ELSEVIER

Speech Communication 21 (1997) 255–272

SPEECH COMMUNICATION

Predictable and random components of jitter

Jean Schoentgen *, Raoul De Guchteneere

Laboratory of Experimental Phonetics, Institute of Modern Languages and Phonetics, CP 110, Université Libre de Bruxelles, 50, avenue F.-D. Roosevelt, B-1050 Brussels, Belgium

Received 3 July 1996; revised 12 December 1996; accepted 20 February 1997

Abstract

The subject of this article is the study of the deterministic and random components of jitter by means of a statistical time series model. Jitter is the small fluctuations in glottal cycle lengths. The purpose of time series analysis is to take into account the fact that glottal cycles are produced sequentially and that relations between neighbouring perturbations exist. The jitter time series model statistically represents the present perturbation as a weighted sum of past perturbations and random noise. The model is fitted to observed jitter time series by means of conventional linear methods. A discriminant analysis of jitter time series extracted from 279 sustained vocoids [a] [i] [u] shows that the jitter features which separately describe the predictable and random components better characterise healthy and dysphonic speakers than a traditional jitter feature. The conclusion is that the relations between neighbouring cycle length perturbations are an aspect of jitter independent of the scatter of the cycle lengths which is described by conventional jitter features.

Zusammenfassung

Die deterministischen und zufälligen Komponenten der kleinen Schwankungen der Dauer der Glottiszyklen wurden untersucht anhand eines statistischen Zeitreihenmodells. Der Zweck der Zeitreihenanalyse war der Zeitfolge der Glottiszyklen und dem Einfluss benachbarter Schwankungen aufeinander Rechnung zu tragen. Das Zeitreihenmodell stellte die Schwankungen der gegenwärtigen Glottiszyklendauer dar als eine Summe von vorherigen Schwankungen und weissem Rauschen. Das Modell wurde beobachteten Zeitreihen von Dauerschwankungen angepasst mit Hilfe von bekannten linearen Methoden. Eine statistische Analyse von 279 Zeitreihen gemessen anhand von gehaltenen [a] [i] [u] Vokalen zeigte dass Schwankungsmerkmale, welche die zufälligen und vorhersagbaren Komponenten isoliert darstellten, gesunde und dysphonische Sprecher besser beschrieben denn herkömmliche Merkmale. Die Schlussfolgerung ist, dass die Zusammenhänge zwischen Schwankungen der Dauer benachbarter Glottiszyklen und die Schwankungsstreuung welche von herkömmlichen Merkmalen dargestellt werden zwei unabhängige Aspekt dieses Phänomens darstellen.

Résumé

L'article concerne l'étude des composantes déterministe et aléatoire des microperturbations de la durée du cycle glottique à l'aide d'un modèle statistique de série chronologique. L'utilisation d'un modèle de série chronologique rend compte des faits suivants: les cycles glottiques sont produits séquentiellement et il existe des liens entre les perturbations de cycles voisins. Le modèle représente la perturbation de la durée du cycle présent comme une somme pondérée des perturbations

* Corresponding author. National Fund for Scientific Research, Belgium. E-mail: jschoent@ulb.ac.be.

PII S0167-6393(97)00008-3

passées et d'un bruit blanc. Le modèle est ajusté à des séries de perturbations observées, à l'aide de méthodes linéaires conventionnelles. Une analyse discriminante de séries extraites de 279 vocoïdes [a] [i] [u] montre que des indices qui décrivent isolément les composantes prédictibles et aléatoires des perturbations caractérisent mieux locuteurs dysphoniques et sains qu'un indice conventionnel. La conclusion est que les relations entre perturbations de cycles voisins constituent un aspect indépendant de la dispersion des microperturbations décrite à l'aide des indices conventionnels. © 1997 Elsevier Science B.V.

Keywords: Jitter; Linear auto-regressive modeling

1. Introduction

Jitter is the small fluctuations in glottal cycle lengths. The subject of this article is the study of the predictable and random components of jitter. A superficial inspection of cycle length data would suggest that jitter is purely random, i.e. best described by means of the scatter of glottal cycle lengths around their average. However, correlation analysis shows that the perturbation of present cycle lengths must be predictable in part from the perturbations of past cycles (Imaizumi, 1986; Hillenbrand, 1988; De Guchteneere and Schoentgen, 1991). In this article, we examine how jitter can be broken down into a predictable and random component and how both components can be used to characterise dysphonic voices.

Jitter has been studied since the beginning of the 1960s. It has sustained the interest of researchers, who believe that jitter can be reliably measured via the speech signal and used to discriminate between healthy and dysphonic speakers. But, since the second half of the 1980s, research has become less clinically and more methodologically oriented. The reasons are that methods for measuring jitter and features summarising it differed considerably from one study to the next. This made both cross-study comparisons and the gathering of reference data difficult. Also, it seemed desirable to examine the influence on jitter measurement of experimental factors such as recording equipment, the type and number of speech items, additive noise, oversampling, and so forth.

One methodological issue that had to be settled was the choice of the type of speech items, i.e. continuous speech, isolated sentences or sustained vowels. Arguments in favour of continuous speech or isolated sentences are that difficulties in abducting or adducting, or laryngeal pathologies leading to an asymmetry between the left and right vocal folds, may be more easily revealed during non-stationary stretches of speech (De Krom, 1994). The jitter of continuous speech was therefore studied by Gubrynowicz et al. (1977, 1980), Askenfeld et al. (1980), Askenfeld and Hammarberg (1980, 1981), and Laver et al. (1984, 1986). These studies were clinically or screening oriented and did not track the influence of other experimental factors on the outcome of jitter measurement.

The pioneering work of Lieberman (1963) was carried out on isolated sentences spoken affirmatively, interrogatively, joyously and so forth. Jitter was characterised by means of a jitter factor which counted all perturbations larger than 0.5 ms. Later, since the jitter factor failed to discriminate between their speakers, Hecker and Kreul (1971) considered the signs of the perturbations instead of their values.

The most frequently used items are sustained vowels. The arguments in favour of sustained vowels are methodological rather than clinical. Indeed during the emission of sustained vowels, extrinsic perturbations of the vibratory pattern are either steady or absent. These perturbations are the outcome of tract/glottis interaction, micro-prosody, voicing and devoicing for phonetic purposes, and intonation. In addition for the purpose of jitter analysis, conventional algorithms are not robust enough to automatically extract the fundamental period or frequency of running speech (De Krom, 1994).

Earlier studies of jitter in sustained vowels were founded on the calculation of a typical value of the absolute differences between the cycle lengths and their averages or, more rarely, the instantaneous frequencies and their averages (Hollien et al., 1973;

Smith et al., 1978; Horii, 1979, 1982). Basically, these studies agreed on the order of magnitude of jitter, i.e. between 0.1% and 1% of the fundamental period (Heiberger and Horii, 1982; Higgins and Saxman, 1989a; Schoentgen, 1989; Deem et al., 1989), and on the decrease of jitter with increasing fundamental frequencies (Hollien et al., 1973).

Koike et al. (Koike and Takahashi, 1972; Koike, 1973; Kitajima et al., 1975; Koike et al., 1977) were among the first to try to separate the influences of non-flat intonation contours, vibrato and jitter on cycle lengths. In order to isolate intrinsic jitter from other sources of variation they calculated absolute perturbations with reference to a running average and divided the average perturbation by the average fundamental period to compensate for influences related to a speaker's fundamental frequency. As a consequence, this compensation also diminished differences between vowel qualities since high vowels are expected to have a slightly higher intrinsic fundamental frequency than low vowels. Later, Davis (1976, 1979) generalised Koike's definition of relative jitter by calculating the running average over intervals of arbitrary extent. We found that all the jitter features which we are aware of and which use a running average are statistically biased. They underestimate jitter by typically $100/k$ percent where k equals the number of glottal cycles entering into the calculation of the running average; k typically amounts to 3 or 5. This means that a jitter of 1% measured using a five-point moving average technique is in fact a 1.2% jitter (Schoentgen and De Guchteneere, 1995).

A majority of the authors conclude from their results that it is possible to discriminate between healthy and dysphonic voices by means of jitter features. Ludlow et al. (1987) disapprove since rarely more than 70% of dysphonic voices can be classified correctly. This was deemed to be insufficient for screening purposes, for example.

