VOQUAL'03, Geneva, August 27-29, 2003

Parameterisation Methods of the Glottal Flow Estimated by Inverse Filtering

Paavo Alku

Lab. of Acoustics and Audio Signal Processing
Helsinki University of Technology, Finland
Paavo.Alku@hut.fi

Abstract

Estimation of the source of voiced speech, the glottal volume velocity waveform, with inverse filtering involves usually a parameterisation stage, where the obtained flow waveforms are expressed in numerical forms. This stage of the voice source analysis, the parameterisation of the glottal flow, is discussed in the present paper. The paper aims to give a review on the different methods developed for the parameterisation and it discusses how these parameters have reflected function of the voice source in various voice production studies.

1. Introduction

Inverse filtering is a widely used method to estimate the source of voiced speech, the glottal volume velocity waveform. A model for the vocal tract transfer function is first computed in inverse filtering. The effects of vocal tract resonances can be cancelled by filtering the signal (either the oral flow recorded at lips or the speech pressure waveform captured by a microphone in free field) through the inverse model of the vocal tract transfer function.

Since the presentation of the idea of inverse filtering by Miller [1] many different versions of this technique have been developed. Inverse filtering methods can be divided into two categories depending on the procedure that is used in recording the input signal. The first category consists of methods that are based on applying a specially designed pneumotachograph mask (the so called Rothenberg's mask [2]) in recording the oral flow at the mouth. The second group of methods is based on processing the speech pressure waveform that has been recorded by a microphone in free field outside the mouth. The methods belonging to this category need to take into account the lip radiation effect [3], that is, the changing of the volume velocity at the lips to a pressure signal outside the mouth. This effect can be estimated for a mid range of frequencies as a derivative. The involvement of the lip radiation effect implies that it is impossible to compute the DC-value of the glottal volume velocity waveform when estimating the glottal source from speech pressure waveforms recorded outside the mouth. If, however, inverse filtering is computed from the oral flow with the Rothenberg's mask, it is possible to calibrate the measurements to obtain real values of the glottal flow including both the DC-component and the AC-component.

Analysis of voice production with inverse filtering comprises usually two stages. In the first one, the estimates of the glottal flow (or its time derivative) are computed. The second stage of the analysis, *the parameterisation stage*, implies quantifying the obtained waveforms with properly selected numerical values. These quantities, the glottal flow parameters, aim to represent the most important features of the original flow waveforms in a compressed numerical form. Hence, the selection of the parameterisation method of the glottal flow is a crucial part in voice production studies because it dictates how effectively the original information embedded in the estimated glottal flows is transferred to the experimenter. In order to successfully quantify voice production with inverse filtering one must know the different alternatives available in the parameterisation stage and know on which features of the glottal flow these measures focus. With this knowledge it is possible to select the "best" numerical measure that reflects the behaviour of the voice source in the experiment in question.

Parameterisation methods of the glottal flow estimated by inverse filtering can be applied in the following, partly overlapping three areas. The first and the most general application area is the categorisation of the source of voice production. This corresponds to dividing speech sounds into various categories according to the different modes of the glottal flow pulse. These modes, in turn, are used in speech communication to vary, for example, vocal loudness [e.g., 4-9] or to carry the emotional stage of the speaker [10, 11]. Even though the applications within this area are mainly in the basic research of voice production, there are emerging applications for the parameterisation methods of the glottal source, for example, in forensic speech research such as in speaker recognition [12]. Secondly, parameterisation of the glottal flow can be applied in diagnosis and treatment of vocal disorders [13, 14]. An area of increasing importance closely related to this is measuring loading of voice [15, 16]. Due to increasing number of employees working in such professions where voice is the main tool of trade, the environmental voice care will become increasingly important, which, in turn, will emphasise the role of the glottal flow parameterisation techniques. Thirdly, parameterisation of the glottal flow can also be applied in speech technology, especially in voice synthesis [17-19].

