Group Delay Function as a Means to Assess Quality of Glottal Inverse Filtering

Paavo Alku, Matti Airas, Tom Bäckström, Hannu Pulakka

Helsinki University of Technology,
Lab. of Acoustics and Audio Signal Processing,
Espoo, Finland
paavo.alku@hut.fi

Abstract

The estimation of glottal flow with inverse filtering requires typically subjective evaluation of the obtained flow waveforms by the experimenter. In this paper, we propose a straightforward method that yields an objective analysis tool to be used in the assessment of the quality of inverse filtered glottal flow estimates. The method computes the negative derivative of the phase response, the group delay function, over a single glottal cycle of the estimated flow. Incorrect settings of the anti-resonances of the inverse filter result in distortion of the group delay function. This distortion is easily detectable from the group delay function and can therefore be used in a straightforward manner to analyse how well the zeros of the inverse filter match the poles of the vocal tract.

1. Introduction

Inverse filtering (IF) is a non-invasive method to estimate the source of voiced speech, the glottal volume velocity waveform, generated by fluctuating vocal folds. The idea behind IF is first to form a computational model for the vocal tract transfer function. The effect of the vocal tract is then cancelled from the produced speech waveform by filtering this through the inverse of the model. The waveform, from which the cancelling of the vocal tract contribution is done, can be either the oral flow recorded in the mouth with a flow mask (e.g. [1]) or the pressure waveform captured by a microphone in free field outside the mouth (e.g. [2, 3]).

Older IF-methods typically used manually adjustable analog circuits in implementation of the inverse model of the vocal tract [4]. In these techniques, the experimenter tuned the anti-resonances of the inverse filter until the output was of an "optimal" shape. The "optimal" result of the inverse filtering analysis was interpreted typically as a waveform which showed a maximally flat horizontal closed phase with minimal remaining formant ripple. Currently used methods are based on digital modelling of the vocal tract, which is usually implemented either completely automatically or semi-automatically [5-9]. Even though adjustments of inverse filter anti-resonances can be done undoubtedly in a much more convenient and accurate manner in these digital IF-methods, they are still based more or less on the same assessment criterion as the older analog techniques: the "optimal" settings of the inverse filter are iterated until the resultant estimate for the glottal flow pulse shows a maximally flat closed phase. Applying this kind of subjective criterion in adjusting the parameter settings of inverse filtering is problematic because it causes the estimation outputs to be dependent on, e.g., the experience and personal opinions of the experimenter. In addition, subjective adjustment of the filter parameters is time-consuming. This, in turn, might seriously limit utilising inverse filtering in such applications that call for analysing extensive amounts of speech data (e.g., measuring effects of vocal loading on individuals with heavy voice usage in realistic environments, such as teachers speaking in classroom).

The quality assessment of glottal flow pulseforms estimated by inverse filtering is addressed in this study. In order to avoid problematic subjective criteria in adjustments of inverse filtering settings, we propose a new method based on the use of the group delay function. The method proposed should be regarded as a general tool which can be combined with most of the known inverse filtering techniques in order to assess objectively the quality of the glottal flow pulseform. In this study, the new approach is combined with one of our previous IF methods [10] in order to demonstrate the performance of the idea in voice source analysis of real speech.

2. Method

This section describes the proposed method. The properties of the group delay function are first dealt with in section 2.1. The application of the group delay function in glottal inverse filtering is then described in section 2.2.

2.1 Group delay function

The Fourier analysis of real data yields a spectrum, which has two components: magnitude and phase. In signal analysis, the phase spectrum given by the Fourier analysis is typically transformed into the group delay function, that is, the negative derivative of the phase spectrum [11]. In most cases, the main object of interest in spectral analysis is the magnitude component; the phase spectrum is considered less important. To overlook the role of the phase spectrum is in general true also in the spectral analysis of speech. There are, however, some studies where phase spectra have been used. These studies show that the phase information (interpreted using the group delay function) provides a potential analysis tool which can be used, for example, in formant [12-14] and epoch extraction [15] of voiced speech. Moreover, there are also studies on applying the phase component of the Fourier transform to speech recognition [16].