Be that as it may, during the 1980s more and more studies appeared which dealt with purely methodological problems. Laver et al. (1982), for example, compared several fundamental frequency extractors in order to determine the most robust algorithm with reference to dysphonic voices (Gold and Rabiner, 1969). One of the first purely methodological studies of jitter and shimmer was by Heiberger and Horii (1982), who reviewed definitions and measurement procedures and proposed a perceptual interpretation of jitter and shimmer. Later, Pinto and Titze (1990) also criticised the ad hoc approach to defining jitter features, and the discrimination performances that had remained feeble. They studied jitter features by means of synthetic signals and recommended a common terminology for well-used features such as the jitter factor (Hollien et al., 1973) and the relative average perturbation (Koike et al., 1977). Cox and Ito (1989) studied the effects of quantization and measurement errors on relative average perturbation. They concluded that the latter gave correct estimates as long as the fundamental frequency remained below 250 Hz and the sampling frequency was at least 20 kHz. Quantization errors appeared to be negligible for quantizers with a 12 bit resolution at least. Titze et al. (1987) were also interested in the effects of sampling frequency and resolution on perturbation measures. They concluded that interpolation was an acceptable method for increasing time resolution, an increase which was required when fewer than 500 samples per cycle were available. They also concluded that low-pass filtering could be carried out as long as the cut-off frequency was at least an octave above the fundamental frequency and that 20 to 30 cycles represented perturbations accurately, at least for healthy speakers. Deem et al. (1989) studied the effect of interpolation on perturbation measurements by means of sinusoidal signals. They preferred to detect zero passages whose positions they refined by means of linear interpolation. They recommended analysis intervals comprising 40 cycles. Baer et al. (1983), and subsequently Horiguchi et al. (1987), compared measurement techniques such as high speed cinematography and electroglottography, or electroglottography and speech signal analysis. They found that the unsteadiness of the electroglottogram baseline adversely influenced jitter measurements. They therefore recommended peak detection. Still other authors tried to quantify intra-speaker variability (Garret and Healey, 1987; Higgins and Saxman, 1989a,b). They concluded that inter-session differences in jitter measurements exist for the same speaker and suggest recording several intra-session vowel samples over several sessions to arrive at a typical perturbation value for a given speaker. Other methodological

studies were concerned with microphone and recorder types (Titze and Winholtz, 1993).

The objective of our article is to address a further methodological complication. Indeed, all the jitter features of which we are aware quantify the scatter of glottal cycle lengths about their average in one way or another. This is also true of features such as the period perturbation quotient, which summarises jitter with reference to a running average the purpose of which is to detrend cycle lengths, trends which otherwise would increase jitter artificially.

The assumption that jitter can be validly assimilated to scatter around an average cycle length implies that the following statistical model is true:

$$P_i = \langle P \rangle + Z_i. \tag{1}$$

P_i is the ith cycle length, $\langle P \rangle$ the average and Z_i the ith sample of a white noise process. This means the autocorrelation coefficients of series Z_i are not significantly different from zero. Model (1) implies that cycle perturbations $P_i - \langle P \rangle$ follow one another as if they had been randomly sampled from a population described by a gaussian distribution.

Model (1) may not be a valid representation of jitter when this is not true, i.e. when neighbouring cycle perturbations are interrelated. Conventional jitter features which assume model (1) to be true would consequently miss those components of jitter that are locally predictable due to correlations between adjacent cycle perturbations. In other words, conventional features only describe jitter exhaustively when the successive glottal cycle lengths can be mixed without changing their statistical properties. However, experiments that we carried out earlier, show glottal cycle lengths to be correlated (De Guchteneere and Schoentgen, 1991).

The handicap of a statistical measure of dispersion describing time series data is illustrated in Table 1, which presents two artificial series of numbers. The first sequence was obtained by perturbing an average value of 6.55 ms randomly while the second was created by perturbing the same average periodically. The calculations of the period perturbation quotient PPQ (formula 4) of both series yielded the same result, i.e. 4.09%. As a measure of dispersion, the PPQ was not able to discriminate between these two time series which were qualitatively different.

Table 1
Period perturbation quotient of randomly (I) and periodically (II) perturbed artificial time series. The time series average is 6.55 ms

	Data				
I. Random	6.53	6.97	6.66	6.99	6.3
PPQ = 4.09%	6.81	6.19	6.09	6.31	6.67
II. Deterministic	6.28	6.82	6.28	6.82	6.28
PPQ = 4.09%	6.82	6.28	6.82	6.28	6.82

In this article we study how predictable and random components of jitter can be separated and used to describe jitter in normal and dysphonic voices.

2. Time series models of jitter

The statistical processing of data which are produced sequentially and which cannot be scrambled without changing their statistical properties must be based on time series models. A time series is a sample of a sequence of random variables and a time series model a mathematical relation between present and past variables that captures the deterministic part of the changes over time of the otherwise random time series data.

The time series analysis of jitter therefore means that a time series model must be chosen and fitted to the observed cycle length perturbations, and that the model's parameters must be tested for statistical significance. The part of the time series data that is not represented by means of the model's coefficients must be described via the residue, which must be tested for "whiteness". Indeed when the residue is "white", i.e. decorrelated, the model is considered to be appropriate since a decorrelated residue indicates that the deterministic component of the data has been correctly represented by the time series model.

The linear auto-regressive model (2) is the most frequently used time series model,

$$P_i - \langle P \rangle = a_1(P_{i-1} - \langle P \rangle) + a_2(P_{i-2} - \langle P \rangle) + \cdots + a_M(P_{i-M} - \langle P \rangle) + Z_i. \tag{2}$$

Here, datum P_i is the ith glottal cycle length, $\langle P \rangle$ the average, coefficients a_j the auto-regressive model coefficients, number M the order, and Z_i the residue.

Once coefficients a_j have been estimated, the residue is calculated by subtracting the right-hand side of the expression from the left-hand side. If coefficients a_j were zero, model (2) would be identical to model (1). The choice of model (2) over other linear time series models is discussed in later sections.

The interpretation of expression (2) is that the present cycle length perturbation is a weighted sum of past perturbations plus a random quantity Z_i – the residue – which is characterizable via its statistical distribution. When model (2) has been successfully fitted to a sequence of cycle length perturbations residue Z_i then represents the purely random component of jitter; this component is best summarised by means of a feature which measures the Z_i's scatter around zero. Coefficients a_j represent the predictable component of jitter, i.e. the part which is normally lost by conventional features or left to bias them in unpredictable ways.

The purpose of what follows is to examine how far model (2) is an appropriate representation of jitter. It is important to note that time series modelling is not an option which may or may not be taken into account. From a statistical point of view at least, scatter is inadequate to represent non-independent data that are produced in a sequence, as is the case with jitter data. Statistical time series models offer to separate the purely random component from the deterministic one, which can be handled and examined separately.

3. Methods

3.1. Corpus

The corpus was comprised of 38 healthy French-speaking subjects (22 males and 16 females), and 45 dysphonic speakers (16 males and 29 females). Dysphonic subjects were patients in the ENT-department of the St Pierre University Hospital Brussels, Belgium. The healthy speakers were from the University's staff or were recruited by advertisements. The average ages were, respectively, 45.8, 49.1, 48.1 and 46.9 years for the healthy male, the healthy female, the dysphonic male and the dysphonic female speakers. The pathologies, diagnosed by the ENT-department doctors, fell under the following headings: œdema of the vocal folds, nodules, hypotonia and asthenia, pseudo-cysts, ulcers, granuloma, congestion of the vocal folds, polyps, chronic laryngitis, hyperkinesia and paralysis.

The recordings were carried out in an isolated booth via a Sennheiser electret microphone mounted on a head-set so as to keep the mouth-microphone distance constant. The electroglottographic signal was recorded by means of a Fourcin laryngograph. The signals were digitised at 48 kHz by means of a PCM Sony audio-processor and stored on video tape. Later, a stable portion of each vowel signal was selected and redigitised at 20 kHz at the resolution of 12 bits and stored on the disk of a Masscomp computer for further processing. Before resampling, the signals were filtered by means of an anti-aliasing filter with a cut-off frequency of 10 kHz.

