View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Radboud Repository

Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/76252

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

ELSEVIER

Speech Communication 22 (1997) 67–79

SPEECH COMMUNICATION

Parabolic spectral parameter – A new method for quantification of the glottal flow

Paavo Alku [a,*], Helmer Strik [b], Erkki Vilkman [c]

[a] *Department of Applied Physics, University of Turku, FIN-20014 Turku, Finland*
[b] *Department of Language and Speech, University of Nijmegen, P.O. Box 9103, NL-6500 HD Nijmegen, The Netherlands*
[c] *Department of Otolaryngology and Phoniatrics, University of Oulu, FIN-90220 Oulu, Finland*

Received 14 November 1996; revised 10 April 1997; accepted 28 April 1997

Abstract

This study presents a new frequency domain parameter, Parabolic Spectral Parameter (PSP), for the quantification of the glottal volume velocity waveform. PSP is based on fitting a parabolic function to the low-frequency part of a pitch-synchronously computed spectrum of the estimated glottal flow. PSP gives a single numerical value that describes how the spectral decay of an obtained glottal flow behaves with respect to a theoretical limit corresponding to maximal spectral decay. By analyzing speech signals of different phonation types the performance of the new parameter is compared to three commonly used time-based parameters and to one previously developed frequency domain method. © 1997 Elsevier Science B.V.

Zusammenfassung

In dieser Studie wird zur Quantifizierung des glottalen Wellenform ein neuer Parameter vorgelegt: der parabolische Spektralparameter (PSP). PSP ist basiert auf einem Näherungsverfahren, in welchem eine parabolische Funktion dem tieffrequenten Anteil eines pitch-synchron berechneten Spektrums des glottalen Luftstroms angenähert wird. PSP resultiert in einem einzigen numerischen Wert, der angibt, wie der sprektrale Abfall des so erhaltenen glottalen Luftstroms sich im Hinblick auf eine theoretisch festgelegte Grenze, die dem maximalen spektralen Abfall entspricht, verhält. Durch Analyse von Sprachsignalen mit unterschiedlichen Phonationstypen wird die Wirkungsweise des neuen Parameters mit der von drei gebräuchlichen zeitabhängigen Parametern und mit einer vorher entwickelten Methode im Frequenzbereich verglichen. © 1997 Elsevier Science B.V.

Résumé

On présente ici un nouveau paramètre de fréquence, PSP (Parabolic Spectral Parameter), pour la quantification de la vélocité de volume de l'onde glottique. PSP est basé sur l'adaptation d'une fonction parabolique à la partie basse-fréquence du spectre pitch-synchrone du flux glottique estimé. PSP donne une valeur numérique qui décrit comment la décroissance spectrale d'un flux glottique obtenu se comporte par rapport à la limite théorique correspondant à la décroissance spectrale

* Corresponding author. Mailing address: Helsinki University of Technology, Acoustics Laboratory, P.O. Box 3000, FIN-02015 TKK, Finland.

0167-6393/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved.
PII S0167-6393(97)00020-4

maximale. Les performances de ce nouveau paramètre, pour l'analyse de signaux de parole caractéristiques de différents types de phonation, sont comparées à celles de trois paramètres d'usage courant, basés sur le temps, ainsi qu'à une méthode fréquentielle développée antérieurement. © 1997 Elsevier Science B.V.

Keywords: Voice production

1. Introduction

In the analysis of voice production inverse filtering has become a technique that is widely applied. Extracting information from the voice source with inverse filtering usually consists of two stages. First, an estimate for the glottal volume velocity airflow is computed either from the speech pressure waveform (e.g., Matausek and Batalov, 1980; Childers and Lee, 1991; Strik and Boves, 1992; Alku and Vilkman, 1994) or from the oral flow recorded by a pneumotachometric mask (frequently called Rothenberg's mask) (e.g., Rothenberg, 1973; Holmberg et al., 1988, 1989; Hertegård et al., 1990, 1992). The second stage of an application of inverse filtering is called the parametrization of the glottal airflow waveform. The goal of the parametrization is to present the obtained glottal waveforms in a compressed form using one or few numerical values that characterize the behaviour of the voice source. Parametrization allows the analysed voices to be categorized according to, for example, the type and loudness of phonation (e.g., Gauffin and Sundberg, 1989) and it can also be applied in the diagnosis and treatment of voice disorders (e.g., Hillman et al., 1990).

A large number of different methods have been developed for the parametrization of the glottal waveform. One of the most widely used methods to characterize the voice source is to apply time-based parameters, i.e., certain ratios between the closed phase, the opening phase and the closing phase of the glottal volume velocity waveform. Applying time-based parameters usually involves computing the following three ratios (e.g., Holmberg et al., 1988): (1) open quotient (OQ), which is defined as the ratio between the length of the glottal open phase and the length of the fundamental period; (2) speed quotient (SQ), which is defined as the ratio between the lengths of the glottal opening and closing phases; and (3) closing quotient (CQ), which is defined as the ratio between the length of the closing phase and the length of the fundamental period. Accurate computation of the time-based parameters is known to be problematic (Holmberg et al., 1988; Dromey et al., 1992). Time instants of glottal opening and closure might be difficult to extract exactly due to formant ripple and noise that is present in the glottal waveforms given by inverse filtering. Even in the absence of formant ripple, computation of OQ and SQ is difficult because of the gradual opening of the vocal folds. As a result of the problems mentioned above, computation of the time-based parameters is sometimes performed by replacing the true time instants of the glottal opening and closure by the time instants when the glottal flow crosses a level which is set to a certain ratio (e.g., 50%) of the difference between the maximum and minimum amplitude of the glottal cycle (Dromey et al., 1992).

Also the first derivative of the glottal airflow waveform has been used for time-domain quantification of voice production (e.g., Price, 1989; Fant, 1993; Sundberg et al., 1993). If inverse filtering is performed on a flow signal recorded by Rothenberg's mask it is possible to calibrate the amplitude level of the obtained glottal waveform to correspond with the real flow that is generated by the airstream passing the vibrating vocal folds (Rothenberg, 1973). The amplitude of the minimum peak of the flow derivative has been applied extensively in the parametrization of voice production (e.g., Gauffin and Sundberg, 1989; Fant, 1993; Sundberg et al., 1993). The negative peak of the differentiated glottal waveform obtained by inverse filtering the oral flow that was recorded by Rothenberg's mask has been shown to correspond closely with the sound pressure level (SPL) (Gauffin and Sundberg, 1989).

An approach which is widely applied in the quantification of voice production is to fit certain mathematical functions to the time domain waveforms given by inverse filtering. Among the developed voice source models one of the most frequently used is the Liljencrants–Fant model (LF-model) (Fant et al., 1985). In the LF-model the first derivative of the

glottal airflow waveform is presented by cosine and exponential functions that are defined by four parameters. The LF-model has been used not only for voice analysis purposes (e.g., Karlsson, 1990; Strik and Boves, 1992) but also for speech synthesis (Carlson et al., 1991). Application of the LF-model and its transformed version has been recently studied by focusing on the frequency domain parametrization of voice production (Fant, 1995).

