Publication V

Paavo Alku, Brad Story, and Matti Airas, "Estimation of the voice source from speech pressure signals: Evaluation of an inverse filtering technique using physical modelling of voice production." *Folia Phoniatrica et Logopaedica*, 58(2), pp. 102–113, 2006.

Original Paper

Folia Phoniatr Logop 2006;58:102–113
DOI: 10.1159/000089611

Folia Phoniatrica et Logopaedica

Estimation of the Voice Source from Speech Pressure Signals: Evaluation of an Inverse Filtering Technique Using Physical Modelling of Voice Production

Paavo Alku[a] Brad Story[b] Matti Airas[a]

[a]Helsinki University of Technology, Espoo, Finland; [b]University of Arizona, Tucson, Ariz., USA

Key Words

Voice production · Physical modelling · Inverse filtering · Glottis

Abstract

Objective: The goal of the study is to use physical modelling of voice production to assess the performance of an inverse filtering method in estimating the glottal flow from acoustic speech pressure signals. ***Methods:*** An automatic inverse filtering method is presented, and speech pressure signals are generated using physical modelling of voice production so as to obtain test vowels with a known shape of the glottal excitation waveform. The speech sounds produced consist of 4 different vowels, each with 10 different values of the fundamental frequency. Both the original glottal flows given by physical modelling and their estimates computed by inverse filtering were parametrised with two robust voice source parameters: the normalized amplitude quotient and the difference (in decibels) between the levels of the first and second harmonics. ***Results:*** The results show that for both extracted parameters the error introduced by inverse filtering was, in general, small. The effect of the distortion caused by inverse filtering on the parameter values was clearly smaller than the change in the corresponding parameters when the phonation type was altered. The distortion was largest for high-pitched vowels with the lowest value of the first formant. ***Conclusions:*** The study shows that the proposed inverse filtering technique combined with the extracted parameters constitutes a voice source analysis tool that is able to measure the voice source dynamics automatically with satisfactory accuracy.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2006 S. Karger AG, Basel
1021–7762/06/0582–0102$23.50/0

Accessible online at:
www.karger.com/fpl

Paavo Alku, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology
PO Box 3000, FI–02015 HUT (Finland)
Tel. +358 9 4515680, Fax +358 9 460224
E-Mail paavo.alku@hut.fi

Introduction

Inverse filtering is a widely used method of voice production analysis, which aims to estimate the source of voiced speech, the glottal volume velocity waveform. The idea in inverse filtering is first to form a computational model for the vocal tract and then to cancel its effect from the produced speech sound by filtering the voice signal through the inverse of the model. The waveform, from which the cancelling of the vocal tract contribution is carried out, can be either the oral flow recorded in the mouth with a flow mask [1] or the pressure waveform captured by a microphone in free field outside the mouth [2]. The idea of estimating the glottal source with inverse filtering was first proposed by Miller [3], who used analogue circuits in implementing the vocal tract model. Later, several versions of inverse filtering were published [4–8]. All current methods are typically based on digital processing of speech and can be implemented either completely automatically or semi-automatically. The main area within which inverse filtering has been used is the basic research of speech production [9–12] and its applications to speech synthesis [13, 14]. However, there is currently increasing interest in inverse filtering methodologies in such areas of speech science as environmental voice care [15, 16] and analysis of the emotional content of speech [17, 18].

A fundamental problem present both in developing new inverse filtering algorithms and in comparing existing methods is the fact that assessing the performance of an inverse filtering technique is complicated. When inverse filtering is used in estimating the glottal flow of natural speech, it is actually *never* possible to assess in detail how closely the obtained waveform corresponds to the true glottal flow generated by the vibrating vocal folds. There are, however, certain voice production analysis methods that have been used together with inverse filtering in order to determine how the inverse filtering method in question reflects the function of the vocal folds. It is possible, for example, to compare the estimated glottal flow given by an inverse filtering method to other information signals extracted from the fluctuating vocal folds by exploiting techniques such as electroglottography [19], high-speed filming [20], videokymography [21] or high-speed digital videoscopy [22]. Beyond a doubt, these methods provide valuable information from the vibration of the vocal folds. They are, however, problematic in evaluating inverse filtering because of the unknown mapping between the glottal volume velocity and the corresponding information signal provided by these voice production analysis techniques, e.g. the time-varying impedance signal in the case of electroglottography and the glottal area function in the case of high-speed imaging techniques. On the other hand, it is possible to assess the accuracy of inverse filtering by using synthetic speech that has been created using a known, artificial waveform of the glottal excitation. This kind of evaluation, however, is not truly objective because speech synthesis and inverse filtering analysis are typically based on similar models of the human voice production apparatus, e.g. the source filter model [23].