Each speaker was instructed to sustain vowels [a], [i] and [u] at a comfortable loudness and pitch level and to avoid getting out of breath. This sequence of three vowels was repeated at least three times in the same order. One vowel segment of each quality was later selected for jitter measurement. The final corpus was therefore made up of 114 vowels (38 speakers, three vowel qualities) produced by the healthy, and 165 vowels produced by the dysphonic subjects (45 speakers, of whom 8 were recorded two or three times, giving a total of 55 recordings times three vowels).

3.2. Oversampling

Given the smallness of jitter, i.e. 0.1–1% of the average glottal cycle lengths, the speech signal had to be sampled at a high enough time resolution. Indeed, measuring jitter with a 10% degree of accuracy would require sampling frequencies between 100 kHz and 1 MHz. On the other hand, bandwidth appears not to be an important issue as far as jitter is concerned (Titze et al., 1987). We therefore decided to low-pass filter the signal at 10 kHz, to sample it at 20 kHz and to oversample it 8 times by means of interpolation, so producing a final time resolution of 6.25 μs.

The problem posed by interpolation was that the inserted samples were only estimates of in-between samples, the exact values of which were unknown.

The quality of the inserted samples was therefore evaluated by means of a statistical method proposed by Hess and Indefrey (1987). Basically, the method consists of analysing the time intervals between pitch markers (i.e. signal peaks or zero-crossings) and the previous recorded (i.e. not interpolated) sample of the speech signal. The statistical distribution of these time intervals must be uniform. Otherwise, since no relation exists between sampling frequency and pitch marker positions, the conclusion would be that the interpolated sample positions had been biased by the interpolation method. The uniformity of the distribution was tested by means of a χ^2-test (Press et al., 1992). Further technical details concerning this statistical quality control of oversampling can be found in (Hess and Indefrey, 1987; Schoentgen and De Guchteneere, 1991a, 1995).

A comparison of different interpolation methods is presented in (Schoentgen and De Guchteneere, 1991b). The comparison shows that parabolic interpolation or intercalation via Fourier transforming are less reliable than interpolation by means of a Finite Impulse Response filter as was suggested by Hess et al. and employed here.

3.3. Contamination by noise

Given the tininess of jitter, the risk that jitter measurements will be biased by aerodynamic, measurement or quantization noise is never nil. We therefore compared the glottal cycle lengths measured on the acoustic speech signal to lengths measured via the corresponding electroglottographic signal. Indeed, since both signals were physically very different, an agreement between the length measurements obtained from both signals was unlikely to be due to chance. Statistical comparisons were carried out via an inter-correlation analysis. A high inter-correlation between the length measurements of different origins indicated that they reflected the underlying glottal cycle lengths correctly. However, when the inter-correlation was small, an examination of the series was carried out in order to detect any outliers that could have adversely influenced the inter-correlation and the recordings were eliminated from the corpus when the low value of the inter-correlation could not be explained by means of isolated abnormal data (which could be corrected). The statistical significance of the inter-correlation coefficients was determined by means of a Student's t-test (Guilford, 1965).

3.4. Algorithm

The algorithm for measuring the glottal cycle lengths involved:

1. The reading from the computer disk of the 20 kHz-sampled acoustic and electroglottographic signals.
2. The low-pass filtering of the glottographic signal at a cut-off frequency included in the interval between 2.5 and 8 kHz followed by numerical differentiation (Demidovich and Maron, 1976); the low-pass filtering of the speech signal at a cut-off frequency included in the interval between 0.5 and 1.5 kHz.
3. The estimation of the average fundamental period by means of the autocorrelation function of the derivative of the glottographic signal; when the estimation failed, the re-estimation of the same quantity by means of the filtered acoustic signal.
4. The gross detection of the signal peaks of the glottographic signal by means of a method proposed by Davis (1976); the detection of the positive going zero-crossings of the acoustic signal.
5. The calculation of the gross cycle lengths by means of the period markers in the form of the signal zero-crossings or peaks; the display of both cycle length series and the optional return to step 4. Indeed, the visual detection of anomalous lengths more often than not pointed to erroneously inserted or omitted period markers. The marker detection program could then be relaunched with other detection criteria, i.e. the length of the search intervals, the search starting time coordinate, etc. This optional step was especially important in the case of the signals produced by the dysphonic speakers because the reliable detection of period markers in these voices was not totally automatizable.
6. The oversampling in the vicinity of the peaks of the glottograph derivative and the zero-crossings of the speech signal; the oversampling was the seven-fold recopy of the value of the last observed sample; the Finite Impulse Response filtering of the resulting "stair-case" signal; the detec-

tion of the signal peaks or zero-crossings of the oversampled and FIR-filtered signal.

7. The calculation of the cycle lengths by means of the finely positioned period markers; a check by means of the Hess's test of the quality of the interpolation; the calculation of the inter-correlation between the length time series obtained via the acoustic and glottographic signals.

3.5. Auto-regressive jitter time series model

Given that glottal cycle length perturbations are partly predictable from past perturbations, their statistical processing appeals to a time series model. As pointed out in the previous section, the linear auto-regressive model (2) is by far the most frequently used statistical time series model. The reasons for choosing this model are as follows.

(i) It can be experimentally shown that neighbouring cycle perturbations are correlated. We therefore assume that the present cycle perturbation (i.e. $P_i - \langle P \rangle$) is influenced by a linear combination of past perturbations.

(ii) The dependence of present on past perturbations via weights a_j is assumed to be linear, so that the weights can be computed by means of conventional linear methods for a given time series.

Weights a_j of auto-regressive model (2) are computed by solving the following system of algebraic equations which have the form expected of a linear regressive model when the present values are regressed on the past values of a sequence (Box and Jenkins, 1976).

$$\rho_i = a_1 \rho_{i-1} + a_2 \rho_{i-2} + \cdots + a_M \rho_{i-M}, \quad i = M+1, \ldots, N. \tag{3}$$

The unknowns are weights a_j; coefficients ρ_i are the autocorrelation coefficients of the observed time series; N is the total number of cycle length perturbations and M the order of model (2).

It can be shown that in the case of a linear auto-regressive model (2) of the order M, weight a_{M+1} is equal to zero and the residue is identical to a purely random process, i.e. no significant correlations exist between the shifted and unshifted series Z_j. In practice, an appropriate order M is determined, for an observed sequence of glottal cycle lengths, via the series' partial autocorrelation coefficients (Gilchrist, 1984). Intuitively speaking, the mth partial autocorrelation coefficient of a series is equal to weight a_M of an Mth order model (2) of the same series. Therefore, the index of the last partial autocorrelation coefficient that is significantly different from zero is the appropriate order for an auto-regressive model of the time series. The corresponding statistical test was the Student's t-test (Mélard, 1990).

3.6. Whiteness of residue Z_i

A check on the relevance of model (2) as a representation of an observed jitter time series is provided by the test of the whiteness of residue Z_i, which is obtained by subtracting the right-hand side from the left-hand side of expression (2). Model (2) must be considered to be appropriate when samples Z_i are uncorrelated, since all the predictable behaviour of the cycle perturbations has been absorbed by the linear combination, so leaving the purely random component to the residue. We made use of three tests for the randomness of samples Z_i, i.e. the Durbin–Watson test and two others which examined the residue's correlation coefficients both individually and globally (Mélard, 1990). The Durbin–Watson statistic is an expression which can only be meaningfully computed for time series data. Its value is comprised between 0 and 4, a value around 2 indicating the absence of correlation (Mélard, 1990).

3.7. Period perturbation quotients

One of the objectives of this article is to compare a conventional descriptor of jitter with a time series description. We chose the period perturbation quotient (4) as the conventional descriptor since it was formulated to deal with trends in the cycle length data and has been frequently used. Its flaw is that it is statistically biased, a bias which could be removed by dividing the local sum by four instead of five. But here we left it in its original form in order to facilitate comparisons with earlier data.

$$\mathrm{PPQ} = \frac{\frac{1}{N-4}\sum_{i=3}^{N-2}\left|\frac{1}{5}\sum_{j=1}^{5} P_{i+j-3} - P_i\right|}{\frac{1}{N}\sum_{i=1}^{N} P_i}. \tag{4}$$

N is the number of lengths and P_i the ith datum in the sequence of cycle lengths.

