Journal of Phonetics (1995) 23, 189–201

Time series analysis of jitter

Jean Schoentgen* and Raoul de Guchteneere

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

Received 17th September 1993, and in revised form 11th July, 1994

This article concerns the time series analysis of jitter. Jitter involves small fluctuations in glottal cycle lengths. Time series analysis is the statistical processing of data that are recorded over a period of time. Conventionally, the amount of jitter in an analysis interval is estimated by a measure of dispersion of the glottal cycle lengths. The problem is that measures of dispersion only describe the fluctuations in the cycle lengths unambiguously when these are statistically independent. This means that the fluctuations are white noise and that changing the order of the cycles does not change their statistical properties. But it can be shown experimentally that neighbouring cycle lengths are not statistically independent because they are correlated. We therefore studied jitter by means of time series analysis methods. These dispense with the assumption that glottal cycle lengths are statistically independent. They make it possible to distinguish between mean- and short-term perturbations and to remove correlations between neighbouring perturbations. We studied dispersion measures of raw and whitened jitter (i.e., jitter from which correlations had been removed). Jitter time series were obtained from vowels [a], [i], [u] sustained by male and female healthy and dysphonic speakers. Results showed that the inter-speaker differences were smaller for whitened jitter than for raw jitter. Inter-speaker variability was reduced because time series analysis separated random from non-random perturbations.

1. Introduction

This article concerns the time series analysis of jitter. Jitter involves small fluctuations of the glottal cycle lengths. Time series analysis is the statistical processing of data that are recorded over a period of time.

Lieberman (1963) was the first to compare the small cycle-to-cycle fluctuations in the fundamental periods of healthy and diseased larynxes. He measured the glottal cycle lengths of speakers producing isolated sentences. Since then, jitter has been measured on sustained vowels (e.g., Koike, Takahashi and Calcaterra, 1977; Davis, 1979; Haji, Horiguchi, Baer and Gould, 1986), isolated sentences (e.g., Hecker and Kreul, 1971; Hammarberg, Fritzell, Gauffin, Sundberg and Wedin, 1980) and continuous speech (e.g., Gubrynowicz, Mikiel and Zarnecki, 1980; Laver, Hiller, Mackenzie and Rooney, 1986). Today, studies of jitter mainly concentrate on

* National Fund for Scientific Research, Belgium.

0095-4470/95/010189 + 13 $08.00/0

sustained vowels. Possible applications are the acoustic description of laryngeal pathologies and improvement of the naturalness of synthetic speech. However, over the last few years, studies of jitter have become less application- and more methodology-oriented.

In sustained vowels, the order of magnitude of jitter is typically 0.1% of the fundamental period. The origin of jitter is presently not fully understood. Pinto and Titze (1990), for example, distinguish between mean- and short-term fluctuations. Some of the mean-term perturbations might be the outcome of the response of some laryngeal muscles to neurological impulses (Baer, 1981; Heiberger and Horii, 1982; Titze, 1991). Circumstantial evidence suggests that the heartbeat also may be a source of mean-term perturbations (Orlikoff and Baken, 1989). Short-term, noise-like fluctuations might be the outcome of an uneven distribution of mucus on the vocal folds, asymmetries between the vocal folds (Isshiki, 1972; Ishizaka and Isshiki, 1976; Wong, Ito, Cox and Titze, 1991), turbulence of the airflow in the glottis, vibrations of the thyroid cartilage (Pesak, 1991), or the coupling between the larynx and the supra-glottal cavities (Horii, 1979).

The reason for concentrating on sustained vowels instead of connected speech is that during the production of connected speech, voluntary perturbations are frequent and their relative contribution to overall jitter cannot be factored out. Indeed, during the generation of connected speech, vocal fold tension and subglottal pressure continually vary to produce prosodic and voicing cues. The vocal tract also retroacts mechanically and acoustically on the larynx. Vocal fold motion and glottal signal are thus continually perturbed by changes in shape of the supra–laryngeal system. As a consequence, average perturbations are greater in connected speech than in sustained vowels (Schoentgen, 1989).

Generally speaking, jitter is evaluated by measuring successive glottal cycle lengths and calculating the average deviation of the length of the current cycle from:

(i) The length of the preceding cycle;
(ii) The average cycle length;
(iii) A running average over, typically, five cycles.

Type (i) is equivalent to type (ii) expressions when these are applied to the discrete derivative of the cycle length sequence (Pinto and Titze, 1990). The purpose of the running average (iii) is to remove any trend due to intonation (Koike, 1973; Koike *et al.*, 1977; Davis, 1979).

