Parameterization of the glottal source with the phase plane plot

Manu Airaksinen, Paavo Alku

Department of Signal Processing and Acoustics, Aalto University, Finland

manu.airaksinen@aalto.fi, paavo.alku@aalto.fi

Abstract

Parameterization of the glottal flow is a process where the glottal flow is represented in terms of a few numerical values. This study proposes a novel parameterization technique called the phase plane symmetry (PPS) parameter that utilizes the symmetrical properties of the phase plane plot. Phase plane is a way to graphically visualize the glottal source in a 2-dimensional space spanned by two amplitude-domain axes. A correctly normalized phase plane plot has also close ties to the normalized amplitude quotient (NAQ) parameter, and it is shown that the inverse NAQ value is represented as a single point in the phase plane plot. The experiments conducted in this study support that PPS is powerful in discriminating between various phonation types and within the same range of robustness as the NAQ parameter.

Index Terms: speech analysis, glottal source, parameterization

1. Introduction

The source signal of voiced speech, the glottal flow, is key in understanding speech production. Various methods have been proposed to estimate the glottal flow from speech [1, 2]. In many cases, it is useful to represent the estimated glottal flow waveform (or its derivative) with a few numerical parameters. Parameterization of the glottal flow is used in various applications ranging from emotion detection [3], expressive speech synthesis [4], and vocoding [5] to basic research in speech production [6].

Glottal flow parameterization can be divided into time, frequency, and amplitude domain parameters. Time-based parameters are typically determined as quotients of the time-lengths of certain sub-sections of the glottal flow pulse (i.e. opening phase, closed phase, or closing phase) and the duration of the fundamental period [7]. Many time-based parameters are also generative in the sense that they are used in a mathematical model of the glottal source. An example of this is the Liljencrants-Fant (LF) model [6]. Frequency-domain (amplitude) parameters are typically used to model the decay of the voice source spectrum either from its harmonics [8, 9] or by taking advantage of the entire spectrum [10, 11]. Frequency-domain parameters are generally non-generative. Finally, amplitude-domain parameters (in the time domain) such as the ac flow, minimum flow, and the negative peak amplitude of the differentiated flow can be used when the inverse filtering is performed using a properly calibrated Rothenberg mask [12]. An exception to this is the normalized amplitude quotient (NAQ) [13] which utilizes amplitude-domain measures of the glottal flow and its derivative in time-domain parameterization.

In this study, a new amplitude-domain method is proposed for the parameterization of the glottal flow. The proposed phase plane symmetry (PPS) parameter is based on the application of the phase plane plot [14], which is a way to graphically visualize the glottal source in a 2-dimensional space spanned by two amplitude-domain axes (glottal flow, time-derivative of the flow) (Fig. 2). The phase plane plot was proposed in [14] as a method to assess the performance of inverse filtering in removing vocal tract resonances. In the current study, the phase plane plot is utilized as a means to express glottal pulse forms of different phonation types in the 2-dimensional space. It will be shown that the symmetry of the phase plane plot depends on the type of phonation and by parameterizing this symmetry with the proposed PPS measure, an effective new glottal source parameter is obtained. In comparison to previous glottal flow parameters, such as closing quotient (CQ) [15] and NAQ [13], PPS shows improved performance in separating phonation types because the novel parameter reflects the behavior of the glottal pulse during the entire glottal cycle and not only its closing phase.

The organization of this paper is as follows. In Section 2, the phase plane plot and its properties are introduced. Section 3 explains the parameterization scheme used for the PPS parameter. In Section 4, PPS is compared to CQ and NAQ by LF pulses with varying phonation types, and the results are reported in Section 5. Finally, conclusions on the proposed PPS parameter are presented in Section 6.