We have seen that residue Z_i is the purely random component of jitter from which any cycle-interdependencies have been removed. A relative, whitened jitter feature can therefore be defined as follows, in accordance with formula (4):

$$\mathrm{PPQw} = \frac{\frac{1}{N-M}\sum_{i=M+1}^{N}|Z_i|}{\frac{1}{N}\sum_{i=1}^{N}P_i}. \tag{5}$$

N is the number of cycles, M the order of auto-regressive model (2) and Z_i the ith datum of the residue.

Taking into account that $Z_i = P_i - a_1 P_{i-1} - a_2 P_{i-2} - \cdots$, expression (5) could be turned into expression (4) by choosing $M = 5$ and putting auto-regressive coefficients a_i equal to $1/5$ and shifting the local sum three samples to the right (i.e. into the future).

3.8. Data analysis

For data analysis we retained the following features: order M of jitter model (2), the first three auto-regressive coefficients a_i, the average fundamental frequency F_0 of the vocoid, the "whitened" period perturbation quotient PPQw, and the conventional period perturbation quotient PPQ. Data analysis was carried out by means of discriminant analysis. Discriminant analysis describes the link between a qualitative feature, with a finite number of modalities, and a set of quantitative descriptors observed in a number of individuals. The modalities of the qualitative feature are here "healthy" and "dysphonic". A discriminant function is a linear combination of the quantitative features, so that healthy and dysphonic individuals are separated in the best possible way (Romeder, 1973; SPSS Inc., 1992). Table 2 summarises the discriminant analyses carried out.

A discriminant analysis using acoustic features was performed to separate the speakers, who were a priori known to be either healthy or dysphonic, into two classes. The assumption was that the more relevant a feature was, the better it would be able to reconstruct the original a priori known classification of the speakers. The objective of the present study is not, therefore, to examine the classification performances of jitter features in a clinical or screening framework but to study a novel method and to compare it, within a corpus of normal and dysphonic speakers, to a conventional feature of jitter.

The statistical significance of the observed discrimination scores was tested as follows. Within the framework of discriminant analysis speakers, a priori known to be either dysphonic or healthy, were separated into two classes by means of acoustic features. The discrimination score was the sum of correctly classified healthy and dysphonic speakers divided by the total number of speakers. This score is statistically significant when it can be shown that it was not achieved by chance alone, i.e. the null hypothesis H_0 could be rejected that the underlying Bernoulli probabilities were the same for the healthy and dysphonic groups. Indeed, the division by means of acoustic features of a group of speakers into two classes, labelled "dysphonic" and "healthy", is a Bernoulli experiment with an underlying probability of their ending up in one class or the other. Of course, only when the inferred Bernoulli probability of being labelled "dysphonic" was different for the groups of dysphonic and healthy speakers could the acoustic features be considered to meaningfully represent the speaker's state of health.

Any row of any of the Tables 3–13 could be considered to be the outcome of a Bernoulli experiment. Any two rows could therefore be statistically compared by means of tests assessing the hypothesis that the observed classification percentages were not significantly different from each other, i.e. that the underlying Bernoulli probabilities were equal. We made use of two tests which evaluated this zero hypothesis at the 5% level, i.e. the risk was 1 in 20 of erroneously rejecting H_0. The tests were the Fisher–Irwin test and a test based on the normal approximation of the Bernoulli distribution (Ross, 1987). The outcomes of the two tests, which evaluated the same hypothesis H_0 but which were computationally very different, never disagreed as far as our data were concerned.

Whole tables were compared by means of the same tests by collapsing the two rows of each table into a single one by adding together the numbers of correctly and incorrectly classified items. In this way, inter-gender, inter-quality and inter-feature

Table 2
Summary of Tables 3 to 13. Symbol a_j is the linear auto-regressive model coefficients and M its order; PPQw the period perturbation quotient of the decorrelated jitter and PPQ of the raw jitter; F_0 is the fundamental frequency. Symbol PCC is the percentage of correctly classified items. An asterisk (*) marks a statistically not significant (at the 5% level) percentage of correctly classified items

Table index	Number of signals	Vocoid quality	Gender	Features	PCC
3	279	[a], [i], [u]	M, F	a_1, a_2, a_3, PPQw, F_0, M	63%
4	279	[a], [i], [u]	M, F	PPQw, M	63%
5	279	[a], [i], [u]	M, F	a_1, PPQw, M	65%
6	279	[a], [i], [u]	M, F	PPQ	51%
7	279	[a], [i], [u]	M, F	PPQw	54%
8	126	[a], [i], [u]	M	a_1, a_2, a_3, PPQw, M	64%
9	153	[a], [i], [u]	F	a_1, a_2, a_3, PPQw, M	59% (*)
10	93	[a]	M, F	a_1, a_2, a_3, PPQw, M	66%
11	93	[i]	M, F	a_1, a_2, a_3, PPQw, M	62%
12	93	[u]	M, F	a_1, a_2, a_3, PPQw, M	62%
13	279	depending on vote	M, F	PPQw, M	68%

comparisons were carried out by grouping together all the signals either produced by the male or female speakers, or corresponding to a given vowel quality, or described by a specified set of features.

4. Results

All the jitter time series retained for discriminant analysis had passed the following preliminary tests. Firstly, Hess's test for oversampling quality, secondly, the intercorrelation test for noise, thirdly, the Student's t-test relating to the order of model (2), and fourthly, three randomness tests of residue Z relating to the overall relevance of model (2). Figs. 1 and 2 show two examples of raw jitter time series together with their residues, i.e. their purely random components. The figures display two kinds of correlation. Fig. 1 shows a negative correlation, due to subharmonics, between adjacent cycles, and Fig. 2 a positive correlation due to local sequences of longer or shorter-than-average cycle lengths. These predictable behaviours have disappeared from the residues.

In order to evaluate the behaviour of the time series features characterising the jitter of vocoids sustained by healthy and dysphonic speakers we carried out discriminant analyses which grouped the jitter time series into two classes labelled "healthy" or "dysphonic". The features were the period perturbation quotient PPQw of residue Z_i, model order M, the first three auto-regressive coefficients a_1, a_2, a_3 and fundamental frequency F_0, or a sub-set. We also performed discriminant analyses by means of conventional period perturbation quotient PPQ calcu-

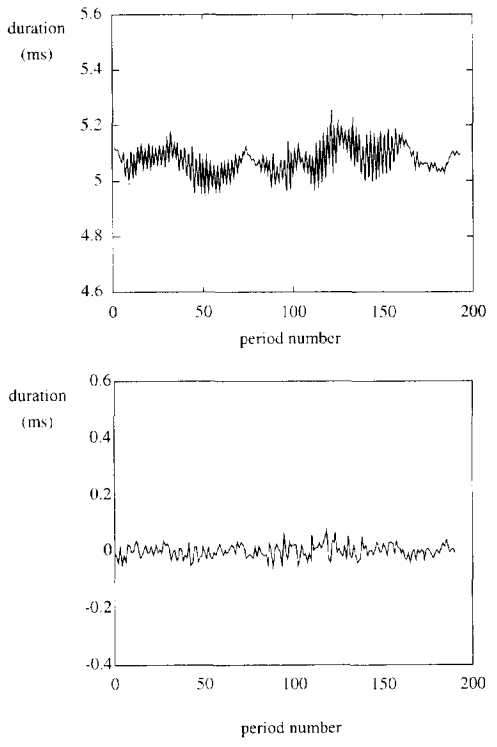

Fig. 1. Raw jitter time series with negative correlations between adjacent perturbations (above) and corresponding decorrelated jitter time series (below).

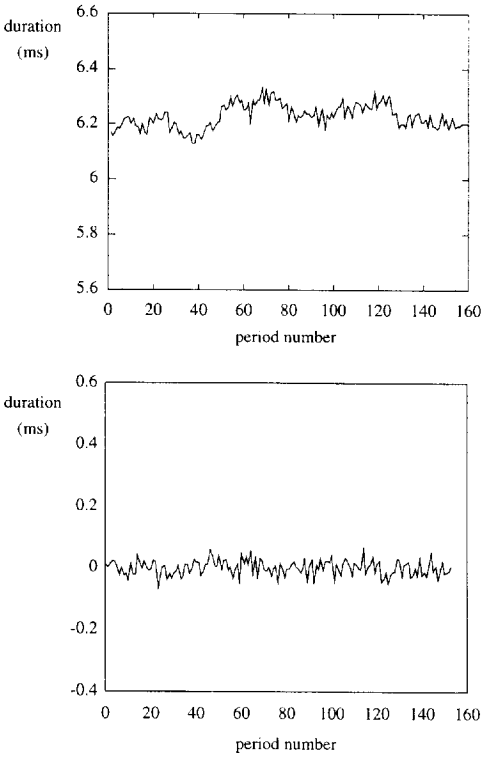

Fig. 2. Raw jitter time series with positive correlations between adjacent perturbations (above) and corresponding decorrelated jitter time series (below).

lated from raw, unprocessed jitter time series in order to compare its discriminatory performance to that achieved by features arrived at via time series modeling.