In the current study, we use a different strategy in order to analyse how accurately an inverse filtering method can estimate the glottal flow. The idea is to use *physical modelling* of the vocal folds and the vocal tract in order to simulate time-varying waveforms of the glottal flow and radiated acoustic pressure. By using the simulated pressure waveform as an input to an inverse filtering method, it is possible to determine how closely the obtained estimate of the voice source matches the sim-

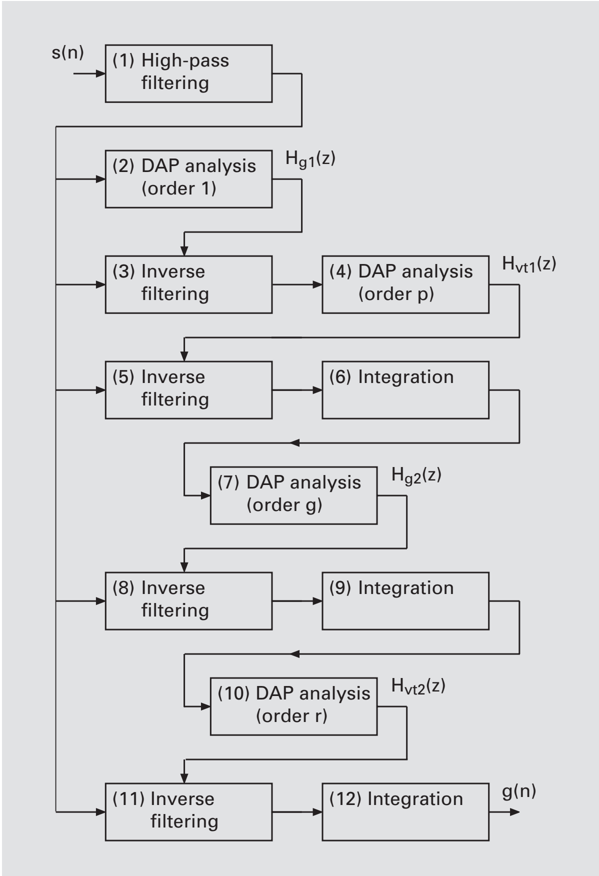

Fig. 1. Block diagram of the IAIF method. The input of the method is the speech pressure signal, denoted by s(n), from which the inverse filtering algorithm estimates the glottal volume velocity waveform, denoted by g(n).

ulated glottal flow. This approach is different from one where synthetic speech excited by an artificial form of the glottal excitation is used, because the glottal flow waveform results from the interaction of the self-sustained oscillation of the vocal folds with subglottal and supraglottal pressures. Hence, the glottal flow is generated by physical law, rather than by a parametric waveform model.

Materials and Methods

Inverse Filtering

The inverse filtering method used is based on the authors' previous experiments in developing automatic methods to estimate the glottal flow from the speech pressure waveforms with the iterative adaptive inverse filtering (IAIF) method [6]. The current method, the flow diagram which is shown in figure 1, is a slightly modified version from previous ones used by the authors. Parametric spectral models that are used in various blocks of the flow diagram are computed with the discrete all-pole modeling (DAP) method [24] instead of the conventional linear predictive analysis. This makes it possible to obtain estimates of the formant frequencies that are less biased by the harmonic structure of the speech spectrum. The description of the various blocks shown in the flow diagram is given below. (Transfer functions of the digital filters are expressed using the z-transform, where z denotes the complex frequency variable [25].)

Firstly (block No. 1), the speech signal is high-pass filtered in order to remove any distorting low-frequency fluctuations captured by the microphone during the recordings. The high-pass filter must be a linear-phase finite impulse response filter, and its cut-off frequency should be adjusted to be smaller than the fundamental frequency (F_0) of the speech sound analysed. In the current study, the cut-off frequency was set to 60 Hz. Secondly (block No. 2), a first-order all-pole filter is computed from the high-pass filtered speech signal in order to acquire a preliminary estimate for the combined effects of the glottal flow and the lip radiation effect on the speech spectrum. This stage yields a first-order all-zero filter, the transfer function of which is denoted by $H_{g1}(z)$ in figure 1. Next (block No. 3), the estimated effects of the glottal flow and lip radiation are cancelled from speech by filtering it with $H_{g1}(z)$. The output is analysed using a pth-order DAP analysis (block No. 4) to obtain a model for the vocal tract filtering, denoted by $H_{vt1}(z)$. Then (block No. 5), the effect of the vocal tract is cancelled from speech by filtering it through the inverse of the obtained pth-order model. A first estimate for the glottal flow is obtained (block No. 6) by cancelling the effect of the lip radiation by integrating the output of block No. 5. The IAIF method next computes (block No. 7) a new estimate, denoted by $H_{g2}(z)$, for the contribution of the glottal flow to the speech spectrum by computing DAP analysis of order g to the first estimate of the glottal excitation obtained. By first cancelling the effect of the estimated glottal contribution (block No. 8) and the lip radiation effect (block No. 9), a new model for the vocal tract filtering is obtained by an rth-order DAP analysis (block No. 10). The final result is obtained by cancelling the effect of the new vocal tract model (block No. 11) and the lip radiation effect (block No. 12).