The corresponding mathematical expressions describe the variability of the glottal cycle lengths. They are measures of dispersion that only describe the spread of a set of lengths unambiguously when these are statistically independent. This means that it must be assumed that fluctuations in cycle lengths are white noise and that changing the order of the cycles does not change their statistical properties.

This assumption is problematic because earlier discussion of the possible origins of jitter suggests that the sequentially produced cycle-to-cycle fluctuations may be interrelated because of mean-term perturbations, for example. This conjecture can be confirmed experimentally. Experiments show that the perturbations of adjacent cycles are, indeed, correlated (De Guchteneere and Schoentgen, 1991). When two neighbouring cycle perturbations are correlated, then the perturbation of the one can be partly predicted from the perturbation of the other. They are not, therefore, statistically independent.

We, therefore, applied time series analysis methods to the study of jitter. Time series analysis methods dispense with the hypothesis that the data are statistically independent. They are thus suitable for representing data that are sequentially produced in time. They make it possible to distinguish between mean- and short-term perturbations and to separate non-random from random perturbations. Accordingly, we report here preliminary results on whitened jitter, i.e., jitter time series from which correlations between neighbouring perturbations have been removed.

2. A probabilistic model of jitter

If jitter were white noise, i.e., if the fluctuations of the glottal cycle lengths were purely random, then expression (1) would be an adequate probabilistic model of jitter. P_i is the length of cycle i, m is the average cycle length, Z_i is white noise of zero mean and constant variance:

(1) $$P_i - m = Z_i.$$

However, a more complicated model is needed when neighbouring fluctuations are correlated. The simplest probabilistic model (2) of jitter that takes correlations into account is a linear generalisation of expression (1). It is the so-called linear auto-regressive model (AR) of order p (Box and Jenkins, 1976). The a_i's are the weightings of the past fluctuations and m the average cycle length. Z_i is the so-called residue. Z_i is white noise of zero mean and constant variance.

(2) $$P_i - m = a_1(P_{i-1} - m) + a_2(P_{i-2} - m) + \ldots + a_p(P_{i-p} - m) + Z_i.$$

Models (1) and (2) differ from each other because of the weighted sum of the past fluctuations. This sum represents possible correlations between neighbouring fluctuations. When the number of weights a_i and their values are correctly chosen, model (2) outputs a time series whose second-order statistical properties (i.e., its correlation coefficients) agree with those of the original series. The method that we used to estimate weights a_i and order p is explained in the Materials and Methods section.

Model (2) is not the only possible linear generalisation of model (1). An alternative representation of $P_i - m$ would consist of a weighted sum of p past samples of the white noise residue Z_i. This model is called a moving average model.

3. Materials and methods

Twenty-two healthy male and 15 healthy female speakers, and 12 male and 19 female dysphonic ones sustained the vowels [a], [i] and [u] at a comfortable pitch and loudness level. The average ages were 47 years for the healthy male speakers, 49 years for the healthy female speakers, 48 years for the dysphonic male speakers, and 50 years for the dysphonic female speakers. The corpus of the dysphonic speakers comprised cases of nodules (5), oedema of the vocal folds (4), hypotonia (4), congestion of the vocal folds (3), paralysis of the left or right vocal fold (2), hypertonia (2), and others (11). Among the latter was a case of asymmetry between the vocal folds.

The recordings of the speech signal were made with a Sennheiser electret microphone mounted on a headset worn by the speaker. The distance between the mouth and the microphone was fixed at approximately 5cm. The electroglottogram was recorded concomitantly with a Fourcin laryngograph. Both signals were digitised via a SONY digital audio processor PCM-701 Es and stored on video tape. The recordings were carried out in a soundproof booth. The speech signals were later pre-filtered at 10 kHz and all the signals were redigitised at 20 kHz with a 12 bit analog-to-digital converter, and a one-second portion was stored on computer disc for further processing.

Since the procedure that we developed for measuring the fluctuations of the cycle lengths has been presented elsewhere (Schoentgen and De Guchteneere, 1991), we will confine ourselves to explaining it only briefly here. It comprises the following stages.

(i) Gross estimation of the average length of the fundamental period;
(ii) Low-pass filtering of the laryngogram at 2 kHz and band-pass filtering of the speech signal at 50 Hz and at a frequency between 500 Hz and 2 kHz (depending on vowel quality);
(iii) Discrete differentiation of the laryngogram;
(iv) Gross detection of the peaks of the derivative of the laryngogram and the zero-crossings of the speech signal;
(v) Discrete differentiation and oversampling of the laryngogram in the vicinity of the peaks, and oversampling of the speech signal in the vicinity of the zero-crossings;
(vi) Fine peak and zero-crossing detection;
(vii) Measurement of the duration between two peaks or two zero-crossings.