Tables 3–12 summarise the results obtained. Each table gives the following information: firstly, the features included in the analysis; secondly, the numbers and percentage of healthy and dysphonic voices correctly and incorrectly classified; thirdly, the correlation between the features and the discriminant function, correlation which is proportional to the contribution of each feature to the overall discrimination performance. A value close to one indicates that the corresponding feature had much influence on the classification result, and a value close to zero the opposite. The signs of the correlation coefficients are relative only. Two correlation coefficients with opposite signs mean that the corresponding features increased and decreased with the state of health of the speakers. Table 2 is a summary of Tables 3–12. It gives the following information: table indices, features, vowel qualities, the number of signals examined, speaker gender and the overall discrimination score in %. The discrimination score marked with an (*) was not significant at the 5% level, i.e. the null hypothesis of the chance occurrence of the differences between the percentages of correctly classified normal and dysphonic speakers could not be rejected at that level. Or, in other words, the hypothesis could not be rejected that the underlying Bernoulli probabilities of ending up in the healthy or dysphonic class were the same for the healthy and dysphonic female speakers.

Tables 3–7 give the results of discriminant analyses of jitter time series (three series per speaker) via different groups of features. As far as our corpora were concerned, discrimination scores peaked around 65%. Among the groups of features, whitened period perturbation quotient PPQw and model order M appeared to be the most discriminatory (Table 4). They characterised the random and deterministic components of jitter, respectively. Adding model coefficient a_1 to this group increased the separation performance by 2% (Table 5).

The pre-eminence of order M and "whitened" perturbation quotient PPQw was also shown via a stepwise analysis, the principle of which was to introduce features one by one according to a selec-

Table 3
Discrimination scores and correlations between discriminant function and features. A high positive or negative correlation indicates that the corresponding feature contributed greatly to the discrimination scores. The experimental conditions and percentage of correctly classified items are summarized in Table 2

Actual classification	Total	Predicted classification	
		dysphonic	healthy
dysphonic	165	83 (50.3%)	82 (49.7%)
healthy	114	20 (17.5%)	94 (82.5%)

Feature	Correlation
a_1	−0.69149
PPQw	0.68854
order	0.68000
F_0	0.40763
a_2	0.36726
a_3	0.26054

Table 4
Discrimination scores and correlations between discriminant function and features. A high positive or negative correlation indicates that the corresponding feature contributed greatly to the discrimination scores. The experimental conditions and percentage of correctly classified items are summarized in Table 2

Actual classification	Total	Predicted classification	
		dysphonic	healthy
dysphonic	165	84 (50.9%)	81 (49.1%)
healthy	114	22 (19.3%)	92 (80.7%)

Feature	Correlation
PPQw	0.72496
Order	0.71597

tion criterion. The criterion was the Wilks statistic, which measures a single feature's intra-group versus its inter-group variability. At each step the feature, the Wilks statistic of which was the smallest, was added to the analysis. The stepwise introduction of new features stopped when the discrimination scores no longer increased anymore. The features so obtained can be considered to be the most discriminatory. In our case, stepwise analysis retained only model order M and "whitened" perturbation quotient PPQw from among the initial six features.

Fig. 3 displays the discriminant function that underlies the results obtained by means of order M and quotient PPQw (Table 4). The signals emitted by the healthy speakers cluster in the lower left-hand corner of the display where order M and perturbation quotient values are low.

Table 5
Discrimination scores and correlations between discriminant function and features. A high positive or negative correlation indicates that the corresponding feature contributed greatly to the discrimination scores. The experimental conditions and percentage of correctly classified items are summarized in Table 2

Actual classification	Total	Predicted classification	
		dysphonic	healthy
dysphonic	165	86 (52.1%)	79 (47.9%)
healthy	114	19 (16.7%)	95 (83.3%)

Feature	Correlation
a_1	−0.70429
PPQw	0.70129
order	0.69259

Table 6
Discrimination scores and correlations between discriminant function and features. A high positive or negative correlation indicates that the corresponding feature contributed greatly to the discrimination scores. The experimental conditions and percentage of correctly classified items are summarized in Table 2

Actual classification	Total	Predicted classification	
		dysphonic	healthy
dysphonic	165	35 (21.2%)	130 (78.8%)
healthy	114	6 (5.3%)	108 (94.7%)

Feature	Correlation
PPQ	1.00000

Finally, Tables 6 and 7 compare the performance of the gross (4) and whitened period perturbation quotients (5). The difference in performance was 3%, but this difference was not statistically significant at the 5% level. Therefore, it was impossible to conclude that decorrelated jitter alone would perform better than gross jitter on any statistically representative corpus of the signal population that we examined. But for the healthy and dysphonic speakers taken separately, the differences between quotients PPQ and PPQw were significant. Indeed, the overall increase of 3%, which was not statistically significant, was the outcome of quotient PPQw performing

Table 7
Discrimination scores and correlations between discriminant function and features. A high positive or negative correlation indicates that the corresponding feature contributed greatly to the discrimination scores. The experimental conditions and percentage of correctly classified items are summarized in Table 2

Actual classification	Total	Predicted classification	
		dysphonic	healthy
dysphonic	165	52 (31.5%)	113 (68.5%)
healthy	114	15 (13.2%)	99 (86.8%)

Feature	Correlation
PPQw	1.00000

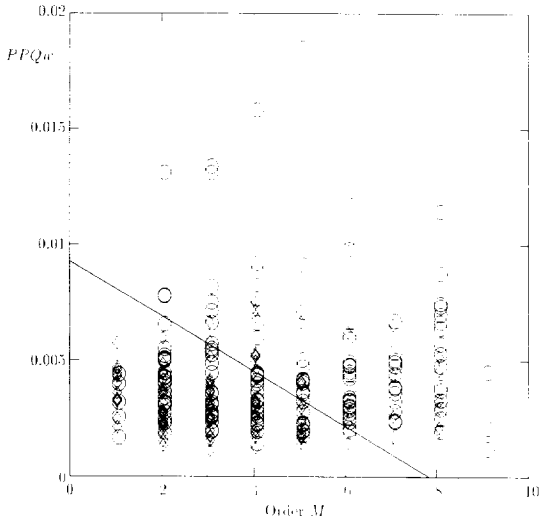

Fig. 3. Discriminant function obtained by means of the discriminant analysis of quotient PPQw of the decorrelated jitter and order M of the linear auto-regressive model obtained from healthy and dysphonic speakers sustaining vocoids [a] [i] [u]; ◇: time series produced by healthy speakers, ○: time series produced by dysphonic speakers. Healthy speakers cluster in the lower left-hand corner where order M and quotient PPQw are small.

significantly better on the dysphonic, and significantly worse on the normal speakers.

However, the classification performances of order M, quotient PPQw and coefficient a_1 together (Table 5) were significantly different from the scores of quotients PPQ or PPQw alone whatever the group of signals, i.e. healthy, dysphonic or combined.

In order to check on a possible uneven influence of vowel quality or speaker gender on the discrimination performances, we carried out a separate analysis for each vowel quality and for each gender (Tables 8–12).