The goal of this paper is to describe the different methods available for the parameterisation of the glottal flow and to discuss, how they have behaved in various studies conducted in the area of voice production. It is assumed that the starting point of the parameterisation, the glottal volume velocity waveform, has been obtained by inverse filtering, which is based either on the application of the flow mask or the speech pressure signal captured in free field. Moreover, parameterisation methods presented are restricted to those, which apply only the estimated glottal flow or its derivative as an input information. Hence, such measures to characterise voice production that apply, for example, sub-glottal pressure (e.g., glottal permittance, [8]) are omitted. In case parameterisation requires the first time-derivative of the flow, it is assumed throughout the text that this sound is computed in the digital time-domain by filtering the flow with a first order FIR with its transfer function equal to $1\text{-}z^{-1}$.

2. Methods to parameterise the glottal flow

A large number of different methods have been developed for the parameterisation of the glottal flow. In the following, these techniques are discussed by dividing the methodologies into two categories depending on whether the parameterisation is performed in the time-domain or in the frequency-domain.

2.1. Time-domain methods

The most straightforward method to parameterise the glottal flow waveform obtained by inverse filtering is to extract certain critical time-spans and amplitude values from the time-domain flow signal or from its first derivative. Fig. 1 shows an example of a glottal flow (upper panel) and its time-derivative (lower panel) obtained by inverse filtering the vowel /a/ produced by a male speaker using normal phonation. Using the notations depicted in the two waveforms we can sum up the following list of parameters that have been used in voice source studies:

Open quotient (OQ) = $(t_o + t_{cl}) / T$ [e.g., 5, 13-16, 20-22]
(Open quotient is sometimes replaced with Closed quotient (CQ) = $t_c / T = 1 - \mathrm{OQ}$ [9, 23-26])
Speed quotient (SQ) = t_o / t_{cl} [e.g., 5, 13-16, 20-22]
Closing quotient (ClQ) = t_{cl} / T [e.g., 5, 13-16, 20-22]
Return quotient (RQ) = t_{ret} / T [23]
Normalized amplitude quotient (NAQ) = $ac / (d_{peak} \cdot T)$ [27, 28]

2.1.1. Time-based parameters

The most widely used time-based parameters are undoubtedly OQ, SQ, and ClQ. The behaviour of these measures in different loudness and pitch conditions was studied in [5]. Glottal flows inverse filtered from the mask recordings were analysed from 25 male and 20 female subjects. The time-based parameters indicated that the glottal flow changed in general towards a more asymmetric shape (that is, OQ decreases and SQ increases) when loudness was altered towards loud. However, in female voices this did not happen between normal and loud voices. Voices from a total of 224 subjects were studied with different voice source parameters in [9]. Subjects, phonating at three intensity conditions, were divided into trained and untrained. It was found that voice training affected few of the measures: ClQ of females was higher for trained than for untrained subjects and SQ of trained males was higher that of untrained males. However, the gender of the speaker caused several statistically significant effects on the parameter values. It was found, for example, that CQ of males was larger than that of females but ClQ was smaller in males than in females. Mean values of OQ, SQ and ClQ obtained in these two large experiments are as follows. In [5], the mean of OQ, SQ and ClQ for males in normal loudness conditions was 0.60, 1.82, and 0.22, respectively. For female speech of normal loudness, Holmberg *et al.* reported the mean of OQ, SQ and ClQ to be 0.76, 1.65, and 0.29, respectively. In [9], slightly different values of the time-based parameters were obtained: the mean value of OQ (computed from the CQ values measured in [9] by equation OQ = 1.0-CQ), SQ and ClQ for (untrained) males in normal intensity was 0.49, 1.52, and 0.20, respectively, and the mean value of OQ, SQ, and ClQ for females was 0.55, 1.36, and 0.24, respectively. In [7], OQ and SQ were analysed from voices of ten female speakers. The speaking task was to increase the vocal intensity in 5 dB increments between the SPL values of 70 dB and 95 dB. It was found that the value of OQ decreased when the intensity was increased. The value of SQ, however, did not show a monotonic behaviour with increasing of the intensity. In stead, SQ first increased but for the loudest voice samples it started to decrease. Differences in glottal sources of female and male subjects were analysed in [23] by using CQ and RQ. It was reported that females tend to have shorter values of CQ and longer values of RQ. To use time-based parameters for objective assessment of vocal disorders was studied in [13]. The authors applied OQ, SQ and ClQ for parameterisation of voiced produced in different loudness and pitch conditions. It was reported that SQ and ClQ are most sensitive to abnormalities of the glottal flow resulting from presence of lesions that add mass to the covers of the vocal folds. The effects of vocal loading caused by prolonged reading was analysed using OQ, SQ, and ClQ in [15] from data produced by 80 subjects. The results showed that vocal loading increased SQ and decreased ClQ for female subjects, which implies that the prolonged reading caused a change towards hyperfunction. The effects of time-based parameters on perception of speech has also been studied [22, 29]. Childers and Lee [29] reported that sensation of vocal effort is closely related to SQ: a large value of SQ corresponds to tense voice quality whereas its small value implies a lax or hypofunctional voice quality. Glottal flow characteristics of singing voices have also been studied using time-based parameters [8, 25, 26]. One of the interesting results in these studies was reported in [26], where its was shown by using CQ that country singers use basically the same phonation type both in singing and in speaking.

