MEAN-TERM PERTURBATIONS OF THE PSEUDO-PERIOD OF THE GLOTTAL WAVEFORM

R. De Guchteneere and J. Schoentgen*

Institute of Phonetics, Free University of Brussels
Av. F.D. Roosevelt, 50, B-1050 Brussels, Belgium
* Research Associate, Belgian Fund for Scientific Research

ABSTRACT

Jitter is defined as the fluctuations from one glottis cycle to the next of the duration of the fundamental period in voiced speech sounds. We propose to study jitter from a time series point of view. Results obtained in a framework of an experiment on sustained vowels show that in a majority of cases adjacent period durations are strongly correlated and that the relationships between neighbouring periods are not the outcome of systematic long-term melodic variations.

1. INTRODUCTION

In this article we study jitter from a time series point of view. By jitter one understands the fluctuations from one cycle to the next in the duration of the fundamental period in voiced speech sounds.

Jitter has been studied for some thirty years (e.g. [4], [3], [8]). Conventionally, the amount of jitter during sustained phonation, for instance, is estimated by measuring successive glottal cycle durations and by computing a suitable dispersion measure over an analysis interval of, typically, fifty periods. Frequently, the differences between individual periods and a local average are taken into account instead of the differences between nearest neighbours. The purpose of the running average is to remove the effects of any long-term trend on the durations of individual cycles. Trends are believed to be the consequence of melodic variations, i.e. a prosodic phenomenon under volitional control.

There are two hidden assumptions in the approach described. Both are unwarranted.

1) The first assumption is that, once any trend has been removed, the differences between adjacent periods are statistically independent. Indeed, it is under this hypothesis alone that a dispersion measure is a sufficient descriptor of jitter. In other words, neglecting the time series aspect of a sequence of cycle durations is only without consequence when periods are interchangeable.

2) The second assumption is related to the first. It consist of supposing that any systematic relationship between neighbouring periods can be effectively removed by smoothing and that, moreover, what is so removed is not worth taking into account since it is simply the consequence of melodic variations.

Both hypotheses can be dropped when the sequence of glottal cycle durations is examined from a statistical time series point of view. Preliminary results we obtained thus show that the assumptions laid out above are indeed unjustified. Differences between neighbouring periods are not statistically independent since random short term fluctuations appear to be superimposed on stochastic mean-term perturbations. These do not seem to be the consequence of a smoothly evolving melodic curve. These observations hold for a great majority of our speakers. Most of them show a positive correlation between the durations of adjacent periods. This means that longer than average durations tend to be followed by longer cycles and that, vice versa, shorter than average durations tend to be followed by shorter cycles.

2. SUBJECTS AND METHODS

Twenty five adult speakers served as subjects for this preliminary study (eight healthy males, five healthy females, five dysphonic males and seven dysphonic females). They were told to sustain three vowels ([a], [i], and [u]), at a comfortable pitch and loudness level, as long and as steadily as they could.

The signals were recorded in a sound-proofed room. The microphone was placed approximately 5 cms from the lips. The laryngograph (or EGG for electroglottograph) signal, which varies proportionally to laryngeal conductance, was recorded simultaneously. The signals were digitized by a two channel SONY PCM audio processor and recorded on video tape. A central one-second portion of the signal of each vowel was redigitized at a 20 kHz sampling frequency with 12 bit resolution and stored for further processing in two files (EGG - and acoustic signal) on the hard disk of a Masscomp 5050 computer.

The algorithm that we designed to measure the duration of individual glottis cycles made use of oversampling to obtain high resolution in time. The measurements were made in two steps :

- Firstly, a gross detection of the important events in the original signals was carried out, i.e. (i) the peaks in the first derivative of the EGG signal, which were assumed to mark the instant of glottal closure, and (ii) the zero crossings in the filtered acoustic signal.
- Secondly, a portion of the signal centred on the main events was oversampled eight times and low-pass filtered; the period markers were then detected with improved accuracy, leading to a theoretical resolution in time of 6.25 μs. A statistical test was used to check oversampling reliability [2].

