

A Toolkit for Voice Inverse Filtering and Parametrisation

Matti Airas, Hannu Pulakka, Tom Bäckström, Paavo Alku

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

Matti.Airas@hut.fi

Abstract

Voice inverse filtering is a process in which the effects of the vocal tract are cancelled from the speech signal to estimate the airflow through the glottis. Although inverse filtering has many applications in both research and clinical examination of voice production, few voice inverse filtering software packages exist. In the present paper, we propose a flexible software package, HUT Voice Source Analysis and Parametrisation Toolkit (Aparat) for use in the MATLAB environment. It implements glottal inverse filtering and several time-based parameters of the voice source in a graphical user interface. The software package is available under an open source licence at <http://www.acoustics.hut.fi/software/aparat/>.

1. Introduction

The importance of the analysis of the voice source is well established in speech science. According to the acoustic theory of speech production, the acoustic source signal produced by the vibrating vocal folds is filtered by the vocal tract to produce the speech output signal [1]. In consequence of this process, the voice source controls various essential features in speech communication. For example, personal characteristics of voice, the voice quality, and much of the emotional content of speech are already embedded in the voice source signal.

Inverse filtering (IF) is a common and useful technique for voice source analysis. The principle of IF is to cancel the effect of the vocal tract from a recorded speech signal to acquire the glottal flow [2, 3]. This may be performed manually by setting the formant frequencies, or automatically by estimating the vocal tract transfer function using suitable DSP algorithms [4, 5]. Inverse filtering is mainly used as a tool in research of the voice source. Furthermore, in the clinical examination of the voice source, glottal inverse filtering conveys information of irregularities caused by vocal nodules or polyps, or voicing changes caused by speaker fatigue [6, 7].

Unfortunately, software for research and application of inverse filtering is rather scarce, as most research groups appear to have implemented their own tools. DeCap is a manual IF software developed by Svante Granqvist [8]. Paul Milenkovich maintains TF32, which is a speech signal analysis software including LPC based inverse filtering [9]. However, these are proprietary software with limited provisions for user modifications. Lee and Childers have also implemented manual inverse filtering in MATLAB environment [10], but their program does not appear to be available.

During the research tasks at Helsinki University of Technology (TKK) Laboratory of Acoustics and Audio Signal Processing, iterative adaptive inverse filtering (IAIF) algorithm and some voice source parametrisation methods such as the normalized amplitude quotient (NAQ) were implemented [5, 11]. A

graphical user interface (GUI) was then developed to ease the inverse filtering of larger amounts of vowel samples. This GUI then evolved to HUT Voice Source Analysis and Parametrisation Toolkit (Aparat) described in this paper. To facilitate inverse filtering research and future algorithm development, it was decided to put this software freely available under an open source licence. This paper describes Aparat from the user's point of view.

2. Inverse filtering

The inverse filtering method used in Aparat is IAIF [12]. The flow diagram of the current version, which is a slightly modified version from the previous ones, is shown in Fig. 1. Most notably, parametric spectral models used in various blocks of the flow diagram are computed with the discrete all-pole modeling (DAP) method [13] instead of the conventional linear predictive analysis. This reduces the bias by the harmonic structure of the speech spectrum in the formant frequency estimates. In block no. 1 of Fig. 1, the speech signal is high-pass filtered using a linear-phase FIR filter to reduce any low frequency fluctuations captured during the recordings. Stages 2–6 form the first glottal flow approximation by making an estimate of the vocal tract transfer function and inverse filtering the signal with that estimate. The first approximation is used as a basis for stages 7–12, which roughly repeat the process of the earlier stages to yield the final glottal flow estimate. A detailed description of the process may be found in [12].

3. Description of the software

Aparat is an inverse filtering framework designed for flexible inverse filtering and parametrisation of large amount of speech samples. Several assisting views are available to visualise the performance of the inverse filtering process.

Aparat has been implemented using MATLAB. While it is currently developed in the Linux environment using MATLAB 7, it has been tested to work also in Windows and with MATLAB version 6.5. MATLAB is a flexible environment for signal processing and speech research, and therefore it was a natural platform choice for Aparat.

