

Publication III

Matti Airas, “TKK Aparat: An Environment for Voice Inverse Filtering and Parameterization.” *Logopedics Phoniatics Vocology*, 33(1), pp. 49–64, 2008.

SPECIAL SECTION

ORIGINAL ARTICLE

TKK Aparat: An environment for voice inverse filtering and parameterization

MATTI AIRAS

Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Helsinki, Finland

Abstract

The study of the glottal flow, the acoustic excitation for voiced speech, provides insight into the voice signal, which is of potential benefit in many disciplines. One common method for estimating the glottal flow is inverse filtering, in which the effects of the vocal tract and the lip radiation are removed from a microphone signal. This paper presents a new inverse filtering and parameterization software package, which is available under an open-source licence. It provides a user-friendly graphical interface for rapid inverse filtering and parameterization, and the algorithms and parameters can be easily re-used in other projects. The system has already proved to be useful in algorithm development, speech science research, as well as in the study of occupational voice.

Key words: *Glottal flow, glottal flow parameterization, software, voice inverse filtering*

Introduction

The volume velocity of air-flow through the glottis, the glottal flow, is the excitation source for voiced speech. The glottal flow is the origin for many voice characteristics, such as the fundamental frequency, and several voice quality features, such as breathiness, harshness, and creakiness. Furthermore, it is an important element in the regulation of vocal intensity and in individual speech characteristics. Hence, being able to analyse the functioning of the voice source is essential in order to understand various acoustical cues used in everyday speech production. The study of the glottal flow gives insight into the voice signal, which is of potential benefit in many disciplines such as speech synthesis, study of vocal expression of emotions, and clinical diagnosis and treatment of voice.

Muscle pressure exerted on the lungs generates an air-flow through the trachea. During voicing, this air-flow is modulated in the larynx by setting the vocal folds into vibration, obstructing the flow periodically so that a quasi-periodic pulse train, the glottal flow, is created. The glottal flow is conceived

deep within the throat in the larynx, surrounded by many sensitive and vital organs and arteries. Due to the location of the larynx, glottal flow cannot be directly measured, since pneumotachographs cannot be readily inserted in the vocal tract without obstructing breathing and the actual voicing process. While several methods have been developed for the estimation of the glottal flow, they typically use the same basic fundamental assumptions: according to Fant's source-filter theory, the voice excitation and the vocal tract are linearly separable, and the voice source signal, i.e. the glottal flow, is filtered by the vocal tract to yield the air-flow at the mouth (1). This air-flow is then converted to a pressure waveform at the lips and propagated as a sound signal. The process is illustrated in the upper row of Figure 1. Due to the separability of the different model components, an estimate of the glottal flow can be acquired by removing the effects of the estimated vocal tract and the lip radiation from a measured air-flow or pressure waveform. This technique is called *inverse filtering*, since in the process the estimated vocal tract and lip radiation filters are

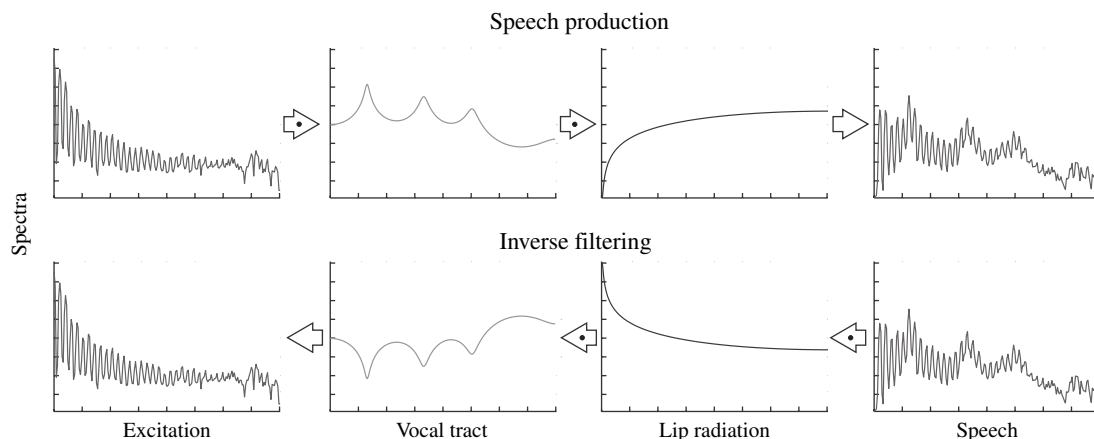


Figure 1. The graphs in the diagram are schematic spectra of the respective signals and filters. The upper row represents the separated speech production model. The lower row represents the corresponding inverse filtering process, in which the lip radiation and vocal tract filters are inverted to acquire an estimate for the glottal flow waveform.

inverted to acquire the glottal flow estimate, as shown in the lower row of Figure 1. It has been shown that in reality the voice source and the vocal tract interact, and the interaction is even vital in supporting the vocal fold vibration. Thus the source-filter theory should be considered a simplification of the actual voice production process (e.g. 2–4); however, despite its theoretical shortcomings, many studies have shown it to be valid in practice (e.g. 5).

Although the source-filter theory was formally published in Fant's book in 1960 (1), inverse filtering was already presented by Miller a year earlier (6). Since then, numerous articles on inverse filtering have been published. Two alternatives exist for the input signal in inverse filtering. Either a flow mask may be used to estimate the actual air-flow out of the mouth (7), or a microphone at a certain distance may be used to measure the speech pressure signal (8). If absolute flow values and measurement of the minimum flow are required, a calibrated flow mask has to be used. However, flow masks have poor frequency responses (linear only up to 1.6 kHz 9), and positioning the mask tightly around the mouth and nose poses restrictions on natural production of speech (9,10). In contrast, quality condenser microphones are commonly available, and their amplitude and phase response characteristics are excellent. Microphones may be placed on a stand at a predetermined distance from a stationary speaker, or they may even be attached to the speaker's head with a headset. Neither of these methods affects natural voice production. Due to these reasons, microphone recordings are widely used. In measurements taking place for example in real working situations, such as vocal loading and studies of occupational voice, flow masks cannot be used at all, thus necessitating the use of microphone recordings.

In the early inverse filtering studies, the vocal tract was estimated by setting the vocal tract formant frequencies by hand, appropriately called manual inverse filtering (6,7). Although manual inverse filtering is still in common use (e.g. 11), it is quite time-consuming, and, due to the manual adjustment of the anti-resonances by the experimenter, it is also subject to the user's personal preference in determining the final shape of the glottal flow waveform. Many different automatic inverse filtering methods have since been proposed. Allen and Curtis (12), Milenkovic (13), and Alku (14) have suggested inverse filtering methods based on linear prediction of the vocal tract. Strube (15), Wong et al. (16), Mataušek and Batalov (17), Ananthapadmanabha and Fant (18), and Plumpe et al. (19) have developed the idea of closed phase covariance analysis, in which the qualities of the vocal tract can be estimated from the closed phase of the glottal flow. Methods exploiting the frequency-domain characteristics of voiced speech have also been developed (e.g. 20), as well as methods utilizing a-priori knowledge of the glottal pulse shape (e.g. 21–23). A comprehensive review of different glottal inverse filtering methods may be found in Walker and Murphy (24).

While the estimated glottal flow is often inspected qualitatively, any quantitative analysis requires parameterization of the glottal flow pulses. There are three main categories of parameterization methods of the glottal flow: time-domain, frequency-domain, as well as model-based methods.

In time-domain methods the so-called critical time instants, for example the instant of the glottal closure, in the glottal flow pulses are marked, and the absolute or relative durations of the phases defined by the critical time instants are measured. Some of the most conventional phases are illustrated

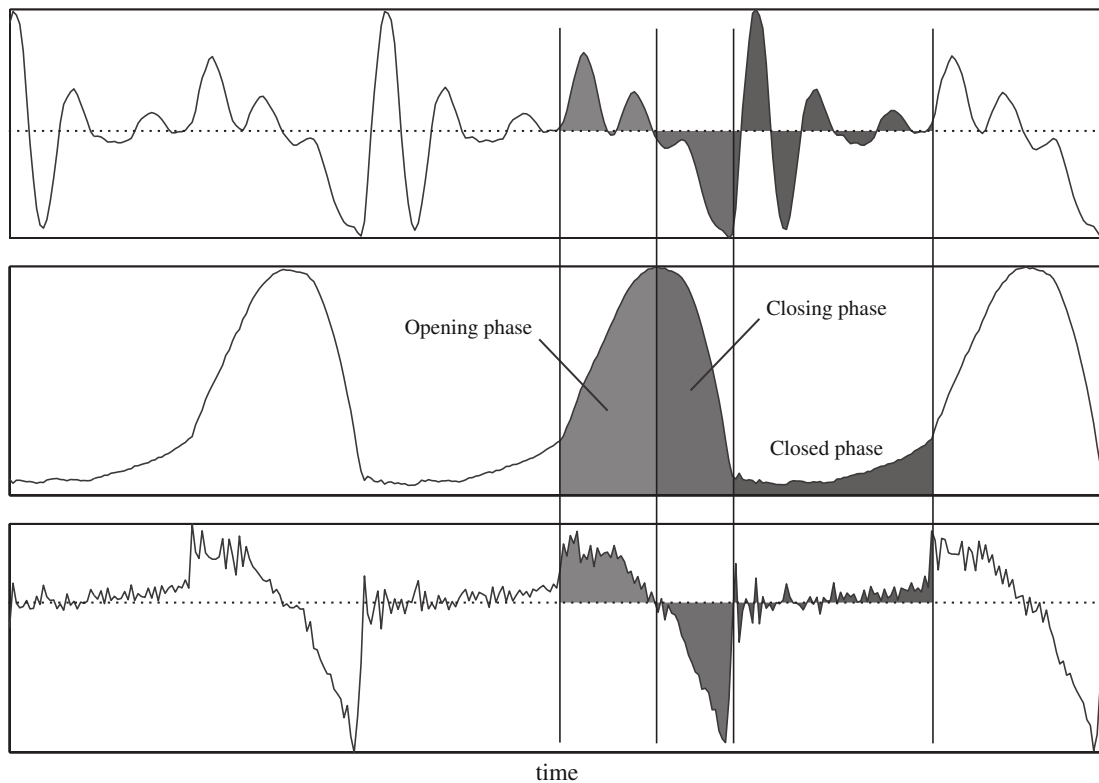


Figure 2. Three periods of a sound-pressure waveform and the respective glottal flow and its derivative. The opening, closing, and closed phases of the glottal flow waveform are highlighted for clarity.