The IAIF method described above has limitations. It is based on straightforward linear modelling of speech production without taking into account, for example, the interaction between the glottal source and the vocal tract. Moreover, the digital model of the vocal tract is a pure all-pole filter, which is not accurate for nasals. Despite these inherent limitations, the proposed technique provides a promising method to estimate the glottal flow especially given the fact that the method can be implemented (if desired) in a completely automatic manner at a reasonable computational cost.

Physical Modelling

Glottal flow and sound pressure waveforms used for testing the inverse filtering technique were generated by means of a computational model of vocal fold vibration and acoustic wave propagation. The body-cover structure of the vocal folds [26] was simulated with a lumped element model, consisting of three masses coupled to one another with stiffness and damping elements [27]. A diagram of the vocal fold model is shown in figure 2. The cover portion of each vocal fold was represented by the two small masses near the midline, m_1 and m_2, which were coupled to each other by the stiffness and damping elements k_c and d_c. These 'cover masses' were also coupled, via k_b and d_b, to a larger mass m_b, representing the vocal fold body. The mechanical constants corresponding to mass, stiffness, damping and vocal fold length were set to produce typical male and female fundamental frequencies. (See Titze and Story [28] for details on setting these constants based on physiological parameters.) For all simulations generated in this study, bilateral symmetry was assumed such that identical vibrations occur within the right and left vocal folds. The vocal fold model was coupled to the pressures in the trachea and the vocal tract in accordance with the aerodynamic and acoustic methods reported in Titze [29]. This coupling generated the driving forces on the vocal folds that allow for self-sustained oscillation, from which glottal area was derived. Glottal flow was determined by the interaction of the glottal area with the time-varying pressures that are present just below and just above the glottis.

Acoustic wave propagation in both the trachea and vocal tract was simulated with a wave reflection model (digital waveguide) that operates synchronously with the model of the vocal folds.

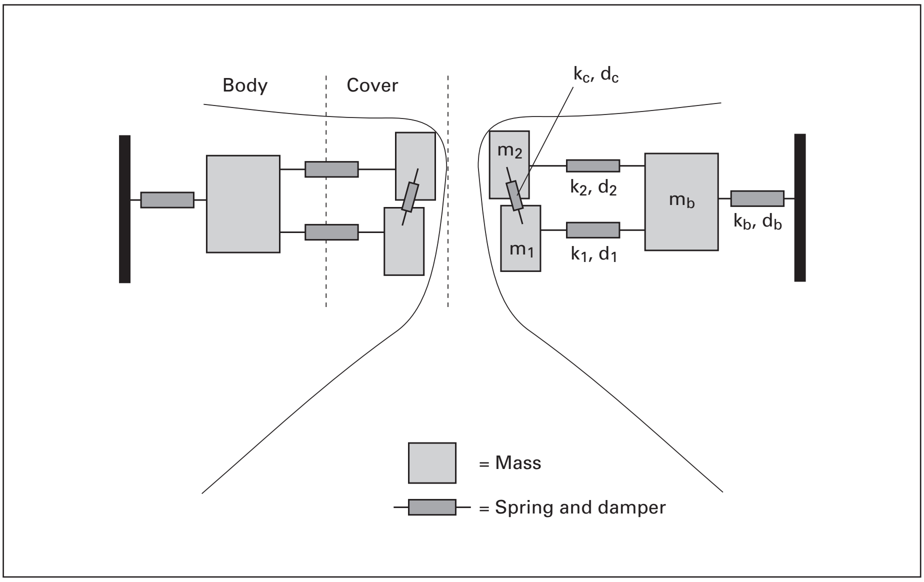

Fig. 2. Diagram of the three-mass model of the vocal fold cover-body structure. The cover is represented by the masses m_1 and m_2, and the body by m_b. The masses are coupled by stiffness and damping elements.

Figure 3a shows an idealization of a small portion of the upper trachea and lower vocal tract approximated as a set of concatenated cylindrical elements, or tubelets. Each tubelet represents the cross-sectional area at a specific distance from the glottis. Within each tubelet there exist forward (p^+) and backward (p^-) travelling wave pressures, calculated at each time sample during a simulation. The pressure components incident upon the first tracheal and vocal tract tubelets, respectively, are calculated as

$$\begin{aligned} p_1'^- &= p_1'^+ - (\rho c/A_1')u \\ p_1^+ &= p_1^- + (\rho c/A_1)u \end{aligned} \tag{1}$$

and can be seen in figure 3a as the two partial pressures travelling away from the glottis, labelled $\leftarrow p_1'^-$ and $p_1^+ \rightarrow$. A_1' and A_1 are cross-sectional areas, ρ is the air density, c is the speech of sound, and u is the glottal air flow at the current time sample determined with the methods in Titze [29]. The forward and backward travelling partial pressures in other tubelets are calculated with scattering equations. For example, pressures resulting from the area discontinuity between tubelets 1 and 2 are determined by