2. Phase plane plot

In the field of system analysis, phase plane analysis is a visual display of certain characteristics of differential equations. Its use for the objective assessment of glottal inverse filtering was first proposed in [14], and the method was later elaborated in [16]. Phase plane analysis for the glottal flow waveform is based
on the assumption that the vocal tract can be modeled as a cascade of second-order resonators [17]. Thus the system can be modeled with the second order harmonic equation:

\[
\frac{d^2 x}{dt^2} + x = 0
\] (1)

In the phase plane \((x, y)\), this system can be analyzed by:

\[
\frac{dx}{dt} = y \quad \text{and} \quad \frac{dy}{dt} = -x
\] (2)

In simple terms this means that the phase plane plot is obtained by plotting the glottal flow \((x)\) into the x-axis and the glottal flow derivative \(\left(\frac{dx}{dt}\right)\) into the y-axis. More detailed mathematical analysis can be found in [14]. Based on the initial assumption, a glottal pulseform with no formant ripple should be cyclic with respect to the fundamental period in the phase plane (see Fig. 1). Formant ripple adds different solutions that are also periodic, and they produce additional loops into the phase plane plot. The size of these loops is proportional to the amplitude of the formant ripple. This is illustrated in Fig. 1.

2.1. Properties of the phase plane

In previous studies, the phase plane plot has been used in assessing the quality of glottal inverse filtering. The focus has been mainly in determining and minimizing the amount and/or total area of formant ripple loops. However, the use of the phase plane is not restricted to the analysis of formant ripple alone. Instead, this 2-dimensional expression can be used, for example, to demonstrate phonation types as shown in Figure 2. As depicted in Figure 2, the overall shape of the phase plane is affected by the phonation type of speech: the shape of the phase plane becomes more symmetric when the phonation type changes form modal to breathy and then to voiced whisper.

The bottom peak of the phase plane, shown in Fig. 2 by ‘X’s for the three phonation types, is of special importance. This point on the phase plane, namely, can be shown to be equal to the inverse of the NAQ value of the corresponding glottal flow: Because NAQ is estimated as the pitch-normalized ratio of maximum flow difference \(f_{ac}\) and the negative derivative peak \(\Delta f_{peak}\):

\[
\text{NAQ} = \frac{f_{ac}}{\Delta f_{peak}} \cdot \frac{1}{T},
\] (3)

the bottom peak of the phase plane will be equal to the inverse of the NAQ value provided that the glottal flow is scaled to the range \([0, 1]\) (resulting in \(f_{ac} = 1\)), and the glottal flow derivative is scaled with the length of the fundamental period \(T\). Hence, it can be deduced that the phase plane plot involves the NAQ parameter in the form of a single point. However, differently from NAQ and the classical time-domain parameters such as CQ, the phase plane is a rich 2-dimensional representation that takes into account the characteristics of the glottal source during the entire glottal cycle and not just a specific sub-section such as the closing phase.

The shape of the phase plane is defined solely by the amplitude characteristics of the glottal flow and its derivative. Time-domain information of the flow signals can be heavily altered to obtain a duplicate overall shape for the phase plane plot: it is only required that the glottal flow derivative as a function of the glottal flow remains the same. This allows for the utilization of decimation and/or interpolation operations within the source signal for the goal of, for example, obtaining uniform-distance samples to the phase plane. Furthermore, the overall shape of the phase plane plot is determined by the low-frequency components of the glottal flow, which allows for the removal of high-frequency noise components by low-pass filtering the glottal flow and its derivative. Finally, the outer edges of the phase plane remain relatively similar in shape in the presence of formant ripple. A non-ideal (in the sense of format-ripple loops) phase plane plot can thus be approximated as ideal by taking into account only the shape of its outer edges. This is illustrated in Fig. 1.

3. Phase plane symmetry parameter

Based on the general properties of the phase plane described in Section 2.1, a method, phase plane symmetry (PPS), was developed to parameterize the symmetry of the phase plane. In order to describe the computation of PPS, let us assume that a glottal flow waveform has been estimated by inverse filtering. Let us denote one cycle of this time-domain waveform by \(g[m]\) and its derivative by \(d[m]\), where \(m = 0, 1, \ldots, M - 1\) and \(M\) is the fundamental period (in samples). When a voice source is initially represented in the 2-dimensional space (e.g. in Fig. 2), the consecutive points \((g[m], d[m])\), \(m = 0, 1, \ldots, M - 1\) are not evenly spaced in the glottal flow axis. Therefore, a new transformed representation is required in order to express \(g[m]\) and \(d[m]\) in a manner that corresponds to points that occur at regular intervals in the x-axis. The transformed representation, denoted by \(gr[n]\) for the glottal flow and by \(dT[n]\) for the flow derivative, is illustrated in Fig. 3 and it is obtained as follows.