Measurement-wise, we mainly concentrated on solving two problems. The first was that measuring jitter requires a time resolution better than 10^{-5} seconds. This is a consequence of jitter typically being in an interval between 0.1 and 1% of the fundamental period. It has, however, been found that the required bandwidth is relatively small (Titze, Horii and Scherer, 1987). In order to save memory space, the signals were, therefore, sampled at a frequency of 20 kHz and numerically oversampled to 160 kHz in the vicinity of the events marking the beginnings and ends of the glottal cycles. The markers were the maxima of the discretely differentiated electroglottogram and the $-/+$ zero-crossings of the speech signal. Oversampling was carried out by inserting 7 additional samples the values of which were equal to the value of the last recorded sample and by differentiating and low-pass filtering the resulting staircase signal at 10 kHz with a Finite Impulse Response filter. Obviously, the oversampled data points so obtained were only estimates of the original, unknown, samples. Hess and Indefrey (1987) proposed a method for indirectly evaluating the fidelity with which data points obtained by oversampling reconstructed original samples. They argued that the distances, measured in number of samples, between the last recorded sample and the sample that marked the beginning of the glottal cycle were uniformly distributed provided that the oversampled signal was not biased by the interpolation method. We therefore checked the uniformity of the distribution of the distances by carrying out a χ^2 test.

A second problem was that the smallness of jitter made its measurements susceptible to bias by noise. Precautions therefore had to be taken to make sure that the measurements of the lengths of the glottal cycles were not unduly contaminated by quantization, measurement, background or intrinsic noise (due to turbulence in the glottis). The noise monitoring was carried out in the following manner. The electroglottogram and speech signal were recorded simultaneously. The cycle lengths were measured on both signals. The quality of the agreement between the two measurements was evaluated by calculating the inter-correlation function and examining the scattergram of the two sets of measurements. When the two series of measurements agreed reasonably well for a given sustained vowel, we concluded that they in fact reflected vocal fold activity and were not unduly contaminated by noise. The reason was that a good agreement between the two length measurements was unlikely to occur by chance alone because the speech signal and laryngogram were physically very different.

Once the series of the glottal cycle lengths had been obtained, the following statistical entities were computed.

(i) The autocorrelation function for lag $k = 1 \ldots 30$;
(ii) The period perturbation quotient;
(iii) The linear auto-regressive model.

We chose the period perturbation quotient (4) because it constitutes an early attempt to remove correlations from jitter time series. The period perturbation quotient (PPQ) is a measure of dispersion that has been used to evaluate the amount of jitter in an analysis interval (Koike, 1973; Koike *et al.*, 1977; Davis, 1979). N was the number of cycles in the analysis interval and P_i the length of the ith cycle. The running average over 5 cycles smoothed the time series locally.

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

Weightings a_i of the auto-regressive model (2) were determined by solving the so-called Yule-Walker equations (Box and Jenkins, 1976). These equations were obtained by making k equal to 1, 2, 3, ... p in expression (3).

$$\text{(3)} \qquad \rho(k) = a_1\rho(k-1) + a_2\rho(k-2) + \ldots + a_p\rho(k-p).$$

$\rho(k)$ was the autocorrelation function of lag k of the time series. A condition for expression (3) to be valid was that jitter should be stationary. This meant that shifting the time origin had no effect on the joint distribution of the fluctuations.

In practice, order p was determined in the following manner. A series of auto-regressive models of increasing order was constructed and the value of highest weighting a_p was examined. For a given time series, the value of a_p generally decreased as p increased. When the value of a_p became statistically negligible, the corresponding p was retained as the appropriate order. The outcome of this procedure also confirmed the appropriateness of the choice of an auto-regressive model over a moving-average one. Indeed, the time series was best represented by an auto-regressive model when the a_p values went rapidly to zero with increasing p

while the correlation function of the time series decreased only slowly with increasing lag k. If the opposite had been true, the time series would have been best represented by a moving average model (Gilchrist, 1984).

For a given time series, the adequateness of the estimates of order p and weightings a_i was also verified *a posteriori* in the following manner. The autoregressive model (2) predicted that residue Z_i was white noise. Its whiteness was therefore checked by carrying out two different χ^2 tests. The zero-hypothesis of the tests was that residue Z_i was white. Whenever the zero-hypothesis had to be rejected, it meant that either the model itself or its order was inadequate.