$$\begin{aligned} p_1^- &= p_1^+ r + p_2^-(1 - r) \\ p_2^+ &= p_1^+(1 + r) - p_2^- r \end{aligned} \tag{2}$$

where r is the reflection coefficient at the junction between A_1 and A_2,

$$r = \frac{A_1 - A_2}{A_1 + A_2} \tag{3}$$

Similar calculations are performed for every tubelet junction within each time sample. Hence, the forward and backward wave components are propagated through the airway in time synchrony with the mechanical vibration of the vocal fold model. The specific formulation of the

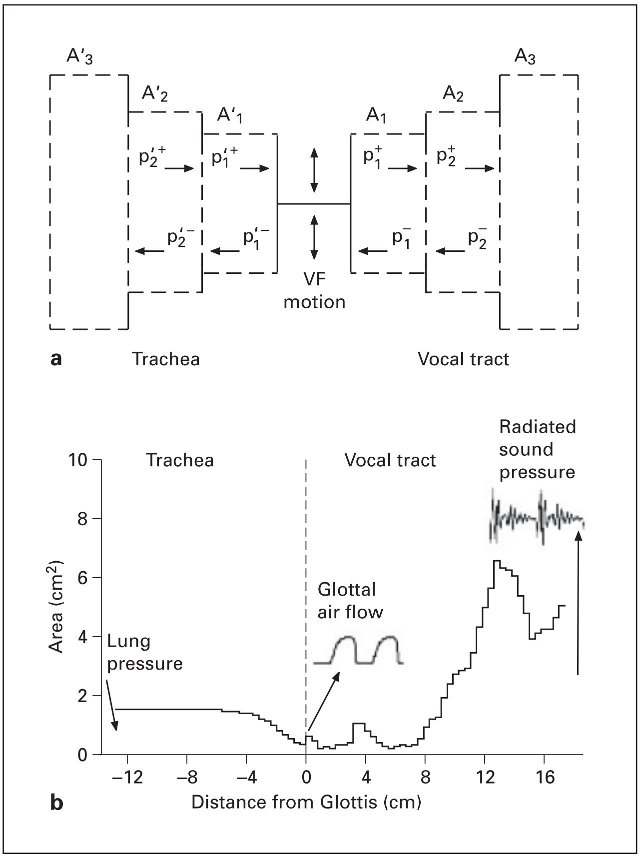

Fig. 3. Example of propagating acoustic pressures with a wave reflection model. **a** Tubular idealization of the upper portion of the trachea and lower portion of the vocal tract with forward (p^+) and backward (p^-) travelling partial pressures. Note that the vocal fold model is assumed to be operating at the point labelled 'VF motion'. **b** Sample area function of the tracheal/vocal tract system. Glottal air flow and radiated sound pressure signals are also shown.

wave reflection model used for this study was based on Liljencrants [30] and Story [31], and included extensive modifications of the reflection coefficient calculation to account for energy losses due to yielding walls, viscosity, reflection at the bronchi and radiation at the lips.

A conceptualization of the complete model is given in figure 3b, where the entire airway system extending from bronchi to lips is shown as an area function, plotted in stair-step fashion to indicate the discrete concatenation of cylindrical elements. The vocal fold model (glottis) is located between the trachea and the vocal tract at 0 cm. The trachea, which extends 12.7 cm in the leftward direction, was set to be a uniform tube with a cross-sectional area of 1.5 cm^2 but tapered to 0.3 cm^2 near the glottis. Extending rightward 17.46 cm is the vocal tract, which is configured in the shape of an adult male /a/ vowel. Shown on the left side of the plot is the lung pressure that drives the model; it was set to be constant at 7,840 dyn/cm^2 (8 cm H_2O). An example glottal flow waveform is indicated above the area function near the glottis (0 cm). Note that the ripples in the waveform are due to formant oscillations in the vocal tract. The sound pressure waveform radiated at the lips is also shown and can be considered analogous to a microphone signal recorded for a speaker.

Glottal flow and pressure waveforms were generated for 4 vowels (/a/, /ae/, /eh/ and /ih/) in both male and female configurations. The area functions were taken from those reported for an adult male in Story et al. [32]. For the simulations of male speech, the area functions were used directly with the exception of the vocal tract length, which was normalized to 17.46 cm. For female speech simulations, the same male area functions were used, but the length was non-uniformly scaled to 14.28 cm. The length scaling factors were based on those reported in Fitch and Giedd [33]. The trachea in all cases was the same as that shown in figure 3b.

In summary, the model is a simplified but physically motivated representation of a speaker. It generates both the signal on which inverse filtering is typically performed (microphone signal) and the signal that it seeks to determine (glottal flow), thus providing an idealized test case for an inverse filtering algorithm.