Parametrization of the glottal flow using a frequency domain approach has been used, for example, by Childers and Lee (1991). They presented a quotient, called harmonic richness factor (HRF), which measures the decay of the voice source spectrum. HRF is defined from the spectrum of the estimated glottal flow as the ratio between the sum of the amplitudes of harmonics above the fundamental and the amplitude of the fundamental. Applying the spectral harmonics of the glottal airflow waveform in the quantification of voice production has also been used by Howell and Williams (1988, 1992), who measured the decay of the voice source spectrum by computing linear regression analysis over the first eight harmonics. Titze and Sundberg (1992) analyzed the spectral decay of the voice source of singers by computing the difference between the amplitude of the fundamental and the second harmonic.

It will be shown in this study that some of the existing parametrization methods of the glottal flow do not always succeed in characterizing different phonation types correctly. Therefore, our goal was to develop a new parameter that could quantify the glottal flow reliably in the case when phonation is changing. In order to avoid the problematic extraction of time-domain waveforms of the glottal source we decided to use a frequency domain approach in developing the new parameter. The new parameter, called Parabolic Spectral Parameter (PSP), is based on the quantification of the spectral decay of the voice source. Parametrizing the glottal flow by measuring the spectral decay is justified since it is known that frequency domain characteristics of the glottal source are affected by the phonation type: the breathier the phonation type the larger the roll-off of the voice source spectrum (e.g., Gauffin and Sundberg, 1989; Childers and Lee, 1991; Fant, 1995). Our goal was also to develop a parameter that gives only a single numerical value for the characterization of the glottal flow. To be able to quantify the glottal flow using a single parameter is practical especially if a large amount of speech material is to be studied. Accurate fit to the voice source spectrum is not possible to be obtained over a wide frequency range (e.g., from 0 to 4 kHz) using only a single parameter. Therefore, in PSP the quantification of the decay of the voice source spectrum is computed over the lowest frequency range of the spectrum that contain spectral values of the largest energy.

2. Speech material and inverse filtering

2.1. Speech material

The speech material that was used in order to compare the performance of PSP with other parametrization techniques consisted of voices produced by five female and five male speakers. None of the subjects had a history of hearing or voice disorders. The age of the subjects varied between 29 and 52 years for female speakers and between 32 and 47 years for males. The speakers produced sustained /a/-vowels lengths of which were at least 2 seconds using breathy, normal and pressed phonation. Subjects were allowed to use their natural fundamental frequency and loudness during the recording. The speakers were first trained to produce the vowels with three different phonation types by mimicking an experimenter. During the recording voice quality was assessed by a phoniatrician who asked the subject to repeat the speaking task until phonation was satisfactory. Recording of the signals was performed in an anechoic chamber using a condenser microphone (Brüel & Kjær 4133 together with a Brüel and Kjær 2636 preamplifier) which was positioned 40 cm from the lips of the speaker. Speech data was saved onto a digital tape using a DAT-recorder (TEAC RD-200T). The sampling frequency in the A/D-conversion was 22.050 kHz.

After recording the speech signals an informal listening test was made in order to find out whether the voices were perceived to belong to the three different phonation types. Three phonation types of each speaker were played to a panel in random order. Each panelist was asked to mark the order of the

phonation type. The panel consisted of five members, all of whom had experience in voice research. The result of this experiment was clear: all the members of the panel sorted the phonation types of each speaker correctly. The judgments of the panel were thus in line with those of the phoniatrician. Hence, we can conclude that all the subjects succeeded in producing the phonation types as required.

2.2. Inverse filtering

Parametrization of the voice source requires estimation of the glottal volume velocity waveform. This is most often done by applying inverse filtering. The obtained glottal waveforms given by an inverse filtering technique are then used as an input for parametrization. Our scheme assumes that glottal pulseforms are obtained on an arbitrary amplitude scale. This implies that all the parametrization techniques discussed in this study can be computed without applying a flow mask in inverse filtering. Estimating the glottal source from the acoustic speech pressure waveform without using a flow mask can be justified for two reasons. Firstly, applying a flow mask might interfere with natural voice production. Secondly, the bandwidth of Rothenberg's mask is not sufficient for reliable quantification of voices that have been produced using high fundamental frequency and pressed phonation (Hertegård and Gauffin, 1992; Alku and Vilkman, 1995).

Parametrization of the glottal flow was analyzed in our study by using the new PSP-approach and four previously-developed quantification techniques. Estimation of the glottal airflow waveforms was performed in the present study with an inverse filtering technique that is described in detail in Alku and Vilkman (1994). This inverse filtering technique applies the acoustic speech pressure waveform that has been recorded in a free field for estimation of the voice source, i.e., no flow mask is required. The developed method is based on modeling the vocal tract transfer function with an all-pole filter which is determined using a sophisticated algorithm called Discrete All-pole Modeling (DAP) (El-Jaroudi and Makhoul, 1991). The DAP-technique is able to estimate formants of the vocal tract more accurately than the conventional linear predictive coding (LPC) which is usually applied in automatic inverse filtering. Hence, the estimated glottal airflow waveforms are less distorted by formant ripples.

The sampling frequency of the signals was first decreased from the original value of 22.050 kHz to 8.0 kHz. In order to avoid aliasing, all the signals were low-pass filtered before the down-sampling with a linear phase FIR-filter, that had its cut-off frequency at 4.0 kHz. Signals were then high-pass filtered with a linear phase FIR-filter using a cut-off frequency of 50.0 Hz in order to remove any possible low frequency air pressure variations picked up during the recordings. Inverse filtering was computed by estimating the vocal tract transfer function with an all-pole filter the order of which was 8, 10 and 12 in the analysis of breathy, normal and pressed phonation type, respectively. Estimation of the glottal flow was computed using a block length of 32 ms together with Hamming windowing. The position of the analysis window was varied in order to find a setting that yielded a waveform with minimum formant ripple.

3. Methods of parametrizing the glottal flow

3.1. Background

Our new frequency domain technique for the parametrization of the voice source is based on the application of a pitch-synchronous spectrum, i.e., a Fourier-transform is computed over a single period of the glottal volume velocity waveform (Oppenheim and Schafer, 1975). The waveform of the pitch-synchronous spectrum does not contain a harmonic structure. Thus, by utilizing the pitch-synchronous spectrum, an inherent difference in methodology exists when compared to previously-developed frequency domain techniques (e.g., Childers and Lee, 1991; Titze and Sundberg, 1992) that quantify the decay of the voice source spectrum by computing the ratio between the level of the harmonics and the level of the fundamental. This implies that the spectrum has to be computed over several fundamental periods, i.e., using a pitch-asynchronous approach. In the analysis of soft voices or in the case of breathy phonation the spectrum of the voice source is characterized by a strong fundamental with extensively damped harmonics (e.g., Fant, 1995). Irregular periodical structure of the glottal flow also causes an