in Figure 2. Furthermore, the amplitude data of the critical time instants may be inspected. The first time-domain parameters, the open and the speed quotient, were introduced by Timcke et al. (25), although they used them to describe the vocal fold opening instead of the glottal flow. The open quotient (OQ) measures the relative portion of the open phase compared to the cycle duration. The speed quotient (SQ) measures the ratio of the duration of the opening phase to the duration of the closing phase.

Since the main excitation of the vocal tract takes place during the time the vocal folds are closing (see Figure 2), parameters focusing on the closing phase of the glottal pulse are most essential in quantifying the function of the voice source. One of the most widely used time-domain parameters, the closing quotient (CIQ), measures the ratio of the duration of the closing phase to the period length. CIQ was apparently first introduced by Mosen and Engbretson (26).

When a mask is used to estimate the air-flow, it is possible to measure absolute air-flow values of the voice source using amplitude-based parameters. These parameters include the peak flow, the minimum flow, and the peak-to-peak flow, as well as the amplitude of the negative peak of the first derivative (e.g. 27,28). Amplitude-based parameters also exist,

which essentially compute features related to the temporal structure of the glottal flow, such as the amplitude quotient (AQ), which is the ratio of the flow peak-to-peak amplitude and the minimum peak of the pulse derivative, and the normalized amplitude quotient (NAQ), in which AQ is normalized by dividing it by the period length (29,30). In contrast to the critical time instants, the exact location of which may often be subject to interpretation, amplitude levels are straightforward to measure both from the glottal flow and its derivative, since no subjective judgement is required to determine the maximum or minimum amplitude instants. This makes the AQ and the NAQ more robust than their time-based counterpart, CIQ. Furthermore, they are independent of the signal scaling, so they can be used with microphone as well as flow mask recordings. The otherwise problematic extraction of the critical time instants can also be avoided by using the so-called quasi-quotients, such as the direct amplitude-domain counterpart of the OQ parameter, the quasi-open quotient (QQQ), in which the open duration is defined as the time during which the flow is above a set level, usually 50% above the minimum flow (e.g. 31). Once again, while the opening may be gradual and difficult to define precisely, the instant at which the signal is half-way between the minimum

and the maximum can be determined without any uncertainty.

Glottal flow parameterization is achieved in the frequency-domain by taking measurements from the power spectrum of the flow signal, as shown in Figure 3. Probably the most straightforward frequency-domain parameter is H1–H2, which is just the difference of the first and second harmonics in decibels (e.g. 32). A somewhat similar measure is the harmonic richness factor (HRF), which is the ratio between the sum of the magnitudes of the harmonics above the fundamental frequency and the magnitude of the fundamental in decibels (33):

$$\text{HRF} = \frac{\sum_{k \geq 2} H_k}{H_1}, \quad (1)$$

where H_k represents the magnitude of the k -th harmonic. An often-overlooked property of these two parameters is that the density of the harmonic series affects them. Thus, their values co-vary with the fundamental frequency. When the fundamental frequency increases, the distance between the harmonics grows and, therefore, the value of H1–H2 increases, and the value of HRF decreases.

Howell and Williams computed harmonic drop-off rates as the slope of the regression line drawn through the first eight harmonics (34). Alku et al. (35) introduced a somewhat similar measure, the parabolic spectral parameter (PSP), which fits a second-order polynomial to the flow spectrum on a logarithmic scale computed over a single glottal cycle. Due to the regression analysis approach, these two parameters are less affected by changes in the fundamental frequency.

The model-based parameterization methods take some mathematical formula that yields artificial waveforms similar to glottal flow pulses and then adjust the model parameters to fit the waveform shape to the measured flow. The waveforms acquired by using these models are easy to modify by changing the model parameters. On the other hand, model-based methods by definition ignore

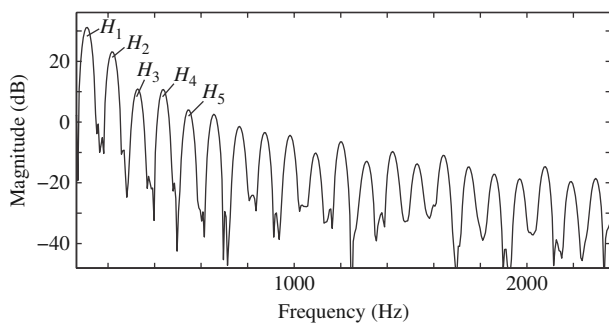


Figure 3. Illustration of a flow spectrum. The levels of the first five harmonics are depicted as H1 to H5.

glottal flow features not included in the model. By far the most used mathematical glottal flow model is the Liljencrants-Fant (LF) model (36). The LF-model is a four-parameter ad-hoc mathematical formulation of the glottal flow pulse derivative. It has been widely used in both voice source analysis and speech synthesis (e.g. 37–41). Parameters also exist that are derived using the waveform assumptions of the LF-model, but which can be computed using amplitude measures of the glottal flow. These parameters include R_{db} , which is equal to NAQ except for an arbitrary scaling coefficient (42), and OQ_a , which approximates to OQ for an ideal LF-pulse (11).

In contrast to the abundance of both inverse filtering techniques as well as parameterization methods developed in the past decades, few publicly available packages for voice source analysis and parameterization currently exist. One reason behind this could be that most research groups have implemented their own tools. One such tool, DeCap, is a manual inverse filtering software developed by Svante Granqvist (43). Paul Milenkovich maintains TF32, which is a speech signal analysis software including linear predictive (LP) inverse filtering (13). However, both of these are proprietary software with limited provisions for user modification. Lee and Childers (44) have also implemented manual inverse filtering as well as many other voice analysis algorithms in the MATLAB™ environment. The software has been published as a supplement to a book (45). Mike Brookes has released VOICE-BOX, which is a speech-processing tool-box for MATLAB (46). It includes inverse filtering routines, but has no graphical user interface for them. ‘Inverse Filter’ and ‘Sky’ are open-source voice inverse filtering and analysis software developed by Kreiman et al. (47). They support interactive and automatic inverse filtering, respectively, and implement some time-based parameters. The inverse filtering method used is the one proposed by Javkin et al. (48). The software is available for the Windows platform only. Praat (49) is a commonly used speech analysis software package suited for generic speech analysis. However, it does not currently have facilities for sophisticated inverse filtering. It has been argued that, e.g., the iterative adaptive inverse filtering (IAIF) algorithm could be implemented in Praat, but due to lack of several low-level algorithms, such as discrete all-pole modelling (DAP), much of the programming would have to be done in C, and the required amount of work would probably be impractically high.

During the inverse filtering-related research activities at the Helsinki University of Technology (TKK) Laboratory of Acoustics and Audio Signal Processing,

the iterative adaptive inverse filtering (IAIF) algorithm (14) together with some voice source parameters, such as the NAQ (30), were implemented in MATLAB. To facilitate inverse filtering of large sets of data, a graphical user interface was soon developed. This software evolved into the TKK Voice Source Analysis and Parameterization Toolkit (Aparat) described in this article.

While most of the individual algorithms in TKK Aparat have been previously published, no other software incorporates such a comprehensive set of inverse filtering parameters and inverse filtering analysis tools in a package immediately usable by voice research professionals and easily applicable in other software.

The author expects TKK Aparat to be useful not only in traditional speech research studies, but also in applied disciplines such as the analysis of voice fatigue in the study of occupational voice. Similar issues have been predicted to become increasingly common in the near future (50). The amount of data to be processed in these new fields can be expected to be much larger than in traditional voice source analysis, but tools for efficient analysis of the glottal flow have been lacking or too complex. TKK Aparat attempts to provide both an accessible user interface and parameterization methods useful in such lines of work.

It was decided to release the software to be freely available under an open-source licence for two reasons. First, the author wishes to encourage participation in further development of voice research software, and TKK Aparat in particular. Second, the software offers implementations of different voice source parameterization algorithms in a single environment, acting as a reference and basis for further algorithm development.

Due to its potential importance in multiple speech-related disciplines, this paper describes TKK Aparat in depth. First, the inverse filtering methods are described and evaluated. Next, the parameterization algorithms are discussed, the validity of the inverse filtering and parameterization algorithms is evaluated, and the user interface of TKK Aparat is described. Finally, conclusions are given.