The operation (5) which obtained the residue from the raw perturbations is inverse filtering, which consists of subtracting the weighted sum of p past fluctuations from the current perturbation $P_i - m$ to obtain the residue Z_i:

$$Z_i = (P_i - m) - a_1(P_{i-1} - m) - \ldots - a_p(P_{i-p} - m). \tag{5}$$

Residue Z_i was the jitter time series from which correlations between neighbouring perturbations had been removed. The perturbation quotient of the residue (i.e., uncorrelated jitter) was calculated according to (6):

$$\mathrm{PPQ} = \frac{\frac{1}{N-p} \sum_{i=p+1}^{N} |Z_i|}{\frac{1}{N} \sum_{i=1}^{N} P_i}. \tag{6}$$

Formula (6) was similar to, but more general than, formula (4). In the following, we refer to formulas (4) and (6) respectively as the period perturbation quotients of the unfiltered (i.e., raw, unprocessed) and filtered jitter time series.

Formula (6) would become nearly identical to formula (4) if, in expression (5), p was made equal to 5 and $a_1, a_2, \ldots, a_5$ equal to 1/5. The only difference would be that in formula (6), P_i is subtracted from the average of the last five samples while in formula (4), P_i is subtracted from the average of the last two samples, the present sample and the future two samples. In other words, in formula (6) the running average would be shifted three samples into the past as compared to formula (4).

4. Results

Fig. 1 shows the distribution of the orders of AR-models of 204 jitter time series ($204 = 3$ vowels $\times$ 68 speakers). One sees that the distribution has a maximum at 3 and that, given the rules for estimating orders, each time series required a model of order 1 at least. This means that no time series was considered to be free of correlations.

In the results section we concentrate mainly on the comparison of the values of the period perturbation quotients of the filtered (i.e., whitened) and unfiltered (i.e., raw) jitter time series. Fig. 2 shows the scattergram of the PPQ values of the two time series. One sees that, as long as their values were small, both perturbation measures were proportional. For greater values, saturation set in. Saturation meant that the unfiltered (raw) time series gave rise to very high perturbation quotient values and that these values were greatly reduced after the time series had been modelled and inverse filtered (i.e., whitened). We observed that this occurred mainly when adjacent glottal cycle lengths were negatively correlated.

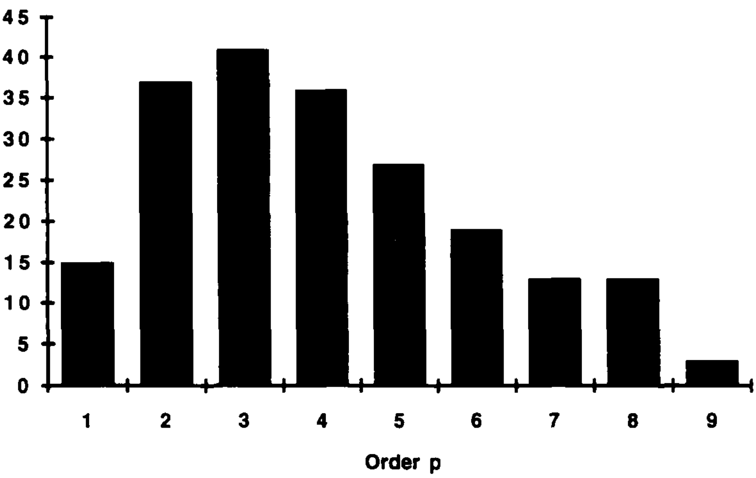

Figure 1. Statistical distribution of the orders of the auto-regressive models of 204 jitter time series. The horizontal axis shows the orders of the models, the vertical axis the number of models of a given order.

Figs 3–6 show the box and whisker plots of the perturbation quotient values broken down according to speaker gender and speaker state of health (healthy or dysphonic). A box and whisker plot displays the extreme values, the upper and lower quartiles and the median of a sample of values. Each figure shows one pair of box plots for each of three vowels ([a], [i] and [u]). The left member of each pair is the unfiltered data and the right member the filtered data.