Studies on the spectral analysis of the glottal flow are almost entirely based on the application of the magnitude component of the Fourier transform. This is understandable because changes in the glottal source are reflected mainly in the envelope of the magnitude spectrum when, for example,

the phonation type or vocal intensity is altered. Studies on the behaviour of the phase spectrum of the glottal source are, however, extremely sparse. To the best of our knowledge, the phase spectrum (or the group delay function) of the glottal flow has been dealt with only in [17-19]. These studies, interestingly, indicate that the phase of the glottal flow, when computed over a single glottal cycle, is almost a linear function of frequency over a wide frequency range with an exception in the lowest frequency range. This implies that the derivative of the phase spectrum, the group delay function, is almost constant as a function of frequency. This is described in Fig. 1 using synthetic glottal flow pulses generated by the LF model [20] by exploiting parameter values reported in [21]. This general, yet poorly known, feature of the group delay function of the glottal flow pulse serves as a basis for the method proposed in this study.

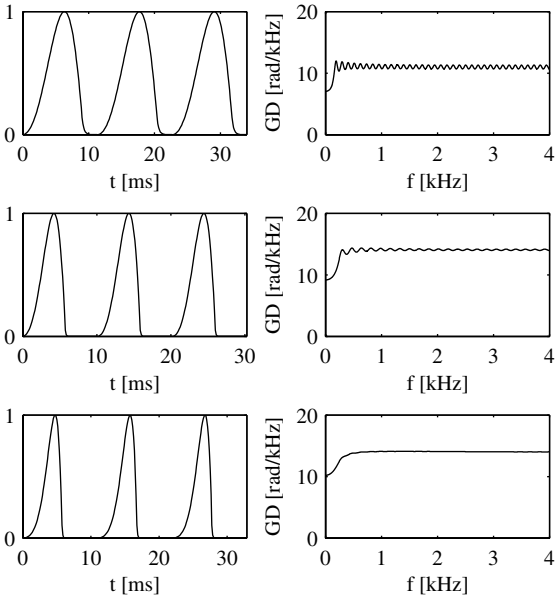

Figure 1: Glottal flow pulseform in the time-domain (left) and its group delay (GD) function (right) computed over a single glottal cycle. The three pulseforms corresponding to voice qualities breathy, modal and creaky were synthesised with the LF model [20] by using parameters reported in [21]. (Unit of the flow arbitrary).

2.2 Proposed method to assess the quality of the glottal flow estimated by inverse filtering

In this section, we describe the idea of the proposed method with the help of a flow diagram (Fig. 2) and a simple example (Fig. 3) based on a synthetic glottal flow generated by the LF-model [20]. The conventional source-tract model of speech production (shown in Fig. 2a) assumes that the speech signal can be described as a cascade of three processes: the glottal flow, the vocal tract and the lip radiation (i.e., the change of the volume velocity at lips into a pressure signal) [22]. Let us assume that a speech signal is created according to this model by estimating the lip-radiation with a (leaky) differentiator. Further, let us also assume that the model is linear, with no interaction between the different parts, and that the vocal tract transfer function is a pure all-pole filter. With these assumptions, computation of the glottal flow implies (as shown in Fig. 2b) filtering the integrated speech pressure signal through an all-zero filter, denoted by $1/V_e(z)$, which cancels the vocal tract resonances. If the zeros of this all-zero filter coincide exactly with the poles of the vocal tract filter, the resultant glottal flow estimate naturally equals the original excitation waveform. If, however, the transfer function of the inverse filter has been estimated incorrectly (i.e., $V_e(z) \neq V_o(z)$), the estimated glottal flow will be the original one filtered through an *pole-zero* filter with its transfer function equal to $H_{pz}(z) = V_o(z) / V_e(z)$. It is known, importantly, that a pole-zero filter cannot have a linear phase response [11]. Consequently, filtering the original glottal flow through $H_{pz}(z)$ corresponds to distorting the (almost) constant group delay function of the original glottal flow by adding to it a non-linear component. This distortion is easily detectable from the group delay function computed over a single glottal cycle of the estimated glottal flow and it forms the core idea of the proposed method.