Inverse filtering

The theoretical background of the inverse filtering in TKK Aparat is set in Fant's source-filter theory, according to which the production of speech can be divided to three separate processes, as shown in Figure 1: the glottal excitation, the vocal tract filtering, and the lip radiation effect (1). The glottal excitation is a pulse train with general low-pass characteristics. The glottal excitation, having a quasi-periodic structure, possesses a spectrum ex-

hibiting a harmonic structure. The amplitudes of the harmonic components are traditionally considered to decrease monotonically at a rate of 12 dB/octave (1). The second effect of the source-filter theory, the vocal tract, can be approximated, at frequencies under 5 kHz, as an acoustic tube with a variable cross-sectional area (51). The actual configuration of the vocal tract results in varying locations of the vocal tract resonances, or the formants. As a rule of thumb, there is one formant for every kilohertz band in the vocal tract transfer function (52). The last process in the source-filter theory is the lip radiation effect, which corresponds to the effect of the coupling of the vocal tract and the effectively infinite surrounding air volume. At low frequencies, the lip radiation effect acts as a differentiator of the signal, contributing a positive slope of approximately +6 dB/octave.

Two inverse filtering methods are implemented in TKK Aparat, both of which utilize the assumptions made in the source-filter theory. The first method is IAIF, the block diagram of which is shown in Figure 4. Although the implementation of IAIF is essentially the same as originally published (14), the linear predictive (LP) modelling of the vocal tract has been replaced with discrete all-pole modelling (DAP), which is based on the minimization of a discrete version of the Itakura-Saito distance between the all-pole spectral envelope sampled at discrete frequencies and spectral amplitudes derived from the short-time Fourier transform (STFT) spectrum (53). In voiced speech, the discrete frequencies are the harmonics of the fundamental frequency. DAP is found to be somewhat more insensitive to the biasing of the formants caused by nearby harmonic peaks.

After the pre-filtering in block 1, the vocal tract estimate is acquired in two phases in blocks 2–10. In steps 11 and 12, the effects of the vocal tract and the lip radiation are removed from the input signal to get the glottal flow estimate. The blocks are described in detail below.

First, in block 1, the original signal $s_0(n)$ is high-pass filtered to eliminate any low-frequency fluctuations captured by the microphone. The filter should be a linear phase finite impulse response (FIR) filter with a cut-off frequency well below the fundamental frequency of the signal (e.g. 60 Hz). This results in signal $s(n)$.

The voice source creates a signal having a declining spectrum slope of approximately -12 dB/octave, while the lip radiation incurs a $+6$ dB/octave high-pass effect on the spectrum. Thus, in block 2, a first-order DAP model $H_{g1}(z)$ is computed for the signal yielding a first-order filter, which forms an estimate for the combined effect of -6 dB/octave of the glottal flow and the lip radiation on the speech

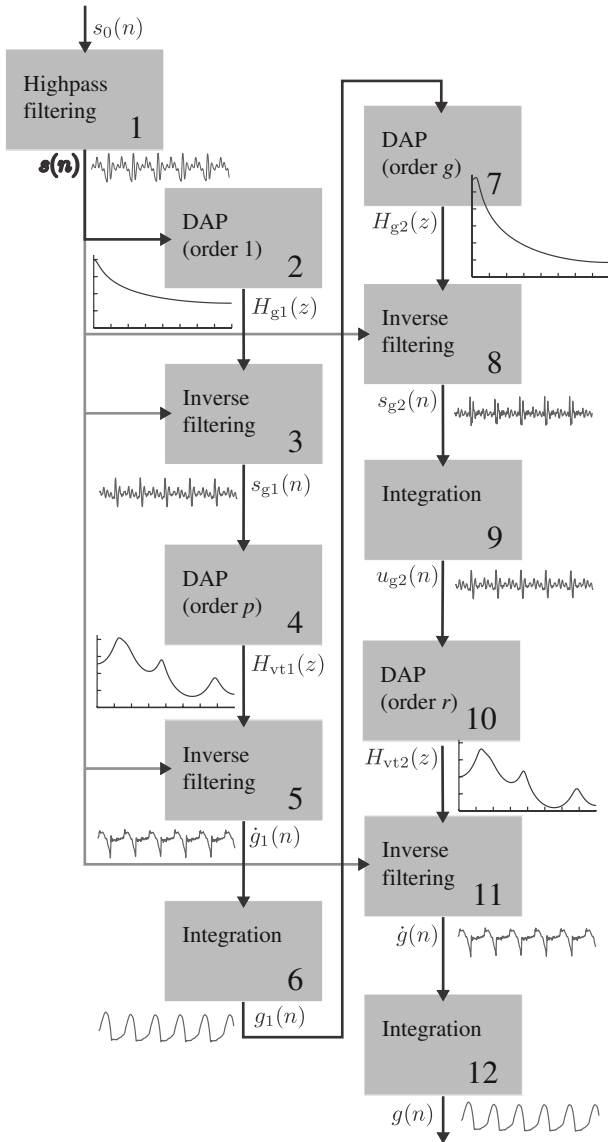


Figure 4. Structure of the iterative adaptive inverse filtering (IAIF) algorithm. Refer to the text for an explanation of different blocks.

spectrum. In block 3, $H_{g1}(z)$ is cancelled from $s(n)$ by inverse filtering, resulting in a pressure signal $s_{g1}(n)$ that only contains the effects of the vocal tract and an impulse-train excitation. In block 4, a p th-order (usually about two times the sampling frequency in kHz) DAP model $s_{vt1}(z)$ is computed. This model is the first estimate for the vocal tract filter, and the effect of the vocal tract is cancelled from the signal $s(n)$ by inverse filtering in block 5. Then, the lip radiation effect is cancelled in block 6 by integrating $\hat{g}_1(n)$, the output of block 5. This concludes the first phase of the vocal tract estimation and also yields a first estimate of the glottal flow, $g_1(n)$.

In an analogous manner to the first phase, blocks 7–10 form a second estimate of the vocal tract filter.

A new estimate of the contribution of the glottal flow to the speech spectrum, $H_{g2}(z)$, is computed. This is done in block 7 by computing a DAP model of order g (usually 2 or 4) from $g_1(n)$. Next, in block 8, the estimated glottal contribution is cancelled by inverse filtering the signal $s(n)$ with $H_{g2}(z)$. Then, the lip radiation effect is removed in block 9 by integrating the signal. Finally, in block 10, a new DAP analysis of order r (usually $r=p$) is computed to acquire a refined model of the vocal tract filter, H_{vt2} . In blocks 11 and 12, a refined estimate of the glottal flow $g(n)$ is computed by removing the effect of the vocal tract by inverse filtering $s(n)$ with H_{vt2} and then integrating the resulting signal to remove the lip radiation effect.

The simpler of the inverse filtering methods implemented in TKK Aparat is a traditional autoregressive modelling-based inverse filtering, dubbed here as *direct inverse filtering* (DIF). DIF consists basically of IAIF blocks 1–6 with a slight reordering of the blocks for improved computational efficiency. The block diagram of DIF is shown in Figure 5. In block 1, the original signal is high-pass filtered in a manner similar to IAIF. In block 2, the pressure signal $s(n)$ is integrated to cancel the lip radiation effect, yielding the flow signal $u(n)$. In a manner similar to block 2 of IAIF, block 3 of DIF is used to compute a first-order DAP model of the signal ($H_g(z)$), which corresponds to the effect of the glottal flow on the spectrum. In block 4, similarly to block 3 of IAIF, the first-order glottal slope filter $H_g(z)$ is cancelled from the signal $s(n)$ by inverse filtering, resulting in signal $s_g(n)$. This signal represents the vocal tract filter excited with an impulse train. In block 5, analogously to the IAIF block 4, a p th-order DAP model is computed to estimate the vocal tract filter. This filter, $H_{vt}(z)$, is then used in block 6 to inverse filter the flow signal $u(n)$ to acquire the final estimate of the glottal flow, $g(n)$.

Parameterization

The glottal flow parameterization process in TKK Aparat is completely automatic, i.e. once the glottal flow is acquired no manual labour is required for parameter point-setting and parameter computation. The details of time and frequency-domain parameters, as well as model-based parameters supported by TKK Aparat, are given below.

Time-domain parameters

Time-domain parameterization involves extracting certain time and amplitude instants from a glottal flow estimate frame. Once the instants are acquired, several different time and amplitude-based

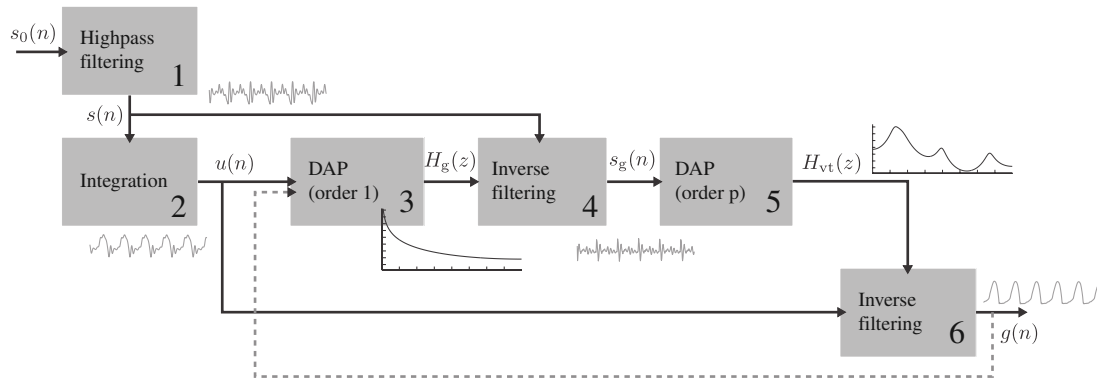


Figure 5. Structure of the direct inverse filtering (DIF) algorithm. Refer to the text for an explanation of the different blocks.

parameters may be computed from these instants. The different critical time instants, such as the glottal opening t_0 and the maximum flow t_{\max} , are illustrated in Figure 6.