When comparing the paired box plots, one sees that the median values of the

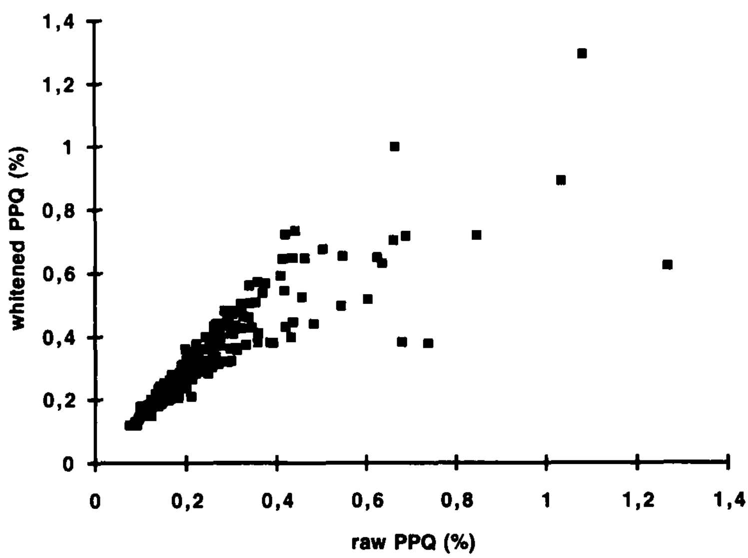

Figure 2. Scattergram of period perturbation quotients of the unfiltered (horizontal axis) and filtered (whitened) jitter time series (vertical axis). The perturbation quotient values are given in percentages. The number of time series is 198 (instead of 204). We removed 6 time series the abnormally high perturbation quotients of which completely distorted the shape of the display.

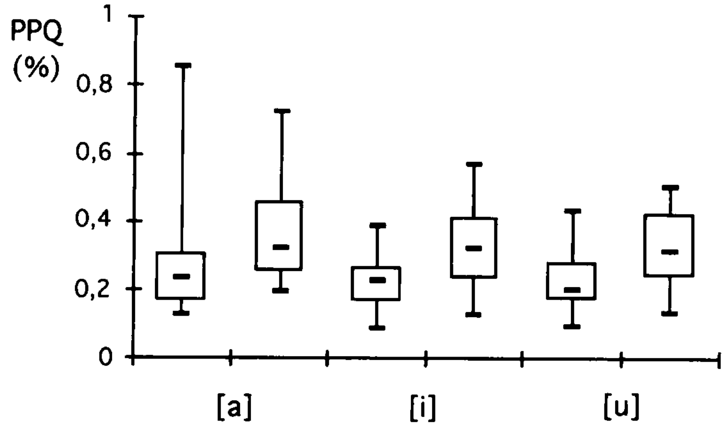

Figure 3. Box and whisker plots of the period perturbation quotient values in percentages (vertical axis) for the healthy male speakers. The horizontal axis shows a pair of box plots for each of the three vowels [a], [i] and [u] respectively. The left member of each pair is the unfiltered and the right member the filtered data.

perturbation quotient were always greater for the filtered than for the unfiltered time series. Instead, when the arithmetic mean (not reproduced here) was examined, the following was observed. When the healthy speakers produced [a], [i], [u] and the dysphonic speakers produced [u], medians and arithmetic means behaved identically. Both were greater for the filtered time series. When the dysphonic speakers produced [i] and [a], medians and arithmetic means behaved differently. The arithmetic mean was lower and the median higher for the filtered time series. This anomaly was the outcome of the influence on the arithmetic mean of a few outliers.

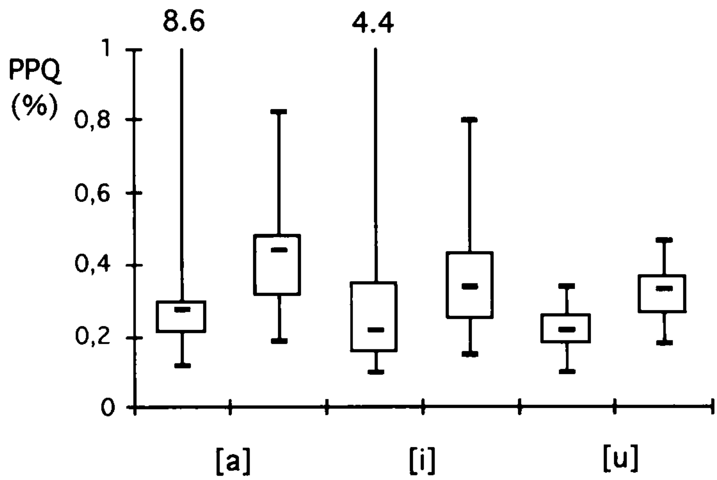

Figure 4. Box and whisker plots of the period perturbation quotient values in percentages (vertical axis) for the dysphonic male speakers. The horizontal axis shows a pair of box plots for each of the three vowels [a], [i] and [u] respectively. The left member of each pair is the unfiltered and the right member the filtered data.