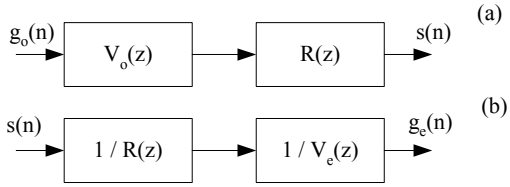

Figure 2:
(a): Production of speech pressure signal according to the source-tract model, (b): estimation of the glottal flow from the speech pressure signal with inverse filtering. Notations: original glottal flow: $g_o(n)$, vocal tract transfer function: $V_o(z)$, lip-radiation effect: $R(z) = 1 - \sigma \cdot z^{-1}$, speech pressure signal: s(n), estimated vocal tract transfer function: $V_e(z)$, estimated glottal flow: $g_e(n)$.

The idea of using the group delay function to assess the performance of inverse filtering is described in the example shown in Fig.3. In this figure, an 8th order all-pole vocal tract model (poles denoted by crosses) and the corresponding inverse filter (zeros denoted by circles) are shown in the z-plane. The original glottal flow and the estimated flow obtained by inverse filtering are depicted as time-domain signals. Graphs in the left side of Fig. 3 correspond to the case when inverse filtering has been computed with correct anti-resonance settings (i.e., $V_o(z) = V_e(z)$). Graphs in the right side of Fig. 3 describe an inverse filtering analysis where one complex conjugate zero pair of $1/V_e(z)$ does not match the corresponding pole pair of $V_o(z)$. Group delay functions computed from one cycle of the estimated glottal flow show clearly different behaviours in the two examples: In the case of correct settings of the inverse filter, the group delay function is more or less a horizontal line, whereas in the case of incorrect inverse filtering settings, there are rapid fluctuations in the group delay function.

In summary, the proposed straightforward method to assess the quality of an estimated glottal flow ($g_e(n)$) obtained by inverse filtering comprises following stages:

1: Compute FFT (with rectangular windowing) over one period of $g_e(n)$. (FFT-size of 2048 is used throughout this study).

2: Compute the group delay function by first differentiating the discrete phase response obtained by FFT and by then multiplying it by -1. Subtraction of successive phase response values shows occasional sharp peaks at

frequencies where the phase function is wrapped to occur between -π and π. Indicate these peaks by using a default absolute level (2.5 was used in this study). Correct the effect of wrapping by adding +/- 2π to the corresponding peaks. (Choose that sign of 2π that yields a smaller absolute value after adding.)

3: If needed, quantify the fluctuations of the obtained group delay function. It is possible to use straightforward measures, such as variance, to express the fluctuations of the group delay functions in numerical forms.

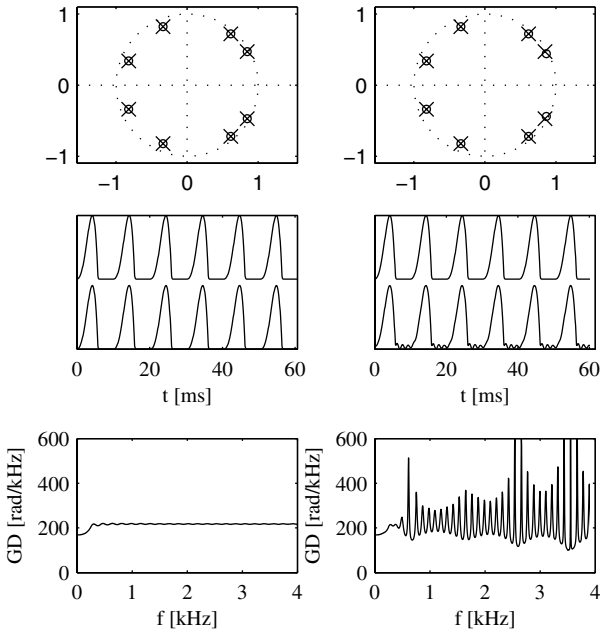

Figure 3:
Left panels: Glottal flow computation with correct inverse filter settings.
Right panels: Glottal flow computation with incorrect inverse filter settings in modelling of the lowest formant.
Vocal tract poles and inverse filter zeros are shown in top panels in the z-domain by crosses and circles, respectively. Middle panels show the original (upper) and estimated (lower) glottal flow as time-domain waveforms. Group delay functions computed over a single period of the estimated glottal flow are shown in bottom panels.