Parameterization is performed on a signal frame, the length of which is usually between 20 and 100 ms, and which contains k consecutive glottal flow periods. First, the signal frame is slightly smoothed to reduce the effect of high-frequency noise on the acquired time instants. The smoothing is performed using a four-tap linear-phase low-pass FIR filter.

The fundamental period length T is acquired by first calculating the f_0 of the signal frame and inverting it. Then, the time instant of the maximum sample value t_{\max}^* of the whole frame (containing multiple glottal flow periods) is retrieved. It is assumed that $t_{\max}^* \in S$, where S is the set of the peak maxima within the frame: $S = \{t_{\max,k}\}$. Therefore, the locations of the other maxima $t_{\max,k}$ can

be acquired by finding the local maxima at time spans of multiples of T before and after t_{\max}^* .

The flow minima $t_{\max,k}$ are sought for after each $t_{\max,k}$, and the peak-to-peak pulse amplitudes $A_{ac,k}$ are calculated as $A_{\max,k} - A_{\min,k}$.

The rest of the time instants are acquired relative to the local period maximum. The fundamental period frames around each of the $t_{\max,k}$ are differentiated. The derivative maximum of the frame $t_{d\max}$ is then sought to the left of t_{\max} and the minimum $t_{d\min}$ to the right. The respective amplitude values, $A_{d\max}$ and $A_{d\min}$, are saved as well.

The closure time instant t_c is estimated by finding the first positive zero-crossing of the flow derivative after $t_{d\min}$. However, due to the more gradual opening of the glottal pulse, the determination of the opening instant is more ambivalent, and two opening instants, the primary and the secondary opening (t_{o1} and t_{o2} , respectively), are estimated (54).

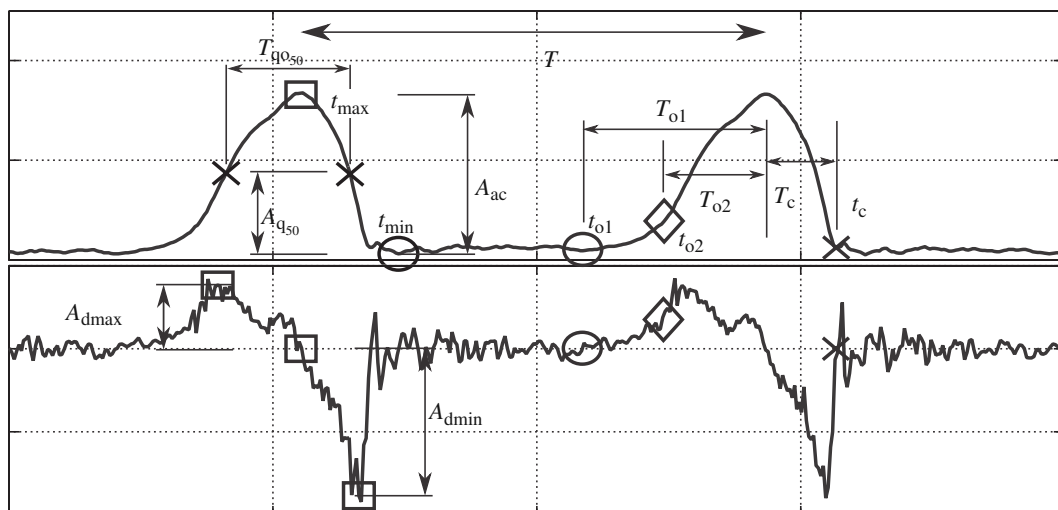


Figure 6. Time and amplitude instants used in calculating the time-domain glottal flow parameters. The upper pane represents the glottal flow estimate and the lower pane the respective derivative. See the text for a detailed description of the different time and amplitude instants and spans.

To detect the primary opening instant, a threshold $A_{o,10\%}$ is defined as 10% (relative to $A_{ac,k}$) above the amplitude of $t_{\min,k}$. The corresponding time instant is acquired, and the frame is then scanned backwards as long as the derivative is positive, or the preceding 5% of the glottal period contains a flow value that is lower than 1% of the flow range below the current scanning position. The latter condition attempts to ensure that the algorithm does not get stuck at a local minimum.

The secondary opening instant is located at the largest local maximum of the smoothed second derivative of the flow in a time window starting at 5% of the glottal cycle duration after $t_{o1,k}$ and extending up to $t_{\max,k}$.

The 50% quasi-opening and -closing time instants t_{qo} and t_{qc} are defined as the points where the amplitude of the curve crosses the 50% of the peak-to-peak amplitude level.

After determining the time and amplitude instants, it is straightforward to compute a variety of parameters. The computed parameters are defined as follows:

$$OQ_1 = \frac{t_c - t_{o1}}{T} \quad (2)$$

$$OQ_2 = \frac{t_c - t_{o2}}{T} \quad (3)$$

$$OQ_a = A_{ac} \left(\frac{\pi}{2A_{d\max}} + \frac{1}{A_{d\min}} \right) f_0 \quad (4)$$

$$QOQ = \frac{t_{qc} - t_{qo}}{T} \quad (5)$$

$$SQ_1 = \frac{t_{\max} - t_{o1}}{t_c - t_{\max}} \quad (6)$$

$$SQ_2 = \frac{t_{\max} - t_{o2}}{t_c - t_{\max}} \quad (7)$$

$$CIQ = \frac{t_c - t_{\max}}{T} \quad (8)$$

$$AQ = \frac{A_{ac}}{A_{d\min}} \quad (9)$$

$$NAQ = \frac{AQ}{T} \quad (10)$$

Frequency-domain parameters

While time-domain parameterization methods are straightforward to apply, even slight non-linearities in the phase response of the recording equipment may adversely affect the quality of the glottal flow estimate. In such cases it might be beneficial to

inspect the frequency-domain properties of the flow estimate. In particular, frequency-domain parameterization of the glottal flow is justified by the fact that the functioning of the voice source in various speech communication situations is reflected by changes in the spectral decay: the breathier the phonation type, the larger the roll-off of the voice source spectrum (e.g. 33,38,55). Multiple frequency-domain voice source parameters exist, which all essentially measure the slope of the spectrum.

The harmonic level difference (H1–H2) is computed simply by acquiring the fundamental and the second harmonic of the amplitude spectrum of the glottal flow waveform in dB and calculating their difference. The harmonic richness factor (HRF) is defined as the ratio of the sum of the amplitude of the higher harmonics and the first harmonic, given in Equation 1. If the higher harmonics are acquired at exact multiples of f_0 ($H_k = kf_0$), even slight inharmonicities and inaccuracies in determining f_0 may completely foil the process. Therefore, in TKK Aparat the harmonics are defined as the local maxima in the frequency regions $kf_0 \pm f_0/2$.

The parabolic spectral parameter (PSP) is based on fitting a parabolic function to the low-frequency part of a pitch-synchronously computed logarithmic spectrum of the glottal flow. The implementation in TKK Aparat closely follows that of the original paper (35).

Model-based parameters

Automatic estimation of LF-parameters has been implemented in TKK Aparat. The fitting method is a modified version of the algorithm proposed by Strik and Boves (56). In the algorithm, initial estimates for the LF-model parameters are sought from the derivative of the glottal flow waveform. These initial estimates are then given to the curve-fitting optimization algorithm, which attempts to make the synthetic LF-model coincide with the actual flow derivative. While Strik and Boves used a two-stage optimization algorithm, the Aparat implementation performs the optimization in a single stage using a subspace trust region least-squares non-linear optimization algorithm as implemented by the MATLAB function `lsqnonlin`. The initial time and amplitude point estimates are given by the time-based parameterization process. The LF-model estimates are computed independently for each period in the flow waveform given by inverse filtering.

Algorithm evaluation

To gain some insight on the quality of the glottal flow estimates acquired by the DIF algorithm, the per-

performances of both IAIF and DIF have been evaluated. Synthetic vowels have been generated using LF-modelled glottal flow waveforms together with artificial vocal tract transfer functions modelled after vowels /a/, /e/, /i/, and /æ/. The four LF-model parameters were set as follows: $T_p=0.45$, $T_e=0.6$, $T_a=0.005$, $T_c=0.65$, and $E_e=1$. The vocal tract transfer functions were generated using the parameters published by Gold and Rabiner (57). The vowels were synthesized using fundamental frequencies of 100 and 200 Hz, representing male and female voices, respectively.

The synthesized vowels were then inverse filtered using TKK Aparat with both IAIF and DIF algorithms. Several glottal flow parameters (OQ_1 , OQ_2 , NAQ, AQ, CIQ, QOQ, SQ_1 , and SQ_2) were acquired from both the synthetic and the inverse filtered glottal flow pulses. Then, the relative differences in the parameter values were compared to assess the magnitude of the changes induced by the two inverse filtering methods. Furthermore, the sum of squared differences of the actual sample values of the glottal flow pulses were also computed.

