *PhD thesis*, Rutgers University 1999.

39. McLaughlin S, Maragos P: **Nonlinear methods for speech analysis and synthesis**. In *Advances in nonlinear signal and image processing, Volume 6*. Edited by Marshall S, Sicuranza G, Hindawi Publishing Corporation 2007:103–.

40. Ockendon JR: *Applied partial differential equations*. Oxford; New York: Oxford University Press 2003.

41. Acheson DJ: *Elementary fluid dynamics*. Oxford; New York: Oxford University Press 1990.

42. Kinsler LE, Frey AR: *Fundamentals of acoustics*. New York: Wiley, 2d edition 1962.

43. Falconer KJ: *Fractal geometry: mathematical foundations and applications*. Chichester; New York: Wiley 1990.

44. Zhao W, Zhang C, Frankel SH, Mongeau L: **Computational aeroacoustics of phonation, part I: Computational methods and sound generation mechanisms**. *Journal of the Acoustical Society of America* 2002, **112**(5 Pt 1):2134–46.

45. Grimmett G, Stirzaker D: *Probability and random processes*. Oxford; New York: Oxford University Press, 3rd edition 2001.

46. Berry DA, Herzel H, Titze IR, Story BH: **Bifurcations in excised larynx experiments**. *J Voice* 1996, **10**(2):129–38.

47. Little M, McSharry P, Moroz I, Roberts S: **Testing the assumptions of linear prediction analysis in normal vowels**. *Journal of the Acoustical Society of America* 2006, **119**:549–558.

48. Tokuda I, Miyano T, Aihara K: **Surrogate analysis for detecting nonlinear dynamics in normal vowels**. *Journal of the Acoustical Society of America* 2001, **110**(6):3207–17.

49. Tokuda I, Tokunaga R, Aihara K: **A simple geometrical structure underlying speech signals of the Japanese vowel a**. *International Journal of Bifurcation and Chaos* 1996, **6**:149–160.

50. Jackson P, Shadle C: **Pitch-scaled estimation of simultaneous voiced and turbulence-noise components in speech**. *IEEE Transactions on Speech and Audio Processing* 2001, **9**(7):713–726.

51. Altmann EG, Kantz H: **Recurrence time analysis, long-term correlations, and extreme events**. *Physical Review E* 2005, **71**(5):–.

52. Ragwitz M, Kantz H: **Markov models from data by simple nonlinear time series predictors in delay embedding spaces**. *Physical Review E* 2002, **65**(5):056201.

53. Lathrop DP, Kostelich EJ: **Characterization of an experimental strange attractor by periodic-orbits**. *Physical Review A* 1989, **40**(7):4028–4031.

54. Hu K, Ivanov PC, Chen Z, Carpena P, Stanley HE: **Effect of trends on detrended fluctuation analysis**. *Physical Review E* 2001, **6401**:–.

55. Bishop CM: *Neural Networks for Pattern Recognition*. Oxford, New York: Clarendon Press; Oxford University Press 1995.

56. Boersma P, Weenink D: **Praat, a system for doing phonetics by computer**. *Glot International* 2001, **5**(9/10):341–345.

57. Boersma P: **Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound**. In *Proceedings of the Institute of Phonetic Sciences, Volume 17*, University of Amsterdam 1993.

58. Titze IR, Liang HX: **Comparison of F0 extraction methods for high-precision voice perturbation measurements**. *Journal of Speech and Hearing Research* 1993, **36**(6):1120–1133.

Figures

Tables

Table 1 - Summary of disordered voice classification results

Summary of disordered voice classification task performance results, for several different combinations of the new measures, the derived irregularity (Irreg) and noise (Noise) components of Michaelis [4], and traditional perturbation measures, jitter (Jitt), shimmer (Shim) and noise-to-harmonics ratio (NHR). The RPDE parameters were the same as for figure 4, and the DFA parameters were the same as for figure 5.