Results

Vowels produced by the physical modelling were inverse filtered with the IAIF method by using the following parameters (fig. 1): p = r = 12, g = 4. The sampling frequency was converted from its original value (44.1 kHz), used in the generation of sounds, to 11.025 kHz. The length of the analysis window was 200 ms, and the number of iterations in the DAP computations was 10. The lip radiation effect (blocks No. 6, 9 and 12 in fig. 1) was cancelled by a first-order all-pole filter with its pole at z = 0.999.

The comparison between the original flows given by physical modelling and those estimated from the pressure signals with inverse filtering was done by quantifying the corresponding flow signals with two voice source parametrisation techniques. The first of these, the normalized amplitude quotient (NAQ), is a time domain parameter, which measures characteristics of the glottal closing phase by using the ratio of the peak-to-peak flow and the negative peak amplitude of the flow derivative (see Alku et al. [34] for details.) As a second parameter, the difference (in decibels) between the levels of the first and second harmonics, denoted by $H_1 - H_2$, was used to estimate the decay of the glottal source spectrum [35]. These two parameters were selected because both of them are easy to extract automatically, thereby avoiding the use of subjective criteria in the computation of the parameters. Moreover, both of the parameters effectively reflect the dynamics of the glottal source with a single numerical value. The two parameters were extracted from individual glottal cycles of the flow pulse forms over a time span of 200 ms, and the final parameter value was computed as a mean of the data obtained.

Two examples of the waveforms obtained are shown in figures 4 and 5. In both of these figures, the original glottal flow, computed by physical modelling, is shown in the upper left panel, and its counterpart, estimated by inverse filtering, is depicted in the lower left one. Corresponding spectra are shown in the right panels. These two figures characterise typical trends found in the majority of the different vowels analysed in testing the inverse filtering method: the estimated glottal flow is in general quite close to the shape of the original flow generated by physical modelling, with the exception of the closed phase of the glottal cycle which is somewhat distorted by a formant ripple.

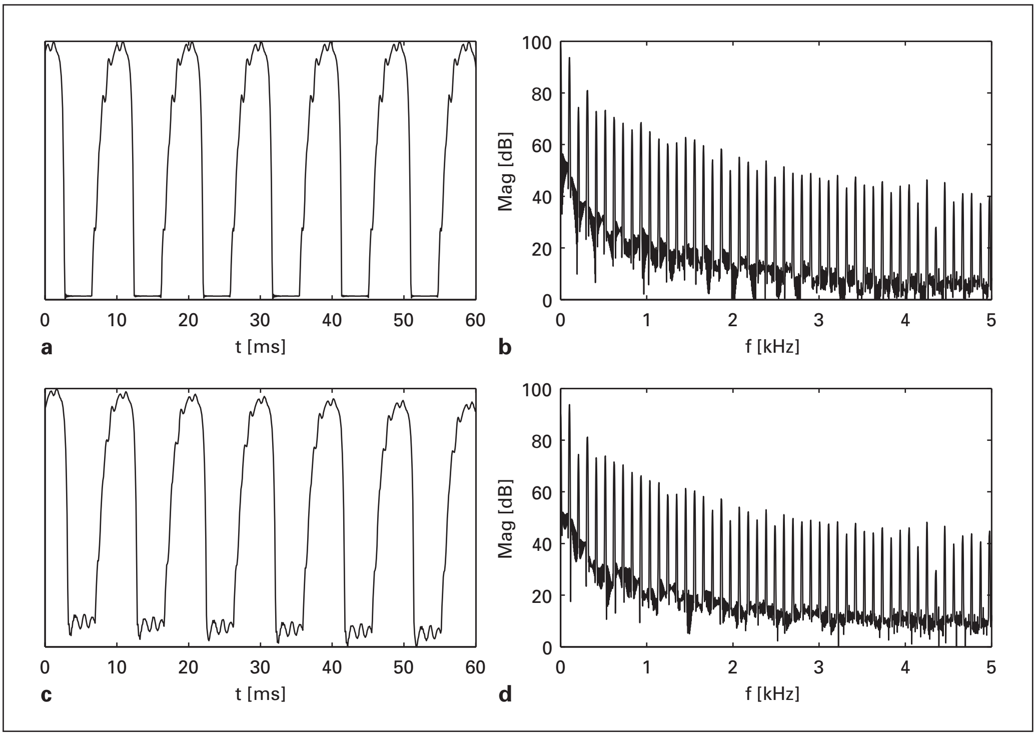

Fig. 4. Examples of flow waveforms and their power spectra (F_0 = 100 Hz). Upper graphs: the original glottal flow computed by physical modelling (**a**) and its power spectrum (**b**). Lower graphs: the flow waveform (**c**) and its power spectrum (**d**) estimated from the speech pressure waveform (vowel /a/) with inverse filtering.

