# A FREQUENCY DOMAIN METHOD FOR PARAMETRIZATION OF THE VOICE SOURCE

*Paavo Alku[1] and Erkki Vilkman[2]*
1: University of Turku, Electronics and Information Technology,
FIN-20520, Turku, Finland
2: University of Oulu, Dept. Otolaryngology and Phoniatrics, FIN-90220, Oulu, Finland

## ABSTRACT

A new frequency domain method for the quantification of glottal volume velocity waveforms is presented in this study. This technique, called Parabolic Spectral Parameter (PSP), is based on fitting a parabolic function to a pitch-synchronously-computed spectrum of the estimated voice source. The PSP-algorithm gives a single numerical value that describes how the spectral decay of an obtained glottal flow behaves with respect to theoretical bounds corresponding to maximal and minimal spectral tilting.

## 1. INTRODUCTION

Inverse filtering is a widely applied technique in the analysis of voice production. Applying 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. 1, 2) or from the oral flow using Rothenberg's mask (e.g. 7, 8). Then the obtained voice source is described in a compressed form using one or few numerical values. This second stage of an inverse filtering survey is called the parametrization of the glottal airflow waveform.

A large number of different methods have been developed for the parametrization of the glottal waveforms. 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 (e.g., 8). Accurate computation of the time-based parameters is known to be problematic due to gradual opening of the vocal folds and also due formant ripple and noise that is often present in the glottal waveforms (3, 8).

The derivative of the glottal airflow waveform has been widely used for quantification of voice production (e.g. 6, 7). The negative peak of the differentiated glottal waveform obtained by inverse filtering the oral flow with Rothenberg's mask has been shown to correspond closely with the sound pressure level (SPL) (7).

Parametrization of the glottal airflow obtained by inverse filtering can also be performed in the frequency domain. Childers and Lee (2) presented a quotient, harmonic richness factor (HRF), which measures the decay of the voice source spectrum by computing the ratio between the sum of the amplitudes of harmonics above the fundamental and the amplitude of the fundamental. Titze and Sundberg (12) analyzed the spectral tilt of the voice source by computing the difference between the amplitude of the fundamental and the second harmonic.

A widely applied technique 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) (5). In the LF-model the derivative of the glottal airflow waveform is presented by cosine and exponential functions that are defined by four parameters.

In the present study we propose a new frequency domain parameter, called the Parabolic Spectral Parameter (PSP), for quantification of the glottal volume velocity waveform obtained by inverse filtering. The derivation of PSP is presented in the next chapter. Chapters 3 and 4 describe the function of PSP in the analysis of natural voices created by different phonation types. Finally, Chapter 5 concludes with characteristics of PSP in the parametrization of the glottal airflow waveform.

## 2. METHOD

Parametrization of the voice source requires the application of an inverse filtering technique to estimate the glottal volume velocity waveform. Our scheme assumes that glottal pulseforms are obtained on an arbitrary amplitude scale, i.e., no flow mask is required.

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 aim was to create a quantification method that gives only a single numerical value to describe voice production. Our technique is based on the application of pitch-synchronous spectrum, i.e. a Fourier-transform is computed over a single period of the glottal volume velocity waveform (9). Applying pitch-synchronous spectrum is a difference in comparison to previously-developed frequency domain techniques (e.g. 2, 11) that apply spectral information which is located on the fundamental and its harmonics, i.e., the spectrum is computed over several fundamental periods 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. For these signals parametrizing the glottal source using spectral harmonics implies that information is extracted from spectral samples that contain noise.

**Algorithm**

The pitch-synchronously-computed spectrum of the glottal pulseform on a logarithmic scale has a close resemblance to parabolic function at low frequencies. By fitting a parabolic function to the pitch-synchronous spectrum it is possible to model the frequency domain behavior of the voice source. Applying the parabolic function is straightforward and effective because this way we are able to use a model which contains only a single parameter whose optimal value can be easily found. Exact modeling of the spectrum of the glottal airflow on a wide frequency range is not possible using parabolic function. However, if matching is done using a frequency range that is narrow enough application of the parabolic function is justified. It is essential to focus this frequency range to include those frequencies that contain the most important information of the voice source.

In the PSP-algorithmn the spectral decay of a voice source is expressed on a normalized scale. This is possible by measuring with the parabolic function the spectral tilt of two hypothetical source waveforms: a DC flow and an ideal impulse. The former has a power spectrum whose shape, when computed by the Fast Fourier Transform (FFT), is given by the square of the sinc-function (9). This forms a spectrum with maximal decay. The latter has a power spectrum which is constant, i.e. the spectral decay is zero. Hence, modeling the source spectrum with the parabolic function makes it possible to normalize the spectral tilt of the voice source computed from natural voice with respect to its theoretical bounds. Therefore, with the PSP-technique we are able to compare glottal sources in terms of their spectral decay even though voices have different fundamental frequencies.