Combination	Subjects	True Positive	True Negative	**Overall**
RPDE/DFA	707	95.4±3.2%	91.5±2.3%	**91.8±2.0%**
Jitt/Shim	685	86.9±6.9%	81.0±4.7%	**81.4±3.9%**
Shim/NHR	684	91.4±5.9%	79.8±4.7%	**80.7±4.0%**
Irreg/Noise	707	78.4±6.2%	90.5±4.9%	**79.3±5.5%**
Jitt/NHR	684	93.2±7.4%	75.0±5.5%	**76.4±4.8%**

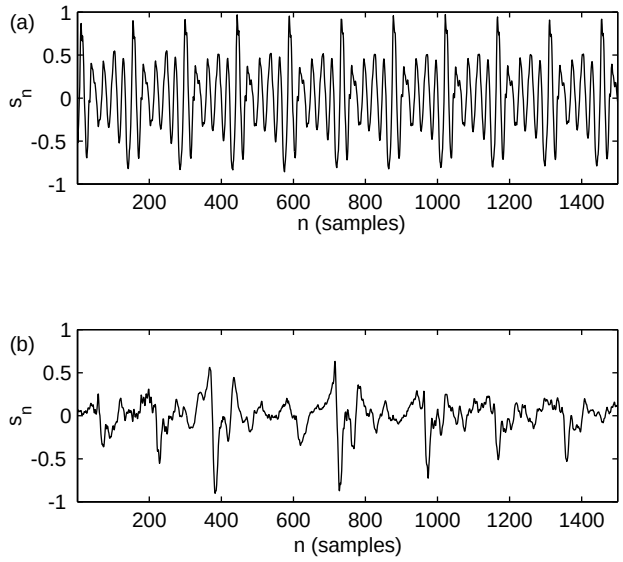

Figure 1: Discrete-time signals from (a) one normal (JMC1NAL) and (b) one disordered (JXS01AN) speech signal from the Kay Elemetrics database. For clarity only a small section is shown (1500 samples).

Additional Files

Additional file 1 – close_ret.c, 6K

Close returns algorithm implemented in C with Matlab MEX interface. Standard ASCII text file format.

Additional file 2 – close_ret.dll, 53K

Close returns algorithm compiled as a DLL for Windows. Standard Windows DLL format.

Additional file 3 – fastdfa_core.c, 9K

Efficient implementation of the detrended fluctuation analysis (DFA) algorithm written in C with Matlab MEX interface. Core C code. Standard ASCII text file format.

Additional file 4 – fastdfa_core.dll, 53K

DFA algorithm core compiled as a DLL for Windows. Standard Windows DLL format.

Additional file 5 – fastdfa.m, 2K

Matlab function wrapper for above DFA algorithm implementation. Standard ASCII Matlab script file format. Type 'help fastdfa.m' at the Matlab command prompt for usage instructions.

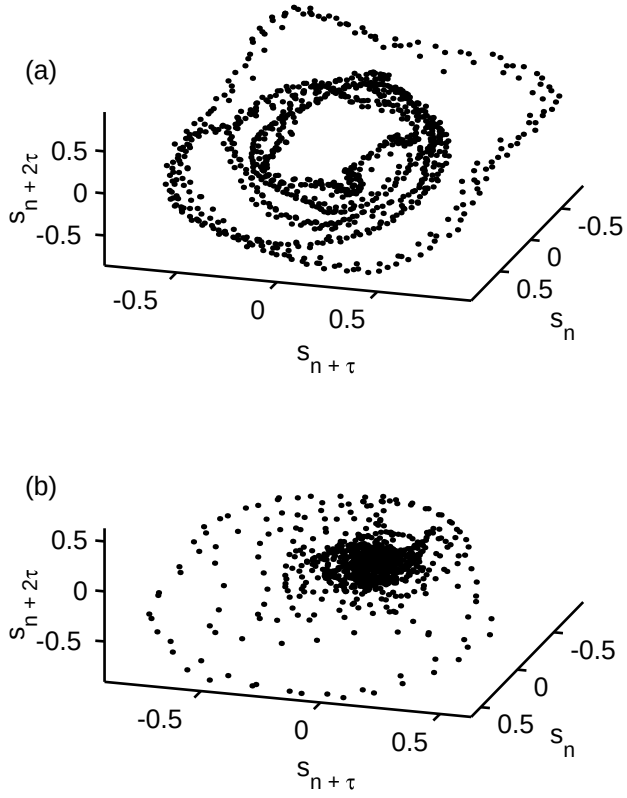

Figure 2: Time-delay embedded discrete-time signals from (a) one normal (JMC1NAL) and (b) one disordered (JXS01AN) speech signal from the Kay Elemetrics database. For clarity only a small section is shown (1500 samples). The embedding dimension is $m = 3$ and the time delay is $\tau = 7$ samples.

