Publication VI

Laura Lehto, Matti Airas, Eva Björkner, Johan Sundberg, and Paavo Alku, "Comparison of two inverse filtering methods in parameterization of the glottal closing phase characteristics in different phonation types.", *Journal of Voice*, 21(2), pp. 138–150, March 2007.

Comparison of Two Inverse Filtering Methods in Parameterization of the Glottal Closing Phase Characteristics in Different Phonation Types

*†Laura Lehto, *Matti Airas, *‡Eva Björkner, ‡Johan Sundberg, and *Paavo Alku

*†*Helsinki, Finland and* ‡*Stockholm, Sweden*

Summary: Inverse filtering (IF) is a common method used to estimate the source of voiced speech, the glottal flow. This investigation aims to compare two IF methods: one manual and the other semiautomatic. Glottal flows were estimated from speech pressure waveforms of six female and seven male subjects producing sustained vole /a/ in breathy, normal, and pressed phonation. The closing phase characteristics of the glottal pulse were parameterized using two time-based parameters: the closing quotient (C1Q) and the normalized amplitude quotient (NAQ). The information given by these two parameters indicates a strong correlation between the two IF methods. The results are encouraging in showing that the parameterization of the voice source in different speech sounds can be performed independently of the technique used for inverse filtering.

Key Words: Inverse filtering—Glottal flow—Closing quotient—Normalized amplitude quotient.

INTRODUCTION

Due to an increasing number of employees working in professions where voice is the main tool of trade, occupational voice research has become an increasingly important area of speech science. To explore voice and its production objectively, several approaches have been used. One of these is inverse filtering (IF), which was developed to estimate the source of voiced speech, that is, the glottal volume velocity waveform, and to examine glottal activity noninvasively. Because the glottal volume velocity is the acoustic source of (voiced) speech, information gained from it is of central interest in the clinical research and treatment of voice problems as well as in prevention of voice disorders.

IF was first presented by Miller in the late 1950s.[1] The idea behind IF is to form a model for the vocal tract transfer function. The effects of

Accepted for publication October 1, 2005.

Presented at the Voice Foundation's 33rd Annual Symposium: Care of the Professional Voice, June 2–6, 2004, Philadelphia, Pennsylvania.

From the *Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Helsinki, Finland; the †Phoniatric Department, ENT Clinic, Helsinki University Central Hospital, Helsinki, Finland; and the ‡Department of Speech, Music and Hearing, Royal Institute of Technology, Stockholm, Sweden.

Supported by the Helsinki University of Technology, the Academy of Finland (project number 201018 and 200859) and the Finnish Cultural Foundation.

Address correspondence and reprint requests to Laura Lehto, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, PO Box 3000 (Otakaari 5A), FIN-02015 HUT, Finland. E-mail: laura.lehto@iki.fi

Journal of Voice, Vol. 21, No. 2, pp. 138–150
0892-1997/$32.00

doi:10.1016/j.jvoice.2005.10.007

vocal tract resonances are then canceled from the produced speech waveform by filtering it through the inverse of the model. The result is an estimate of the glottal flow represented as a time-domain waveform.

IF methods can be classified into manual, semiautomatic, and automatic. In particular, the older techniques of IF typically used manually adjustable analog circuits in implementation of the inverse model of the vocal tract.[1] Manual methods permit the experimenter to manipulate formant frequencies and bandwidths precisely to yield the optimal settings for the vocal tract model from analog or digital input. Instead of adjusting formant bandwidths and center frequencies, the user of semiautomatic methods can change, for example, the order of the digital all-pole model of the vocal tract. This means that the IF method is given a constraint to use a certain maximum number of resonances in modeling the vocal tract. By using this information, the underlying algorithm then automatically defines the formant settings. It should be noted that some studies[2] consider manual IF synonymous with interactive IF, but this is not an unambiguity, because semiautomatic methods also require some user contribution. In automatic IF methods, the user typically first adjusts certain initial parameter values, after which the method estimates the voice source without any subjective user adjustments.

In IF analysis, the input can be either an oral flow or a free field speech pressure signal. The oral flow signal is recorded with a pneumotachograph mask, also known as Rothenberg's mask.[3] Use of the mask is advantageous, as it can obtain both the ac and the dc information of the underlying glottal flow pulse. However, the mask limits the frequency range of the voice source analysis[4] and, moreover, might confine the subject's natural way of phonation.[5] Microphone recordings allow a fully noninvasive approach to capture free voice production.[6] This requires the use of high-quality equipment (eg, the choice of microphone and amplifiers) and decent recording conditions (eg, control of background noise, microphone distance).

Certain parameters are needed for quantitative presentation of results, so that the true information gained from the IF procedure may be exploited. These glottal flow parameters aim to represent the most important features of the original flow waveforms in a compressed numerical form. Many different methods have been developed for the parameterization. They can be categorized, for example, depending on whether the parameterization is performed in the time domain or in the frequency domain. Time-domain methods include time-based parameters (quotients measuring critical time spans of the glottal pulse) and amplitude-based parameters (absolute amplitude values of the flow and its derivative). The most commonly used time-based parameters are open quotient (OQ), speed quotient (SQ), and closing quotient (ClQ).[7–12] The amplitude-based parameters typically extracted are minimum flow (also called the dc offset), the ac flow, and the negative peak amplitude of the flow derivative (d_{min}), also called maximum airflow declination rate.[7,10,12–14] It is also possible to define time-based parameters from amplitude measures by using, for example, the amplitude quotient (AQ) and its normalized version, the normalized amplitude quotient (NAQ).[15] The frequency-domain methods measure the spectral decay of the voice source and typically exploit information located at harmonics of the glottal flow spectrum. One of the most widely used parameters of this kind is the amplitude difference between the first and the second harmonics (H1-H2).[16]