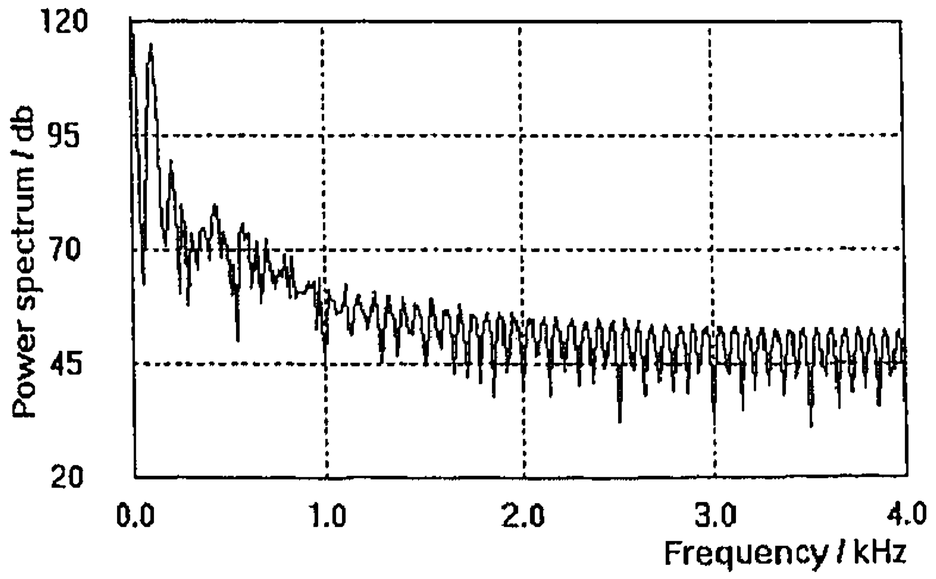

Fig. 1. Pitch-asynchronous spectrum of a glottal waveform in breathy phonation.

increased dynamic range between the level of the fundamental and the levels of the upper harmonics. Parametrizing the glottal source using spectral harmonics in the case when the spectrum has a very strong fundamental with respect to other spectral components implies that information is extracted from frequency samples that contain noise. This phenomenon is described in Fig. 1, which shows a pitch-asynchronous spectrum computed from a glottal airflow waveform that was produced by a male subject using a very breathy phonation type. Applying noisy information of extensively-damped harmonics makes parametrization of this voice source unreliable. As an example of a parameter that is based on the application of spectral harmonics we computed the value of HRF (Childers and Lee, 1991) for the glottal waveform the spectrum of which is shown in Fig. 1. The computation was done twice by shifting the position of the analysis window by 25 ms. The value of HRF on a linear scale was 0.05 for the first experiment and 0.07 for the second. In other words, the HRF-value increased by 40% when analyzing the same signal in two segments that were located only about three fundamental periods from each other. This occurred even though the speech signal was produced by a healthy subject using sustained phonation and constant F_0. Hence, large variation in the value of HRF was not caused by changes in subject's speech production but merely by the method that was used in the extraction of the frequency domain information.

3.2. Parabolic spectral parameter

PSP is described in the following sections. Section 3.2.1 first presents a sub-routine that is needed in order to compute the optimal parabolic match to a given power spectrum. A detailed algorithm is then given in Section 3.2.2 for the computation of PSP.

3.2.1. Parabolic matching

The glottal volume velocity waveform is of a low-pass nature (e.g., O'shaughnessy, 1987). The starting point in developing the new frequency domain parameter was to exploit the strongest frequency domain components of the glottal source extracted from the pitch-synchronously computed low-pass spectrum. By analyzing a large number of glottal airflow waveforms produced by different speakers and phonation types, we found that the pitch-synchronously computed spectrum of the glottal pulseform on a logarithmic scale has a close resemblance to the parabolic function at low frequencies. By fitting a parabola to the pitch-synchronous spectrum it is possible to model accurately the power spectrum of the voice source at low frequencies. The steepness of the parabola is determined by one parameter. Hence, by applying a parabolic function we are able to model the spectral decay of the voice source with a single parameter the optimal value of which can be easily found. It is important to notice that by using a parabolic function it is not possible to accurately match the spectrum of the glottal airflow over a wide frequency range. However, if matching is done using a frequency range that spans the monotonically decreasing low pass spectrum of the glottal source from zero to an upper limit we are able to model the frequency characteristics of one glottal pulse accurately with a parabolic function. (This implies, as shown by examples of Fig. 4, that the parabolic function is fit to the "main lobe" of the pitch-synchronous low pass spectrum of the voice source.) The selection of the upper limit of the frequency range is determined adaptively for each glottal waveform that is to be analysed. This adaptive computation increments the width of the frequency range over which the parabolic matching is done until the error between the original spectrum and its parabolic model exceeds a certain limit.

In addition to the properties mentioned above PSP has one feature which, as far as the authors know, is not used in any of the previously-developed parametrization methods. Namely, by using parabolic modeling it is possible, as will be described in the

next paragraph, to take into account the effect of the fundamental frequency on the spectral decay of the glottal source. Let us, however, first discuss, how F_0 is related to the decay of the voice source spectrum by analysing two synthetic glottal pulses shown in Fig. 2. The pulse of Fig. 2(a) corresponds to a typical glottal source produced by a male speaker ($F_0 = 100$ Hz), whereas the pulse of Fig. 2(b) depicts a synthetic version of a glottal flow created typically by a female subject ($F_0 = 200$ Hz). Both of the glottal pulses correspond to normal phonation. (Time-based parameters OQ, SQ and CQ of the first synthetic pulse were set to the same values as in the second pulse.) Pitch-synchronous spectra of the glottal pulses are shown in Fig. 3. It can be clearly seen from Fig. 3 that the spectral decays of the glottal flows of Fig. 2 are different: the spectrum of the glottal pulse corresponding to the male subject decays faster than the voice source spectrum of the female speaker. This phenomenon results from the fact that the waveform of the glottal flow in the case of female speech changes more rapidly than in the case of the male voice due to different lengths of the glottal cycle. If quantification of the phonation type is based on measuring the absolute spectral decay of the voice source spectrum, the two glottal pulses of