The results of the difference analyses are given in Tables I and II. In the case of the female /i/ vowel, neither inverse filtering method was able to precisely place the first formant, located close to f_0 . This is indicated by the sum of squares (SSQ) values, which are considerably higher for the female /i/ than for other vowels. Figure 7 illustrates the relative changes in the parameter values between the inverse filtered and original synthetic glottal flow pulses. By comparing the IAIF and DIF charts as well as the SSQ columns of Tables I and II, it becomes obvious that IAIF is able to represent the original waveform better. However, at least for parameters such as NAQ and AQ, the differences are modest, indicating that the slight decrease in inverse filtering quality in DIF compared to IAIF would be acceptable, if computational and implementation simplicity is considered important. However, when quality and

reliability are the prime factors, IAIF should still be preferred over DIF.

In addition to the evaluation described above, the IAIF algorithm and some of the parameters have been evaluated in earlier studies as well. The validity of the inverse filtering procedure as well as the calculation of NAQ and CIQ parameters in TKK Aparat have been tested by Lehto et al. (58). In their paper, manual inverse filtering using DeCap was compared to the IAIF method. The results were parameterized using CIQ and NAQ and compared statistically. Even though not explicitly mentioned in their article, two different implementations of the IAIF algorithm were used, one of which was TKK Aparat. Although statistically significant differences were found between the inverse filtering methods, the different inverse filtering methods exhibited a strong correlation between the different methods. It remained unclear whether the statistical differences were caused by the methodological differences or by variations in experimenter preferences. However, Lehto et al. (58) concluded that the discrepancies caused by the use of different inverse filtering methods are, in general, reasonably small. In another study, Alku et al. (5) have examined the IAIF method using simulated vowels created by a physical model of sound production. The synthetic and inverse filtered glottal pulses were parameterized using NAQ and compared. In their study, the waveforms and NAQ values were found to be close to each other, further justifying the use of the methods.

Due to the lack of reference implementations, no quantitative validity checking of parameters other than NAQ or CIQ has been performed. However, it has been verified that their values fall within the range given by publications discussing them.

Aparat user interface

The user interface of TKK Aparat has been designed to allow for rapid processing of large amounts of

Table I. Relative difference of parameters acquired from iterative adaptive inverse filtering (IAIF) inverse filtered synthetic vowels and their respective original Liljencrants-Fant (LF) model glottal flow pulses. The last column represents the sums of squared differences of the original and inverse filtered glottal flow pulses.

Vowel	f_0 (Hz)	OQ_1	OQ_2	NAQ	AQ	CIQ	QOQ	SQ_1	SQ_2	SSQ
a	100	0.01	0.01	0.00	0.00	-0.10	0.02	0.19	0.21	41
e	100	-0.02	-0.05	0.02	0.02	-0.04	0.02	0.03	-0.02	38
i	100	0.08	-0.05	0.01	0.01	-0.04	0.02	0.17	-0.03	40
æ	100	0.01	0.09	0.03	0.03	0.07	0.02	-0.07	0.05	42
a	200	0.12	0.18	0.04	0.04	0.34	0.02	-0.24	-0.21	29
e	200	0.01	0.02	0.04	0.04	0.12	0.00	-0.14	-0.16	18
i	200	0.43	-0.11	0.05	0.05	0.08	0.26	0.48	-0.30	259
æ	200	-0.07	-0.03	0.00	0.00	-0.07	0.01	0.00	0.08	12
Abs. mean		0.09	0.07	0.02	0.02	0.11	0.05	0.17	0.13	60

Table II. Relative difference of parameters acquired from direct inverse filtering (DIF) inverse filtered synthetic vowels and their respective original Liljencrants-Fant (LF) model glottal flow pulses. The last column represents the sums of squared differences of the original and inverse filtered glottal flow pulses.

Vowel	f_0 (Hz)	OQ ₁	OQ ₂	NAQ	AQ	CIQ	QOQ	SQ ₁	SQ ₂	SSQ
ɑ	100	-0.09	0.02	-0.01	-0.01	-0.19	0.00	0.17	0.41	43
e	100	-0.03	0.00	0.00	0.00	-0.09	0.01	0.09	0.16	51
i	100	-0.10	0.06	0.02	0.02	-0.09	0.03	-0.01	0.26	58
æ	100	0.04	0.03	0.01	0.01	0.11	0.00	-0.09	-0.13	50
ɑ	200	-0.27	-0.11	0.13	0.13	-0.18	0.02	-0.17	0.14	47
e	200	0.15	-0.13	0.05	0.05	-0.13	0.00	0.48	0.00	77
i	200	0.27	0.56	0.02	0.02	-0.08	0.58	0.56	1.24	532
æ	200	0.04	-0.14	-0.03	-0.03	-0.17	0.04	0.36	0.05	92
Abs. mean		0.12	0.13	0.04	0.04	0.13	0.09	0.24	0.30	119

pre-segmented voice files. The typical workflow in such use is described below. Arabic numerals in the text refer to the workflow items in Figure 8. Roman numerals refer to other items.

First, the wave file listing of the current working directory (1.) is shown in the main window. The file listing is visible at all times to facilitate rapid selection of working items. When an item is selected in the file listing, the file is loaded, and the waveform is displayed in the signal pane (2.) of the signal window. A 50 ms window in the middle of the signal is

automatically selected and inverse filtered using the parameter settings in the main window (4., 5., 6.).

The selection may be moved by clicking in the signal pane. Dragging the signal pane creates a new selection. Alternatively, the selection size may be adjusted by entering the desired length either in milliseconds or samples in the selection details (3.) of the signal view.

Depending on the inverse filtering method selected, the inverse filtering parameters (4., 5., 6.) may need to be adjusted to acquire an optimal glottal

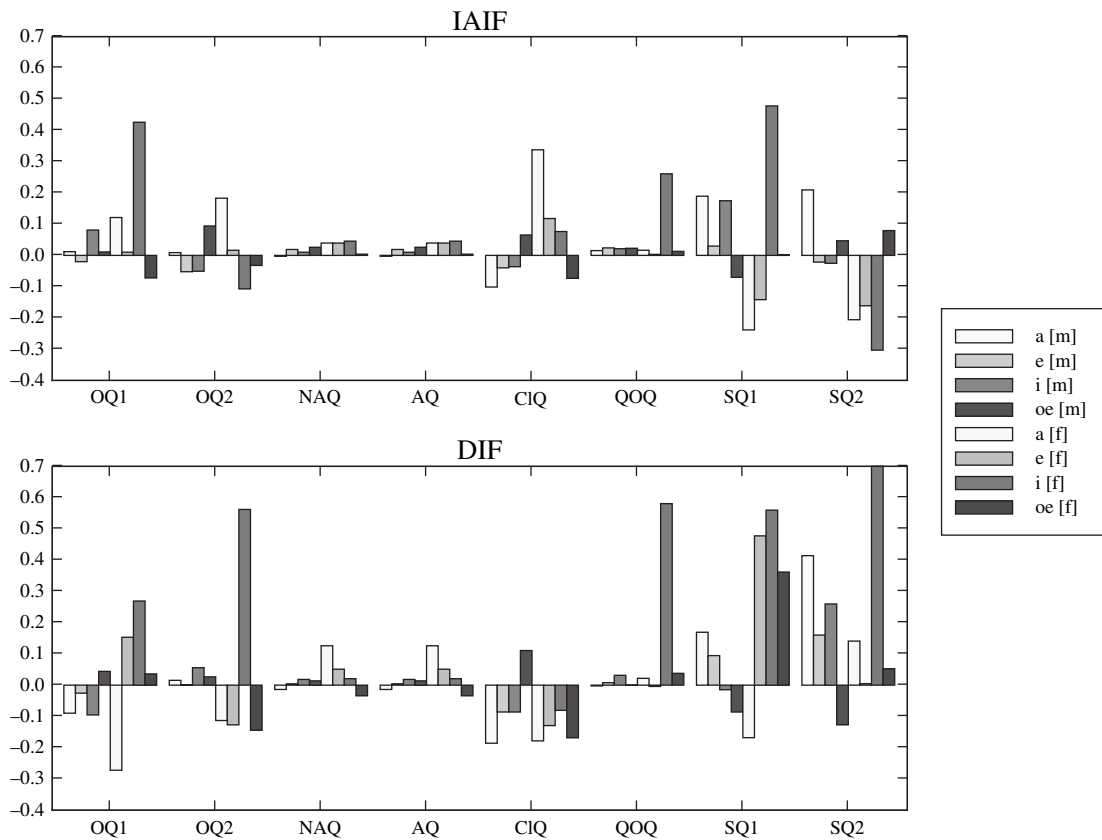


Figure 7. A bar chart of relative differences of parameters acquired from inverse filtered synthetic vowels and their respective original Liljencrants-Fant (LF) model glottal flow pulses.