The algorithm was applied simultaneously to both the EGG and the acoustic signals. Tests carried out so far have shown that the algorithm performs satisfactorily : the comparison of the period values measured shows that both signals agree on most of the fine detail of the period-to-period fluctuations [7].

The algorithm also provides possibilities for graphical visualization (series of the period values, trend, differences between instantaneous period values and running average, statistical distributions, etc...), and a battery of statistical tests. So far we have implemented five different tests (four out of five verify the statistical independence of consecutive period fluctuations, i.e. our null hypothesis) :

1) The comparison to a gaussian distribution of the distribution of the microfluctuations.
2) The run test for randomness.
3) The comparison of the statistical distributions of adjacent local deviations.
4) The Pearson's moment product correlation coefficient of adjacent period durations.
5) The rank correlation coefficient of adjacent period durations.

3. RESULTS AND DISCUSSION

We have summarized in table 1 the results of serial correlation tests carried out on period sequences obtained from male and female speakers. They show that a great majority of vowel signals give rise to a positive correlation between adjacent period durations. Typical period sequences are shown in Figure 1. Figure 1a displays the period time series of five male and Figure 1b the time series of five female speakers. The first sequence in figure 1b presents a case of a negative correlation between neighbouring periods; all the other sequences present positive correlations.

The mechanisms underlying the production of jitter are not yet fully understood. Neurological and cardiac mechanisms, which have been shown to contribute to jitter [1], [5], would lead us to expect perturbations of the fundamental period straddling several cycles. Indeed, in an enumeration of possible candidate mechanisms, Pinto and Titze [6] distinguish between short-term and long-term contributors. Among the former they include the irregular distribution of mucus on the vocal folds, asymmetries in vocal fold geometry, turbulence, and the coupling between the glottis and the vocal tract. They count the neurological factors among the long-term aspects. What this list suggests is the existence of two time scales on the level of which independent factors are active. This point of view is not contradicted by our preliminary findings.

On the other hand it cannot be excluded that statistical models can be shown to exist which describe the cycle duration time series purely in terms of a deterministic component driven by a purely random signal. The need to distinguish between short-term and long-term perturbations could thus be obviated.

Table 1

Results of the Pearson's moment product and the rank correlation test for healthy and dysphonic speakers. Displayed are the number of signals showing positive correlation, no correlation or negative correlation between adjacent period durations.

	Speech signal						EGG signal					
	Pearson			Rank			Pearson			Rank		
	+	0	-	+	0	-	+	0	-	+	0	-
[a]												
Healthy sp. (13)	10	1	2	12	0	1	10	1	2	11	1	1
Dysphonic sp. (12)	9	2	1	11	0	1	9	0	3	11	1	0
TOTAL (25)	**19**	**3**	**3**	**23**	**0**	**2**	**19**	**1**	**5**	**22**	**2**	**1**
[i]												
Healthy sp. (13)	5	3	5	13	0	0	5	3	5	13	0	0
Dysphonic sp. (12)	9	1	2	11	1	0	10	1	1	12	0	0
TOTAL (25)	**14**	**4**	**7**	**24**	**1**	**0**	**15**	**4**	**6**	**25**	**0**	**0**
[u]												
Healthy sp. (13)	11	2	0	13	0	0	11	2	0	13	0	0
Dysphonic sp. (12)	8	1	3	10	1	1	10	0	2	12	0	0
TOTAL (25)	**19**	**3**	**3**	**23**	**1**	**1**	**21**	**2**	**2**	**25**	**0**	**0**

4. REFERENCES

[1] Baer, T. (1980) : Vocal jitter : A neuromuscular explanation. Transcripts of the Eighth Symposium of the Care of the Professional Voice, Voice Foundation, New-York, pp 19-22.

[2] Hess, W. and Indefrey, H. (1987) : Accurate time-domain pitch determination of speech signal by means of a laryngograph, Speech Comm., 6, pp 55-68.