Parametrization of the glottal airflow waveform with the PSP-technique contains the following main stages:

(1.) Compute an estimate for the glottal airflow by applying an inverse filtering technique.

(2.) Cut one glottal cycle of the obtained pulseform. In order to avoid spurious peaks appearing in the spectrum, the cutting should span one period of the glottal flow between two consecutive time instants of glottal closure. The length of the period (in seconds) that was cut corresponds to the fundamental period and is denoted by T in the following.

(3.) Adjust the minimum amplitude of the glottal cycle to zero. The obtained period of the glottal flow is in the following denoted by $g_p(n)$.

(4.) Compute the pitch-synchronous power spectrum for $g_p(n)$ using the FFT-algorithm (9). Express the spectrum on dB-scale. The spectrum should be computed with rectangular windowing. The length of the window should be at least 512 samples to provide enough spectral samples at low frequencies. The obtained power spectrum is denoted in the following by S(f).

(5.) Compute the DC-level of the spectrum i.e. S(0). Search for the frequency when the spectrum falls below 20 dB of S(0). This frequency, denoted by $f_{lim}$, determines the frequency range which is considered to contain the most important information of the voice source.

(6.) Form a parabolic function $Y(f) = a_p f^2 + S(0)$. By iterating the value of from -1.0 to 0.0 using a step size of 0.001, find the value of $a_p$ that minimizes the following mean square error:

$$E = | Y(f) - S(f) |^2 , 0 \leq f \leq f_{lim}$$

The obtained optimal value of $a_p$ is denoted by $a_{opt}$ in the following.

(7.) Repeat stages 4-6 for a signal with a constant amplitude and length equal to T. The obtained value for parameter $a_p$ given by stage 6 is denoted $a_{opt,hyp}$ and serves as a measure for the maximal spectral decay. The minimal theoretical spectral decay corresponds to $a_{opt}$ equals to zero.

(8.) Finally, the PSP-value is obtained by normalizing the $a_{opt}$-value which was determined from the analyzed glottal airflow with $a_{opt}$-value of the corresponding hypothetical flow with maximal spectral decay:

$$PSP = a_{opt} / a_{opt,hyp}$$

## 3. EXPERIMENTS

**Speech material and inverse filtering**

The speech material that was used in order to test the performance of PSP consisted of voices produced by five female and five male speakers. The speakers produced a sustained /a/-vowel using breathy, normal, and pressed phonation types. The pitch was kept constant throughout the recording. Subjects were allowed to use their natural fundamental frequency and loudness during the recording. Recording of the signals was performed in an anechoic chamber using a condenser microphone (Brüel&Kjær 4133) which was held 40 cm from the lips of the speaker.

Estimation of the glottal airflow waveforms was performed in the present study with an inverse filtering technique that is described in (1). 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 of the vocal tract transfer function with an all-pole filter which is determined using a sophisticated algorithm, called Discrete All-pole Modeling (DAP) (4).

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.

**Other methods of parametrization**

In order to compare the proposed PSP-method with other parametrization techniques the obtained glottal waveforms were characterized by one time-based parameter and by one frequency domain parameter. Time-domain quantification was performed using the closing quotient (CQ), i.e. the ratio between the glottal closing phase and the lenght of the fundamental period (8). The parameter that we used in the frequency domain quantification of the voice source was the harmonic richness factor (HRF) (2).

## 4. RESULTS

The obtained values for all three analyzed parameters are given in Tables 1, 2, and 3 for breathy, normal, and pressed phonation, respectively. Parameter values obtained for all the individual speakers are shown in the tables together with F0-information.

The main trend according to which value of CQ changed when phonation was altered from breathy to pressed was in line with previous studies (e.g. 7, 8). This implies that the shape of the glottal source for both female and male speakers was most symmetric in breathy phonation. When the phonation type was changed towards pressed, the length of the closing phase decreased. All ten speakers except one female (F5) showed a monotonic decrease in their CQ-values when phonation was changed from breathy towards pressed.

The obtained values of HRF were generally in line with previous studies (2) according to which changing the phonation type corresponds in the frequency domain to changing the spectral tilt 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).

Changing the phonation type by affecting the spectral tilt of the voice source was clearly shown in the obtained values of PSP. A large value of PSP corresponds to fast spectral decay of the voice source, whereas a small PSP-value describes a glottal source which consists of more information in higher frequencies. The obtained data show that PSP-value decreased monotonically for all ten subjects when phonation was altered from breathy to pressed. Hence, PSP was the only one among the analyzed parameters that gave quantitative information which was perfectly in line with the subject's task to change phonation along the axis breathy-normal-pressed.

## 5. SUMMARY AND CONCLUSIONS

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. The PSP-algorithm is based on the application of the 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 that contains those samples whose level is within 20 dB of the DC-value of the voice source spectrum. The PSP-method then computes, using a parabolic function, maximal and minimal spectral tilting of a glottal source whose length of the fundamental period equals to the lenght of one period of the glottal flow given by inverse filtering. The final PSP-value is a number that expresses the spectral decay of the analyzed voice source with respect to its maximal theoretical tilting.