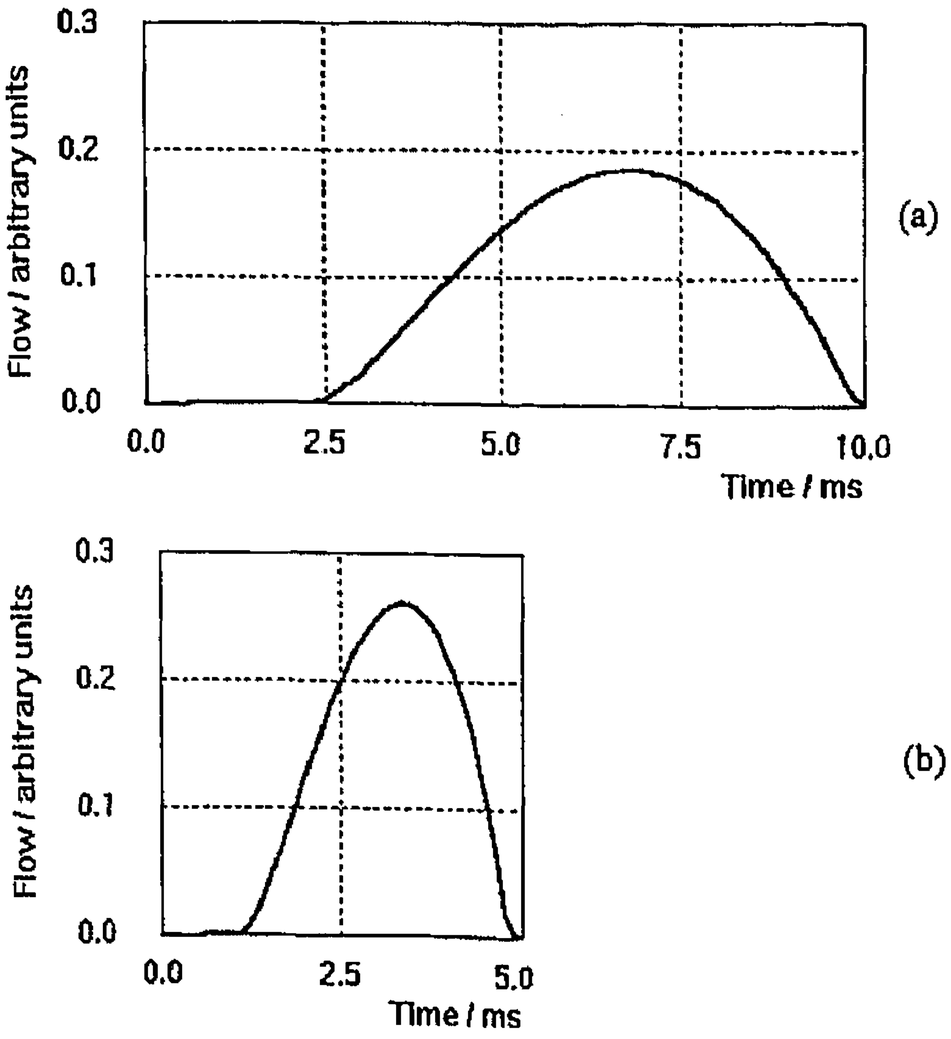

Fig. 2. One period of a synthetic glottal flow in normal phonation. (a) $F_0 = 100$ Hz, (b) $F_0 = 200$ Hz.

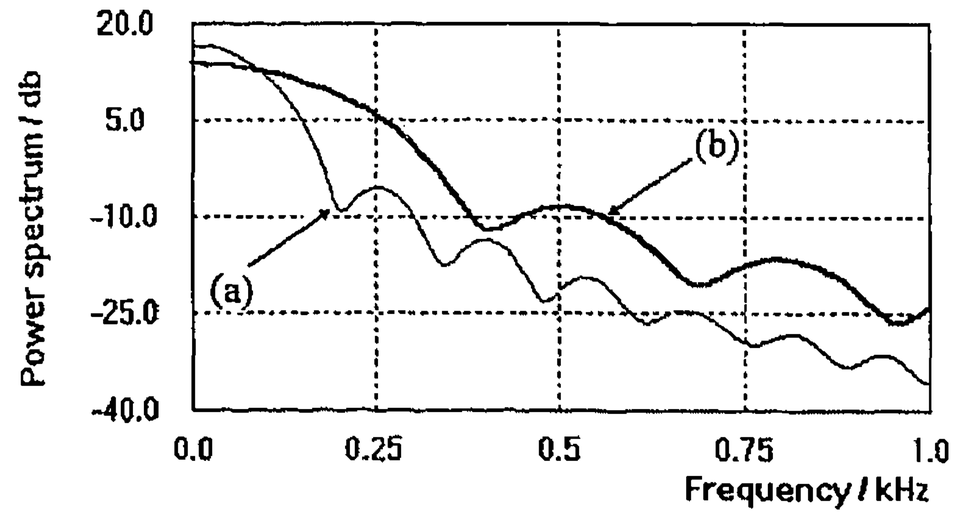

Fig. 3. Spectra of the glottal pulses shown in Fig. 2. Curve (a): $F_0 = 100$ Hz; curve (b): $F_0 = 200$ Hz.

Fig. 2 will be categorized to belong to different phonation types. In our opinion this is an incorrect interpretation of spectral information since the shape of the two glottal pulses of Fig. 2 is the same and both pulses correspond to normal phonation. The difference between these two glottal pulses is in their lengths of the glottal cycle. Therefore, the following question arises: Would it be possible to normalize the effect of the fundamental frequency in quantifying the spectral decay of the glottal flow?

The effect of the fundamental frequency on the spectral decay of the glottal flow is taken into account in our new parametrization technique by measuring with the parabolic function not only the spectral decay of the glottal flow obtained from natural speech but also the spectral decay of two hypothetical source waveforms: a DC flow and an ideal impulse. The lengths of both of these hypothetical glottal sources equal the length of the glottal pulse that is to be analysed. DC-flow has a power spectrum the shape of which, when computed by the Fast Fourier Transform (FFT), is given by the square of the sinc-function (Oppenheim and Schafer, 1975). This forms a spectrum with maximal decay. An ideal impulse has a power spectrum which is constant, i.e., the spectral decay is zero. The spectral decay of a glottal pulse computed from natural speech has to be somewhere between the extreme limits given by the two hypothetical glottal sources. When the fundamental frequency of the glottal flow increases (i.e., the length of the period in the glottal flow decreases) the following occurs for the spectra of the hypothetical glottal sources: the spectrum of the impulse remains the same, but the spectral decay of the

DC-flow (i.e., sinc^2-function) decreases. In other words the maximal spectral decay given by the DC-flow is dependent on the fundamental frequency. Therefore, instead of measuring the absolute spectral decay of the glottal flow it is justified to measure the relative spectral decay by normalizing the decay of the voice source spectrum computed from natural speech with respect to the maximal spectral decay given by the hypothetical source (DC-flow) of the same fundamental frequency. Hence, with the PSP-approach we are able to compare glottal sources in terms of their spectral decay even though the voices may have different fundamental frequencies. PSP is therefore a fundamental frequency invariant analysis method.

In the computation of the proposed PSP-parameter a parabolic function is matched to a discrete spectrum of the glottal volume velocity waveform by applying the mean square error criterion. Optimization of the parabolic function is derived as follows.