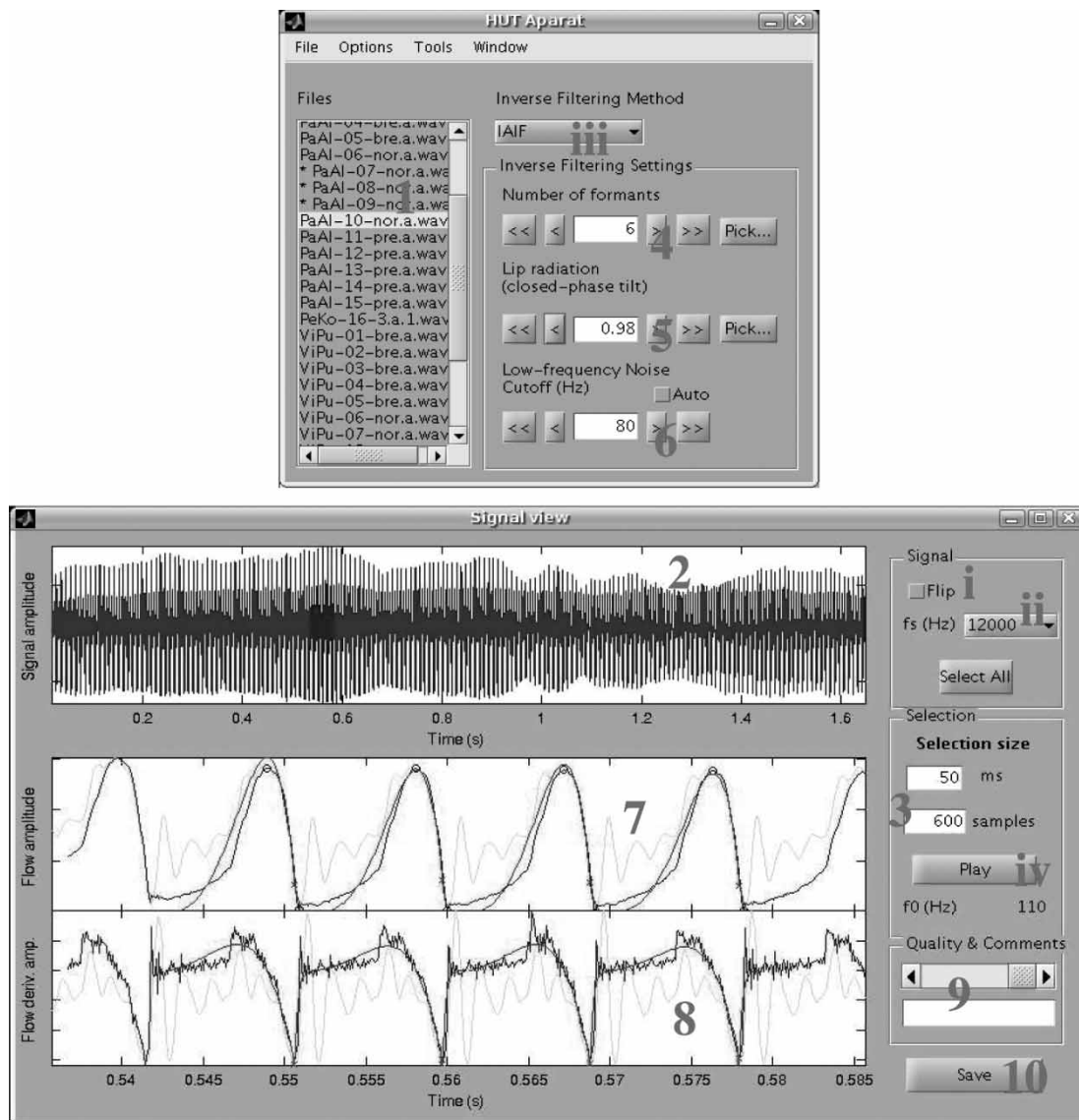


Figure 8. Main dialog and the signal view of Aparat. Refer to the text for the meanings of the bold numeric labels.

flow estimate. For IAIF and DIF, the *number of formants* (4.) is recommended to be set to about $1/\text{kHz}$ of the signal bandwidth, i.e. for a signal with a sampling frequency of 12 kHz the bandwidth is 6 kHz, and the initial guess of the number of formants is recommended to be set to 6. The *lip radiation* (5.) corresponds to the coefficient of the integrator that is the digital filter used to cancel the effect of lip radiation. The value affects the integration coefficient of the flow waveform and is best found experimentally, with values ranging from 0.98 to 1.0. Both the number of formants and lip radiation values may also be selected visually by clicking the Pick button. Then a new window appears showing multiple glottal flow waveforms inverse filtered using different values of the respective parameter. The desired parameter value can be selected by clicking the appropriate waveform.

After finishing the inverse filtering parameter tuning, a subjective quality evaluation and text comments may be set in the *Meta data* (5.) panel of the signal view. When the data are saved, these meta-data are stored among other variables.

Data visualization

Time-domain views of the original signal as well as the glottal flow estimate and its derivative are shown automatically in the signal view. The two lower panes also show in a light grey colour the respective flow and its derivative without inverse filtering, i.e. the original signal only integrated with the lip radiation coefficient and a time-derivative thereof.

TKK Aparat is also able to plot the power spectra of the different signals, shown in Figure 9.

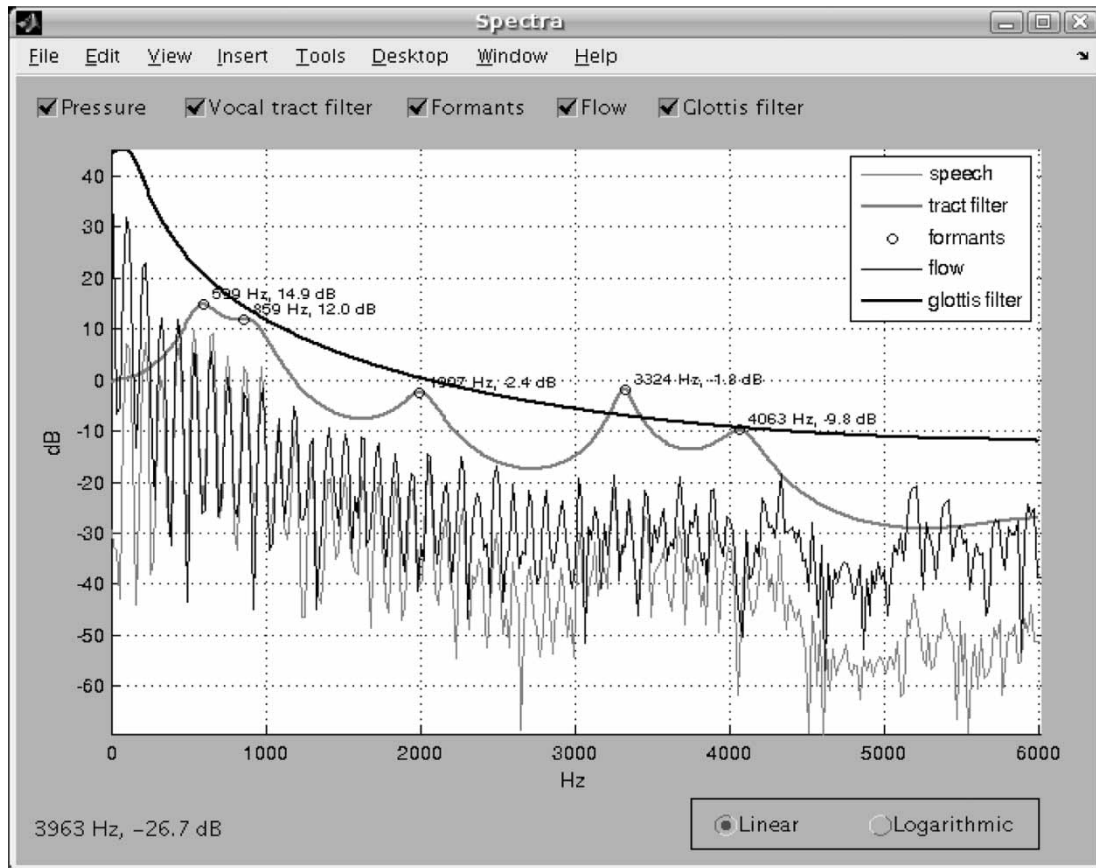


Figure 9. View showing the spectra of a speech signal, the calculated glottal flow, and the used vocal tract filter.

Other data visualization methods in TKK Aparat include a z -plane view of the tract filter, a vocal tract view, and a phase-plane view. The z -plane is a pole-zero plot of the vocal tract filter estimate, giving insight into the performance of the inverse filtering algorithm. The vocal tract view shows a plot of the cross-sectional diameters of the tube model derived from the vocal tract filter (59). The phase-plane view shows an xy -plot with the glottal flow samples on the x -axis and the corresponding samples of the flow derivative as the y -axis. It may be used to assess the quality of the inverse filtering (60).

Inverse filtering quality evaluation

To facilitate modification of inverse filtering parameters, TKK Aparat provides many methods for inspecting the quality of the inverse filtering performance. The most obvious is the time-domain display of the flow and the flow derivative in the signal window. According to the evidence in the literature, the glottal flow pulses have an abrupt closure with a maximally flat closed phase, although the exact shape of the flow pulses will depend on the operator. Furthermore, there should be no residue of any formant resonances, i.e. the flow pulse should contain

only a single peak, and the derivative should preferably have a distinct negative peak from which the signal gradually approaches the x -axis with little ringing. To help with the comparison of the original signal, the respective signals without the formant inverse filtering are shown on the signal windows in light grey.

The spectrum window is able to give more insight on the success of the inverse filtering process. The spectrum of the glottal flow (the higher thin curve in Figure 9) is known to have a monotonically decreasing peak magnitude with all formant resonances removed. The logarithmic view of the spectra might help here, since removal of the lowest resonances is of greatest importance in order to get an accurate estimate of the true glottal flow.

Finally, the phase-plane view together with related metrics may be used to assess the inverse filtering quality either in supervised or completely automatic inverse filtering (60). Sub-cycles in the phase-plane plot indicate the presence of residual formant ripple in the glottal flow estimate.