The authors believe that the new parameter is useful especially when analyzing glottal airflow waveforms with greatly different spectral characteristics. The PSP-algorithm takes advantage of information in the frequency domain using those data samples that are most important in characterizing the spectral behavior of the voice source. This is an improvement in comparison to conventional time-based parameters whose value is 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 defining when glottal opening and closure takes place. In comparison to previously-developed frequency domain methods the PSP-algorithm has two improvements. Firstly, application of information on extensively damped harmonics is avoided, which makes the analysis of voices with large spectral decay more accurate. Secondly, the PSP-technique makes possible a comparison of glottal flows in terms of their spectral decay, even though the fundamental frequency of voices is different.

| Speaker | F0 | CQ | HRF | PSP |
|---|---|---|---|---|
| F1 | 182 | 0.48 | 0.16 | 0.38 |
| F2 | 192 | 0.42 | 0.20 | 0.37 |
| F3 | 240 | 0.44 | 0.14 | 0.34 |
| F4 | 211 | 0.38 | 0.14 | 0.32 |
| F5 | 163 | 0.29 | 0.35 | 0.23 |
| M1 | 109 | 0.51 | 0.13 | 0.34 |
| M2 | 96 | 0.49 | 0.20 | 0.42 |
| M3 | 89 | 0.38 | 0.18 | 0.33 |
| M4 | 108 | 0.43 | 0.25 | 0.37 |
| M5 | 111 | 0.44 | 0.23 | 0.42 |

**Table 1:** Obtained parameters in breathy phonation for five female (F1-F5) and five male (M1-M5) speakers. (F0 = fundamental frequency, CQ = closing quotient, HRF = harmonic richness factor, PSP = parabolic spectral parameter). The unit of F0 is Hz, all others in dimensionless units.

| Speaker | F0 | CQ | HRF | PSP |
|---|---|---|---|---|
| F1 | 176 | 0.26 | 0.56 | 0.25 |
| F2 | 183 | 0.26 | 0.46 | 0.26 |
| F3 | 233 | 0.36 | 0.36 | 0.26 |
| F4 | 195 | 0.29 | 0.39 | 0.16 |
| F5 | 174 | 0.29 | 0.59 | 0.17 |
| M1 | 107 | 0.24 | 0.74 | 0.16 |
| M2 | 96 | 0.28 | 0.48 | 0.19 |
| M3 | 89 | 0.30 | 0.42 | 0.23 |
| M4 | 106 | 0.28 | 0.51 | 0.19 |
| M5 | 112 | 0.25 | 0.64 | 0.19 |

**Table 2:** Obtained parameters in normal phonation for five female (F1-F5) and five male (M1-M5) speakers. (F0 = fundamental frequency, CQ = closing quotient, HRF = harmonic richness factor, PSP = parabolic spectral parameter). The unit of F0 is Hz, all others in dimensionless units.

| Speaker | F0 | CQ | HRF | PSP |
|---|---|---|---|---|
| F1 | 174 | 0.23 | 0.67 | 0.14 |
| F2 | 195 | 0.22 | 0.99 | 0.08 |
| F3 | 235 | 0.32 | 0.74 | 0.15 |
| F4 | 192 | 0.26 | 0.64 | 0.15 |
| F5 | 186 | 0.28 | 0.66 | 0.15 |
| M1 | 111 | 0.18 | 1.58 | 0.07 |
| M2 | 94 | 0.18 | 0.95 | 0.13 |
| M3 | 94 | 0.25 | 0.76 | 0.15 |
| M4 | 107 | 0.24 | 0.69 | 0.16 |
| M5 | 118 | 0.24 | 0.58 | 0.16 |

**Table 3:** Obtained parameters in pressed phonation for five female (F1-F5) and five male (M1-M5) speakers. (F0 = fundamental frequency, CQ = closing quotient, HRF = harmonic richness factor, PSP = parabolic spectral parameter). The unit of F0 is Hz all others in dimensionless units.

## 6. REFERENCES


1. Alku, P., and Vilkman, E. “Estimation of the glottal pulseform based on discrete all-pole modeling,” Proc. Int. Conf. on Spoken Language Processing, Yokohama, Japan, Sept. 18-22, 1619-1622, 1994.

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

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

4. El-Jaroudi, A., and Makhoul, J. “Discrete all-pole modeling,” IEEE Trans. Signal Proc. 39 (2), 411-423, 1991.

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

6. Fant, G. "Some problems in voice source analysis," Speech Communication 13, 7-22, 1993.

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

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

9. Oppenheim, A.V., and Schafer, R.W. *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs), 1975.

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

11. Sundberg, J., Titze, I., and Scherer, R "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, 1993.

12. Titze, I.R., and Sundberg, J. “Vocal intensity in speakers and singers,” J. Acoust. Soc. Am. 91 (5), 2936- 2946, 1992.