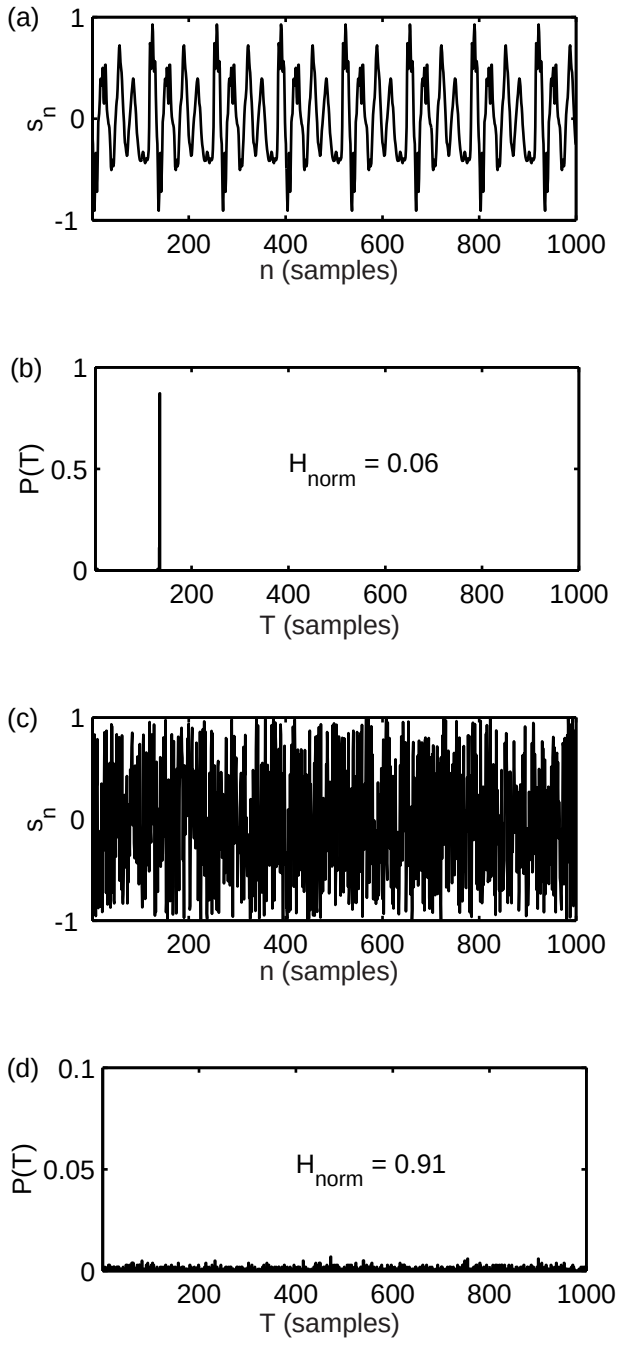

Figure 3: Demonstration of results of time-delayed state-space recurrence analysis applied to a perfectly periodic signal (a) created by taking a single cycle (period $k = 134$ samples) from a speech signal and repeating it end-to-end many times. The signal was normalised to the range $[-1, 1]$. (b) All values of $P(T)$ are zero except for $P(133) = 0.1354$ and $P(134) = 0.8646$ so that $P(T)$ is properly normalised. This analysis is also applied to (c) a synthesised, uniform i.i.d. random signal on the range $[-1, 1]$, for which (d) the density $P(T)$ is fairly uniform. For clarity only a small section of the time series (1000 samples) and the recurrence time (1000 samples) is shown. Here, $T_{\max} = 1000$. The length of both signals was 18088 samples. The optimal values of the recurrence analysis parameters were found at $r = 0.12$, $m = 4$ and $\tau = 35$.