The computation of the time-based parameters described in above requires extracting the critical time instants, e.g., glottal opening and closure, from the waveform given by inverse filtering. Exact determination of these time instants is sometimes problematic due to formant ripple and noise present in the waveform. Even in the absence of formant ripple or ambient noise, computation of OQ and SQ is difficult because of the gradual opening of the vocal folds. Therefore, 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 [7, 30].

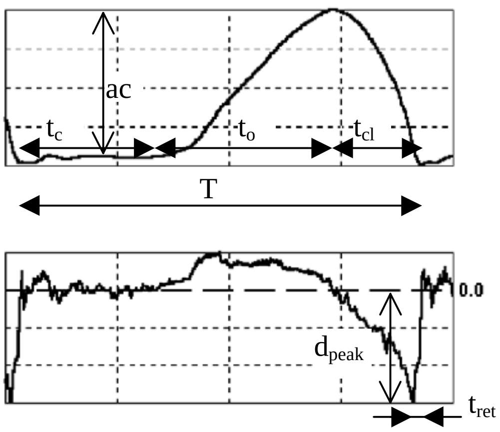

Figure 1: One cycle of the glottal flow (upper panel) and its time-derivative (lower panel) obtained by inverse filtering the vowel /a/ produced by a male speaker using normal phonation. Time spans: fundamental period (T), closed phase (t_c), opening phase (t_o), closing phase (t_{cl}), return phase (t_{ret}), i.e., the time during which the derivative returns back to the zero level after the instant of the negative peak. Amplitude values: AC-flow (ac) and the negative peak amplitude of the derivative (d_{peak}).

In order to make the extraction of the time-based parameterisation easier, a new parameter called NAQ (Normalized Amplitude Quotient) was presented in [27]. Computation is this parameter is done from two amplitude-domain values, the AC-amplitude of the flow and the negative peak amplitude of the glottal flow derivative (see Fig. 1). It is worth noticing that these two amplitude measures are the extreme values of the flow and its derivative and, therefore, they are straightforward to be extracted. It can be shown that the ratio of these two amplitude-domain values is a time-domain quantity, which is interpreted by Fant as "the projection on the time axis of a tangent to the glottal flow at the point of excitation, limited by ordinate values of 0 and the AC-amplitude of the flow" [31]. In [27, 28] it was shown that there is a very high correlation between NAQ and ClQ. However, NAQ, especially in the case of normal and pressed phonation types, has been shown to be much more robust against noise present in the glottal flows obtained by inverse filtering [27].

2.1.2. Amplitude-based parameters

When inverse filtering is done with a properly calibrated Rothenberg's mask it is possible to obtain valuable amplitude-based information from the glottal flow and its derivative. Figure 2 shows the amplitude-domain measures that are typically extracted from the glottal flow (upper panel) and its derivative (lower panel). The most widely used amplitude parameters are minimum flow (also called the DC-offset), the AC-flow and the negative peak amplitude of the flow derivative (also called the maximum airflow declination rate). The physiological explanation for the minimum flow is either an air leakage through a continuously unclosed part of the glottis or a vertical movement of the vocal folds [32]. In [5], it was reported that the (non-zero) minimum flow existed almost always in the obtained glottal flows. Furthermore, the increase of this amplitude value was significant between normal and soft phonations, while this did not occur between normal and loud phonations. Interestingly, there was no gender differences in the value of the minimum flow within the different loudness conditions. Reduction of the minimum flow with vocal intensity was also reported as a general trend in [9]. This phenomenon is probably related to the closing of the posterior part of the glottis. Differently from [5], the study by Sulter and Witt showed for male subjects an increase in the value of the minimum flow when intensity was changed from normal to loud. (This difference between the two studies might be due to a larger dynamics of SPL values studied in [9].) The minimum flow of singing voices was shown to behave differently than various other glottal flow parameters since it did not follow systematically the sound pressure level [25]. Mean values for the minimum flow in normal loudness conditions were 0.12 l/s and 0.09 l/s for males and females, respectively, in [5]. In [9], similar values were reported: the mean of the minimum flow was 0.11 l/s for males and 0.10 l/s for females.