The numerical results describing the similarity between the two flow signals are given in tables 1 and 2 for NAQ and $H_1 - H_2$, respectively. The data given in table 1 are expressed using a ratio of the NAQ value obtained from the original glottal flow to that extracted from the estimated one. This ratio, denoted by NAQ_{rat}, should be equal to unity when the estimation of the glottal flow has succeeded perfectly. The data in table 2 are expressed using a difference between the $H_1 - H_2$ value obtained from the original glottal and the corresponding $H_1 - H_2$ value extracted from the flows given by inverse filtering. This difference, denoted by $H_1 - H_{2dif}$, should be equal to zero in case estimation of the glottal flow has succeeded perfectly. The results on NAQ show that the error caused by inverse filtering was in general small: the mean value of NAQ_{rat} was 0.99 (range: 0.86–1.16; standard deviation: 0.06). The mean of the absolute value of $H_1 - H_{2dif}$ was 1.18 dB (range: –1.88 to 4.60; standard deviation of the absolute value: 1.40). The value of NAQ_{rat} was, in general, consistent without depending either on the fundamental frequency or the vowel, i.e. the characteristics of the vocal tract. The value of $H_1 - H_{2dif}$, on the other hand, indicated that the accuracy of inverse filtering deteriorated for voices of higher F_0. In addition, the performance of inverse filtering, as quantified with $H_1 - H_{2dif}$, reflected

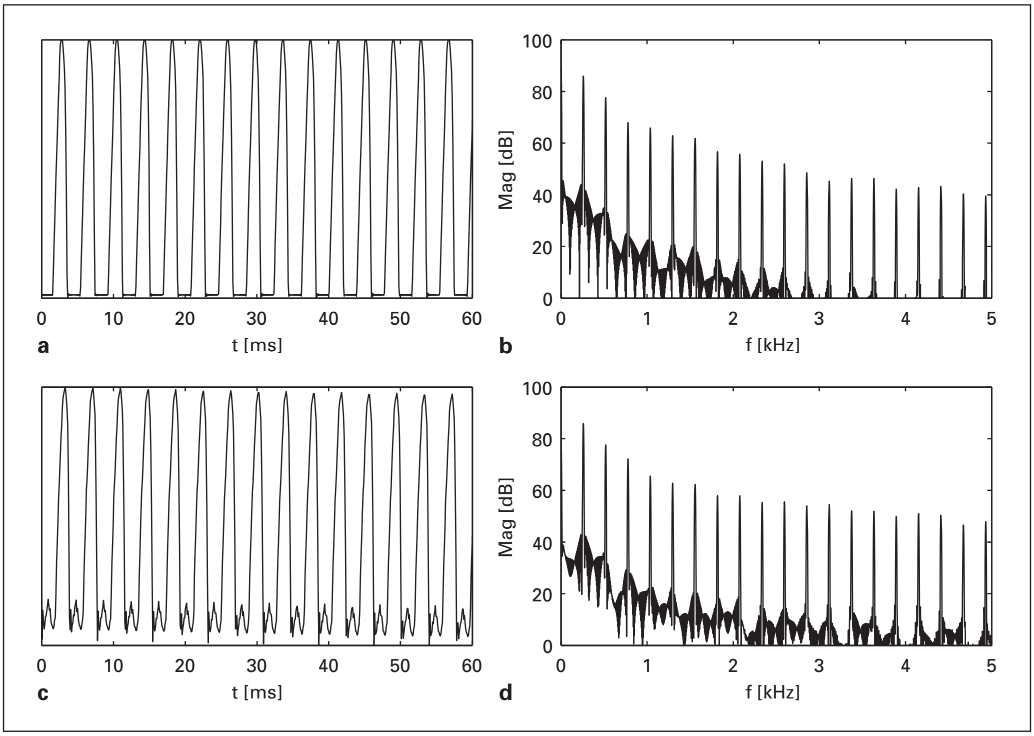

Fig. 5. Examples of flow waveforms and their power spectra (F_0 = 260 Hz). Upper graphs: the original glottal flow computed by physical modelling (**a**) and its power spectrum (**b**). Lower graphs: the flow waveform (**c**) and its power spectrum (**d**) estimated from the speech pressure waveform (vowel /ae/) with inverse filtering.

the differences in the formant settings of the underlying vowels: the value of $H_1 - H_{2dif}$ averaged over all 10 F_0 values equalling 0.44, 1.15, 1.16 and 2.06 dB for vowels /a/, /ae/, /eh/ and /ih/, respectively.

