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 telephony-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 SPEKTO 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 inflection, 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 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 dividing the hoarse speakers into two subgroups: speakers who suffer (a) from larynx and (b) from hyperfunctional 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, /e, /a, /a/ and /o, /o/ and, as one sample of the same vowel from connected speech were digitized in the MEDAV SPEKTO 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 degradation 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, /e, /a, /a/ and /o, /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-

Figure 1. Procedure of the new jitter-algorithm working in the sustained production of the vowel /a/
It was established that most of the values for the normal speakers range between 0.1 and 0.5% RAP. 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.

The jitter-algorithm was found to be published in: Forensic Linguistics, 2 (1), 000-000.

REFERENCES