Let us denote the discrete spectrum to be modelled by $X(k)$. The parabolic function that is used to estimate $X(k)$ is denoted by $Y(k)$. The expression for the parabolic function is as follows: $Y(k) = ak^2 + b$. The square of the error, denoted by E, between the original spectrum and the parabolic function is minimized over a discrete frequency interval where

$$E = \sum_{k=0}^{N-1} \left(X(k) - ak^2 - b \right)^2. \tag{1}$$

The optimal parabolic function is obtained by setting the derivatives of E with respect to a and b to zero:

$$\frac{\partial E}{\partial a} = \sum_{k=0}^{N-1} -2\left(X(k) - ak^2 - b \right)k^2 = 0, \tag{2}$$

$$\frac{\partial E}{\partial b} = \sum_{k=0}^{N-1} -2\left(X(k) - ak^2 - b \right) = 0. \tag{3}$$

Eq. (3) yields the following expression for the optimal value of parameter b:

$$b = \frac{1}{N} \sum_{k=0}^{N-1} \left(X(k) - ak^2 \right). \tag{4}$$

By substituting b from Eq. (4) to Eq. (2) the following expression is obtained for the optimal value of parameter a:

$$a = \frac{N \sum_{k=0}^{N-1} X(k)k^2 - \left[\sum_{k=0}^{N-1} X(k) \right] \left[\sum_{k-0}^{N-1} k^2 \right]}{N \sum_{k=0}^{N-1} k^4 - \left[\sum_{k=0}^{N-1} k^2 \right]^2}. \tag{5}$$

According to Eq. (5) the optimal value of parameter a depends on N, i.e., the frequency range over which the parabolic function matches the pitch-synchronous spectrum of the glottal waveform. In order to treat glottal sources of different spectral characteristics equally with the parabolic function, the value of N is computed adaptively for each signal. The procedure is based on expanding the frequency range in which the matching of the voice source spectrum with the parabolic function is performed until the normalized error NE between the parabolic function and the voice source spectrum exceeds a certain limit. Preliminary experiments revealed that 0.01 is a suitable value for this limit [1]. The adaptive computation of N together with solving the optimal value of parameter a is computed by the following iteration:

1. Start with an initial value of N that equals 3. (This is the smallest number of data samples that can yield a non-zero value for the energy of the error expressed in Eq. (1))

[1] In the PSP-computation the parabolic function is used for matching two kinds of spectra (see Section 3.2.2): (1) those computed from the glottal volume velocity waveforms obtained from real speech using, for example, inverse filtering, and (2) those computed from hypothetical glottal sources with maximum spectral delay, i.e., DC-flows. The latter has a power spectrum given by the sinc^2-function (Oppenheim and Schafer, 1975). The matching of the logarithm of the sinc^2-function with the parabolic function can be accurate only over the monotonically decreasing main lobe of the sinc-function, i.e., over a frequency interval from 0 to F_0. Hence, the lower the fundamental frequency the narrower the frequency range of the hypothetical glottal source in which the use of parabolic matching is justified. By creating a hypothetical glottal source that corresponds to a male voice with a very low fundamental frequency ($F_0 = 80$Hz), we matched an optimal parabolic function to the corresponding spectrum over the main lobe. The NE-value was approximately 0.04. Hence, by using the limit of 0.01 we can be sure that parabolic matching will occur inside the main lobe of the sinc^2-function for all hypothetical glottal sources that are needed in the computation of PSP.

2. Compute the optimal values of a (Eq. (5)) and b (Eq. (4)).
3. Compute the normalized error (NE) by using the given value of N and the obtained optimal values of a and b:

$$\mathrm{NE} = \frac{\sum_{k=0}^{N-1}\left(X(k) - ak^2 - b\right)^2}{\sum_{k=0}^{N-1}X(k)^2}. \quad (6)$$

4. If NE < 0.01 increment N by one and go back to stage 2, otherwise exit the iteration.

The value of a that is obtained when the iteration above terminates is used in PSP to measure the decay of spectrum $X(k)$.

3.2.2. Computation of parabolic spectral parameter

By using the sub-routine that was presented at the end of Section 3.2.1 we can now define the method used to compute PSP. Parametrization of the glottal airflow waveform with PSP contains the following main stages:

1. Compute an estimate for the glottal volume velocity waveform.
2. Cut one glottal cycle of the obtained pulseform. In order to avoid spurious peaks appearing in the FFT-spectrum calculated later in stage 5, the cutting should span one period of the glottal flow from a time instant of the minimum amplitude within a glottal cycle to the corresponding time instant in the next glottal cycle. A straightforward way to do this is to span the cutting between two consecutive time instants of glottal closure. The cutting of the period should be done with care so that the amplitude in the beginning and at the end of the cut period is the same. The length of the period (in seconds) that was cut corresponds to the fundamental period and is denoted by T_0 in the following.
3. Adjust the minimum amplitude of the glottal flow to zero by subtracting from the flow its minimum value. The obtained period of the glottal flow is in the following denoted by $g_p(n)$.
4. Scale the energy of $g_p(n)$ to unity.
5. Compute the pitch-synchronous power spectrum for $g_p(n)$ using the FFT algorithm (Oppenheim and Schafer, 1975). Express the spectrum on the dB-scale. The spectrum should be computed with rectangular windowing. In order to provide enough spectral samples at low frequencies, the size of the FFT should be increased by zero-padding signal $g_p(n)$. The experiments reported in this study were computed using the FFT-size of 2048 samples. The power spectrum given by the FFT is denoted in the following by $X(k)$.
6. By using $X(k)$ compute the optimal value for the parabolic parameter a as explained in Section 3.2.1.
7. Repeat stages 4–6 for a signal with the amplitude equal to unity and length equal to T_0. (i.e., the hypothetical DC-flow). The obtained value for the parameter a given by stage 5 is denoted $a_{\max}$ and serves as a measure for the maximal spectral decay.
8. Finally, the PSP-value is obtained by normalizing the measured spectral decay of the analysed glottal flow with respect to the theoretical maximal spectral decay. Normalization is computed by dividing the a-value which was determined from the analyzed glottal flow with the a-value determined from the DC-flow (i.e., $a_{\max}$):

$$\mathrm{PSP} = \frac{a}{a_{\max}}. \quad (7)$$

Notice that the a-value of the second hypothetical glottal source, the impulse, does not have to be computed because it is always equal to zero. Hence, the difference between the maximal and minimal spectral decay when quantified using the parabolic parameter a is as follows: $a_{\max} - a_{\min} = a_{\max} - 0 = a_{\max}$. Therefore the normalization procedure described by Eq. (7) actually implies dividing the spectral decay of the analysed glottal flow by the difference between the maximal and minimal spectral decay that can be achieved with a given length of the fundamental period. It should also be noted that both the spectrum of the glottal flow computed from a natural vowel and the spectrum of the corresponding hypothetical DC-flow are of a low-pass nature. Hence, the parabolic modeling yields for both of them a negative a-value. Therefore, the value of PSP, being a ratio of two negative numbers, will be positive.

Fig. 4 depicts two examples of the PSP-computation that were obtained from glottal waveforms with largely different spectral characteristics. A pitch-synchronous spectrum of the voice source that was