3. Experiments with real speech

In order to analyse how the proposed method works with real speech data, we conducted preliminary experiments with an inverse filtering method described in [10]. In this method, the effect of the glottal source on the speech spectrum is first cancelled using a library of FIR-filters, whose spectra simulate frequency characteristics of different glottal pulses. The filter library used in [10] comprises 30 FIRs. After cancelling the glottal contribution and the lip radiation effects, the method estimates the vocal tract using the Discrete All-pole Modeling [23]. Consequently, a vowel segment to be analysed yields 30 candidates for the glottal flow and the user has to select which of these is the best one.

As speech material we used an utterance (vowel /a/) produced by a female speaker using a fairly high fundamental frequency (F0 = 445 Hz). The glottal flow waveform of this utterance was estimated with the method described in [10]. The method proposed in the present paper was then computed for the glottal flow candidates obtained. Fig. 4 shows three examples of these glottal flow estimates along with the corresponding group delay functions. This example describes how the proposed method clearly gives a much smoother group delay function for the glottal flow (bottom panel), which also, according to the subjective criteria, is the most reliable candidate for the "real" glottal flow (i.e., the one with a long flat closed phase). In addition, the candidate with a remarkable ripple in its closed phase (top panel) shows fluctuations in its group delay function. Interestingly, the third candidate (middle panel), which does not show formant ripple but rather almost an entire absence of the closed phase, also indicates non-linear behaviour in the corresponding group delay function. All in all, the group delay functions yield results according to which the "best" glottal flow candidate, based on the subjective criteria, can be found objectively by using the proposed method.

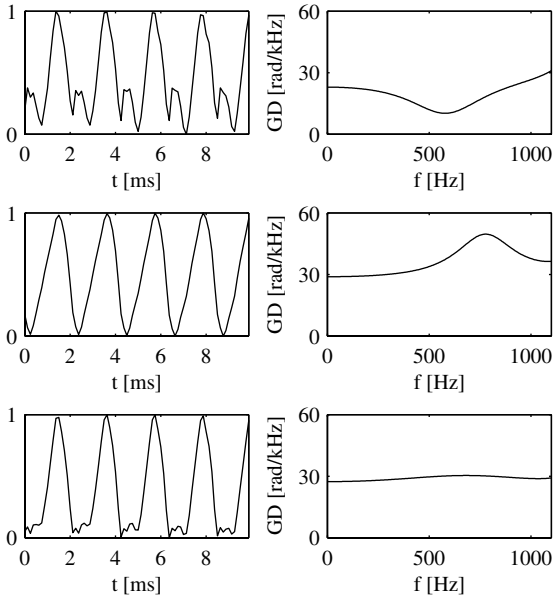

Figure 4: Examples of estimated glottal flow candidates (left panels) computed from real speech (female speaker, vowel /a/) using the inverse filtering method described in [10]. Group delay (GD) functions, computed over a single period of each glottal flow candidate, are shown in the right side. (Unit of the flow arbitrary).

4. Conclusions

We have proposed a straightforward method to assess the quality of glottal flow waveforms computed by inverse filtering. This method computes the negative derivative of the phase response, the group delay function, over a single glottal cycle of the estimated flow. Incorrect settings of the anti-resonances of the inverse filter result in the distortion of the group delay function of the original glottal flow. This distortion is easily detectable from the group delay function and can therefore be used in a straightforward manner to analyse how well the settings of the anti-resonances of the inverse filter match the poles of the vocal tract transfer function. According to our preliminary results, the proposed

method shows promising results in the analysis of both synthetic and real vowel sounds. Hence, the method proposed constitutes a potential tool that can be used in the development of automatic inverse filtering methodologies.

5. Acknowledgements

This study was supported by the Academy of Finland (projects no. 200859 and 205962).

6. References

[1] Rothenberg, M., "A new inverse-filtering technique for deriving the glottal airflow waveform during voicing", *J. Acoust. Soc. Amer.,* 53(6):1632-1645, 1973.