[3] Kasuya, K., Kobayashi, Y., Kobayashi, T., Ebihara, S. (1983) : Characteristics of pitch period and amplitude perturbations in pathologic voice, IEEE-ASSP Conference, Boston, pp 1372-1375.

[4] Lieberman, Ph. (1963) : some acoustic measures of the fundamental periodicity of normal and pathologic larynges, J. Acoust. Soc. Am., 35, 3, pp 344-353.

[5] Orlikoff, R.F. and Baken, R.J. (1989) : Fundamental frequency modulation by the heartbeat: preliminary results and possible mechanisms, J. Acoust. Soc. Am., 85, pp 888-893.

[6] Pinto, N.B. and Titze I.R. (1990) : Unification of perturbation measures in speech signals, J. Acoust. Soc. Am., 87, 3 , pp 1278-1289.

[7] Schoentgen, J. and De Guchteneere R. (1990) : An algorithm for the measurement of jitter, Proceedings of the ESCA Workshop on Speaker Characterization, Edinburgh, pp 175-180.

[8] Titze, I.R., Horii, Y., Scherer, R.C. (1987) : Some technical considerations in voice perturbation measurements, J. Speech Hear. Res., 30, pp 252-260.

Fig. 1a

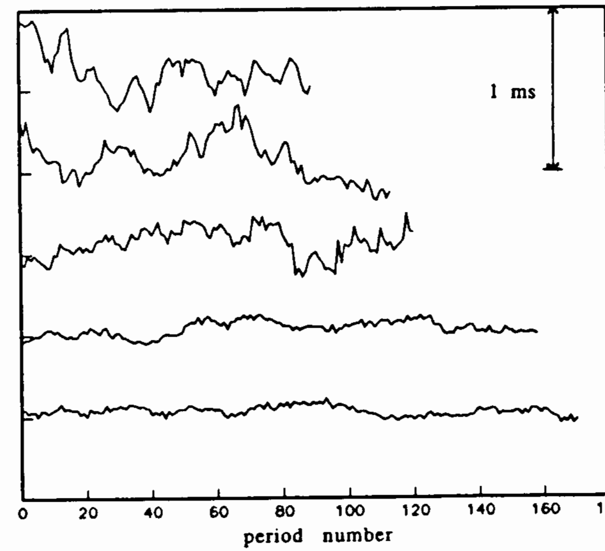

Fig. 1b

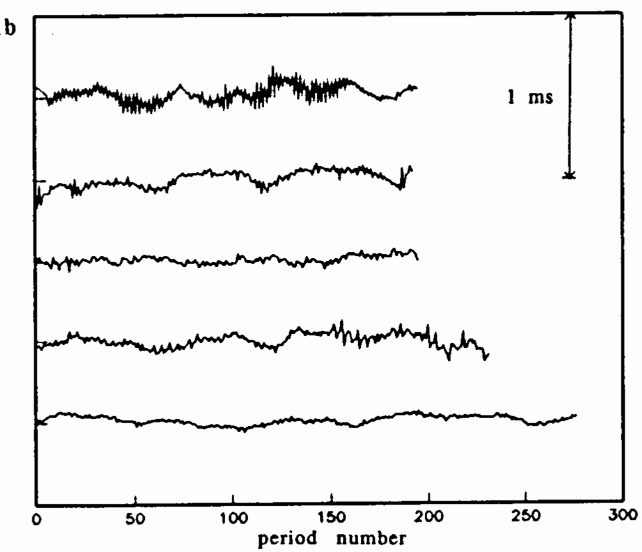

Figure 1

Period time series measured on a one second analysis interval for five male (fig. 1a) and five female (fig. 1b) speakers. The vertical axis is labelled in milliseconds. The average of each sequence has been offset by a constant amount, in order to avoid overlap. The horizontal axis gives the period number. The average fundamental frequencies are respectively equal to 89, 117, 123, 161 and 173 Hz in fig. 1a and to 197, 197, 200, 236 and 279 Hz in fig. 1b (from above to below).The first sequence in figure 1b presents a case of a negative correlation between neighbouring periods; all the other sequences present positive correlations.