The distortion caused by inverse filtering can be assessed by comparing the obtained errors in the NAQ and $H_1 - H_2$ values to the variation of the same parameters when extracted from real speech in different phonation types. For the NAQ value, it is possible to use a previous study by Alku et al. [34], who showed that the ratio of the mean parameter value extracted in pressed phonation, divided by that extracted in normal phonation, equalled 0.80 and 0.69 for female and male speakers, respectively. Similarly, the mean NAQ value in normal phonation divided by that extracted in breathy phonation was shown to be equal to 0.68 for females and 0.46 for males. For the $H_1 - H_2$ parameter, it was demonstrated in a study by Huffman [36] that the difference of the spectral tilt between breathy and non-breathy voices was approximately 7 dB. These data suggest that the error introduced by inverse filtering is clearly smaller than the changing of the parameter value caused by altering the phonation type.

Table 1. Comparison of the original glottal flows and the estimated ones with a time-domain parameter, the NAQ

F_0, Hz	/a/	/ae/	/eh/	/ih/
100	0.9955	1.0095	1.0512	1.0377
115	0.9060	1.0140	0.9062	0.9605
130	0.9364	1.0132	0.9364	0.9649
145	1.0059	0.9320	0.9725	0.9974
200	0.9598	1.0558	1.0290	1.0788
210	0.9822	0.9988	1.0497	1.0250
230	0.9519	0.9368	1.0993	1.0152
255	0.9916	1.0704	1.1627	1.0062
310	1.0060	0.8646	0.9133	0.9856
410	0.9461	0.9221	1.1072	–

Inverse filtering was computed from 4 vowels (/a/, /ae/, /eh/, /ih/) at 10 different values of fundamental frequency (F_0) produced by physical modelling. The data are expressed using a ratio (denoted by NAQ_{rat}) between the NAQ value of the original flow and that extracted from the flow estimate given by inverse filtering. (Estimation of the glottal flow for the vowel /ih/ with F_0 = 410 Hz was distorted to such an extent that this voice sample was not included in the parametrisation.)

Table 2. Comparison of the original glottal flows and the estimated ones with a frequency domain parameter, the level difference of the first and second harmonic ($H_1 - H_2$)

F_0, Hz	/a/	/ae/	/eh/	/ih/
100	0.1374	–0.1683	–0.4031	–0.4240
115	0.1417	–0.1344	–0.1349	–0.4401
130	0.0847	–0.1754	–0.3253	–0.5460
145	0.0070	–0.0943	–0.1071	–0.5147
200	–0.2394	1.0351	1.0831	2.2467
210	–0.1235	1.9820	0.0664	3.3108
230	–0.1506	1.3918	–0.2853	4.5985
255	0.2112	0.0657	1.9202	3.8995
310	0.7806	4.5226	4.3788	2.5497
410	2.5021	–1.8849	2.8522	–

Inverse filtering was computed from 4 vowels (/a/, /ae/, /eh/, /ih/) at 10 different values of fundamental frequency (F_0) produced by physical modelling. The data are expressed using a difference (denoted by $H_1 - H_{2dif}$) between the $H_1 - H_2$ value obtained from the original flow and the corresponding $H_1 - H_2$ value extracted from the flow estimate given by inverse filtering. (Estimation of the glottal flow for the vowel /ih/ with F_0 = 410 Hz was distorted to such an extent that this voice sample was not included in the parametrisation).

Conclusions

Evaluation of inverse filtering methods is problematic because direct measurements of the glottal flow are difficult, if not impossible. In addition, using synthetic speech as test material does not make a fully objective evaluation possible because voice synthesis and inverse filtering are typically based on the same voice production models.

The present study aimed to avoid these fundamental limitations by using vowels produced with physical modelling in the evaluation of inverse filtering. The glottal flows estimated by an inverse filtering technique were compared to the original flows computed by physical modelling by parametrising both the time-based and frequency-based main features of the glottal waveforms using a parameter that is easy to extract. The results were encouraging in showing that combining the inverse filtering technique with the robust parametrisation yielded by NAQ and $H_1 - H_2$ produced, in general, only small differences between the original flow and its estimated version. This was true even though the voices analysed comprised a wide range of different F_0 values. Even in the cases when the estimated voice source was distorted during the glottal closed phase by formant ripple, the value of the parameters extracted from the estimated flow remained close to those extracted from the original glottal source. It is, however, necessary to emphasise that the position of the lowest formant will have a large effect on the performance of the method: for vowels like /i/, where the first formant is located near 300 Hz, the IAIF method is not recommended for use.

Acknowledgements

This study was supported by the Academy of Finland (projects No. 200859 and 205962) and the National Institutes of Health (grant No. R01 DC04789).