Many studies in the field of voice research have exploited a combination of IF and parameterization. Different phenomena of voice production have been studied by concentrating on issues like phonation type,[17] intensity,[8] voice quality,[18] emotions,[19] pitch,[7,12] disturbed voice functions,[10,20–25] singing styles,[16,26–28] and vocal loading.[9,29,30] In addition, some studies have discussed IF from a methodological point of view.[6,31] Given the prevalence of IF in the field of voice science, it is surprising that the differences between IF methods have not yet been studied extensively. To the best of our knowledge, there are only two previous studies comparing IF methods. Hertegård et al[24] and Södersten et al[32] have compared manual and automatic IF methods. Both studies used the "Inverse" program for the automatic analysis of the glottal flow IF.[33] The automatic function means that the program continuously adjusts the inverse filter to the signal based on changes in the formant frequencies and

bandwidths.[32] The automatic program could be operated also semi-interactively, but in both Hertegård et al[24] and in Södersten et al,[32] this option was used sparsely. For the manual IF, Hertegård et al[24] used the INA[34] program and Södersten et al[32] performed IF during the recording using the Glottal Enterprises System. The Rothenberg flow mask was used in both studies when recording the flow samples. The subjects repeated the syllable string [ba: pa:pa:pa:p] three times at three loudness levels (normal/neutral, soft (not whispery), loud). The pitch was not strictly controlled, but the subjects were encouraged to phonate as close to habitual pitch as possible.

The study of Hertegård et al[24] used voice samples of 28 patients (9 women, 19 men) with spindle-shaped glottal insufficiency (SGI). The parameters in focus were peak flow, minimum flow, ac flow, mean flow, peak flow, glottal resistance, flow derivative, first formant (F1), OQ20% (the duty cycle of the flow waveform measured as the open quotient at 20% of the ac flow), sound pressure level (SPL), and subglottal pressure (Ps). They found no significant differences between the two IF methods in regard to the glottal airflow values and the estimates of glottal closure from flow glottograms. Södersten et al[32] used 17 normal female subjects in their study. The parameters studied were fundamental frequency (F0), SPL, peak flow, peak-to-peak flow (ie, ac flow), minimum flow, and maximum derivative (ie, d_{min}). There was a high level of agreement between the two IF methods sampled across loudness levels for the glottal flow parameters peak flow, minimum flow, peak-to-peak flow, and the maximum derivative.

The aim of this study is to compare manual (manual adjustment of formant frequencies) and semiautomatic IF methods. We were especially interested in analyzing whether glottal closing phase characteristics show larger variation when parameterized by manual IF method compared with semiautomatic IF. There are three major differences between this study and the two previous ones.[24,32] First, this study analyzes speech pressure signals instead of the flow signals used by Hertegård et al[24] and Södersten et al.[32] Second, the parameters also differ: Instead of extracting flow parameters as in Hertegård et al[24] and Södersten et al,[32] this study focuses on the parameterization of the time-domain behavior of the glottal closing phase by using two robust time-based parameters: the ClQ and the NAQ. Third, instead of loudness levels, three different phonation types (breathy, normal, and pressed) are examined to have a large dynamics of glottal pulse characteristics in the comparison of IF methodologies.

MATERIALS AND METHODS

Recordings

Six women and seven men participated in the recordings. They were between 27 and 42 years of age. None of the subjects had a history of any voice problem. The material recorded for the purposes of this study consisted of three strings of five /a:/ vowels produced in breathy, normal, and pressed manners. The vowel /a:/ was chosen because of its high first formant to minimize source–filter interactions and effects from yielding of the vocal tract walls.[24]

The recordings were made in the anechoic chamber at Helsinki University of Technology's Laboratory of Acoustics and Audio Signal Processing. The recording session was supervised by three expert instructors who were in the chamber with the subject. The subjects were trained to produce the different phonations, and the experts simultaneously determined whether any given sample was an accurate representation of the desired phonation type. The subjects were asked to repeat the phonations if necessary.

A Brüel & Kjær 4188 condenser microphone [frequency range from 8 to 12500 Hz (± 2 dB)] was placed at a distance of 40 cm from the subject's mouth. The microphone was connected to a Sony DTC-690 DAT recorder (Sony Corporation, Tokyo, Japan) through a preamplifier (Brüel & Kjær 2138 Mediator, Brüel & Kjær, Nærum, Denmark). The DAT recorder used a standard sampling rate of 48 kHz. Phase correction, as applied in older IF studies with analog recordings (eg, Holmes[35]), was not needed due to the use of high-quality phase-linear recording equipment. To prevent signal degration, the recorded signals were digitally transferred from DAT tapes to a computer. The frequency of the signals was downsampled to 22.05 kHz. The middle sample (the third of 5) of each phonation

type was analyzed. Finally, the analysis window was selected to cover 10 glottal cycles starting from 100 ms from the beginning of the sample.

IF procedure

The acoustical pressure waveforms were inverse filtered with the two techniques. The analyses were performed independently by six experimenters, three of which used manual IF and the other three semiautomatic IF. The manual IF was performed by three experimenters working at the Department of Speech, Music and Hearing at the Royal Institute of Technology (Kungliga Tekniska Högskolan, KTH) in Stockholm, Sweden. The semiautomatic IF was performed by three experimenters at the Laboratory of Acoustics and Audio Signal Processing at Helsinki University of Technology, Espoo, Finland. All experimenters were experienced users of the corresponding IF program.

The manual IF method used in this study was the custom-made *Decap* program (Svante Granqvist, Department of Speech, Music and Hearing, KTH). In this program, the user can manipulate formant frequencies and bandwidths by means of the computer cursor. The program displays the resulting waveform and the spectra of the input and filtered signals in real time. The criteria for correct IF when tuning the filter frequencies and bandwidths were a maximally flat horizontal closed phase for the flow waveform and a minimal remaining formant ripple. These criteria are commonly used in various studies.[3,36] The form of the spectrum of the flow pulse was also taken into account: A smooth envelope of the source spectrum was pursued as a result of the IF.

