JITTER-MEASUREMENTS FROM TELEPHONE-TRANSMITTED SPEECH

Isolde Wagner
Bundeskriminalamt - FB Sprechererkennung, Wiesbaden, Germany

ABSTRACT

The present study investigates the validity of a new jitter-algorithm on telephone-transmitted speech samples which are particularly degraded by band-width limitation. The algorithm has been developed for specific use in forensic speaker identification and allows for the quantification of hoarseness not only from isolated sustained vowels but also from vowels in connected speech. The results of a pilot study are presented here. They show that the algorithm is valid to differentiate speakers with certain kinds of pathological hoarseness from speakers with normal voices.

INTRODUCTION

In order to quantify voice qualities which are perceived as hoarse, special attention has been paid to a phenomenon which refers to the temporal irregularities of the vibration process of the vocal folds. The phenomenon is called jitter. It is defined as the involuntary short-term variation of the voice fundamental frequency (fo) from one cycle to the next, in contrast to the voluntary and controlled long-term variation of fo which is the physical correlate of sentence intonation.

While several studies have proposed methods to allow for jitter measurements in high quality recordings of isolated sustained vowels [1,2,3,4,5,6 for example], there seems to be no reliable method of measuring jitter in connected or degraded speech. These problems, however, arise in forensic speaker identification (SI), where (a) non-cooperative speakers have to be examined who are not inclined to produce sustained vowels, and (b) the majority of the speech samples to be analysed are telephone-transmitted with a pass-band between about 300 to 3400 Hz. Therefore a new jitter-algorithm was developed by the Forensic Science Laboratory of the Bundeskriminalamt (Federal Criminal Police Office) and the University of Trier which was designed to yield reliable results even under forensic conditions.

JITTER-ALGORITHM

The new algorithm differs from previous ones by allowing jitter measurements either from sustained vowels or from vowels in connected speech, irrespective of the underlying sentence intonation. It has been implemented on a MEDAV SPEKTRO 3000 computer system and consists of two analytical procedures: (a) a new fo analysis method which is based on frequency demodulation procedures providing high resolution fo values, and (b) a new method for the computation of jitter taking up the basic idea of relative average perturbation (RAP) as suggested by Koike [4].

The two procedures have been explained in detail in an earlier study [7], however, the specificity of the new algorithm is described in more detail here. It consists in treating the high resolution fo contour as a multidimensional vector. A second, auxiliary vector is derived from the first using a method of approximation with a third order polynomial function - a contour with one point of inflexion, serving as a reference vector of the fo contour. In Figure 1, the second window from the bottom gives an example how the procedure works with jitter in a portion of 120 ms duration of the sustained vowel /a:/ produced by a hoarse male speaker at an average fo of 88 Hz. The steps represent the high resolution fo contour including short-term variations; the smooth curve represents the long-term variations derived by a third order polynomial function which describes the intonation contour. The deviation of the actual values from the polynomial function is calculated and the result is shown in the lower right corner in terms of RAP. The value is 2.8655 %.

EXPERIMENT

In order to test the validity of the algorithm based on speech samples which are degraded by band-pass filtering and thus do not contain the fundamental in the signal, the study uses a harmonic rather than the fo as a multidimensional vector in the procedure. The results of the jitter measurements obtained in this way are compared to the results of measurements on the basis of high quality recordings.

SUBJECTS AND MATERIAL

The material consists of recordings of seven male German speakers with normal voices and seven speakers with pathological hoarseness of various origins deviding the hoarse speakers into two subgroups: speakers who suffer (a) from *hypo-* and (b) from *hyper*functional dysphonia. The recordings were made under sound treated conditions using high quality equipment. Subjects were required to produce different types of sustained vowels, both in isolation and in a /mVm/-context, and also various sequences of connected speech.