References

1 Rothenberg M: A new inverse-filtering technique for deriving the glottal airflow waveform during voicing. J Acoust Soc Am 1973;53:1632–1645.
2 Wong DY, Markel JD, Gray AH Jr: Least squares glottal inverse filtering from the acoustic speech waveform. IEEE Trans Acoust Speech Signal Processing 1979;27:350–355.
3 Miller RL: Nature of the vocal cord wave. J Acoust Soc Am 1959;31:667–677.
4 Veeneman DE, BeMent S: Automatic glottal inverse filtering from speech and electroglottographic signals. IEEE Trans Acoust Speech Signal Processing 1985;33:369–377.
5 Milenkovic P: Glottal inverse filtering by joint estimation of an AR system with a linear input model. IEEE Trans Acoust Speech Signal Processing 1986;34:28–42.
6 Alku P: Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Speech Commun 1992;11:109–118.
7 Krishnamurthy AK: Glottal source estimation using a sum-of-exponentials model. IEEE Trans Signal Proc 1992;40:682–686.
8 Fröhlich M, Michaelis D, Strube HW: SIM – Simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals. J Acoust Soc Am 2001;110:479–488.
9 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.
10 Gauffin J, Sundberg J: Spectral correlates of glottal voice source waveform characteristics. J Speech Hear Res 1989;32:556–565.

11 Sulter AM, Wit HP: Glottal volume velocity waveform characteristics in subjects with and without training, related to gender, sound intensity, fundamental frequency, and age. J Acoust Soc Am 1996;100:3360–3373.
12 Fant G: The voice source in connected speech. Speech Commun 1997;22:125–139.
13 Klatt DH, Klatt LC: Analysis, synthesis, and perception of voice quality variations among female and male talkers. J Acoust Soc Am 1990;87:820–857.
14 Childers DG, Ahn C: Modeling the glottal volume-velocity waveform for three voice types. J Acoust Soc Am 1995;97:505–519.
15 Vilkman E, Lauri ER, Alku P, Sala E, Sihvo M: Loading changes in time-based parameters of glottal flow waveforms in different ergonomic conditions. Folia Phoniatr Logop 1997;49:234–246.
16 Vintturi J, Alku P, Lauri ER, Sala E, Sihvo M, Vilkman E: The effects of post-loading rest on acoustic parameters with special reference to gender and ergonomic factors. Folia Phoniatr Logop 2001;53:338–350.
17 Laukkanen AM, Vilkman E, Alku P, Oksanen H: Physical variations related to stress and emotional state: a preliminary study. J Phonet 1996;24:313–335.
18 Gobl C, NiChasaide A: The role of voice quality in communicating emotion, mood and attitude. Speech Commun 2003;40:189–212.
19 Baken RJ: Electroglottography. J Voice 1992;6:98–110.
20 Baer T, Löfqvist A, McGarr NS: Laryngeal vibrations: a comparison between high-speed filming and glottographic techniques. J Acoust Soc Am 1983;73:1304–1308.
21 Svec JG, Schutte HK: Videokymography: high-speed line scanning of vocal fold vibration. J Voice 1996;10:201–205.
22 Granqvist S, Lindestad PÅ: A method of applying Fourier analysis to high-speed laryngoscopy. J Acoust Soc Am 2001;110:3193–3197.
23 Fant G: The Acoustics Theory of Speech Production. The Hague, Mouton, 1960.
24 El-Jaroudi A, Makhoul J: Discrete all-pole modeling. IEEE Trans Signal Proc 1991;39:411–423.
25 Oppenheim AV, Schafer RW: Discrete-Time Signal Processing. New Jersey, Prentice-Hall, 1989.
26 Hirano M: Morphological structure of the vocal cord as a vibrator and its variations. Folia Phoniatr Logop 1974;26:89–94.
27 Story BH, Titze IR: Voice simulation with a body-cover model of the vocal folds. J Acoust Soc Am 1995;97:1249–1260.
28 Titze IR, Story BH: Rules for controlling low-dimensional vocal fold models with muscle activities. J Acoust Soc Am 2002;112:1064–1076.
29 Titze IR: Regulating glottal airflow in phonation: application of the maximum power transfer theorem to a low dimensional phonation model. J Acoust Soc Am 2002;111:367–376.
30 Liljencrants J: Speech Synthesis with a Reflection-Type Line Analog; dissertation, Royal Institute of Technology, Stockholm, 1985.
31 Story BH: Physiologically Based Speech Simulation Using an Enhanced Wave-Reflection Model of the Vocal Tract; dissertation, University of Iowa, Iowa City, 1995.
32 Story BH, Titze IR, Hoffman EA: Vocal tract area functions from magnetic resonance imaging. J Acoust Soc Am 1996;100:537–554.
33 Fitch TW, Giedd J: Morphology and development of the human vocal tract: a study using magnetic resonance imaging. J Acoust Soc Am 1999;106:1511–1522.
34 Alku P, Bäckström T, Vilkman E: Normalized amplitude quotient for parameterization of the glottal flow. J Acoust Soc Am 2002;112:701–710.
35 Titze I, Sundberg J: Vocal intensity in speakers and singers. J Acoust Soc Am 1992;91:2936–2946.
36 Huffman M: Measures of phonation type in Hmong. J Acoust Soc Am 1987;81:495–504.