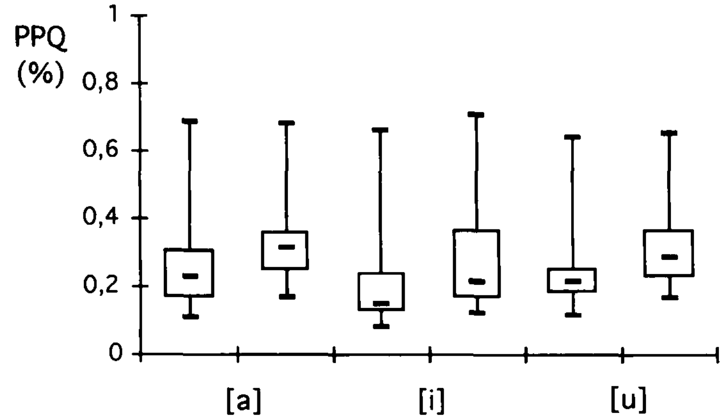

Figure 5. Box and whisker plots of the period perturbation quotient values in percentages (vertical axis) for the healthy female speakers. The horizontal axis respectively shows a pair of box plots for each of the three vowels [a], [i] and [u]. The left member of each pair is the unfiltered and the right member the filtered data.

The interquartile range is the difference between the first and third quartile. It is a measure of dispersion. For 10 out of 12 pairs of box plots, the interquartile range of the period perturbation quotient values was greater for the filtered than for the unfiltered time series (Figs 3–6). Instead, one observed for 9 out of 12 pairs of box plots that the difference between the maximum and the third quartile of the perturbation quotient values was smaller for the filtered than for the unfiltered time series.

With the exception of the unfiltered time series of vowel [i] sustained by male speakers, the median values of perturbation quotients (4) & (6) were greater for the

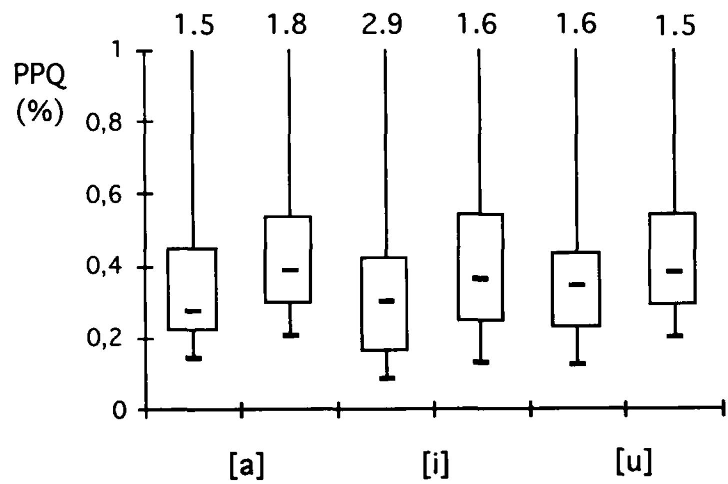

Figure 6. Box and whisker plots of the period perturbation quotient values in percentages (vertical axis) for the dysphonic female speakers. The horizontal axis respectively shows a pair of box plots for each of the three vowels [a], [i] and [u]. The left member of each pair is the unfiltered and the right member the filtered data.

dysphonic than for the healthy speakers. Differences, though, were small especially between healthy and dysphonic male speakers. Similarly, differences between the maxima and third quartiles of the perturbation quotient values (4) and (6) tended to be greater for the dysphonic than for the healthy speakers. For the female dysphonic speakers, both the interquartile ranges and the differences between the maxima and third quartiles were greater.

With the exception of the unfiltered time series of vowel [u], the median values of perturbation quotients (4) and (6) were greater for the healthy male than for the healthy female speakers (Figs 3 and 5). This indicated that the average absolute perturbations in the male speakers were clearly greater than those in the female speakers. Indeed, one must not forget that period perturbation quotients (4) and (6) were relative measures of perturbation. For the dysphonic speakers, inter-gender differences were not systematic. This was expected considering that the laryngeal disorders were not the same in male and female speakers.

For the healthy speakers, the median values of perturbation quotients (4) and (6) were slightly greater for low vowel [a] than for high vowels [i] and [u]. For the dysphonic speakers, the trend was the same with the exception of the unfiltered jitter time series produced by the female dysphonic speakers. This indicated that the average absolute perturbations of the glottal cycle lengths were greater for low vowel [a]. Indeed, one must not forget that (4) and (6) were relative measures of perturbation and that for all our corpora the median fundamental frequencies were smaller for the low vowel [a] than for the high vowels [i] and [u].