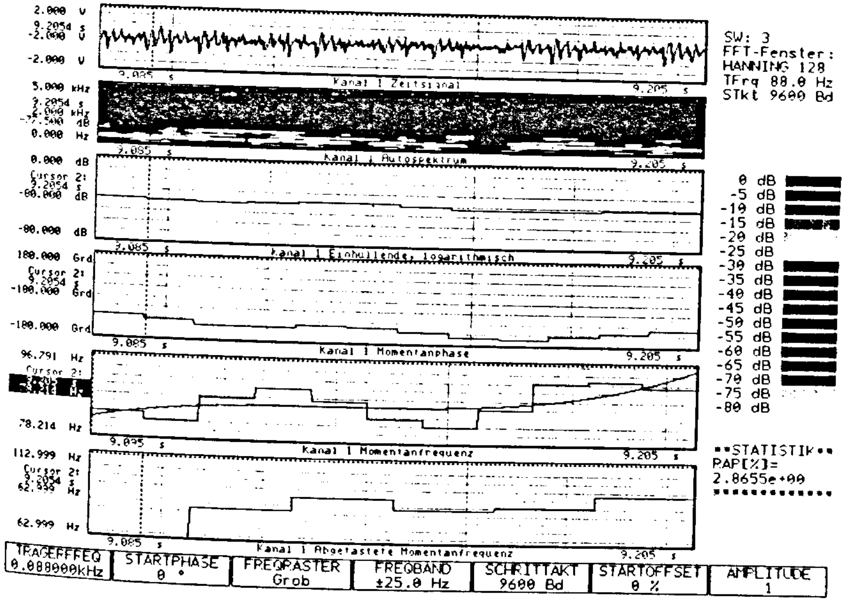
ANALYSIS PROCEDURE

Recordings of both of the two sustained productions of the vowels /e/, /ε/, /a/, and /o/, and one sample of the same vowel from connected speech were digitized in the MEDAV SPEKTRO 3000 computer system in a two channel mode, where subsequently one of the two channels was band-pass filtered from 300 to 3400 Hz, thus simulating the degrading characteristic to telephone-transmission. Jitter-measurements were made using the fo from the channel containing the high quality recording, and the *third* harmonic (h3) from the filtered channel, because even in low male voices, it can be safely assumed to be within the range of telephone-transmission.

RESULTS

Because it was observed that jitter values vary with the duration of the measured portion of the vowel and because of the fact that in connected speech vowel durations of more than 120 ms are rare, portions of 120 ms were used for all measurements. Furthermore, it was found that the four different vowels /e/, /ε/, /a/, and /o/ did not yield any systematical differences in jitter values. Therefore, these vowels were pooled in the study.

For the purpose of comparing between high quality and degraded speech samples distributions of high and low jitter values were investigated on measure-



SW: 3
FFT-Fenster:
HANNING 128
TFrq 88.0 Hz
STkt 9600 Bd

STATISTIK
RAP[%]=
2.8655e+00

TRAGERFFREQ 0.088000kHz | STARTPHASE 0 ° | FREQRASTER Grob | FREQBAND ±25.0 Hz | SCHRITTAKT 9600 Bd | STARTOFFSET 0 % | AMPLITUDE 1

Figure 1. Procedure of the new jitter-algorithm working in the sustained production of the vowel /a:/

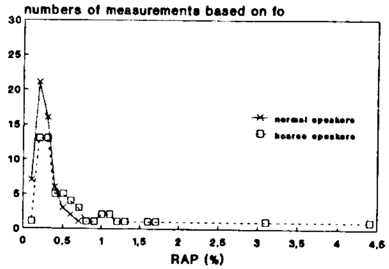


Figure 2a. Distribution of jitter values: measurements based on fo of sustained vowels

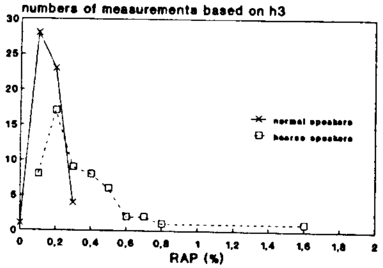


Figure 3a. Distribution of jitter values: measurements based on h3 of sustained vowels

ments based both on fo and h3 for the hoarse and normal speakers and for both vowel conditions separately. Mean jitter values, as well as minimum and maximum values were computed for the different types of measurements for the normal speakers and the two groups of hoarse speakers.

Jitter-measurements based on fo

Figure 2a shows the distribution of jitter values for measurements based on fo for the two *sustained vowel* productions. The full line represents the values for the normal speakers, the broken line the values for the hoarse speakers.

It was established that most of the values for the normal speakers range between 0.2 to 0.3% RAP. Values of more than 0.5% are rare. The means are 0.3%, the minimum values 0.1% and the maximum values 0.7% RAP for both of the two sustained productions. The values for the hoarse speakers, however, show a very large range of between 0.1 and 4.4% RAP. But, whereas the values for the hypofunctional group are well within the range of the normal speakers, the values for the hyperfunctional group are clearly higher, and values of less than 0.5% RAP are very rare. The mean values are 1.0 and 1.1% RAP for the two sustained productions, however, the variation within this group is large.