The AC-amplitude of the glottal flow is known to correlate with SPL and its value increases with intensity, both for normal speakers and patients with voice problems [32]. There is also a strong relationship between the AC-amplitude and the amplitude of the source spectrum fundamental [6]. In [5], the value of AC-amplitude of male speech was found to be significantly larger than that of female speech in all loudness conditions analysed. This finding was corroborated by the study of Sulter and Witt, where it was also shown that males used a larger increment of the AC-flow when raising vocal intensity [9]. The larger value of the AC-flow of males is explained most likely by the larger glottal area function. However, a correspondence between the AC-amplitude and the amplitude of the vocal fold vibration was only partly corroborated in [33]. In their study, it was shown that in pressed phonation with ventricular vocal fold adduction the AC-amplitude was lower than expected as compared to the measured amplitude of the vocal fold vibration. The mean value of the AC-amplitude in normal loudness conditions was 0.26 l/s and 0.14 l/s for males and females, respectively, in [5]. In [9], the mean of the AC-amplitude in normal intensity was 0.57 l/s and 0.26 l/s for males and females, respectively.

It is also possible to parameterise the amplitude-based features of the glottal flow by combining the AC-flow and the minimum flow [34]. The ratio between the minimum flow and the AC-amplitude has been shown to correlate with perceived breathiness [35].

Since the main excitation of the vocal tract occurs during the glottal closing phase [36], it is reasonable to focus the parameterisation of the glottal flow near the time instant of the main excitation. The most widely used parameter focusing on the amplitude features of the glottal closing phase is the negative peak amplitude of the flow derivative [e.g., 5, 6, 8, 9, 31]. This amplitude measure has been shown to correlate strongly with SPL [6]. It has also been found that the negative peak amplitude of the flow derivative shows abnormally high values for voices with organic manifestations of vocal hyperfunction [13]. Typical values obtained for this amplitude-domain parameter are as follows: in [5] the mean value of the d_{peak} in normal loudness was 280 l/s^2 and 164 l/ s^2 for males and females, respectively. In [9], the mean value of the negative peak amplitude of the flow derivative of males was 1026 l/ s^2 and that of females was 504 l/ s^2. The large differences in the values of d_{peak} between the two studies depend according to [9] on the speaking task (a vowel was use in [5] but a CVC sequence in [5]), on the analysis bandwidth (1.6 kHz was used in [9] but 0.9 kHz in [5]) and on the intensity range (loud voices were on average 15 dB louder in [9]). (These reasons are also valid for the differences between [5] and [9] in the reported values of the mean AC-amplitude.)

All the parameterisation methods described so far are based on extracting certain time-based or amplitude-based measures from the most important instants of the flow or from its derivative. It is also possible to parameterise the voice source by searching for an artificial waveform that matches the original flow (or its derivative) obtained by inverse filtering. In other words, one aims to look for a compressed set of parameters that model the entire flow (or its derivative) instead of measuring isolated critical instants of the waveform. This approach implies selecting a pre-defined mathematical function to model the glottal flow (or its derivative) and a procedure, which optimises the parameters of the function in order to get the best possible match between the waveform given by inverse filtering and its artificial counterpart. Among the developed voice source models one of the most widely used is the Liljencrants-Fant model (LF-model) [37]. In this model, the flow derivative is presented by cosine and exponential functions that can be defined by four parameters. The LF-model was used together with automatic inverse filtering of continuous speech followed by automatic fitting of the model parameters in [38]. The authors report that the parameterisation works in general satisfactorily, but problems encounter in defining the parameter T_a of the LF model for waveform computed from consonants. (In the LF model, T_a is defined as the projection of the tangent of the flow derivative at the instant of the negative peak to the time axis. Extraction of this parameter from natural utterances has been reported to be difficult in various studies [e.g., 39].) Childers and Ahn [40] analysed the voice source behaviour by fitting the LF model on glottal flows computed from three voice types: model, vocal fry and breathy. The most important parameter to separate the three types was OQ. It is also possible to fit the voice source over a single glottal cycle by using a polynomial [18]. In this case, however, the model parameters have no physiological correspondence and therefore the approach is useful only for synthesis or coding applications.