The semiautomatic IF method used in this study was the iterative adaptive inverse filtering (IAIF) method.[37] The method consists of two stages: First, a preliminary estimate of the glottal flow is computed. A low-order all-pole filter is then fitted to this rough estimate of the voice source to model the contribution of the glottal flow in the speech spectrum. An estimate of the vocal tract is then obtained by canceling the estimated glottal contribution and the effect of lip radiation. To improve the estimation of formant frequencies for high-pitch voices, the IAIF method models the vocal tract by using an effective technique, discrete all-pole modeling,[38] instead of the widely user conventional linear prediction. The IAIF method consists of two attributes that the user can affect: the order of the vocal tract model and the position of the zero of the first-order FIR filter that is used to model the lip radiation effect. The user adjusts these quantities until the outcoming estimate of the glottal flow shows a maximally long and ripple-free closed phase.

Examples of pulse forms computed by both of the IF methods are shown in Figure 1. This figure includes results obtained by inverse filtering the same speech sound (male speaker, normal phonation) by all six experimenters. It is worth noticing that both IF methods are based on the all-pole modeling of the vocal tract transfer function. Hence, they are well suited in the analysis of non-nasalized vowels.

Parameterization

The glottal flow waveforms estimated by both IF methods were parameterized by two time-based parameters: the ClQ and the NAQ (Figure 2). These parameters are among the most robust time-based parameters,[15] because their extraction does not involve the problematic determination of time-instant of the glottal opening. Studies by Alku et al[15] and Bäckström et al[39] have shown that there is a high correlation between NAQ and ClQ.

ClQ is defined as the ratio between the durations of glottal closing phase and the fundamental period. Correspondingly, NAQ is defined as the ratio of the ac flow amplitude to the negative peak amplitude of the flow derivative, normalized by the period length. It is worth noting that these two amplitude measures are the extreme values of the flow and its derivative, and therefore, they are straightforward to extract. It can be shown that the ratio between the ac flow amplitude and the negative peak amplitude of the flow amplitude is a time-domain quantity that represents a subsection of the glottal closing phase.[15,40] This quantity is interpreted by Fant[40] 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.''

The quantities needed for the computation of ClQ and NAQ were extracted by analyzing three signals—the microphone signal, glottal flow, and

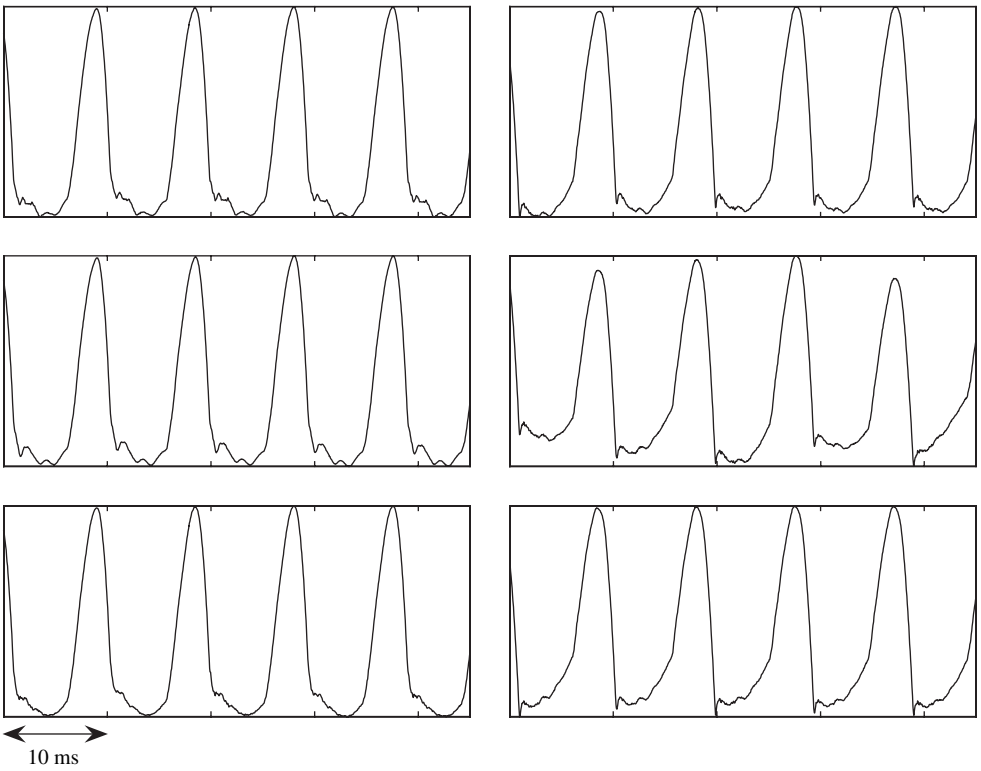

FIGURE 1. Different inverse-filtered glottal pulses. On the left-hand side, glottal pulses inverse filtered with the manual method; on the right-hand side, glottal pulses inverse filtered with the semi-automatic method. Same sample (male speaker, normal phonation) in all panels.

its derivative—over a time-window whose length was equal to the one used in IF (Figure 3). First, the fundamental frequency F0 was computed from the microphone signal using the YIN algorithm by de Cheveigne and Kawahara.[41] The average period length T_0 was defined as the inverse of the fundamental frequency. Then, the maximum amplitude A_{max} of the glottal flow was obtained. The corresponding time instant t_{max} is known to be the instant of peak flow in one glottal period inside the analysis window. The other glottal peaks are known to be approximately at distances of $\pm T_0$, $\pm 2\ T_0$, and so forth from the first peak. Thus, the instants of maximum flow in the other glottal periods of the analysis window were obtained by searching for the local maxima around these locations. After acquiring the peak flow time instants t_{max} and the corresponding flow values A_{max}, the other time instants needed for computation of ClQ and NAQ could be found. Within the period beginning at t_{max}, the minimum of the first derivative d_{min} and its time instant t_{dmin}, as well as the period minimum amplitude A_{min}, were determined. The first positive zero-crossing after t_{dmin} was chosen as the instant of the glottal closure t_c. The closing time (T_c) was then defined as $T_c = t_c - t_{max}$. Thus, ClQ is acquired as

$$ClQ=\frac{T_c}{T_0}=\frac{(t_c-t_{min})}{T_0}. \quad (1)$$

Given A_{min} and A_{max}, the maximal flow amplitude f_{ac} can be defined as $f_{ac} = A_{max} - A_{min}$. This yields AQ:

$$AQ=\frac{f_{ac}}{d_{min}}=\frac{A_{max}-A_{min}}{d_{min}}. \quad (2)$$

When the AQ is normalized by the average period length T_0, the NAQ is acquired:

$$NAQ=\frac{AQ}{T_0}=\frac{f_{ac}}{T_0 d_{min}}=\frac{A_{max}-A_{min}}{T_0 d_{min}}. \quad (3)$$