The results of these analyses are given in Tables 10–12 for the vowel qualities. It will be seen that

Table 8
Discrimination scores and correlations between discriminant function and features. A high positive or negative correlation indicates that the corresponding feature contributed greatly to the discrimination scores. The experimental conditions and percentage of correctly classified items are summarized in Table 2

Actual classification	Total	Predicted classification	
		dysphonic	healthy
dysphonic	60	28 (46.7%)	32 (53.3%)
healthy	66	13 (19.7%)	53 (80.3%)

Feature	Correlation
a_1	−0.74064
PPQw	0.61858
a_2	0.61671
order	0.53536
a_3	0.08036

Table 9
Discrimination scores and correlations between discriminant function and features. A high positive or negative correlation indicates that the corresponding feature contributed greatly to the discrimination scores. The experimental conditions and percentage of correctly classified items are summarized in Table 2

Actual classification	Total	Predicted classification	
		dysphonic	healthy
dysphonic	105	58 (55.2%)	47 (44.8%)
healthy	48	15 (31.3%)	33 (68.8%)

Feature	Correlation
PPQw	0.74886
order	0.58697
a_1	−0.52083
a_3	0.23306
a_2	−0.16940

Table 10
Discrimination scores and correlations between discriminant function and features. A high positive or negative correlation indicates that the corresponding feature contributed greatly to the discrimination scores. The experimental conditions and percentage of correctly classified items are summarized in Table 2

Actual classification	Total	Predicted classification	
		dysphonic	healthy
dysphonic	55	29 (52.7%)	26 (47.3%)
healthy	38	6 (15.8%)	32 (84.2%)

Feature	Correlation
a_1	−0.74428
order	0.69997
PPQw	0.69738
a_3	0.46684
a_2	0.16592

Table 11
Discrimination scores and correlations between discriminant function and features. A high positive or negative correlation indicates that the corresponding feature contributed greatly to the discrimination scores. The experimental conditions and percentage of correctly classified items are summarized in Table 2

Actual classification	Total	Predicted classification	
		dysphonic	healthy
dysphonic	55	26 (47.3%)	29 (52.7%)
healthy	38	6 (15.8%)	32 (84.2%)

Feature	Correlation
order	0.67995
PPQw	0.60745
a_1	−0.50698
a_2	0.27997
a_3	0.19440

discriminatory performance varied only slightly with vocoid quality: the discrimination score was slightly higher for vowel [a] than for vowels [i] and [u]. But inter-vocoid differences were not significant at the 5% level whatever the vowel quality or speaker group (dysphonic, healthy or combined).

Another manner of indirectly examining the possible influence of vowel quality on the discrimination scores was to determine the class of each speaker via a vote based on the three vowel qualities: a speaker was classed among the dysphonics if a majority (i.e. at least 2 out of 3) of the vowel signals were classed as dysphonic. The results of the vote are given in Table 13 when order M and whitened perturbation PPQw were taken into account. The percentage of correctly grouped speakers is 68%. In other words, the voting did not improve the discrimination score. Only the classification of the healthy speakers was slightly more precise (89% versus 81% without a vote). The classification performance obtained for dysphonic speakers remained the same. No vocoid could therefore be considered to be better at making distinctions between normal and dysphonic speakers. Indeed, had one of the qualities have systematically led to a better- or worse-than-average discrimination score, discrimination by vote would have achieved better marks since the least discriminatory vowel quality would have been ignored. This was not the case.

Table 12
Discrimination scores and correlations between discriminant function and features. A high positive or negative correlation indicates that the corresponding feature contributed greatly to the discrimination scores. The experimental conditions and percentage of correctly classified items are summarized in Table 2

Actual classification	Total	Predicted classification	
		dysphonic	healthy
dysphonic	55	28 (50.9%)	27 (49.1%)
healthy	38	8 (21.1%)	30 (78.9%)

Feature	Correlation
a_1	−0.79096
PPQw	0.73107
order	0.61396
a_3	0.43692
a_2	0.35173

For the sexes, the analysis results are given in Tables 8 and 9. The discrimination scores between the normal and dysphonic jitter time series of the male and female speakers were 64% and 59%. The former score was significant at the 5% level while the latter was not. But the differences between the sexes were not significant at the 5% level whatever the speaker group: dysphonics, healthy or combined. The group coming closest to significance was the healthy group. The differences in the scores attained by the male and female speakers can presumably be traced to the different ratios of healthy and dysphonic speakers in the male and female corpora (cf. Section 5).

A last point is that the experimental results given in Tables 3–9 were arrived at by means of a corpus which, for each speaker, contained three jitter time

Table 13
Discrimination scores and correlations between discriminant function and features. The experimental conditions and percentage of correctly classified items are summarized in Table 2

Actual classification	Total	Predicted classification	
		dysphonic	healthy
dysphonic speakers	55	29 (52.7%)	26 (47.3%)
healthy speakers	38	4 (10.6%)	34 (89.4%)

series (one per vowel quality) obtained at the same recording session. A posteriori, it was found that inter-quality differences were negligible. Consequently, it might be argued that, for a given speaker, the perturbation time series were not independent and that statistical significance had been subsequently boosted artificially. We therefore carried out the same tests again with one third of the time series, i.e. one time series per speaker instead of three. The results were as follows. Of course, the non-significant results remained so. All the significant discrimination scores in Table 2 remained significant, as did the difference between raw perturbation quotient PPQ and combined features PPQw, M and a_1. Only the formerly significant differences between quotients PPQ and PPQw vanished.

5. Discussion

As far as we know, the problem of the statistical interdependence of the magnitudes of glottal cycle perturbations has not yet been quantitatively dealt with in the literature. Perturbation measures such as the Relative Average Perturbation or the Period Perturbation Quotient only take into account the dispersion of cycle durations around a global or running average computed over a few cycles. The running average's role is to distinguish short-term perturbations from long-term "perturbations" due to a non-flat intonation contour, for example.

Generally speaking, three categories of linear time series models were available to represent jitter data. These were the linear auto-regressive, the moving average and the mixed auto-regressive and moving average models (Gilchrist, 1984). The choice of the auto-regressive model was made on the basis of a rule which states that the data are best described by means of a linear auto-regressive model when the autocorrelation function of a time series is significantly different from zero for many lags, whereas its partial autocorrelation coefficients go to zero rapidly. When autocorrelation and partial autocorrelation coefficients behave in the opposite way, time series data are best described via a moving average model. On the basis of our observation of these two kinds of correlation coefficients we decided to model jitter via a linear auto-regressive model. Indeed, the results that we present elsewhere show that our jitter time series could be described by means of auto-regressive models with orders within the range of 1 to 9 with a maximum at 3 (Schoentgen and De Guchteneere, 1995). This shows that jitter time series can in fact be rendered by linear auto-regressive models. Otherwise, the experimentally determined model orders would have been greater.

Representing jitter by means of auto-regressive model (2) circumvents the choice of an arbitrary smoothing window to carry out running averages since the selection of the "best" auto-regressive model order is carried out for each jitter sequence individually. The somewhat wayward distinction between random short-term perturbations and the running average is replaced by a mathematical separation between fluctuations that are purely random – the residue of model (2) – and a deterministic component – the weighted sum of the past perturbations in formula (2). Statistically speaking, a non-flat intonation contour would manifest itself as a time series trend. A strong trend would lead to model residues that could not be decorrelated even with a high-order model. If this was the case, detrending would have to be carried out beforehand by differentiating the time series (Pinto and Titze, 1990).

After constructing an AR-model for each sequence of cycle length perturbations we carried out a linear discriminant analysis in order to examine the model's ability to distinguish between normal and dysphonic voices and to compare its discriminatory performance with the conventional period perturbation quotient within the same corpus of normal and dysphonic voices. To repeat what we underline elsewhere in the text, the purpose of the present study is not to predict the discrimination performance of jitter features within the framework of a screening task, for instance, but to study the behaviour, via a reasonable corpus, of features separately describing the deterministic and the purely random components of jitter, i.e. coefficients a_j and order M on the one hand, and the perturbations of decorrelated residue Z_i on the other.

A first analysis took into account model order M, the first three auto-regressive coefficients a_j, whitened perturbation PPQw and average fundamental frequency F_0. The percentage of correctly classified vocoids was 63.4%. The relevance of this per-

centage will be discussed below. The three factors that contributed most and in similar proportions to the discrimination performance were first model coefficient a_1, order M and quotient PPQw. Alone, model order M and period perturbation quotient PPQw achieved almost the same score as when combined with coefficient a_1 (Tables 3–5).

The correlation of individual features with the discriminant function is an indicator of their discriminatory power. It emerges that the correlations of model order M and model coefficient a_1 had opposite signs; this was generally true. The explanation is as follows. The sum of the coefficients of the linear auto-regressive model of a stationary auto-regressive process of the same order must be less than 1. Thus, when the model order (i.e. the number of coefficients) increased, the values of the individual coefficients decreased accordingly.