5. Discussion

The median and interquartile ranges of the period perturbation quotient values were smaller for the unfiltered than for the filtered jitter time series. A possible explanation is the following. The absolute value term in the numerator of formula (4) can be rewritten as follows:

$$(7) \qquad |\bullet| = \left| \frac{P_{i-2} - P_i}{5} + \frac{P_{i-1} - P_i}{5} + \frac{0}{5} + \frac{P_{i+1} - P_i}{5} + \frac{P_{i+2} - P_i}{5} \right| .$$

The middle term is always zero because it is the subtraction of P_i from itself. In other words, expression (7) is a sum of four terms that is divided by five. As a consequence, formula (4) systematically underestimates jitter. The perturbation quotient values were therefore greater for the filtered than for the unfiltered jitter time series because this bias was absent from formula (6). The bias did not occur because the length of the current cycle was subtracted from the lengths of past cycles only. Also, in formula (6), the number of weightings and their values were not arbitrarily chosen, but adapted to each time series.

In contrast to medians and interquartile ranges, arithmetic means and differences between maxima and third quartiles of the perturbation quotients tended to be smaller for the filtered than for the unfiltered time series in the case of the dysphonic speakers (Figs 4 and 6). The disagreement stemmed from the fact that arithmetic means and total ranges were affected by outliers which did not affect medians and interquartile ranges. Where did the outliers come from? We observed that the values of the perturbation quotient were abnormally high (i.e., were outliers) when the perturbations of adjacent glottal cycles were negatively correlated. The cycle lengths then jumped between small and large values. Formula (4) counted these systematic deviations among the random cycle–to–cycle fluctuations and, consequently, yielded higher values for the period perturbation quotient. In contrast, the AR–model of the time series modelled these periodic variations and, consequently, inverse filtering removed them from the time series. Formula (6) therefore gave rise to lower perturbation quotient values. The removal of the outliers also explained the decrease of the difference between the maximum and the third quartile of the perturbation quotients of the filtered time series. Since time series with negative correlations were more frequent among the dysphonic than among the healthy speakers, the decrease was more marked for dysphonic speakers.

To sum up the above discussion, filtered jitter time series (i.e., jitter from which correlations between adjacent perturbations had been removed) gave rise to statistical distributions of perturbation quotients which were shifted upwards towards somewhat higher values, but which were more compact around the mean because the total ranges were smaller. Time series analysis compressed total ranges because it separated random from non–random perturbations.

As far as inter-gender and inter-vowel differences are concerned, results obtained here confirmed what had been observed elsewhere (Horii, 1979). Perturbations were greater for male speakers and for low vowel [a]. These differences were not always noticeable in relative perturbation measures because normalisation by the average fundamental period blurred differences between the males and the females and between the low and the high vowels because the males had a longer average fundamental period and the low vowels a longer intrinsic fundamental period.

As far as inter-condition differences are concerned, we expected the perturbations to be higher for the dysphonic than for the healthy speakers. Dysphonic speakers did exhibit greater median perturbation quotient values, yet the differences were small especially between healthy and dysphonic male speakers. Also, the statistical distributions of the values of the perturbation quotient (6) of the filtered time series overlapped a great deal. This was possibly due to the removal of non-random perturbations by inverse filtering, but also to greater measurement precision and a better control of the influence of noise on cycle length measurements. In an earlier study of jitter, we measured somewhat higher perturbation quotient values for the dysphonic speakers by means of signal processing methods that were less refined than those employed here (Schoentgen, 1989). Another factor that might explain the overlap of the perturbation measures in healthy and dysphonic speakers is that pathologies leading to an imbalance in the tensions or masses of the vocal folds were scarce in our corpora. It can be predicted with some confidence that these kinds of pathologies increase jitter.

6. Conclusion

We studied jitter from a time series point of view. The conclusions were as follows. Firstly, the conventional period perturbation quotient (4) of the raw, unprocessed jitter time series underestimated the amount of jitter because of a running average that was centred on the current cycle. Secondly, the inter–speaker variability of the period perturbation quotient (6) of the whitened jitter time series (i.e., time series from which correlations had been removed) was reduced because the time series analysis separated random from non–random perturbations. Thirdly, improved time resolution and better noise control were other factors that reduced differences in jitter between healthy and dysphonic speakers. Fourthly, trends in the shape of higher absolute fluctuations for male speakers and low vowel [a] were as expected for inter-gender and inter-vowel differences. However, the trends were not clear-cut for relative perturbation measures because the normalisation by the fundamental period blurred the differences between high and low vowels and male and female speakers.

We thank Professor D. Hennebert, O. R. L. Department of the University Clinic St. Pierre, Brussels for providing speech samples of dysphonic speakers. The present work was carried out in the framework of the New Initiatives in Fundamental Research programme of the Université Libre de Bruxelles.