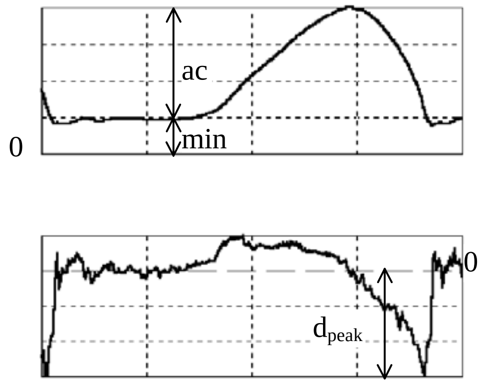

Figure 2: One cycle of the glottal flow (upper panel) and its time-derivative (lower panel) obtained by inverse filtering the vowel /a/ produced by a male speaker using normal phonation. Amplitude values: Minimum flow (min), AC-flow (ac) and the negative peak amplitude of the derivative (d_{peak}).

2.2. Frequency-domain methods

Changing the shape of the time-domain waveform of the glottal flow, for example, in the axis breathy-normal-pressed corresponds in the frequency-domain to altering the decay of the power spectrum of the voice source. Therefore, the methods developed to parameterise the glottal flow in the frequency-domain typically measure the spectral decay in one form or another. The spectrum can be the ordinary FFT-spectrum computed either pitch-asynchronously over several fundamental periods, i.e., with a harmonic structure, or pitch-synchronously over a single glottal cycle. It is also possible to measure the spectral decay by using all-pole modelling in computation of the source spectrum.

Parameterisation of the decay of the voice source spectrum using a pitch-asynchronously computed spectrum applies information located on F0 and its multiple integers, that is, the harmonics. Childers and Lee [29] presented a quotient, called harmonic richness factor (HRF), which 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. With this quotient vocal fry was characterised by a high value (2.1 dB), modal voices yielded a medium value (-9.9 dB) and breathy phonation a larger spectral decay (-16.8 dB). Applying the spectral harmonics of the glottal airflow waveform in the quantification of voice production has also been used by Howell and Williams [41, 42], who measured the decay of the voice source spectrum by computing linear regression analysis over the first eight harmonics. Titze and Sundberg [43] analysed the spectral decay of the voice source of singers by computing the difference between the amplitude of the fundamental and the second harmonic. This measure, usually denoted by H1-H2, is demonstrated in Fig. 3 by two examples. By analysing singing voices, it was shown in [25] that H1-H2 had a large correlation with CQ. A measure of the voice source spectrum based on pitch-synchronously computed spectrum was presented in [44]. This measure, the Parabolic Spectral Parameter (PSP), matches a second order polynomial to the flow spectrum computed over a single glottal cycle. The behaviour of the spectrum is measured by the coefficient of the second order polynomial. The final value of PSP is defined by normalising the obtained coefficient with a similar coefficient defined from a hypothetical source waveform of the same period length but with a maximal spectral decay. In [44], PSP was shown to differentiate three phonation types (breathy, normal, pressed) effectively.