Parameterization

Whenever a signal is inverse filtered in TKK Aparat, the time- and amplitude-based parameters

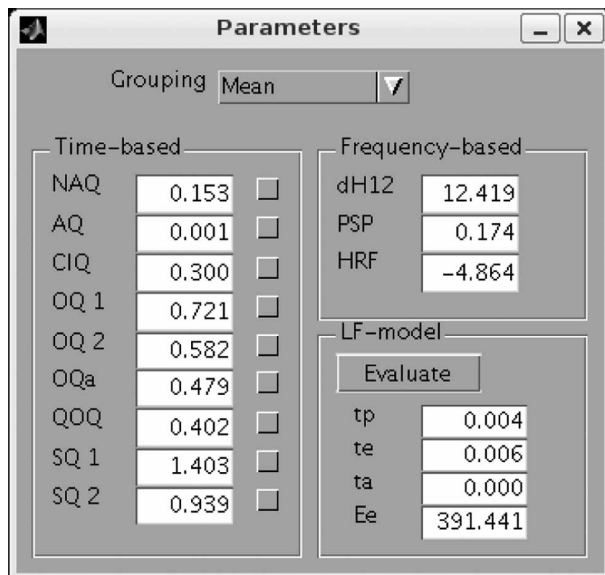


Figure 10. View showing the parameters computed from a glottal flow estimate.

are automatically computed. The data are shown in the parameter window, shown in Figure 10. At the top of the parameter window is a selection box in which the grouping function may be selected with which the parameters are summarized.

The time and amplitude points of the time-based parameters may be shown in the signal window by marking the check box next to the respective parameter.

While all other parameters are calculated automatically after inverse filtering, the LF-model fitting must be explicitly performed due to its computational complexity. The resulting waveforms are also shown in the signal window.

The parameterization results are, among other data, exported to the MATLAB work-space as variable `aprt`, which allows for easy interactive experimenting with the data. Furthermore, when the file is saved, the parameter data are stored in a `mat` file among other data. The saved files are stored in the regular binary MATLAB file format. The files include the full data of the model, including the original signal segment, inverse filtered signal segment, computed parameters, and inverse filtering meta data.

A common task in parameterization of multiple files is to combine, or to aggregate, the parameter values of multiple vowels for later analysis. This is directly supported in Aparat: all parameter data of different files are combined into a single text file, which may easily be imported to different statistical computation or spreadsheet software for further processing.

Usability testing

In order to validate and improve the usability of TKK Aparat, usability tests were conducted. Five volunteers participated in the testing. While the number may seem small for tests typically conducted in speech science, use of as few as 3–6 participants is a well established practice in usability testing (61). The participants were selected so that they form a representative sample of the potential user group, with test users having speech engineering, phonetics, as well as speech therapy and phoniatrics backgrounds.

In the usability test, the participants performed various tasks comprising five common usage scenarios. The first task was to visually inspect the different user interface elements and describe their assumed purposes. The purpose of the task was to find inconsistencies and unintuitive features in the user interface and the terminology used. The second task was to inverse filter a single /a/ vowel. The intent was to observe how naïve users succeed in basic inverse filtering tasks. The third task was to inspect the spectra and find the locations of the formants of the vowel. The task was designed to assess the accessibility of the menu structure and hidden windows and that of the spectrum view. In the fourth scenario, the users were asked to observe the parameter values and the locations of the opening instants of the pulses. Furthermore, the users were asked to perform an LF-model fitting. The purpose of the task was to find any inconsistencies or problems in the parameterization interfaces. The final scenario was to rapidly inverse filter and parameterize a handful of files with pre-defined sampling rate and window length settings. Efficiency in the use of the user interface was observed, along with the adopted usage habits.

During the performance of the tasks, the actions were registered both as a screen capture and a sound recording. The test users were asked to think aloud to gather as much information of their views as possible. Any discontent or difficulties in completing the tasks were carefully noted and analysed. Before and after the test, a short interview was conducted to guide the participant and assess his or her opinions regarding the software and its use. To acquire unbiased opinions, the interviews were conducted in a neutral and non-leading manner.

An average of 19.4 usability problems or suggestions were observed for each user. The amount and nature of problems appeared to depend mostly on the user's experience, with more experienced users reporting more problems and suggestions but performing the tasks much more fluidly than the

inexperienced ones, who tended to get genuinely stuck in the issues.

The first task (visual inspection) resulted in multiple labelling and terminology clarifications to better match the test users' expectations. The second task (inverse filtering of a single file) resulted in modification of the inverse filtering parameter selection sliders with discrete buttons. Furthermore, many changes in different text box default values were made, and some confusing buttons were removed. Both the spectrum viewing and parameterization tasks resulted only in some minor terminology changes. In response to the results of the final task (inverse filtering of multiple vowels), more explicit user interface feedback was added. These changes are believed to address the most obvious usability issues in TKK Aparat and to ensure that anyone with basic knowledge of the theory of inverse filtering should be able to pick the software up without any major difficulties. Observations during the tests and interviews suggested that the most obvious concerns were successfully alleviated, and user satisfaction improved consistently throughout testing.

Conclusions

In this paper, a freely available voice inverse filtering and parameterization software, TKK Voice Source Analysis and Parameterization Toolkit, or TKK Aparat for short, has been described. The system estimates the glottal volume velocity waveform from an acoustical speech pressure signal. The glottal flow is automatically parameterized using the most common time- and frequency-domain parameters. Furthermore, parameter fitting with the LF-model may be performed. The software is usable for algorithm development, speech science research, as well as for clinical study of voice.

TKK Aparat has already been used in several research projects. Airas and Alku (62) used TKK Aparat to inverse filter and parameterize a large amount of vowels segmented from emotional, continuous speech. In the work by Lehto et al. (58), the inverse filtering results of TKK Aparat and another IAIF implementation have been compared to manual inverse filtering. Pulakka (54) analysed the human voice production process using inverse filtering, electroglottography, and high-speed imaging in his Master's thesis. The inverse filtering portions of his work were performed using TKK Aparat. Cabral and Oliveira (63) have used Aparat to analyse voice segments for emotional speech synthesis. Furthermore, there is an on-going project at TKK Laboratory of Acoustics and Audio Signal Processing and the Finnish Institute of Occupational Health in

which the effects of dust exposure on voice are studied using inverse filtering with TKK Aparat.

TKK Aparat is developed in the MATLAB environment, which may prove, due to the high cost of the software, problematic to some interested users. Fortunately, many research facilities already have site licences for MATLAB, considerably reducing the problem. Furthermore, the MATLAB Compiler allows for the creation of stand-alone packages of MATLAB applications, including TKK Aparat. In this manner, it is possible to fully use TKK Aparat even without access to MATLAB, while only losing the ability to interactively experiment with the signals in the MATLAB environment—something people without prior MATLAB expertise hardly would do in any case.

The functionality of TKK Aparat is also available as MATLAB functions, usable independently of the graphical user interface. This permits the use of the algorithms in other projects as well. For example, it would be straightforward to construct a script which automatically inverse filters and parameterizes audio files using these functions.

Several free mathematical software packages exist, such as Octave, Scilab, and RLaB, which are largely compatible with MATLAB. TKK Aparat, however, is dependent on MATLAB's object-oriented programming model as well as signal processing and graphical user interfaces, which generally are not implemented in the free software packages. Unfortunately, this precludes the use of the free mathematical software packages with TKK Aparat.

As an open-source software, TKK Aparat is available free of charge. The latest version can be accessed at: <http://aparat.sourceforge.net>. Proficient users are able to access the underlying source code and modify the functionality as well as utilize the implemented algorithms directly in other software projects.

TKK Aparat provides significant improvement over existing inverse filtering software by integrating multiple inverse filtering algorithms and a wide range of glottal flow parameters in a single refined graphical user interface. As such, TKK Aparat has already proven useful in multiple speech research tasks and shows potential in related areas such as in the study of voice fatigue, in which copious amounts of vowel samples have to be processed in rapid succession.

Acknowledgements

This research was supported by the Academy of Finland (project number 200 859) Kaupallisten ja teknillisten tieteidien tukisäätiö KAUTE, and the Emil Aaltonen Foundation. The Graduate School of Language Technology in Finland.