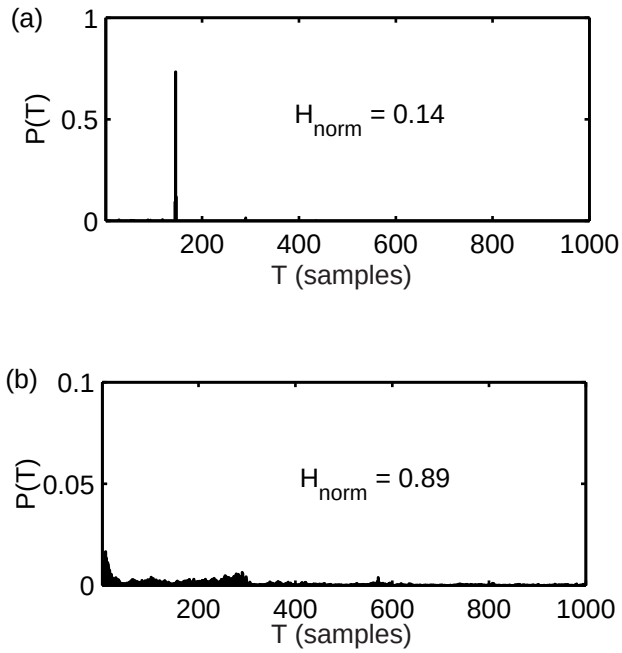

Figure 4: Results of RPDE analysis carried out on the two example speech signals from the Kay database as shown in figure 1. (a) Normal voice (JMC1NAL), (b) disordered voice (JXS01AN). The values of the recurrence analysis parameters were the same as those in the analysis of figure 3. The normalised RPDE value $H_{\rm norm}$ is larger for the disordered voice.

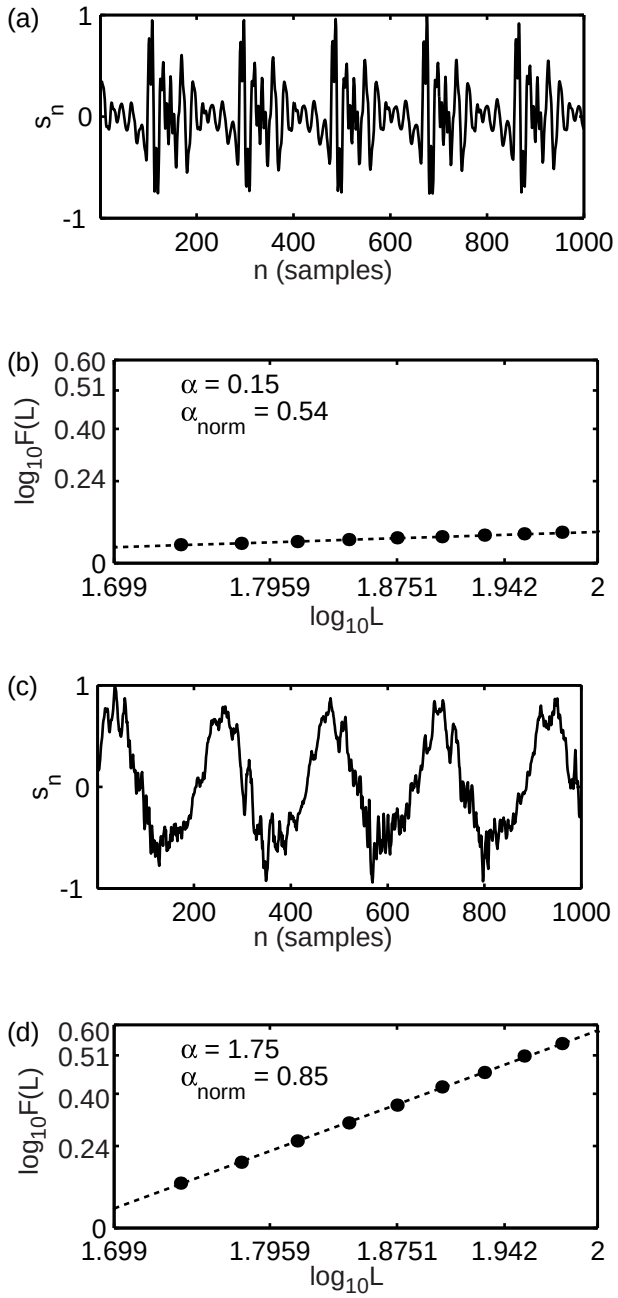

Figure 5: Results of scaling analysis carried out on two more example speech signals from the Kay database. (a) Normal voice (GPC1NAL) signal, (c) disordered voice (RWR14AN). Discrete-time signals s_n shown over a limited range of n for clarity. (b) Logarithm of scaling window sizes L against the logarithm of fluctuation size $F(L)$ for normal voice in (a). (d) Logarithm of scaling window sizes L against the logarithm of fluctuation size $F(L)$ for disordered voice in (b). The values of L ranged from $L = 50$ to $L = 100$ in steps of five. In (b) and (d), the dotted line is the straight-line fit to the logarithms of the values of L and $F(L)$ (black dots). The values of α and the normalised version α_{norm} show an increase for the disordered voice.