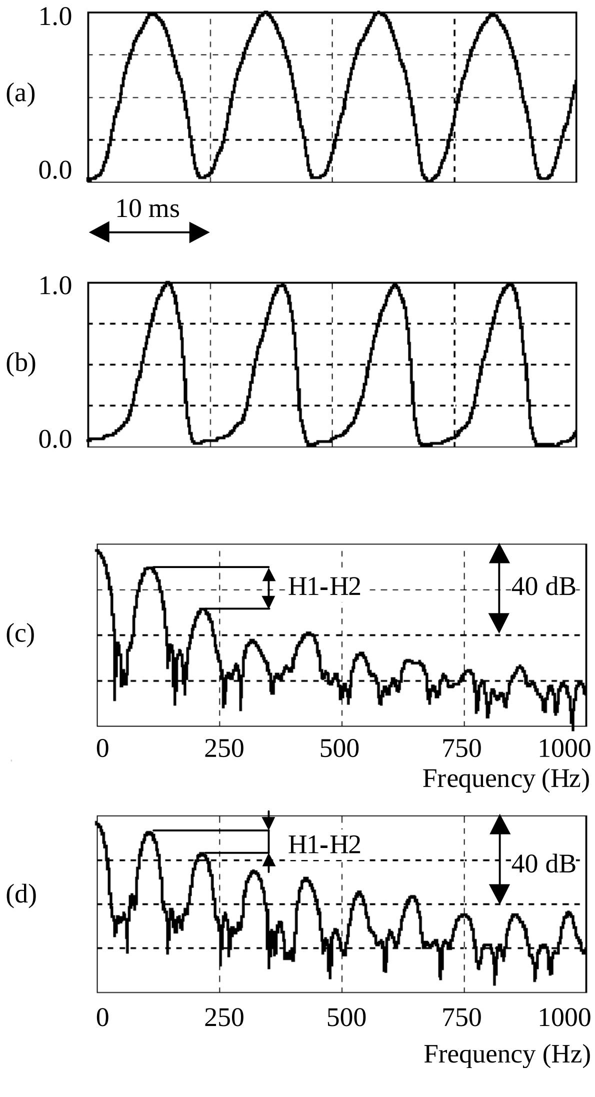

Figure 3: Glottal flows estimated by inverse filtering the acoustical speech pressure waveforms pronounced by a male speaker (vowel /e/). Time-domain waveforms of the glottal flow are shown in breathy and pressed phonation by panels (a) and (b), respectively (y-axis arbitrary). Spectra of the glottal flows in breathy and pressed phonation are shown by panels (c) and (d), respectively. Different spectral decay of the two phonations is quantified by H1-H2: the value of H1-H2 is 18.4 dB and 9.6 dB in panel (c) and panel (d), respectively.

3. Conclusions

Analysis of voice production with inverse filtering is practically always combined with the parameterisation of the obtained glottal waveforms. Therefore, in planning a voice production study it is essential to select the "best" possible technique among the various candidates in order to parameterise the outputs of the inverse filtering stage. There is hardly any single method that would outperform all the others in every respect: once the original time-domain pulseforms are expressed using any parameterisation method, which is in the form of a single numerical value, a major part of the original information embedded in the waveforms will be thrown away. The key issue is therefore how to select a parameter that would waste a minimum amount of that information, which on the focus in the study in question. The following issues might help a voice source researcher in the selection of the "best" method.

Many important features in speech communication, such as vocal loudness and phonation type, are reflected in the changes of the glottal closing phase. Therefore, the role of the closing phase should be emphasised in the time-domain parameterisation of the glottal flow. In terms of the time-based parameters this implies emphasising the role of ClQ and NAQ, because they take into account the pulseform only during the closing phase. These parameters, in particular the recently developed NAQ, are also the most robust among the time-based parameters, because their extraction does not involve determining the problematic time-instant of the glottal opening. Hence, if only a single time-based parameter is to be selected in the parameterisation of the glottal flow, it is recommended to prioritise either ClQ or NAQ. Similarly, if a single amplitude-based parameter is to be selected, it is recommend to use the negative peak amplitude of the glottal flow derivative, because it also reflects changes in the most important part of the glottal cycle, the closing phase. It is worth emphasising that obtaining truthful information about the voice source during the rapidly changing closing phase calls for using an analysis bandwidth that is wide enough [45]. In particular, if the derivative of the glottal flow is taken advantage of in the parameterisation and if the speech data consists of voices with small spectral decay (e.g., loud voices or speech produced with a pressed phonation type) it is recommended to use a bandwidth with is at least 4 kHz.

There are many large studies on voice production [e.g., 5, 9], where parameterisation of the glottal flow has been computed with several different time-domain methods, but without applying any frequency-domain technique. This lack of interests in the application of the frequency-domain methods is surprising given the fact that changing the glottal flow pulse in natural voice production will most likely correspond to changing the spectral decay of the voice source spectrum. Especially in those studies, where changing of the glottal pulse is studied in different intensity conditions, it would be feasible to involve at least one frequency-domain method in addition to the widely used time-domain techniques.