References

1. Fant G. Acoustic theory of speech production. The Hague, Netherlands: Mouton; 1960.
2. Flanagan JL, Meinhart DIS. Source-system interaction in the vocal tract. *J Acoust Soc Am*. 1964;36:2001–2.
3. Rothenberg M. Acoustic Interaction Between the Glottal Source and the Vocal Tract. In: Kenneth N. Stevens, Minoru Hirano, editors. *Vocal Fold Physiology*. 15–19 January, 1980. Tokyo: University of Tokyo Press; 1980. p. 305–28.
4. Childers DG, Wong C-F. Measuring and modeling vocal source-tract interaction. *IEEE Trans Biomed Eng*. 1994;41:663–71.
5. Alku P, Story B, Airas M. Estimation of the voice source from speech pressure signals: Evaluation of an inverse filtering technique using physical modelling of voice production. *Folia Phoniatr Logop*. 2006;58:102–13.
6. Miller RL. Nature of the vocal cord wave. *J Acoust Soc Am*. 1959;31:667–77.
7. Rothenberg M. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *J Acoust Soc Am*. 1973;53:1632–45.
8. Ananthapadmanabha TV. Acoustic analysis of voice source dynamics. *STL-QPSR*. 1984;1–4.
9. Hertegård S, Gauffin J. Acoustic properties of the Rothenberg mask. *STL-QPSR*. 1992;33:9–18.
10. Rothenberg M. Measurement of airflow in speech. *J Speech Hear Res*. 1977;20:155–76.
11. Gobl C, Chasaide AN. Amplitude-based source parameters for measuring voice quality. In: Christophe d'Alessandro, Klaus R. Scherer, editors. *Proc ISCAVOQUAL'03 Workshop on Voice Quality: Functions, Analysis and Synthesis*, 27–29 August, 2003. Geneva, 2003. p. 151–6.
12. Allen JB, Curtis TH. Automatic extraction of glottal pulses by linear estimation. *J Acoust Soc Am*. 1974;55:396.
13. Milenkovic P. Glottal inverse filtering by joint estimation of an AR system with a linear input model. *IEEE Trans Acoust*. 1986;34:28–42.
14. Alku P. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. *Speech Commun*. 1992;11:109–18.
15. Strube HW. Determination of the instant of glottal closure from the speech wave. *J Acoust Soc Am*. 1974;56:1625–9.
16. Wong DY, Markel JD, Gray Augustine H Jr. Least squares glottal inverse filtering from the acoustic speech waveform. *IEEE Trans Acoust*. 1979;27:350–5.
17. Mataušek MR, Batalov VS. A new approach to the determination of the glottal waveform. *IEEE T Acoust Speech*. 1980;28:616–22.
18. Ananthapadmanabha TV, Fant G. Calculation of true glottal flow and its components. *Speech Commun*. 1982;1:167–84.
19. Plumpe M, Quatieri T, Reynolds D. Modeling of the glottal flow derivative waveform with application to speaker identification. *IEEE Trans Speech Audi Process*. 1999;7:569–86.
20. Arroabarren I, Carlosena A. Glottal spectrum based inverse filtering. In: 8th European Conference on Speech Communication and Technology (EUROSPEECH 2003 - INTERSPEECH 2003), 1–4 September 2003. Geneva, 2003. p. 57–60.
21. Kasuya H, Maekawa K, Kiritani S. Joint estimation of voice source and vocal tract parameters as applied to the study of voice source dynamics. In: 14th International Congress of Phonetic Sciences, August 1–7 1999. San Francisco, USA. 1999. vol 3, p. 2505–12.
22. 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;110:479–88.
23. Akande O, Murphy P. Estimation of the vocal tract transfer function with application to glottal wave analysis. *Speech Commun*. 2005;46:15–36.
24. Walker J, Murphy P. A Review of Glottal Waveform Analysis. In: Yannis Stylianou, Marcos Faúndez-Zanuy, Anna Esposito, editors. *Progress in Nonlinear Speech Processing, WNSP (Workshop on Nonlinear Speech Processing)*, 20–23 September, 2005 LNCS 4391. Berlin, Germany: Springer Verlag; 2007. p. 1–21.
25. Timcke R, von Leden H, Moore P. Laryngeal vibrations: measurements of the glottic wave. I. The normal vibratory cycle. *AMA Arch Otolaryngol*. 1958;68:1–19.
26. Monsen RB, Engebretson AM. Study of variations in the male and female glottal wave. *J Acoust Soc Am*. 1977;62:981–93.
27. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *J Acoust Soc Am*. 1988;84:511–29.
28. Hertegård S, Gauffin J, Karlsson I. Physiological correlates of the inverse filtered flow waveform. *J Voice*. 1992;6:224–34.
29. Alku P, Vilkman E. Amplitude domain quotient of the glottal volume velocity waveform estimated by inverse filtering. *Speech Commun*. 1996;18:131–8.
30. Alku P, Bäckström T, Vilkman E. Normalized amplitude quotient for parametrization of the glottal flow. *J Acoust Soc Am*. 2002;112:701–10.
31. Laukkanen A-M, Vilkman E, Alku P, Oksanen H. Physical variations related to stress and emotional state: a preliminary study. *J Phonetics*. 1996;24:313–35.
32. Titze IR, Sundberg J. Vocal intensity in speakers and singers. *J Acoust Soc Am*. 1992;91:2936–46.
33. Childers DG, Lee CK. Vocal quality factors: Analysis, synthesis, and perception. *J Acoust Soc Am*. 1991;90:2394–410.
34. Howell P, Williams M. The contribution of the excitatory source to the perception of neutral vowels in stuttered speech. *J Acoust Soc Am*. 1988;84:80–9.
35. Alku P, Strik H, Vilkman E. Parabolic spectral parameter—A new method for quantification of the glottal flow. *Speech Commun*. 1997;22:67–79.
36. Fant G, Liljencrants J, Lin Q-G. A four-parameter model of glottal flow. *STL-QPSR*. 1985;1–13.
37. Gobl C. A preliminary study of acoustic voice quality correlates. *STL-QPSR*. 1989;9–21.
38. Fant G. The LF-model revisited. transformations and frequency domain analysis. *STL-QPSR*. 1995;119–56.
39. Childers D, Ahn C. Modeling the glottal volume-velocity waveform for three voice types. *J Acoust Soc Am*. 1995;97:505–19.
40. Fant G. The voice source in connected speech. *Speech Commun*. 1997;22:125–39.
41. Gobl C, Chasaide AN. The role of voice quality in communicating emotion, mood and attitude. *Speech Commun*. 2003;40:189–212.
42. Fant G, Kruckenberg A, Liljencrants J, Båvegård M. Voice source parameters in continuous speech. transformation of LF-parameters. In: Third International Conference on Spoken Language Processing (ICSLP 94), 18–22 September 1994. Yokohama, Japan, 1994. p. 1451–4.
43. Granqvist S, Hertegård S, Larsson Hans, Sundberg J. Simultaneous analysis of vocal fold vibration and transglottal airflow: exploring a new experimental setup. *J Voice*. 2003;17:312–30.
44. Lee M, Childers DG. Manual glottal inverse filtering algorithm. In: IASTED International Conference on Signal and

- Image Processing (SIP'96), 1–4 November, 1996. Orlando, Florida, USA: ACTA Press; 1996. p. 34–7.
45. Childers DG. *Speech Processing and Synthesis Toolboxes*. New York, USA: John Wiley & Sons, Inc; 1999.
 46. Brookes M. *Voicebox: Speech processing toolbox for matlab*. 2005. Available at: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>.
 47. Kreiman J, Gerratt BR, Antoñanzas-Barroso N. Analysis and synthesis of pathological voice quality. (Software manual, May 2007). Available at: <http://www.surgery.medsch.ucla.edu/glottalaffairs/>.
 48. Javkin HR, Antoñanzas-Barroso N, Maddieson I. Digital inverse filtering for linguistic research. *J Speech Hear Res.* 1987;30:122–9.
 49. Boersma P, Weenink D. Praat: doing phonetics by computer (version 4.5.15) (Computer program). Available at: <http://www.praat.org/> (Accessed 22 February 2007).
 50. Vilkmán E. Occupational safety and health aspects of voice and speech professions. *Folia Phoniatr Logop.* 2004;56:220–53.
 51. Atal BS, Hanauer SL. Speech analysis and synthesis by linear prediction of the speech wave. *J Acoust Soc Am.* 1971;50:637–55.
 52. Rabiner LR, Schafer RW. *Digital Processing of Speech Signals*. Englewood Cliffs, NJ, USA: Prentice-Hall. 1978.
 53. El-Jaroudi A, Makhoul J. Discrete all-pole modeling. *IEEE Trans Signal Process.* 1991;39:411–23.
 54. Pulakka H. Analysis of human voice production using inverse filtering, highspeed imaging, and electroglottography. Master's thesis. Helsinki University of Technology, Espoo, Finland, 2005.
 55. Gauffin J, Sundberg J. Spectral correlates of glottal voice source waveform characteristics. *J Speech Hear Res.* 1989;32:556–65.
 56. Strik H, Boves L. Automatic estimation of voice source parameters. In: *Third International Conference on Spoken Language Processing (ICSLP 94)*, 18–22 September, 1994. Yokohama, Japan, 1994. p. 155–8.
 57. Gold B, Rabiner L. Analysis of digital and analog formant synthesizers. *IEEE Transactions on Audio and Electroacoustics.* 1968;16:81–94.
 58. Lehto L, Airas M, Björkner E, Sundberg J, Alku P. Comparison of two inverse filtering methods in parameterization of the glottal closing phase characteristics in different phonation types. *J Voice.* 2007;21:138–50.
 59. Markel JD, Gray AH. *Linear prediction of speech*, 2nd ed. Berlin, Germany: Springer Verlag; 1980.
 60. Bäckström T, Airas M, Lehto L, Alku P. Objective quality measures for glottal inverse filtering of speech pressure signals. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, 18–23 March, 2005. Philadelphia, PA, USA, 2005, p. 897–900.
 61. Sinkkonen I, Kuoppala H, Parkkinen J, Vastamäki R. *Professional Psychology of Usability*. USA: IT Press. 2003.
 62. Airas M, Alku P. Emotions in vowel segments of continuous speech: Analysis of the glottal flow using the normalized amplitude quotient. *Phonetica.* 2006;63:26–46.
 63. Cabral JP, Oliveira LC. Emovoice: a system to generate emotions in speech. In: *Ninth International Conference on Spoken Language Processing (INTERSPEECH 2006 - ICSLP)*, 17–21 September, 2006. Pittsburgh, USA, 2006. p. 1798–801.