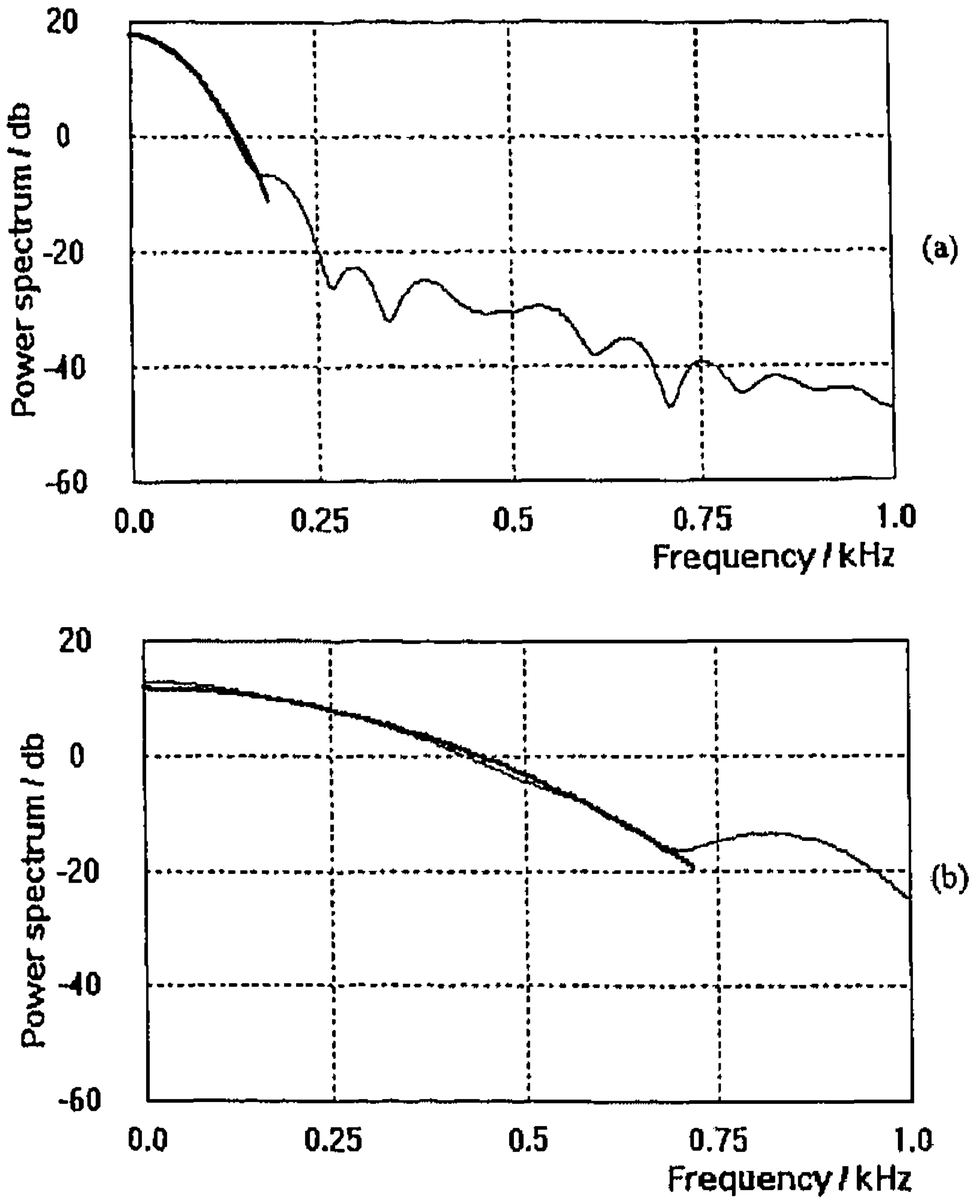

Fig. 4. Pitch-synchronous spectrum of a glottal waveform (thin line) and the optimal parabolic match (thick line). (Inverse filtering was computed for all the signals using a bandwidth of 4 kHz. Examples in this figure are shown for the frequency range of 1 kHz in order to distinguish the two curves.) (a) Male speaker, breathy phonation, $F_0 = 90$ Hz; (b) female speaker, pressed phonation, $F_0 = 190$ Hz.

produced by a male speaker using breathy phonation is shown in Fig. 4(a) (thin curve). The spectral decay of this glottal source is large, which is shown by the parabolic function (thick curve in Fig. 4(a)) that decreases rapidly. Fig. 4(b) shows a pitch-synchronous spectrum (thin curve) which was computed from a glottal waveform obtained from a female voice with pressed phonation. Slow spectral decay of this voice source can now be seen from the parabolic function (thick curve in Fig. 4(b)) that is clearly of a smaller steepness in comparison to Fig. 4(a). Notice that the parabolic function spans a different frequency range in Fig. 4(a,b).

3.3. Other methods of parametrization

In order to compare PSP with other parametrization techniques the obtained glottal waveforms were characterized by three time-based parameters and by one frequency domain parameter. Time-domain quantification was performed using the open quotient (OQ), the speed quotient (SQ) and the closing quotient (CQ) (e.g., Holmberg et al., 1988). The parameter that we used in the frequency domain quantification of the voice source was the harmonic richness factor (HRF) (Childers and Lee, 1991). (Among the frequency domain methods that are based on the application of the spectral harmonics we selected HRF and not, for example, methods used by Titze and Sundberg (1992) or Howell and Williams (1988, 1992). The reason why we selected HRF is that this parameter was presented by Childers and Lee (1991) together with an experiment that was quite similar to ours. In addition, HRF takes into account a larger number of spectral harmonics than, for example, the method used by Titze and Sundberg (1992).)

In order to compute values for the time-based parameters the obtained glottal waveforms were analyzed by an experimenter who marked time instants that were required for computation of OQ, SQ and CQ (i.e., time of glottal opening, time of maximal flow and time of glottal closure). Among these critical time instants extraction of the glottal opening is the most problematic. Our procedure for the determination of the glottal opening was to first look for that time instant after the previous glottal closure when the flow clearly started to ascend. For some of the signals the shape of the glottal waveform was very smooth during the glottal closed phase and there was no clear indication of the opening instant. For these signals the glottal opening was determined by searching for the first amplitude value after the glottal closure which was at least 5% of the difference between the maximum and minimum amplitude of the glottal cycle. This procedure decreases variability of OQ-values that is caused by subjective criteria of the experimenter in defining the time-instant of glottal opening. The limit of 5% was used because we wanted to measure the "true" time-based parameters and not the so called quasi-parameters that apply much larger (e.g., 50%) amplitude limits in defining critical time-instants (Dromey et al., 1992).

4. Results

The obtained values for all five analyzed parameters are given in Tables 1 and 2 for female and male

Table 1
Obtained parameters for female speakers (F_0: fundamental frequency, OQ: open quotient, SQ: speed quotient, CQ: closing quotient, HRF: harmonic richness factor, PSP: parabolic spectral parameter), the unit of F_0 is Hz, whereas the rest of the parameters are given in dimensionless units

Speaker	Phonation	F_0	OQ	SQ	CQ	HRF	PSP
F1	breathy	182	1.0	1.10	0.48	0.16	0.41
	normal	176	0.91	2.44	0.26	0.56	0.36
	pressed	174	0.81	2.50	0.23	0.67	0.16
F2	breathy	192	0.98	1.37	0.42	0.20	0.36
	normal	183	0.77	1.97	0.26	0.46	0.21
	pressed	195	0.61	1.78	0.22	0.99	0.07
F3	breathy	240	0.97	1.20	0.44	0.14	0.35
	normal	233	0.85	1.38	0.36	0.36	0.21
	pressed	235	0.87	1.70	0.32	0.74	0.13
F4	breathy	211	0.91	1.42	0.38	0.14	0.32
	normal	195	0.83	1.83	0.29	0.39	0.16
	pressed	192	0.74	1.91	0.26	0.64	0.10
F5	breathy	163	0.82	1.79	0.29	0.35	0.21
	normal	174	0.84	1.90	0.29	0.59	0.19
	pressed	186	0.86	2.08	0.28	0.66	0.16

speakers, respectively. Parameter values obtained for all the individual speakers are shown in the tables together with F_0-information. OQ, SQ, CQ and PSP are expressed as mean values which were computed over five consecutive glottal periods. HRF was computed using a pitch-asynchronous Fourier-transform with Hamming-windowing that spanned the same five glottal periods. The main findings are discussed in the following by dividing the parameters into three groups: the time-based parameters, the frequency domain parametrization with HRF, and the frequency domain parametrization with PSP.