The use of the flow mask us undoubtedly useful, because it makes possible obtaining the AC-flow and minimum flow values of the glottal source. These values, in turn, are crucial in understanding the behaviour of the voice source from the point of view of basic research, but they also have an important role in clinical applications. However, the mask also limits the frequency range to approximately 1.5 kHz [46] and therefore distorts especially the analysis of voices produced in loud phonation. Moreover, the use of the mask might violate natural speech production. There are also applications, where it is completely impossible to use inverse filtering based on the flow mask, because the recordings have to be done in a completely non-invasive manner (e.g., measuring loading of voice during the working day in ordinary offices or schools). In these studies the inverse filtering must be based on the free field recordings and it is not possible to obtain real amplitude-domain information from the glottal source. However, the use of robust time-domain methods (e.g., ClQ or NAQ) together with a frequency-domain technique to measure the spectral decay of the voice source (e.g., H1-H2 or PSP) makes possible quantification of the glottal source without the need for the absolute flow values.

4. Acknowledgements

This study was supported by the Academy of Finland (project no. 200859).

5. References

[1] Miller, R.L., "Nature of the vocal cord wave", J. Acoust. Soc. Am., 31(6):667-677, 1959.

[2] Rothenberg, M., "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing", J. Acoust. Soc. Am., 53(6):1632-1645, 1973.

[3] Flanagan, J.L., *Speech Analysis, Synthesis, and Perception*, Springer, New York, 1972.

[4] Monsen, R.B. and Engebretson, A.M., "Study of variations in the male and female glottal wave", J. Acoust. Soc. Am., 62(4):981-993, 1977.

[5] 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.

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

[7] 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.

[8] 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.

[9] Sulter, A.R. and Wit, H.P., "Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age", J. Acoust. Soc. Am., 100(5):3360-3373, 1996.

[10] Cummings, K. and Clements, M.A., "Analysis of the glottal excitation of emotionally styled and stressed speech", J. Acoust. Soc. Am., 98(1):88-98, 1995.

[11] Gobl, C. and Ni Chasaide, A., "The role of voice quality in communicating emotion, mood and attitude", Speech Comm., 40:189-212, 2003.

[12] Plumpe, M.D., Quatieri, T.F., and Reynolds, D.A., "Modeling of the glottal flow derivative waveform with application to speaker identification", IEEE Trans. Speech and Audio Proc., 7(5):569-586, 1999.

[13] Hillman, R.E., Holmberg, E.B., Perkell, J.S., Walsh, M., and Vaughan, C., "Objective assessment of vocal hyperfunction: An experimental framework and initial results", J. Speech Hear Res., 32:373-392, 1989.

[14] Hillman, R.E., Holmberg, E., Perkell, J.S., Walsh, M., and Vaughan, C., "Phonatory function associated with hyperfunctionally related vocal fold lesions", J. Voice, 4(1):52-63, 1990.

[15] Lauri, E-R., Alku, P., Vilkman, E., Sala, E., and Sihvo, M., "Effects of prolonged oral reading on time-based glottal flow waveform parameters with special reference to gender difference", Folia Phoniatr Logop., 49:234-246, 1997.

[16] Vilkman, E., Lauri, E-R., Alku, P., Sala, E., and Sihvo, M. "Loading changes in time-based parameters of glottal flow waveforms in different ergonomic conditions", Folia Phoniatr Logop., 49:247-263, 1997.

[17] Carlson, R., Granström, B., and Karlsson, I., "Experiments with voice modelling in speech synthesis", Speech Comm., 10:481-489, 1991.

[18] Childers, D.G. and Hu, H.T., "Speech synthesis by glottal excited linear prediction", J. Acoust. Soc. Am., 96(4):2026-2036, 1994.

[19] Childers, D.G., "Glottal source modeling for voice conversion", Speech Comm., 16:127-138, 1995.

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

[21] Alku, P. and Vilkman, E., "A comparison of glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers", Folia Phoniatr Logop., 48:240-254, 1996.

[22] Scherer, R.C., Arehart, K.H., Guo, C.G., Milstein, C.F., and Horii, Y., "Just noticeable differences for glottal flow waveform characteristics", J. Voice, 12(1):21-30, 1998.

[23] Price, P.J., "Male and female voice source characteristics: Inverse filtering results", Speech Comm., 8:261-277, 1989.