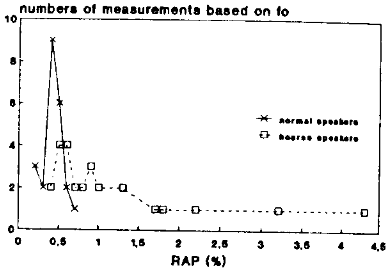


Figure 2b. Distribution of jitter values: measurements based on fo of vowels from connected speech

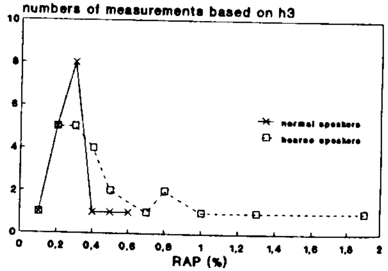


Figure 3b. Distribution of jitter values: measurements based on h3 of vowels from connected speech

Figure 2b shows the distribution of jitter values for measurements based on fo of the vowels from *connected speech*. It emerged that jitter values were higher than in the sustained vowel productions for all groups of speakers, and both of the two groups of hoarse speakers differ from the group of the normal speakers. Whereas the range of the normal speakers' jitter values amounts to 0.2 to 0.7% RAP with a mean of 0.4%, the hoarse speakers exhibit jitter values ranging from 0.4 to 4.3% RAP with a mean of 0.7% for the hypofunctional and a mean of 1.6% for the hyperfunctional group.

Jitter-measurements based on h3

Figure 3a shows the distribution of jitter values for measurements based on h3 of the *sustained vowel* productions. It is evident that the jitter values of all groups of speakers are lower than the comparable values obtained from measurements on the basis of fo. Nearly all of the jitter values (51 out of 56) of the normal speakers are distributed between 0.1 and 0.2% RAP for the two sustained productions. The minimum value is 0.0%, the maximum value is 0.3% RAP. The hoarse speakers, however, exhibit larger variation in the jitter values ranging from 0.1 to 0.5% RAP for the hypofunctional and from 0.1 to 1.6% RAP for the hyperfunctional group. The means are 0.2 and 0.3%, respectively, for the hypo- and 0.4 and 0.5% RAP for the hyperfunctional group for the two sustained productions.

Figure 3b shows the distribution of jitter values for measurements based on h3 of the vowels from *connected speech*. As for the sustained vowels it was observed that all jitter values obtained from measurements based on h3 are lower than those based on fo. However, the values of vowels taken from connected speech are higher than those based on sustained vowels. The normal speakers show values from 0.1 to 0.6% RAP with a mean of 0.3%, and the hypofunctional group yields values which are well within the range of the normal speakers. The values of the hyperfunctional group, however, range from 0.2 to 1.9% RAP with a mean of 0.8%.

DISCUSSION

The jitter-algorithm was found to differentiate speakers with certain kinds of pathological hoarseness from normal speakers through measurements based on h3 as well as on fo and on both types of vowel production. Whereas speakers who suffer from hypofunctional dysphonia show jitter values which are more or less within the range of those of normal speakers, speakers who suffer from hyperfunctional dyphonia exhibit values which are much higher and more widely distributed. The boundary line between these hoarse speakers and the normal speakers is about 0.5% for the measurements based on fo and about 0.3% RAP for the measurements based on h3. Jitter values based on vowels in connected speech were found to be slightly higher than those based on sustained vowels. This can possibly be explained by the influence of coarticulation on the vowels in connected speech. All these findings correspond with the results of a previous study by the present author [7], where jitter-measurements were made on the basis of fo from the vowel /a/ only. The observation that jitter values obtained from measurements based on fo are systematically higher than those based on h3 is considered to be related to the working principle of the algorithm.

It can be concluded that the algorithm is in principle able to measure jitter in connected speech which is degraded by the particularly band-pass filtering of telephone-transmission. However, before the procedure can be established in forensic SI, more research is needed.

REFERENCES

[1] Baken, R.J. (1990), 'Irregularity of vocal period and amplitude: A first approach to the fractal analysis of voice', *Journal of Voice*, 4 (3), 185-197.

[2] Hollien, H.; Michel, J. & Doherty, E.T. (1973), 'A method for analysing vocal jitter in sustained phonation', *Journal of Phonetics*, 1, 85-91.

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

[4] Koike, Y. (1973), 'Application of some acoustic measures for the evaluation of laryngeal dysfunction', *Studia Phonologica*, 7, 17-23.

[5] Lieberman; P. (1963), 'Some acoustic measurements of the fundamental periodicity of normal and pathologic larynges', *The Journal of the Acoustical Society of America*, 35 (3), 344-353.

[6] Ludlow, C. Coulter, D. & Gentges, F. (1983), 'The differential sensivity of measures of fundamental frequency perturbation to laryngeal neoplasms and neuropathologies', in D. M. Bless and J. H. Abbs (eds), *Vocal fold physiology: Contemporary research and clinical issues*, San Diego: Colledge Hill, 381-392.

[7] Wagner, I. (1995), 'A new jitter algorithm to quantify hoarseness: an exploratory study', to be published in: *Forensic Linguistics*, 2 (1), 000-000.