Table 2
Obtained parameters for male speakers (F_0: fundamental frequency, OQ: open quotient, SQ: speed quotient, CQ: closing quotient, HRF: harmonic richness factor, PSP: parabolic spectral parameter), the unit of F_0 is Hz, whereas the rest of the parameters are given in dimensionless units

Speaker	Phonation	F0	OQ	SQ	CQ	HRF	PSP
M1	breathy	109	0.98	0.93	0.51	0.13	0.30
	normal	107	0.88	2.65	0.24	0.74	0.19
	pressed	111	0.42	1.30	0.18	1.58	0.08
M2	breathy	96	0.99	1.0	0.49	0.20	0.39
	normal	96	0.72	1.61	0.28	0.48	0.17
	pressed	94	0.59	2.37	0.18	0.95	0.13
M3	breathy	89	0.89	1.36	0.38	0.18	0.30
	normal	89	0.89	1.99	0.30	0.42	0.27
	pressed	94	0.89	2.51	0.25	0.76	0.11
M4	breathy	108	0.94	1.20	0.43	0.25	0.40
	normal	106	0.79	1.80	0.28	0.51	0.16
	pressed	107	0.81	2.37	0.24	0.69	0.13
M5	breathy	111	1.0	1.25	0.44	0.23	0.43
	normal	112	0.93	2.70	0.25	0.64	0.17
	pressed	118	0.80	2.33	0.24	0.58	0.16

4.1. Time-based parameters

The main trend according to which values of the time-based parameters changed when phonation was altered from breathy to pressed was in line with previous studies (e.g., Gauffin and Sundberg, 1989). This implies that the "average shape" of the glottal source for both female and male speakers was symmetric with a very short closed phase in breathy phonation. When the phonation type was changed towards pressed, the length of the closed phase increased and at the same time the length of the closing phase decreased. However, there was a large variation between different speakers in the way time-based parameters changed when phonation was altered. This was true especially for OQ and SQ, that take advantage of information given by the glottal opening. As an example, the value of OQ for male speaker M1 showed an extensive decrease from 0.98 to 0.42 when phonation was changed from breathy to pressed. However, the same change of phonation type caused the value of OQ for female subject F5 to increase from 0.82 to 0.86. OQ-values changed (decreased) monotonically for three female subjects and for three male subjects when phonation was altered from breathy to pressed. In the case of SQ a monotonic change (increase) of the parameter value occurred for four female subjects and for three male subjects. Among the time-based parameters the value of CQ showed the most consistent way to follow changes in the phonation type: all ten speakers except one female (F5) showed a monotonic decrease in their CQ-values.

4.2. Harmonic richness factor

The obtained values of HRF were generally in line with previous studies (Childers and Lee, 1991) according to which changing the phonation type

corresponds in the frequency domain to changing the spectral decay of the glottal excitation. The spectral decay was largest, i.e., the value of HRF was smallest, for all the subjects in the case of breathy phonation. The value of HRF changed (increased) monotonically when phonation was changed from breathy to pressed for all the subjects except one male speaker (M5).

4.3. Parabolic spectral parameter

The obtained PSP-values show that the speakers decreased the decay of their voice source spectrum when the phonation type was changed from breathy to pressed. When the spectral decay of the voice source is large the matching of the lower part of the spectrum gives a parabolic function which is steep as can be seen in Fig. 4(a). This implies that the PSP-value is large. However, a voice source spectrum that has larger high frequency components yields a parabolic function with a smaller decay, which is shown in Fig. 4(b). Consequently, the PSP-value will be smaller. The obtained data shows that the PSP-value decreased monotonically for all ten subjects when phonation was altered from breathy to pressed. Hence, PSP was the only one among the five computed parameters that gave quantitative information which was perfectly in line with the subject's task to change the phonation type.

The data required for the computation of PSP consists of one glottal cycle of the glottal volume velocity waveform. In order to analyse the stability of the new parameter from one glottal cycle to another we made an additional experiment where the PSP-computation was repeated for ten consecutive glottal periods. This computation was performed for the voices of each of the ten speakers and included the three phonation types, i.e., altogether this analysis consisted of thirty cases. Mean values and standard deviations were then computed over the obtained ten PSP-values for all the thirty speech segments. The stability of the PSP-parameter was analysed by computing the coefficient of variation (ν), i.e., the ratio between the standard deviation and the mean value. The obtained ν-values showed that the PSP-parameter behaved reliably from one period to another: ν-values varied between 1.1% and 8.4% and the mean of ν-values was 4.1%. Stability of the PSP-parameter was further emphasized by the following finding that was true for *all* ten speakers: *each* of the PSP-values computed from ten glottal periods in breathy phonation were larger than *any* of the PSP-values computed from ten periods in normal phonation, which in turn, were *all* larger than *any* of the PSP-values in pressed phonation.

5. Summary and discussion

In this paper we have studied a new frequency domain method, Parabolic Spectral Parameter, for parametrization of the glottal volume velocity waveforms that have been obtained by inverse filtering acoustic speech pressure signals. PSP is based on the extraction of information on speech production using a pitch-synchronous spectrum which is obtained by calculating the Fast Fourier Transform over one period of the glottal flow. The obtained power spectrum is expressed on a logarithmic scale. A parabolic function is then matched to the power spectrum over a frequency range where the normalized error energy between the original spectrum and its parabolic model is less than 0.01. Hence, the spectral decay of the glottal excitation can be quantified by a single numerical value. Computation of theoretical limits for maximal and minimal spectral decay of a glottal source with a known length of the fundamental period is also included in PSP. The final PSP-value is a number that expresses, using a parabolic function, the spectral decay of the analyzed voice source with respect to its maximal theoretical decay.

The authors believe that the new parameter is useful especially when analyzing glottal airflow waveforms with greatly different spectral characteristics. PSP takes advantage of information in the frequency domain using data samples of the largest energy in characterizing the low-pass behavior of the spectrum of the voice source. This is an inherent improvement in comparison to conventional time-based parameters (especially OQ and SQ), values of which are dependent on the extraction of a few time instants. The exact location of the critical time instants that are required for computation of time-based parameters is sensitive not only to the noise and formant ripple that is often present in the estimated glottal flows but also to subjective criteria in defin-

ing when glottal opening and closure takes place. In analyzing the glottal flows of different phonation types it should be noticed that changing the voice register primarily affects the spectral decay of speech (Childers and Lee, 1991). Therefore, it is justified to apply a frequency domain quantification approach in the parametrization of voice production when phonation type is changing. In comparison to previously-developed frequency domain methods PSP has two advantages. Firstly, application of information on extensively damped harmonics is avoided, which makes the analysis of voices with large spectral decay more accurate. Secondly, the new parameter allows a comparison of glottal flows in terms of their spectral decay, even in cases when the fundamental frequency of voices is different.