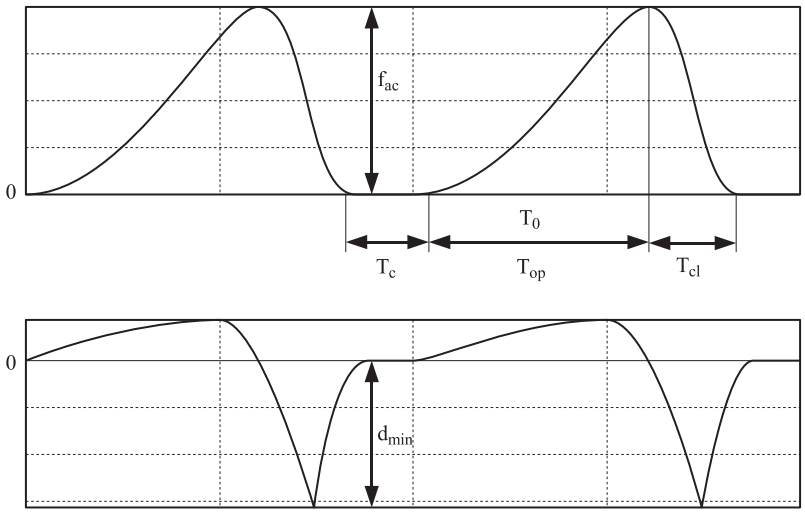

FIGURE 2. Schematic description of the computation of parameters ClQ and NAQ. f_{ac} : maximal flow amplitude; d_{min} : negative peak amplitude of the flow derivative; T_0 : length of the glottal cycle; T_c : closed phase of the glottal cycle; T_{op} : opening phase of the glottal cycle; T_{cl} : closing phase of the glottal cycle.

$$ClQ=\frac{T_{cl}}{T_0} \quad NAQ=\frac{f_{ac}}{d_{min}T_0}$$

The final parameter value in each sample was computed by taking the mean value of all analyzed 10 periods for both ClQ and NAQ.

Statistical analyses

The normality of the data was tested both using Q–Q plots as well as using the Shapiro–Wilk test for normality. The distributions were clearly skewed for both the ClQ and the NAQ. Therefore, parametric statistical tests were not used in the study.

To show that the ClQ and NAQ values computed by both manual and semiautomatic programs were independent of the experimenter, we used the Kruskal–Wallis test, which is a nonparametric equivalent of the one-way analysis of variance. The paired Wilcoxon signed rank test was used to assess group median paired differences between different methods, because it is a nonparametric equivalent of the paired *t* test. Before applying the Wilcoxon signed rank test, the ClQ values were square root transformed and the NAQ values were log transformed, because the test assumes that the population distribution is symmetric. These transforms were found to correct the skewedness of the parameter distributions.

Pearson's product-moment correlation was used to examine the level of association between parameter values acquired using different IF methods. Although 95% confidence intervals were calculated due to unfulfilled normality assumptions, they should be considered only suggestive in nature. Linear regression was used to estimate the nature of parameter differences between different IF methods.

Different phonation types were included in the voice samples to create large dynamics into time-domain behavior of the glottal closing phase. However, the effect of the IF procedure on different phonation types was not statistically tested because of the small amount of samples.

RESULTS

The Kruskal–Wallis test showed that, in both IF methods, the experimenter had no statistically significant effect on the ClQ and NAQ. Therefore, results obtained for each IF method were computed by averaging over the corresponding experimenters. The means and minimum and maximum values for the ClQ and NAQ are shown in Tables 1 and 2, respectively, for both IF methods. The tables also

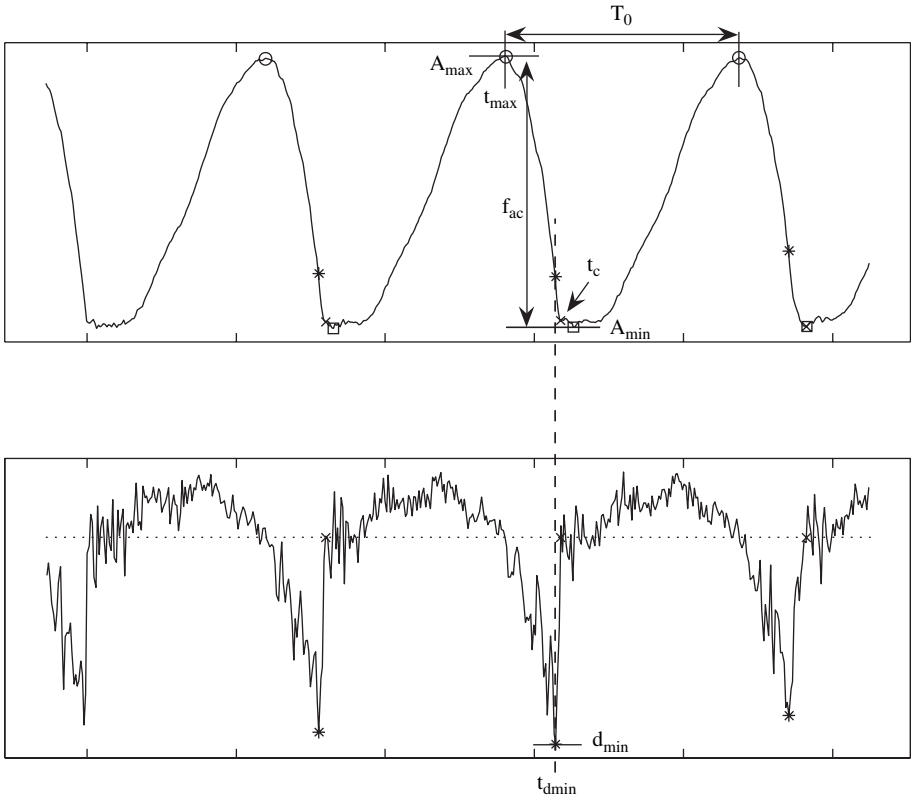

FIGURE 3. Description of the extraction of time instants and amplitude values needed in the computation of ClQ (Equation 1) and NAQ (Equations 2 and 3). T_0 : total length of the glottal cycle; t_{max}: period beginning; A_{max}: maximum amplitude; A_{min}: minimum amplitude; f_{ac} : maximal flow amplitude; t_c: glottal closure; d_{min} : negative peak amplitude of the flow derivative; t_{dmin}: time instant of the negative peak amplitude of the flow derivative.