References

Baer, T. (1981) Investigation of the phonatory mechanism. In *Proceedings of the Conference on the Assessment of Vocal Pathology, ASHA Reports* **11** (C. L. Ludlow and M. O. Hart, Eds), pp. 38–47, Rockville: The American Speech-Language-Hearing Association.

Box, G. & Jenkins, G. (1976) *Time series analysis, forecasting and control.* San Francisco: Holden-Day.

Davis, St. B. (1979) Acoustic characteristics of normal and pathological voices. In *Speech and language, advances in basic research and practice* (N. J. Lass, Ed.) **1**, pp. 271–335, New York: Academic Press.

De Guchteneere, R. & Schoentgen, J. (1991) Mean-term perturbations of the pseudo-period of the glottal waveform. In *Proceedings of the twelfth congress of phonetic sciences*, pp 354–357, Aix-en-Provence: University of Provence.

Gilchrist, W. (1984) Statistical modelling. New York: Wiley.

Gubrynowicz, R., Mikiel, W. & Zarnecky, P. (1980) An acoustic method for the evaluation of the state of the larynx source in cases involving pathological changes, *Archives of Acoustics*, **5**(1), 3–30.

Haji, T., Horiguchi, S., Baer, T. & Gould, W. (1986) Frequency and amplitude perturbation analysis of electroglottograph during sustained phonation, *Journal of the Acoustical Society of America*, **80**(1), 58–62.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J. & Wedin, L. (1980) Perceptual and acoustic correlates of abnormal voice qualities, *Acta Otolaryngologica*, **90**, 441–451.

Hecker, M. H. L. & Kreul, E. J. (1971) Descriptions of the speech of patients with cancer of the vocal folds. Part I: Measures of fundamental frequency, *Journal of the Acoustical Society of America*, **49**(4), 1275–1282.

Heiberger, V. L. & Horii, Y. (1982) Jitter and shimmer in sustained phonation. In *Speech and language, advances in basic research and practice*, **7**, (N. J. Lass, Ed.), pp. 299–332, New York: Academic Press.

Hess, W. & Indefrey, H. (1987) Accurate time-domain pitch determination of speech signals by means of a laryngograph, *Speech Communication*, **6**, 55–68.

Horii, Y. (1979) Fundamental frequency perturbation observed in sustained phonation, *Journal of Speech and Hearing Research*, **22**, 5–19.

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

Ishizaka, K. & Isshiki, N. (1976) Computer simulation of pathological vocal-cord vibration, *Journal of the Acoustical Society of America*, **60**(5), 1193–1198.

Koike, Y. (1973) Applications of some acoustic measures for the evaluation of laryngeal dysfunction, *Studia Phonologica*, **VII**, 17–23.

Koike, Y., Takahashi, H. & Calcattera, T. C. (1977) Acoustic measures for detecting laryngeal pathology, *Acta Otolaryngologica*, **84**, 105–117.

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

Lieberman, Ph. (1963) Some acoustic measures of the fundamental periodicity of normal and pathologic larynges, *Journal of the Acoustical Society of America*, **35**(3), 344–353.

Orlikoff, R. F. & Baken, R. J. (1989) Fundamental frequency modulation of the human voice by the heartbeat: preliminary results and possible mechanisms, *Journal of the Acoustical Society of America*, **85**(2), 888–893.

Pesak, J. (1991) Complex nature of the seemingly simple vocal fold cycle. *Proceedings of the twelfth international congress of the phonetic sciences*, **2**, 70–72.

Pinto, N. B. & Titze, I. R. (1990) Unification of perturbation measures in speech signals, *Journal of the Acoustical Society of America*, **87**(3), 1278–1289.
Schoentgen, J. (1989) Jitter in sustained vowels and isolated sentences produced by dysphonic speakers, *Speech Communication*, **8**, 61–79.
Schoentgen, J. & De Guchteneere, R. (1991) An algorithm for the measurement of jitter, *Speech Communication*, **10**, 533–538.
Titze, I. R., Horii, Y. & Scherer, R. C. (1987) Some technical considerations in voice perturbation measurements, *Journal of Speech and Hearing Research*, **30**, 252–260.
Titze, I. R. (1991) A model for neurological sources of aperiodicity in vocal fold vibration, *Journal of Speech and Hearing Research*, **34**, 460–472.
Wong, D., Ito, M. R., Cox, N. B. & Titze, I. R. (1991) Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases, *Journal of the Acoustical Society of America*, **89**(1), 383–394.