[24] Iwarsson, J., Thomasson, M., and Sundberg, J., "Effects of lung volume on glottal voice source", J. Voice, 12(4):424-433, 1998.

[25] Sundberg, J., Andersson, M., and Hultqvist, C., "Effects of subglottal pressure on professional baritone singers' voice sources", J. Acoust. Soc. Am., 105(3):1965-1971, 1999.

[26] Sundberg, J., Cleveland, T.F., Stone, R.E, Jr., and Iwarsson, J., "Voice source characteristics in six premier country singers", J. Voice, 13(2):168-183, 1999.

[27] Alku, P., Bäckström, T., and Vilkman, E., "Normalized amplitude quotient for parameterization of the glottal flow", J. Acoust. Soc. Am., 112(2):701-710, 2002.

[28] Bäckström, T., Alku, P., and Vilkman, E., "Time-domain parameterization of the closing phase of glottal airflow waveform from voices over a large intensity range", IEEE Trans. on Speech and Audio Proc., 10(3):186-192, 2002.

[29] Childers, D.G. and Lee, C.K., "Vocal quality factors: Analysis, synthesis, and perception", J. Acoust. Soc. Am., 90(5):2394-2410, 1988.

[30] Sapienza, C.M., Stathopoulos, E.T., and Dromey, C., "Approximations of open quotient and speed quotient from glottal airflow and EGG waveforms: Effects of measurement criteria and sound pressure level", J. Voice, 12(1):31-43, 1998.

[31] Fant., G., "The voice source in connected speech", Speech Comm., 22:125-139, 1997.

[32] Hertegård, S., "Vocal fold fibration as studied with flow inverse filtering", Academic Dissertation, Dept. of Logopedics and Phoniatrics, Karolinska Institutet, Huddinge University Hospital, Sweden, 1994.
[33] Hertegård, S. and Gauffin, J., "Glottal area and vibratory patterns studied with simultaneous stroboscopy, flow glottography and electroglottography", J. Speech Hear. Res., 38:85-100, 1995.
[34] Isshiki, N., "Vocal efficiency index", In K.N. Stevens and M. Hirano (Eds.), Vocal Fold Physiology, Tokyo: University of Tokyo Press, 193-203, 1981.
[35] Fritzell, B., Hammarberg, B., Gauffin, J., Karlsson, I., and Sundberg, J., "Breathiness and insufficient vocal fold closure", J. Phonetics, 14:549-553, 1986.
[36] Fant, G., *Acoustic Theory of Speech Production*, the Hague: Mouton, 1960.
[37] Fant, G., Liljencrants, J., and Lin, Q., "A four-parameter model of glottal flow", Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Sweden, 4:1-13, 1985.
[38] Strik, H. and Boves, L., "On the relation between voice source parameters and prosodic features in connected speech", Speech Comm., 11:167-174, 1992.
[39] Fröhlich, M., Michaelis, D., and Strube, H.W., "SIM-Simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals", J. Acoust. Soc. Am., 110(1):479-488, 2001.
[40] Childers, D.G. and Ahn, C., "Modeling the glottal volume-velocity waveform for three voice types", J. Acoust. Soc. Am., 97(1):505-519, 1995.
[41] Howell, P. and Williams, M., "The contribution of the excitatory source to the perception of neutral vowels in stuttered speech", J. Acoust. Soc. Am, 84(1):80-89, 1988.
[42] Howell, P. and Williams, M., "Acoustic analysis and perception of vowels in children's and teenagers' stuttered speech", J. Acoust. Soc. Am., 91(3):1697-1706, 1992.
[43] Titze, I. and Sundberg, J., “Vocal intensity in speakers and singers”, J. Acoust. Soc. Am., 91(5):2936-2946, 1992.
[44] Alku, P., Strik, H., and Vilkman, E., "Parabolic Spectral Parameter - A new method for quantification of the glottal flow", Speech Comm., 22:67-79, 1997.
[45] Alku, P. and Vilkman, E., "Effects of bandwidth on glottal airflow waveforms estimated by inverse filtering", J. Acoust. Soc. Am., 98(2):763-767, 1995.
[46] Hertegård, S. and Gauffin, J., "Acoustic properties of the Rothenberg mask", Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Sweden, 2-3:9-18, 1992.