show the coefficient of variation (cv) for each measure, ie, the ratio between the standard deviation and mean in percentage. The results turned out as expected: Both parameters gave small mean values for pressed phonation and larger values for breathy phonation. This finding is in line with previous studies of ClQ and NAQ.[15]

In the following, the statistical analysis on the effect of the IF method is discussed separately for the ClQ and NAQ.

TABLE 1. *Values of ClQ Computed in All Three Phonation Types by the Manual and the Semiautomatic IF Method*

	Men				Women			
ClQ	**mean**	**min**	**max**	**cv**	**mean**	**min**	**max**	**cv**
Breathy								
Manual	0.39	0.31	0.54	16.5%	0.41	0.35	0.45	6.5%
Semiaut.	0.36	0.30	0.44	11.0%	0.32	0.26	0.39	13.1%
Normal								
Manual	0.24	0.17	0.31	18.5%	0.34	0.24	0.44	18.7%
Semiaut.	0.23	0.17	0.30	16.8%	0.27	0.22	0.33	10.2%
Pressed								
Manual	0.20	0.13	0.31	34.9%	0.18	0.11	0.41	41.3%
Semiaut.	0.20	0.12	0.31	32.5%	0.18	0.14	0.24	15.9%

Abbreviation: cv, coefficient of variation (ie, standard deviation divided by mean).

TABLE 2. *Values of NAQ Computed in All Three Phonation Types by the Manual and the Semiautomatic IF Method*

	Men				Women			
NAQ	**mean**	**min**	**max**	**cv**	**mean**	**min**	**max**	**cv**
Breathy								
Manual	0.18	0.09	0.30	30.3%	0.20	0.16	0.26	12.5%
Semiaut.	0.17	0.11	0.27	29.5%	0.17	0.12	0.19	12.7%
Normal								
Manual	0.10	0.07	0.14	22.3%	0.15	0.09	0.18	16.1%
Semiaut.	0.10	0.07	0.14	21.1%	0.12	0.08	0.16	17.5%
Pressed								
Manual	0.07	0.05	0.12	28.4%	0.08	0.06	0.11	16.5%
Semiaut.	0.07	0.05	0.12	24.1%	0.07	0.05	0.10	19.1%

Abbreviation: cv, coefficient of variation (ie, standard deviation divided by mean).

The effect of the IF method on ClQ

The data of all subjects and all phonation types were pooled for each IF method. A paired Wilcoxon signed rank test was then carried out to determine whether the group medians differ from one another. The results showed that the IF method had a statistically significant effect on the ClQ ($P = 0.0005$). However, a strong correlation of 0.90 was found for the ClQ between the methods (95% confidence interval 0.81–0.95). The slope of the regression line was 1.24. The result is described in Figure 4.

The effect of gender was analyzed by the Wilcoxon signed rank test. In this test, the different phonation types were once again pooled together. It was found that the IF method does not have a statistically significant effect on the ClQ for men ($P = 0.06$). However, for women, the IF method showed a statistically significantly effect on the ClQ ($P = 0.003$).

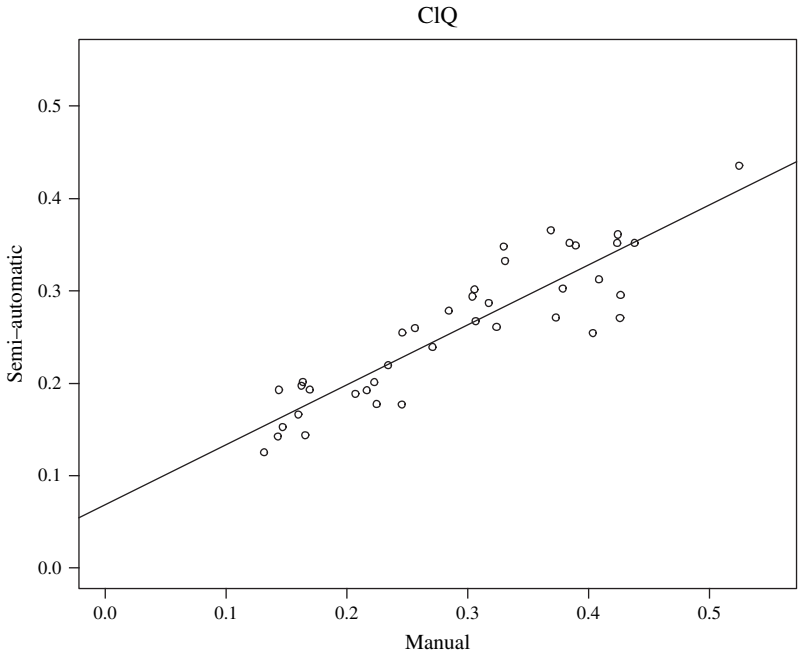

FIGURE 4. Correlation between semiautomatic and manual IF methods for the ClQ. Correlation coefficient $r = 0.90$.

The effect of the IF method on NAQ

To find out whether the group medians for the NAQ differ from each other, a paired Wilcoxon signed rank test was carried out by pooling all phonation types for both IF methods. As a result, the IF method showed a statistically significant effect on the NAQ value ($P = 0.00004$). Again, the correlation between the manual and semiautomatic IF method was very high, 0.96 (95% confidence interval 0.92–0.98). The slope of the regression line equaled 1.13. The result is illustrated in Figure 5.