[2] Wong, D.Y., Markel, J.D. and Gray, A.H., Jr., "Least squares glottal inverse filtering from the acoustic speech waveform", *IEEE Trans. Acoust. Speech and Signal Proc.,* 27(4):350-355, 1979.

[3] Alku, P., "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering", *Speech Comm.*, 11:109-118, 1992.

[4] Miller, R.L., "Nature of the vocal cord wave", *J. Acoust. Soc. Amer.*, 31:667-677, 1959.

[5] Veeneman, D.E. and BeMent, S., "Automatic glottal inverse filtering from speech and electroglottographic signals", *IEEE Trans. Acoust. Speech and Signal Proc.*, 33(2):369-377, 1985.

[6] Milenkovic, P., "Glottal inverse filtering by joint estimation of an AR system with a linear input model", *IEEE Trans. Acoust. Speech and Signal Proc.,* 34(1):28-42, 1986.

[7] Krishnamurthy, A.K., "Glottal source estimation using a sum-of-exponentials model", *IEEE Trans. Signal Proc.*, 40(3):682-686, 1992.

[8] Gobl, C., Monahan, P., Fitzpatrick, L. and NiChasaide, A., "Automatic source-filter decomposition: A knowledge-based approach", Proc. of the Speech Maps Workshop, Vol. 2, paper 9, 1994.

[9] Fröhlich, M., Michaelis, D. and Strube, W., "SIM-simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals", *J. Acoust. Soc. Amer.,* 110(1):479-488, 2001.

[10] Alku, P. and Vilkman, E., "Estimation of the glottal pulseform based on Discrete All-Pole modeling", Proc. Int. Conf. on Spoken Language Processing, 3:1619-1622, Yokohama, Japan, 1994.

[11] Oppenheim, A.V. and Schafer, R.W., *Discrete-time Signal Processing*, Prentice-Hall, New Jersey, 1989.

[12] Reddy, N.S. and Swamy, M.N.S., "High-resolution formant extraction from linear-prediction phase spectrum", *IEEE Trans. Acoust. Speech and Signal Proc.*, 32(6):1136-1144, 1984.

[13] Yegnanarayana, B., "Formant extraction from linear-prediction phase spectrum", *J. Acoust. Soc. Amer.,* 63(5):1638-1640, 1978.

[14] Murthy, H.A. and Yegnanarayana, B., "Formant extraction from group delay function", *Speech Comm*., 10:209-221, 1991.

[15] Yegnanarayana, B. and Veldhuis, R., "Extraction of vocal-tract system characteristics from speech signals", *IEEE Trans. Speech and Audio Proc*., 6(4):313-327, 1998.

[16] Schlüter, R. and Ney, H., "Using phase spectrum information for improved speech recognition performance", Proc. Int. Conf. on Acoust. Speech and Signal Proc, 133-136, Salt Lake City, USA, 2001.

[17] Gardner, W. and Rao, B., "Noncausal all-pole modeling of voiced speech", *IEEE Trans. Speech and Audio Proc*., 5(1):1-10, 1997.

[18] Doval, B. and d'Alessandro, C., "Spectral correlates of glottal waveform models: An analytic study", Proc. Int. Conf. on Acoust. Speech and Signal Proc, 2:1295-1298, Munich, Germany, 1997.

[19] Bozkurt, B. and Dutoit, T., "Mixed-phase speech modeling and formant estimation, using differential phase spectrums", Proc. Isca workshop: Voice quality: Functions, analysis and synthesis (VOQUAL'03), 21-24, Geneva, Switzerland, 2003.

[20] Fant, G., Liljencrants, J. and Lin, Q., "A four-parameter model of glottal flow", STL-QPRS, 4:1-13, Royal Institute of Technology, Sweden, 1985.

[21] Gobl, C., "A preliminary study of acoustic voice quality correlates", STL-QPRS, 4:9-21, Royal Institute of Technology, Sweden, 1989.

[22] Fant, G., *The Acoustics Theory of Speech Production*, Mouton, the Hague, 1960.

[23] El-Jaroudi, A. and Makhoul, J., "Discrete all-pole modeling", *IEEE Trans. Signal Proc*., 39:411-423, 1991.