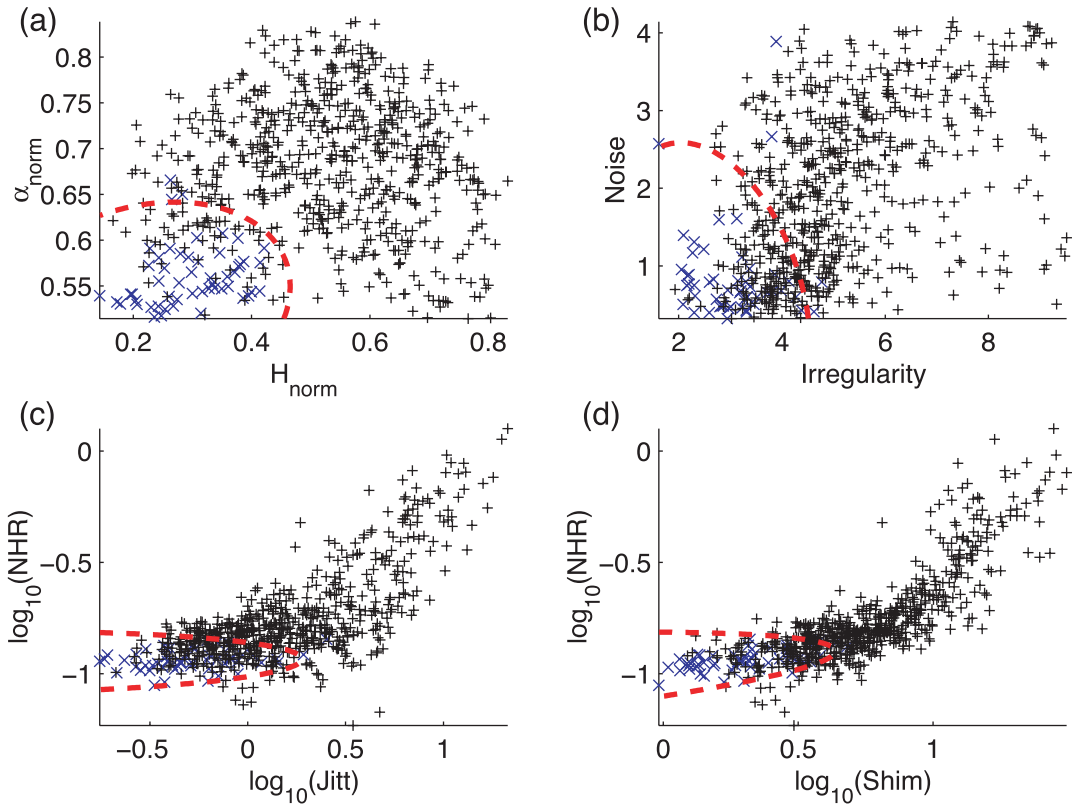

Figure 6: "Hoarseness" diagrams illustrating graphically the distinction between normal (blue 'x' symbols) and disordered (black '+' symbols) on all speech examples from the Kay Elemetrics dataset, for (a) the new measures return period density entropy (RPDE) (horizontal axis) and detrended fluctuation analysis (DFA) (vertical axis), (b) for the irregularity (horizontal) and noise (vertical) components of Michaelis [4], (c) for classical perturbation measures jitter (horizontal) and noise-to-harmonics ratio (NHR) (vertical) and (d) shimmer (horizontal) against NHR (vertical). The red dotted line shows the best normal/disordered classification task boundary over 1000 bootstrap trials using quadratic discriminant analysis (QDA). The values of the RPDE and DFA analysis parameters were the same those in the analysis of figures 3 and 5 respectively. The logarithm of the classical perturbation measures was used to improve the classification performance with QDA.