First, to get evenly spaced samples on the x-axis, \(gr[n] = 0, 1, \ldots, N - 1\) is defined as an isosceles triangle ranging from \([0, 1, \ldots, 0]\) with a length of \(N\) samples. \(N\) can be chosen to be of an arbitrary length, and is not tied to \(M\) of the original \(g[m]\), as long as it can sufficiently represent the overall shape of the phase plane. In this study, \(N = 256\) was used. Next, to obtain \(dT[n]\), interpolation is required between the values of \(d[m]\), as the original \(g[m]\) is not expected to contain samples with exactly the same values as \(gr[n]\). The interpolation is done so that for each sample of \(gr[n]\), the indeces \(m_{TOP,n}\) and \(m_{BOT,n}\) corresponding to the closest value of \(g[m]\) with positive and negative \(d[m]\) values, respectively, are stored. Then, \(dT[n]\) is formed as:

\[
d_T[n] = \begin{cases} 
   d[m_{TOP,n}] & \text{if } n \leq N/2 \\
   d[m_{BOT,n}] & \text{if } n > N/2 
\end{cases}
\] (4)

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{phase_plot.png}
\caption{Phase plane plots for modal (dashed line), breathy (solid line) and voiced whisper (dash-dotted line). Bottom peak, which is equal to the inverse NAQ value of the corresponding phonation type, is marked with ‘X’. Waveforms created with the LF model according to the parameters reported in [18].}
\end{figure}
After this has been performed for all \( n = 0, 1, \ldots, N - 1 \), anti-aliasing low-pass filtering is applied to \( d_T[n] \) to complete the interpolation.

At this point, the data of interest (that is the shape of the phase plane plot) is located in \( d_T[n] \) (Fig. 3, bottom right). Its plot is essentially the same as the phase plane plot in shape, but its negative bottom half is mirrored by the y-axis. Here, an observation can be made about the symmetry of the obtained plot: As discussed in Section 2.1, soft phonation has the effect of moving the bottom peak of the phase plane plot towards the center of the x-axis. For \( d_T[n] \), this means that the plot reminds a cycle of a sine function (see Fig. 3). However, pressed phonation moves the peak towards left, and likewise towards right in \( d_T[n] \). Hence, \( d_T[n] \) has sine-like first half, but the second half is less sine-like and therefore contains also strong cosine components (see Fig. 3).

In order to quantify the observation described above, PPS computes relative energies of the sine and cosine components of \( d_T[n] \). The discrete Fourier transform (DFT) of \( d_T[n] \) is defined as:

\[
X[k] = \sum_{n=0}^{N-1} d_T[n] \cdot e^{-j2\pi kn/N}, \tag{5}
\]

meaning that each bin of \( X[k] \) will be a complex number with its real part consisting of \( X[k] \)'s correlation with the cosine term and its imaginary part consisting of \( X[k] \)'s correlation with the sine term. The energies of the sine and cosine components can thus be obtained by

\[
E_{\cos} = \sqrt{\sum_k \text{Re}\{X[k]\}^2}, \quad E_{\sin} = \sqrt{\sum_k \text{Im}\{X[k]\}^2}. \tag{6}
\]

The component energies are expected to be proportional to the overall amplitudes of the interpolated derivative vector, whose extreme value is the inverse NAQ. Also, \( E_{\cos} \) is expected to be significantly smaller than \( E_{\sin} \), because the overall shape of \( d_T[n] \) is based on a sine wave. Thus the suggested combination of these parameters is their geometric mean \( E_{\cos}E_{\sin} \), which preserves the amplitude proportionality, and also reflects on the relative change of the significantly smaller \( E_{\cos} \) component instead of its absolute change. Finally, to get the proposed PPS parameter to the same domain with the NAQ parameter (meaning that large values of the parameter correspond to soft excitation and small values correspond to pressed excitation), the inverse value of the geometric mean is taken. Thus, the final form of the PPS parameter can be expressed as:

\[
PPS = \frac{1}{\sqrt{E_{\cos}E_{\sin}}}. \tag{7}
\]

In summary, the proposed PPS parameter for a glottal flow waveform is computed using the following steps:

1. Generate the transformed glottal flow waveform \( g_T[n] \).
2. According to \( g_T[n] \), interpolate the glottal flow derivative \( d_T[n] \) (weighted by the fundamental period \( M \)) so that the end result \( d_T[n] \) produces a duplicate phase plane plot to the original.
3. Compute the DFT of \( d_T[n] \).
4. Compute \( E_{\cos} \) and \( E_{\sin} \) of the DFT.
5. Compute the inverse geometric mean of \( E_{\cos} \) and \( E_{\sin} \) to obtain the PPS parameterization.