Since PSP generates only a single numerical value in parametrization of the glottal flow it cannot, of course, give a detailed quantification of the glottal source. In PSP the performance of the single numerical value is focused on those frequencies in the lower part of the voice source spectrum that have the largest energy. This implies, for example, that PSP is not able to distinguish between two glottal sources that have exactly the same spectral shape at low frequencies but different spectral characteristics at higher frequencies. This might be the case, for example, if two glottal pulses have the same length of the fundamental period and their length of the glottal open phase is also equal, but skewing of the pulses is different (i.e., the two pulses have the same values of F_0 and OQ but different values of SQ and CQ). It should also be noticed that PSP cannot be used to synthesize the time domain waveform of the glottal flow, which is the case in some parametrization techniques (e.g., the LF-model) that use more than one parameter in quantification of the voice source. Hence, PSP serves as a tool in voice analysis but cannot be applied directly in speech synthesis.

According to the present study PSP is a promising objective method to quantify the glottal source in the case when the phonation type is changed. An interesting area of future studies is to find out how information obtained by PSP correlates with subjective measurements of, for example, phonatory press and amount of breathiness in speech signals (Hillenbrand and Houde, 1996). We also believe that PSP could be used in studies where effects of eigonomic and environmental factors are analysed on voice loading (e.g., Ohlsson and Löfqvist, 1987; Vilkman et al., 1997). Analysis of pathological voices is also one possible area of future studies where PSP could be used. However, it should be noticed that if the periodical structure of voiced speech is poor, which often is the case in pathological voices, variation of PSP-values from one glottal period to another will most likely increase.

References

Alku, P., Vilkman, E., 1994. Estimation of the glottal pulseform based on discrete all-polmodeling. In: Proc. Internat. Conf. on Spoken Language Processing, Yokohama, Japan, 18–22 September 1994, pp. 1619–1622.

Alku, P., Vilkman, E., 1995. Effects of bandwidth on glottal airflow waveforms estimated by inverse filtering. J. Acoust. Soc. Amer. 98 (2), 763–767.

Carlson, R., Granström, B., Karlsson, I., 1991. Experiments with voice modelling in speech synthesis. Speech Communication 10 (5–6), 481–489.

Childers, D.G., Lee, C.K., 1991. Vocal quality factors: Analysis, synthesis, and perception. J. Acoust. Soc. Amer. 90 (5), 2394–2410.

Dromey, C., Stathopoulos, E.T., Sapienza, C.M., 1992. Glottal airflow and electroglottographic measures of vocal function at multiple intensities. J. Voice 6 (1), 44–54.

El-Jaroudi, A., Makhoul, J., 1991. Discrete all-pole modeling. IEEE Trans. Signal Process. 39 (2), 411–423.

Fant, G., 1993. Some problems in voice source analysis. Speech Communication 13 (1–2), 7–22.

Fant, G., 1995. The LF-model revisited. Transformations and frequency domain analysis. Speech Transmission Laboratory Quarterly Progress and Status Reports, No. 2–3, Royal Institute of Technology, Stockholm, Sweden, pp. 119–156.

Fant, G., Liljencrants, J., Lin, Q., 1985. A four-parameter model of glottal flow. Speech Transmission Laboratory Quarterly Progress and Status Reports, No. 4, Royal Institute of Technology, Stockholm, Sweden, pp. 1–13.

Gauffin, J., Sundberg, J., 1989. Spectral correlates of glottal voice source waveform characteristics. J. Speech Hear. Res. 32, 556–565.

Hertegård, S., Gauffin, J., 1992. Acoustic properties of the Rothenberg mask. Speech Transmission Laboratory Quarterly Progress and Status Reports, No. 2–3, Royal Institute of Technology, Stockholm, Sweden, pp. 9–18.

Hertegård, S., Gauffin, J., Sundberg, J., 1990. Open and covered singing as studied by means of fiberoptics, inverse filtering, and spectral analysis. J. Voice 4 (3), 220–230.

Hertegård, S., Gauffin, J., Karlsson, I., 1992. Physiological correlates of the inverse filtered flow waveform. J. Voice 6 (3), 224–234.

Hillenbrand, J., Houde, R.A., 1996. Acoustic correlates of breathy

vocal quality: Dysphonic voices and continuous speech. J. Speech Hear. Res. 39, 311–321.

Hillman, R.E., Holmberg, E.B., Perkell, J.S., Walsh, M., Vaughan, C., 1990. Phonatory function associated with hyperfunctionally related vocal fold lesions. J. Voice 4 (1), 52–63.

Holmberg, E.B., Hillman, R.E., Perkell, J.S., 1988. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. J. Acoust. Soc. Amer. 84 (2), 511–529.

Holmberg, E.B., Hillman, R.E., Perkell, J.S., 1989. Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch. J. Voice 3 (4), 294–305.

Howell, P., Williams, M., 1988. The contribution of the excitatory source to the perception of neutral vowels in stuttered speech. J. Acoust. Soc. Amer. 84 (1), 80–89.

Howell, P., Williams, M., 1992. Acoustic analysis and perception of vowels in children's and teenagers' stuttered speech. J. Acoust. Soc. Amer. 91 (3), 1697–1706.

Karlsson, I., 1990. Voice source dynamics of female speakers. In: Proc. Internat. Conf. on Spoken Language Processing, Kobe, Japan, 19–22 November 1990, Vol. 1, pp. 69–72.

Matausek, M.R., Batalov, V.S., 1980. A new approach to the determination of the glottal waveform. IEEE Trans. Acoust. Speech Signal Process. 28 (6), 616–622.

Ohlsson, A.-C., Löfqvist, A., 1987. Work-day effects on vocal behaviour in switchboard operators and speech therapists. Scand. J. Logop. Phoniatr. 12, 70–79.

Oppenheim, A.V., Schafer, R.W., 1975. Digital Signal Processing. Prentice-Hall, Englewood Cliffs, NJ.

O'shaughnessy, D., 1987. Speech Communication, Human and Machine. Addison-Wesley, Reading, MA.

Price, P.J., 1989. Male and female voice source characteristics: Inverse filtering results. Speech Communication 8 (3), 261–277.

Rothenberg, M., 1973. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. J. Acoust. Soc. Amer. 53, 1632–1645.

Strik, H., Boves, L., 1992. On the relation between voice source parameters and prosodic features in connected speech. Speech Communication 11 (2–3), 167–174.

Sundberg, J., Titze, I., Scherer, R., 1993. Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. J. Voice 7 (1), 15–29.

Titze, I., Sundberg, J., 1992. Vocal intensity in speakers and singers. J. Acoust. Soc. Amer. 91 (5), 2936–2946.

Vilkman, E., Lauri, E.-R., Alku, P., Sala, E., Sihvo, M., 1997. Loading changes in time-based parameters of glottal flow waveforms in different ergonomic conditions. Folia Phoniatrica et Logopaedica, in press.