The effect of gender on the NAQ was tested by the Wilcoxon signed rank test, which showed, as with the ClQ, that the difference was not significant for men ($P = 0.36$) and was statistically significantly for women ($P = 0.000008$).

DISCUSSION

In the area of occupational voice research, there will be a growing need to monitor and analyze voice production in realistic environments, such as a teacher speaking in a classroom. It is self-evident that only noninvasive methods can be used for this purpose. In addition, occupational voice care typically calls for analyzing extensive amounts of speech data because monitoring vocal loading, for example, requires analyzing voice production changes that take place over a long time. IF constitutes a conceivable method that, at least in principle, fulfills both of these requirements; it can be used to analyze glottal functions from noninvasive recordings in a manner that makes analysis of extensive data amounts possible with reasonable experimenter contribution. Toward this goal, this study compared two different IF methods, one manual and one semiautomatic, to find out whether they would give sufficiently similar results. Ours differs in three ways from the only previous studies within the field.[24,32] The current study (1) analyzed speech pressure signals instead of flow signals, (2) the results were concerned with the ClQ and the NAQ instead of emphasis on absolute flow values, and (3) three different phonation types (breathy, normal, pressed) were examined instead of loudness levels.

A major part of the previous IF studies have used flow recordings. However, when measuring, for example, voice loading changes throughout the working day in realistic situations, the use of a flow mask would be far too invasive and would therefore

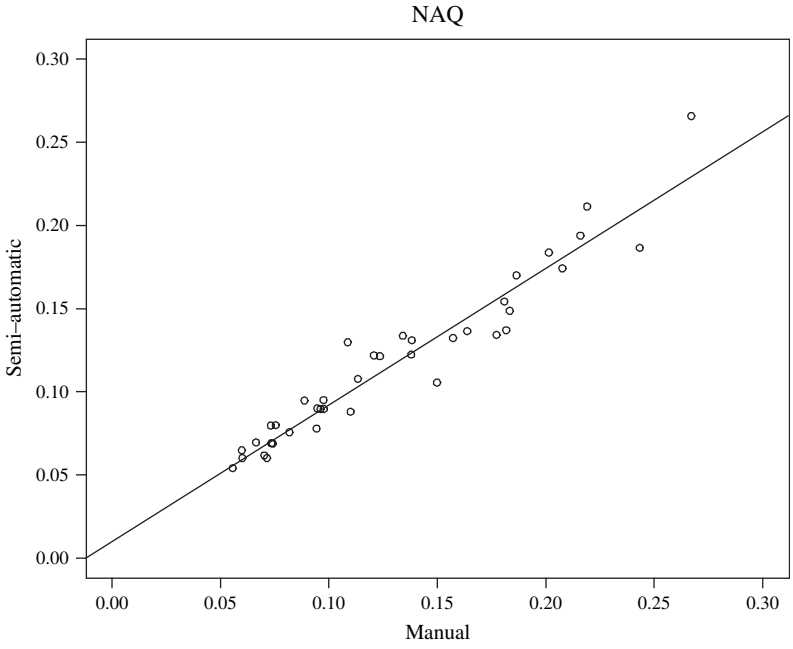

FIGURE 5. Correlation between semiautomatic and manual IF methods for the NAQ. Correlation coefficient $r = 0.96$.

be impractical. Orr et al[5] compared IF from flow and microphone signals from 61 nonpathological subjects (16 men and 45 women). Microphone and flow recordings of the syllable /pæ/ were inverse filtered by using an automatic pitch synchronous IF method.[5] The parameters SQ, OQ, H1-H2, and a measure of spectral slope were extracted from the glottal waveform. The results showed that the presence of a Rothenberg's mask used for the flow recordings had a significant effect on the parameters that were examined. These results might be explained by the subjects' inconsistent voicing strategies, a large within-speaker variation, and the acoustic effects of the flow mask. Studies by Hillman et al[10] and Holmberg et al[7] argue that the flow mask offers a noninvasive possibility to measure air flow. However, if voice measurements are to become a new routine as a part of occupational voice research, the psychological effect of the mask should also be taken into consideration.

The two previous studies on the comparison of IF methodologies[24,32] analyzed F0, SPL, and glottal flow amplitude parameters extracted from recordings made by means of a Rothenberg's mask. In Hertegård et al,[24] the air flow values (including peak flow, minimum flow, maximum flow, and negative peak amplitude of the flow derivative) computed with the automatic IF were 2.2–5.9% lower and in Södersten et al,[32] 2.7–7.7% lower than those estimated by the manual IF. This difference was within the acceptable limits of differences 5–10% set by Rothenberg and Nezelek[42] for clinical purposes, and they point out that normal voices can vary to such a degree or even more in a sentence or at different recording times. For pathological voices, the variation can be even larger. In the study of Hertegård et al,[24] the variation of the glottal parameters was large even when extracted using the same IF method. It was suggested that this might be caused by the larger variation of different voice source characteristics among the SGI patients studied than for normal voice patients in Södersten et al.[32] The current study investigated voice samples of normal speakers. IF works best in this kind of material with steady-state vowels for speakers with low F0 and a constant mode of phonation. In the case of more "complicated" signals (high F0, natural running speech, nonmodal phonation), there are more challenges.[2] These challenges need to be encountered if IF is to become a widely used research method. However, when comparing manual and (semi-)automatic IF methods, Södersten et al[32] point out that the automatic procedure does not require articulation to remain as steady as was needed with the manual IF method. The automatic procedure can automatically change the inverse filter to fit the signal and can change the formants during the phonations. This is advantageous when investigating voice samples from untrained subjects and patients, for example.

In this study, three different phonation types (breathy, normal, pressed) were examined so that a board variety of glottal functions could be used in assessing the functionality of IF and the parameterization. The results turned out to be as expected: ClQ and NAQ both give smaller mean values for the pressed phonation and larger values for the breathy phonation. This finding is in line with previous studies of ClQ and NAQ.[15] There was a statistically significant difference between the two IF methods for both of the parameters when all phonation types were pooled. However, the results also show that there was a strong correlation between the IF methods. The discrepancy between statistically significant differences and good correlation can be explained by the fact that the parameter values were systematically larger for the manual than for the semiautomatic method, as shown by the regression lines in Figures 4 and 5.