4. Experiments

Three experiments were conducted in order to evaluate PPS in voice source parameterization. First, distributions of PPS, NAQ and CQ were compared by using synthetic vowels of different phonation types and \( F_0 \) values. Four phonation types (creaky, modal, breathy, and voiced whisper) were created according to [18] by using the LF pulse as an excitation. \( F_0 \) was varied from 80 Hz to 260 Hz with an increment of 10 Hz. Vocal tract was modeled as in [19] to synthesize six vowels ([a], [e], [i], [o], [u], and [ae]). Distributions were examined from the glottal sources in two cases as follows. Case (a) corresponded to the ideal inverse filtering (i.e. estimated glottal flow derivative equaled the
LF pulse used in the sound synthesis). In case (b), parameterization was computed from glottal flows estimated with a practical inverse filtering method. As a practical inverse filtering method, a recent technique described in [20] was used.

Second, the robustness of PPS was compared to that of NAQ and CQ by studying the absolute value of the relative error of the parameter value in cases where the LF waveform is corrupted with varying degrees of formant ripple to simulate error in glottal inverse filtering. The error used was a varying percentile error in the formant frequencies of the ideal vocal tract all-pole filter. The percentiles used were ±(2%–10%) with 2% increments.

Finally, the parameter was demonstrated for real speech by studying vowels produced in three phonation types (breathy, modal, and pressed). Speech data (vowel [a]) were produced by one male and one female speaker of Finnish. The recordings were conducted in an anechoic chamber under the supervision of a trained phonetician. In case a test subject did not produce an utterance with a correct type, the phonetician asked the test subject to repeat the utterance until the phonation was satisfactory. The data was sampled with 16 kHz, and inverse filtering was performed with the QCP method [20].

5. Results

The boxplots of the parameter distributions are shown in Fig. 4. The figure illustrates, both in case (a) and (b), that PPS is superior to NAQ particularly in discriminating breathy and voiced whisper. When inverse filtering was involved, the parameter distributions became wider. Compared to the ideal case, CQ values changed most.

The results for the second test are presented in Fig. 5. They illustrate the respective parameters’ robustness to the increase in formant ripple. The NAQ parameter shows the best robustness. For PPS, it can be observed that the parameter values are smaller when relative formant error is positive than when it is negative. This is because the negative-biased formant errors produce counter-clockwise loops in the phase plane plot that have a larger impact on its overall shape.

Representative examples computed from real speech are presented in Fig. 6. It can be seen that the corresponding PPS values for male and female speakers are similar. The differences in the values between phonations are similar to the differences in Fig. 4. The values are also systematically approximately 0.05 units higher than those computed for the synthetic LF pulses. This is caused by the characteristics of the glottal flow inverse filtered from real speech which show less abrupt waveform changes at the instant of glottal closure, which decreases the maximum amplitude of the differentiated flow.

6. Conclusions

This study presented a new amplitude-domain glottal source parameterization method called the phase plane symmetry (PPS) parameter. PPS utilizes the symmetrical properties of the glottal source within the phase plane plot [14], which are dependent on the type of phonation used (Fig. 2). PPS was evaluated with a set of synthetic vowels with different phonation types to assess its parameter distributions and its robustness to inverse filtering errors. The results show that compared to the NAQ and CQ parameters, PPS is better in discriminating between different phonation types. The robustness of PPS is slightly weaker to formant ripple errors than NAQ, which can be considered as a minor drawback for the parameter. However, the results presented in Fig. 4(b) suggest that with practical high-quality inverse filtering, the parameter distribution is tighter for PPS than for NAQ or CQ.

Further work on the parameter include the assessment of its deviations and robustness on a real speech database with consistent and differing phonation qualities, its robustness to noise, and assessing how well the parameter maps the differences between perceived phonation qualities.

7. Acknowledgements

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 287678 and from the Academy of Finland (project no. 256961).
8. References