Let us consider the discrimination scores which were 65% at best. If clinical tasks had been the objective of our endeavours (which they were not), the following comments would have been in order. In the field of medical application such as screening, for example, the prevalence of the dysphonic speakers, i.e. their numbers in the population as a whole would have to be taken into account. Our corpus contains more dysphonic than normal speakers, which is contrary to what would be expected in the general population, where only a few percent belong to the dysphonic group.

The jargon of medical statistics defines the percentage of dysphonic speakers or signals classed as such as the sensitivity of an analysis and the percentage of healthy speakers or normal signals classed as such as its specificity. The expected percentage of correctly classified speakers (PCC) would therefore have been equal to the following expression:

$$\mathrm{PCC} = \mathrm{prevalence} \times \mathrm{sensitivity} + (100 - \mathrm{prevalence}) \times \mathrm{specificity}.$$

Given a specificity of 82% and a sensitivity of 50% (Table 3) and assuming a prevalence of 2%, the percentage of correctly classified speakers would have been 81%. But, the percentage of correctly classified dysphonic speakers would still be around 50% (Tables 3–13). In other words, given the assumption of the representativity of our corpora, as far as the mix of pathologies is concerned, the jitter-related features appear to be reasonably specific but only moderately sensitive. It could be that moderate sensitivity was related to the relative rarity in our corpus and, presumably, in the population as a whole, of pathologies leading to asymmetries of the masses or tensions of the right and left vocal folds. Those pathologies are believed to induce higher than normal jitter given the irregular vibratory pattern which is sometimes the consequence of such a condition (Isshiki, 1972; Ishizaka and Isshiki, 1976).

Prevalence also helps explain the difference in the discrimination scores of male and female subjects (Tables 8 and 9). The higher proportion of dysphonic to normal female speakers in our corpus appears to bias the discrimination score. The recalculation of the percentage of correct classification by means of "male" proportions of 0.52 and 0.48, for example, increases the female PCC from 59% to 62% ($0.48 \times 55.2 + 0.52 \times 68.8$).

Be that as it may, the purpose of the present study was not to examine the feasibility of screening, but to compare the intra-corpus performance of new and conventional methods of characterising jitter. A possible explanation of the absence of inter-vowel differences is discussed in (Schoentgen and De Guchteneere, 1995).

Results show that besides the gain in interpretability related to the statistically correct processing of period sequences, sensitivity was increased and specificity slightly decreased as a consequence of the possibility of taking the deterministic component of jitter into account (Table 5 versus Table 6). Indeed, in our corpus, the disappointingly low sensitivity (21%) of raw feature PPQ and its almost perfect specificity (95%) was due to the presence of quite a few jitter time series with negative correlations which artificially boosted the raw PPQ. As a consequence, only a minority of pathological time series were detected and almost no normal speaker misclassified (Table 6). Removing negative correlations from the time series increased the sensitivity of perturbation factor PPQw to 32% and lowered specificity to 87% (Table 7). Adding features a_1 and M finally increased specificity to 52% and lowered sensitivity from 87% to 83% (Table 5).

Typically, half of the dysphonic speakers were correctly classified. This percentage is presumably representative of the clientele of the ENT-department

where we made our recordings. Another mix of dysphonics would have given rise to other discrimination scores. One possibility for examining the practical relevance of the statistically significant differences between conventional feature PPQ and time series features would be to record homogeneous corpora of dysphonic voices and test if pathologies exist that are revealed by deterministic component M or a_1 of the time series model and not by features of random component Z_i.

For example we separately analysed eight cases of nodules giving rise to 24 jitter time series. The results were as follows. Whitened perturbation quotient PPQw alone achieved an overall discrimination score of 50%. The perturbations of dysphonic and healthy speakers were statistically comparable. But, order M alone achieved a discrimination score of 64% since a trend existed towards higher orders for the cases of nodules. Both scores, though, were not statistically significant given the low number of signals.

Another possible application is as follows. Some authors have speculated that, at least in some cases, the observed cycle length perturbations might be chaotic. This would mean that the perturbations were deterministic, but not predictable. This issue is difficult to settle experimentally since the signature of chaos is difficult to find in relatively short (i.e. a few hundred cycles) and noisy time series. Linear correlations, which are accounted for by model (2), are not a signature of chaos. A possible angle would therefore be to remove any linear correlations from the jitter time series, and to search in a whitened jitter time series Z_i for any remaining, unexplained relations between neighbouring perturbations. If no such relations can be found, it might be concluded that the perturbations are not chaotic in origin since chaos would necessarily require the presence of nonlinear relations between perturbations. Using several methods, we therefore searched in residue Z for any remaining relations between neighbouring perturbations unaccounted for by model (2). In the great majority of our time series no remaining relations could be found; for several of the series abnormal relations were due to outliers, i.e. locally aberrant cycle lengths, and a very few nonlinear relations only could not be explained by either (Schoentgen and De Guchteneere, 1996). It may therefore be provisionally concluded that chaotic vocal fold vibrations are, if they exist, rare or confined to special pathological conditions, a conclusion that is in keeping with the results of simulation studies which show that chaotic vibrations go together with biologically unlikely values of vocal fold control parameters (Awrejcewicz, 1991; Wong et al., 1991).

6. Conclusion

Linear autoregressive time series models were fitted to jitter time series. The purpose was to separately represent random fluctuations and relations between near cycle length perturbations. Negative correlations due to sub-harmonics and positive correlations due to local sequences of longer (or shorter) than average perturbations could in this way be removed from the jitter time series and examined apart. The most relevant features were the decorrelated jitter quotient and the orders of the time series models, i.e. the number of past perturbations that influence present perturbations. Comparisons showed that the decorrelated jitter, normalised by the average cycle length, and the model order were not significantly different for male and female speakers or distinct vowel qualities. Dysphonic speakers, however, tended to produce larger decorrelated jitter and larger model orders than normal speakers.

Acknowledgements

We thank Professor Denis Hennebert for permitting the recordings of dysphonic voices in his ENT-department and Professor Guy Mélard for his advice concerning the statistical issues.

References

A. Askenfeld and B. Hammarberg (1980), "Speech waveform perturbation analysis", *Speech Transmission Laboratory – Quarterly Progress and Status Report*, Vol. 4, pp. 40–49.

A. Askenfeld and B. Hammarberg (1981), "Speech waveform perturbation analysis revisited", *Speech Transmission Laboratory – Quarterly Progress and Status Report*, Vol. 4, pp. 49–68.

A. Askenfeld, J. Gauffin, J. Sundberg and P. Kitzing (1980), "A comparison of contact microphone and electroglottograph for the measurement of vocal fundamental frequency", *J. Speech Hearing Res.*, Vol. 23, pp. 258–273.

J. Awrejcewicz (1991), *Bifurcation and Chaos in Coupled Oscillators* (World Scientific, Singapore).

T. Baer, A. Löfqvist and N.S. McGarr (1983), "Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques", *J. Acoust. Soc. Amer.*, Vol. 73, pp. 1304–1308.

G.E.P. Box and G.M. Jenkins (1976), *Time Series Analysis Forecasting and Control* (Holden-Day, San Francisco).

N.B. Cox and M.R. Ito (1989), "Quantization and measurement errors in the analysis of short-time perturbations in sampled data", *J. Acoust. Soc. Amer.*, Vol. 86, pp. 42–54.

S.B. Davis (1976), *Computer Evaluation of Laryngeal Pathology based on Inverse Filtering of Speech*, SCRL Monograph, Vol. 13 (Santa Barbara, California).

S.B. Davis (1979), "Acoustic characteristics of normal and pathological voices", in: *Speech and Language: Advances in Basic Research and Practice*, Vol. 1 (Academic Press, New York), pp. 271–335.

R. De Guchteneere and J. Schoentgen (1991), "Mean-term perturbations of the pseudo-period of the glottal waveform", in: *Proc. 12th Internat. Congress of Phonetic Sciences*, Aix-en-Provence, 19–24 August 1991, pp. 354–357.

G. De Krom (1994), Acoustic correlates of breathiness and roughness experiments on voice quality, OTS Dissertation series, Onderzoeksinstituut voor Taal en Spraak, Utrecht,.

J.F. Deem, W.H. Manning, J.V. Knack and J.S. Matesich (1989), "The automatic extraction of pitch perturbation using microcomputers: Some methodological considerations", *J. Speech Hearing Res.*, Vol. 32, pp. 689–697.