Both parameters indicated that there was no significant difference for male voices, whereas for female speakers, results from the IF methods differed significantly. The result reflects the IF of male voice being typically more straightforward than that of female speech. This, in turn, can be explained by the spectral differences in the speech sounds produced by the two genders; in the case of high-pitched female speech, there is a sparse harmonic structure in the speech spectrum that may distort accurate estimation of formants in IF.

The correlation between the two IF methods was found to be slightly lower for ClQ than for NAQ. This might be explained by the ClQ calculation formula: To determine the closing quotient, the beginning and the end of the closing phase must be defined precisely. According to Figure 6, it can be

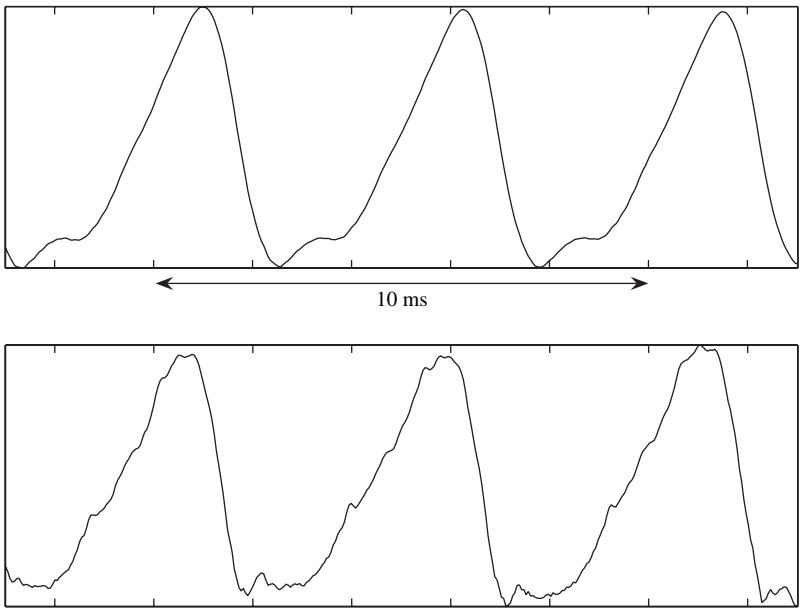

FIGURE 6. An example of a glottal pulseform computed by the manual (upper panel) and the semiautomatic (lower panel) IF method. Same sample (male speaker, normal phonation) in both pictures.

concluded that especially in a case of a smooth waveform, or in case of a waveform with formant ripple, the precise definition of these measures is difficult. NAQ is a more stable parameter because it measures closing phase characteristics from two easily detectible amplitude values, the ac amplitude of the flow and the negative peak amplitude of the glottal flow derivative.

It can be speculated that the differences between IF methods in this study might not be solely due to methodological differences: All experimenters were trained in using the corresponding program. Therefore, the small variation between the users of the two methods might also depend on research traditions. The wave shape of an ideal glottal pulseform resulting from IF might be interpreted differently by different ''schools.'' Another explanation might also be that with manual IF, there are more potential outcomes to choose from than for the semiautomatic IF program. However, the current results and those obtained in previous investigations[24,32] comparing manual and (semi-)automatic IF are congruent and encouraging in showing that discrepancies caused by the use of different IF methods are, in general, reasonably small.

It is worth noticing that the material used in this study was recorded in an ideal anechoic environment and consisted of sustained vowels produced by healthy speakers using average female and male F0. In addition, the analyses were performed only for the phoneme /a/, which is known to be the vowel with the highest first formant,[24] and therefore, its vocal tract contribution can be more easily separated from the glottal source than that of other utterances such as the vowel /i/. In contrast, if IF is to be exploited in field recordings, the realistic environment brings along many challenges. For example, continuous speech contains nasalized vowels and large variation in segment durations, both of which decrease the accuracy of IF techniques. Other properties of spontaneous speech that are problematic for IF analyses are high-pitched sounds and pathological voice qualities. Severe background noise will also affect the accuracy of IF. However, the current study shows that it is possible to obtain similar estimates of the voice source by using two different methods, both of which apply the microphone pressure signal of the vowel /a/ recorded from various speakers. This encourages us to continue developing IF methodologies that

can cope with more challenging speech material. It is possible, for example, to combine speech recognition to IF and to run inverse filtering only to those sections of continuous speech where the accuracy of IF is known to be at its best.

CONCLUSIONS

High correlation was found between a manual and a semiautomatic IF method when glottal closing phase characteristics were parameterized with time-domain quotients ClQ and NAQ from different phonation types. Manual IF showed a slightly larger variation in the parameter values. The result of this study can be considered encouraging in showing that automatic IF can be developed in the future to meet the needs of extensive speech data analysis.

REFERENCES

1. Miller RL. Nature of the vocal cord wave. *J Acoust Soc Am*. 1959;31:667–677.
2. Gobl C. *The voice source in speech communication [Doctoral thesis]*. Stockholm, Sweden: Royal Institute of Technology; 2003.
3. Rothenberg M. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *J Acoust Soc Am*. 1973;53:1632–1645.
4. Hertegård S, Gauffin J. Acoustic properties of the Rothenberg mask. *Speech Transmission Laboratory, Quarterly Progress and Status Report*. Stockholm, Sweden: Royal Institute of Technology; 1992. 2–3, 9–18.
5. Orr R, Cranen B, de Jong F. An investigation of the parameters derived from the inverse filtering of flow and microphone signals. In: *Proceedings of the ISCA Workshop on Voice Quality: Functions, Analysis and Synthesis (VOQUAL03)*. Geneva, Switzerland: ISCA; 2003:35–40.
6. Wong D, Markel J, Grey A. Least squares glottal inverse filtering from the acoustic speech waveform. *IEEE Trans Acoust, Speech Signal Proc*. 1979;27:350–355.
7. Holmberg E, Hillman R, Perkell J. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *J Acoust Soc Am*. 1988;84:511–529.
8. Dromey C, Stathopoulos E, Sapienza C. Glottal airflow and EGG measures of vocal function at multiple intensities. *J Voice*. 1992;6:44–54.
9. Lauri E-R, Alku P, Vilkman E, Sala E, Sihvo M. Effects of prolonged oral reading on time-based glottal flow waveform parameters with special reference to gender differences. *Folia Phoniatr Logop*. 1997;49:234–246.
10. Hillman R, Holmberg E, Perkell J, Walsh M, Vaughan C. Objective assessment of vocal hyperfunction: an experimental framework and initial results. *J Speech Hear Res*. 1989;32:373–392.
11. Scherer R, Arehart K, Guo C, Milstein C, Horii Y. Just noticeable differences for glottal flow waveform characteristics. *J Voice*. 1998;12:21–30.
12. Sulter AM, Wit HP. 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*. 1996;100:3360–3373.
13. Isshiki N. Vocal efficiency index. In: Stevens KN, Hirano M, eds. *Vocal Fold Physiology*. Tokyo: University of Tokyo Press; 1981:193–203.
14. Gauffin J, Sundberg J. Spectral correlates of glottal voice source waveform characteristics. *J Speech Hear Res*. 1989;2:556–565.
15. Alku P, Bäckström T, Vilkman E. Normalized amplitude quotient for parameterization of the glottal flow. *J Acoust Soc Am*. 2002;112:701–710.
16. Sundberg J, Thalén M, Alku P, Vilkman E. Estimating perceived phonatory pressedness in singing from flow glottograms. *J Voice*. 2004;18:56–62.
17. Alku P, Vilkman E. A comparison of glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers. *Folia Phoniatr Logop*. 1996;48:240–254.
18. Price PJ. Male and female voice source characteristics: inverse filtering results. *Speech Comm*. 1989;8:261–277.
19. Gobl C, NiChasaide A. The role of voice quality in communicating emotion, mood and attitude. *Speech Comm*. 2003;40:189–212.
20. Gomez P, Godino JI, Rodriguez F, et al. Evidence of vocal cord pathology from the mucosal wave cepstral content. In: *Proc IEEE Int Conf Acoust Speech Signal Proc (ICASSP '04)*. 2004;5:437–440.
21. Colton R, Brewer D, Rothenberg M. Evaluating vocal fold function. *J Otolaryngol*. 1983;12:291–294.
22. Fritzell B, Hammarberg B, Gauffin J, Karlsson I, Sundberg J. Breathiness and insufficient vocal fold closure. *J Phonet*. 1986;14:549–553.
23. Hammarberg B, Fritzell B, Gauffin J, Sundberg J. Acoustic and perceptual analysis of vocal dysfunction. *J Phonet*. 1986;14:533–547.
24. Hertegård S, Lindestad P-Å, Gauffin J. A comparison between manual and automatic flow inverse filtering for patients with spindle-shape glottis during phonation. *Scand J Log Phon*. 1994;19:117–134.
25. Hertegård S, Gauffin J. Insufficient vocal fold closure as studied by inverse filtering. In: Gauffin J, Hammarberg B, eds. *Vocal Fold Physiology*. San Diego, CA: Singular Publishing; 1991:243–250.
26. Sundberg J, Titze I, Scherer R. Phonatory control in male singing: a study of the effects of subglottal pressure, fundamental frequency, and mode of phonation of the voice source. *J Voice*. 1993;7:15–29.

27. Sundberg J, Kullberg Å. Voice source studies of register differences in untrained female singers. *Log Phon Vocol*. 1999;24:76–83.
28. Björkner E, Sundberg J, Cleveland T, Stone E. Voice source differences between registers in female musical theatre singers. *J Voice*. In press.
29. Vintturi J, Alku P, Lauri E-R, Sala E, Sihvo M, Vilkman E. Objective analysis of vocal warm-up with special reference to ergonomic factors. *J Voice*. 2001;15:36–53.
30. Vilkman E, Lauri E-R, Alku P, Sala E, Sihvo M. Loading changes in time-based parameters of glottal flow waveforms in different ergonomic conditions. *Folia Phoniatr Logop*. 1997;49:247–263.
31. Alku P, Vilkman E, Laukkanen A-M. Parameterization of the voice source by combining spectral decay and amplitude features of the glottal flow. *J Speech Lang Hear Res*. 1998;41:990–1002.
32. Södersten M, Håkansson A, Hammarberg B. Comparison between automatic and manual inverse filtering procedures for healthy female voices. *Log Phon Vocol*. 1999;24:26–38.
33. Imaizumi S. *Inverse. A Custom-Made Manual*. Stockholm, Sweden: Department of Speech Communication and Music Acoustics, Royal Institute of Technology; 1990.
34. Liljencrantz J. *INA. Custom-Made Program. Manual*. Stockholm, Sweden: Department of Speech Communication and Music Acoustics, Royal Institute of Technology; 1990.
35. Holmes J. Low-frequency phase distortion of speech recordings. *J Acoust Soc Am*. 1975;58:747–749.
36. Gauffin-Lindqvist J. Studies of the voice source by means of inverse filtering. *Speech Transmission Laboratory, Quarterly Progress and Status Report*. 1965;2:8–13.
37. Alku P. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. *Speech Comm*. 1992;11: 109–118.
38. El-Jaroudi A, Makhoul J. Discrete all-pole modeling. *IEEE Trans Signal Proc*. 1991;39:411–423.
39. Bäckström T, Alku P, Vilkman E. Time domain parameterization of the closing phase of the glottal airflow waveform from voices over large intensity range. *IEEE Trans Speech Audio Proc*. 2002;10:186–192.
40. Fant G. The voice source in connected speech. *Speech Comm*. 1997;22:125–139.
41. de Cheveigne A, Kawahara H. YIN, a fundamental frequency estimator for speech and music. *J Acoust Soc Am*. 2002;111:1917–1930.
42. Rothenberg M, Nezelek K. Airflow-based analysis of vocal function. In: Gauffin J, Hammarberg B, eds. *Vocal Fold Physiology.* San Diego, CA: Singular Publishing; 1991: 139–148.