B.P. Demidovich and I.A. Maron (1976), *Computational Mathematics* (Mir Publishers, Moscow).

K.L. Garret and E.C. Healey (1987), "An acoustic analysis of fluctuations in the voices of normal adult speakers across three times of day", *J. Acoust. Soc. Amer.*, Vol. 82, pp. 58–62.

W. Gilchrist (1984), *Statistical Modelling* (Wiley, New York).

B. Gold and L. Rabiner (1969), "Parallel processing techniques for estimating pitch periods of speech in the time domain", *J. Acoust. Soc. Amer.*, Vol. 46, pp. 442–448.

R. Gubrynowicz, W. Mikiel and P. Zarnecki (1977), "Evaluation de l'état pathologique des cordes vocales d'après l'analyse des variations du fondamental", in: *Actes des 8es Journées d'Etude sur la Parole*, Aix-en-Provence, 25–27 May 1977, pp. 21–27.

R. Gubrynowicz, W. Mikiel and P. Zarnecki (1980), "An acoustic method for the evaluation of the state of the larynx source in cases involving pathological changes", *Archives of Acoustics*, Vol. 5, No. 1, pp. 3–30.

J.P. Guilford (1965), *Fundamental statistics in Psychology and Education* (McGraw-Hill, New York) 4th edition.

M.H.L. Hecker and E.J. Kreul (1971), "Descriptions of the speech of patients with cancer of the vocal folds; Part 1: Measures of fundamental frequency", *J. Acoust. Soc. Amer.*, Vol. 49, pp. 1275–1282.

V.L. Heiberger and Y. Horii (1982), "Jitter and shimmer in sustained phonation", in: *Speech and Language: Advances in Basic Research and Practice*, Vol 7 (Academic Press, New York), pp. 299–332.

W. Hess and H. Indefrey (1987), "Accurate time-domain pitch determination of speech signal by means of a laryngograph", *Speech Communication*, Vol. 6, No. 1, pp. 55–68.

M.B. Higgins and J.H. Saxman (1989a), "A comparison of intrasubject variation across sessions of three vocal frequency perturbations indices", *J. Acoust. Soc. Amer.*, Vol. 86, pp. 911–916.

M.B. Higgins and J.H. Saxman (1989b), "Variations in vocal frequency perturbation across the menstrual cycle", *J. Voice*, Vol. 3, pp. 233–243.

J. Hillenbrand (1988), "Perception of aperiodicities in synthetically generated voices", *J. Acoust. Soc. Amer.*, Vol. 83, pp. 2361–2371.

H. Hollien, J. Michel and E.T. Doherty (1973), "A method for analyzing vocal jitter in sustained phonation", *J. Phonetics*, Vol. 1, pp. 85–91.

S. Horiguchi, T. Haji, T. Baer and W.J. Gould (1987), "Comparison of electroglottographic and acoustic waveform perturbation measures", in: T. Baer, C. Sasaki and K. Harris, eds., *Laryngeal Function in Phonation and Respiration* (College-Hill Publication, Boston), pp. 509–518.

Y. Horii (1979), "Fundamental frequency perturbation observed in sustained phonation", *J. Speech Hearing Res.*, Vol. 22, pp. 5–19.

Y. Horii (1982), "Jitter and shimmer differences among sustained vowel phonations", *J. Speech Hearing Res.*, Vol. 25, pp. 12–14.

S. Imaizumi (1986), "Acoustic measures of roughness in pathological voice", *J. Phonetics*, Vol. 14, pp. 457–462

K. Ishizaka and N. Isshiki (1976), "Computer simulation of pathological vocal-cord vibration", *J. Acoust. Soc. Amer.*, Vol. 60, pp. 1193–1198.

N. Isshiki (1972), "Imbalance of the vocal cords as a factor for dysphonia", *Studia Phonologica*, Vol. 6, pp. 38–44.

K. Kitajima, M. Tanabe and N. Isshiki (1975), "Pitch perturbation in normal and pathologic voice", *Studia Phonologica*, Vol. 9, pp. 25–32.

Y. Koike (1973), "Application of some acoustic measures for the evaluation of laryngeal dysfunction", *Studia Phonologica*, Vol. 1, pp. 17–23.

Y. Koike and H. Takahashi (1972), "Glottal parameters and some acoustic measures in patients with laryngeal pathology", *Studia Phonologica*, Vol. 6, pp. 45–50.

Y. Koike, H. Takahashi and T.C. Calcatera (1977), "Acoustic measures for detecting laryngeal pathology", *Acta Otolaryngologica*, Vol. 84, pp. 105–117.

J. Laver, S. Hiller and R. Hanson (1982), "Comparative performance of pitch detection algorithm on dysphonic voices", in: *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Paris, 3–5 May 1982, pp. 192–195.

J. Laver, S. Hiller and J. Mackenzie (1984), "Acoustic analysis of vocal fold pathology", *Proc. Institute of Acoustics*, Vol. 6, pp. 425–430.

J. Laver, S. Hiller, J. Mackenzie and E. Rooney (1986), "An acoustic screening system for the detection of laryngeal pathology", *J. Phonetics*, Vol. 14, pp. 517–524.

P. Lieberman (1963), "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges", *J. Acoust. Soc. Amer.*, Vol. 35, pp. 344–353.

C.L. Ludlow, C.J. Bassich, N.P. Connor, D.C. Coulter and Y.J. Lee (1987), "The validity of using phonatory jitter and shimmer to detect laryngeal pathology", in: T. Baer, C. Sasaki and K. Harris, eds., *Laryngeal Function in Phonation and Respiration* (College-Hill Publication, Boston), pp. 492–508.

G. Mélard (1990), *Méthodes de Prévision à Court Terme* (Editions de l'Université Libre de Bruxelles, Bruxelles).

N. Pinto and I. Titze (1990), "Unification of perturbation measures in speech signals", *J. Acoust. Soc. Amer.*, Vol. 87, pp. 1278–1289.

W.H. Press, S.A. Teukolsky, W.T. Vetterling and B. Flannery (1992), *Numerical Recipes in C* (Cambridge University Press, Cambridge, 2nd ed.).

J.M. Romeder (1973), *Méthodes et Programmes d'Analyse Discriminante* (Dunod, Paris).

S. Ross (1987), *Introduction to Probability and Statistics for Engineers and Scientists* (Wiley, New York), pp. 230–234.

J. Schoentgen (1989), "Jitter in sustained vowels and isolated sentences produced by dysphonic speakers", *Speech Communication*, Vol. 8, No. 1, pp. 61–79.

J. Schoentgen and R. De Guchteneere (1991a), "An algorithm for the measurement of jitter", *Speech Communication*, Vol. 10, Nos. 5–6, pp. 533–538.

J. Schoentgen and R. De Guchteneere (1991b), "Analyse des microfluctuations des cycles glottiques", in: *Actes du Séminaire Traitement et Représentation du Signal de Parole*, Le Mans, pp. 108–111.

J. Schoentgen and R. De Guchteneere (1995), "Time series analysis of jitter", *J. Phonetics*, Vol. 23, Nos. 1/2, pp. 189–201.

J. Schoentgen and R. De Guchteneere (1996), "Searching for nonlinear relations in whitened jitter time series", in: *Proc. Internat. Conf. on Spoken Language Processing*, Philadelphia, 3–6 October 1996, pp. 753–756.

B.E. Smith, B. Weinberg, L.L. Feth and Y. Horii (1978), "Vocal roughness and jitter characteristics of vowels produced by œsophageal speakers", *J. Speech Hearing Res.*, Vol. 21, pp. 240–249.

SPSS Inc. (1992), *SPSS for Windows Advanced Statistics* (SPSS Inc., New York).

I.R. Titze and R. Winholtz (1993), "Effect of microphone and placement on voice perturbations measures", *J. Speech Hearing Res.*, Vol. 36, pp. 1177–1190.

I.R. Titze, Y. Horii and R.C. Scherer (1987), "Some technical considerations in voice perturbation measurements", *J. Speech Hearing Res.*, Vol. 30, pp. 252–260.

D. Wong, M.R. Ito, N.B. Cox and I.R. Titze (1991), "Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases", *J. Acoust. Soc. Amer.*, Vol. 89, pp. 383–394.