The workflow for Aparat is shown in Fig. 2. First, the speech signal is recorded or imported using some external means. As the signal quality is of great importance in inverse filtering, recording should be performed in non-reverberant conditions with a high-quality condenser microphone. The recorded signals are then segmented to separate the voiced portions of interest. Although simple selection of single vowels may be performed in Aparat, more complex tasks such as segmentation of continuous speech should be handled using other software designed for such tasks. In the next stage, the signals are in-

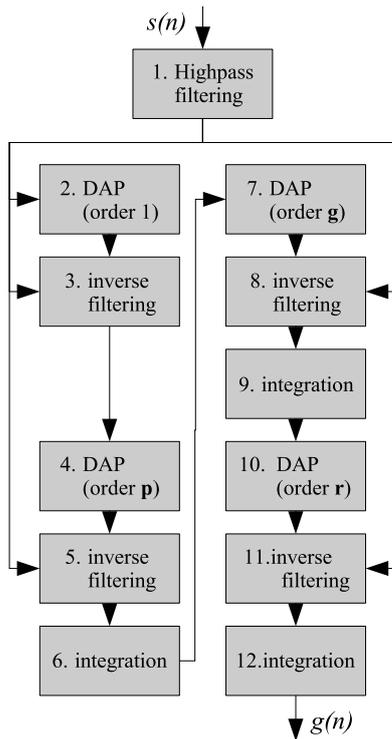


Figure 1: The block diagram of the IAIF method for the estimation of the glottal excitation $g(n)$ from the speech signal $s(n)$.

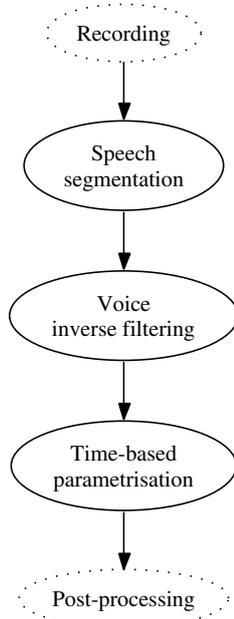


Figure 2: The intended workflow for Aparat. The parts of the workflow not supported by Aparat are drawn with a dotted line.

verse filtered. This is a semi-automatic process, in which the user observes the signal view and tunes the IF parameters accordingly. This process takes 5–15 seconds for a given sample of reasonable quality. During the inverse filtering stage, some meta-data such as the subjective evaluation of IF quality can also be given. When the IF stage is completed, a stable portion of the signal can be marked and automatically parametrised using several time-based parameters. This data is automatically exported to the MATLAB workspace and may be saved at a single click of a button. After this, the data may be post-processed either using the variables exported to the MATLAB workspace or by writing a MATLAB script to process the saved data files.

4. Use of the software

When Aparat is run, the main dialogue and the signal view are shown. The windows are illustrated in Fig. 3. By default, the files in the current directory are shown in the file listing (no. 1 in Fig. 3), but the directory may be changed by selecting File, Open Directory. By clicking a file, it is opened and the waveform is displayed in the upper pane of the signal view. The desired channel of a stereo signal may also be selected, together with the orientation of the signal (2). The IF window size may be set either by dragging the selection in the waveform view or by setting a fixed-size window in the main dialogue (3). When a selection is made, its fundamental frequency given by the YIN algorithm [14] can be seen in (4). It is also possible to play back the selection there. In (5), the IF parameters may be set. If the Auto button is pressed, the IF parameters are determined automatically according to an algorithm described in [15]. Alternatively, the DAP model order (labelled “max no. of formants”), the integration coefficient (labelled “Lip radiation”), and the high-pass filter cutoff frequency may be defined either by setting the sliders or by entering numeric values [12]. Finally, the cut-off frequency of high-pass filtering is adjusted automatically using the fundamental frequency if the check box is marked. The glottal flow acquired by inverse filtering with the given parameters is shown in the lower pane of the signal view in Fig. 3. In (6), meta-data about the sample may be entered. The slider can be used to grade the sample, and the text box may be used to enter notes about the sample. Finally, by pressing the Save button (7), the inverse filtering data is saved to a standard MATLAB file with the same base name as the sample wave file.

There exist several tools to assist the tuning of the inverse filtering parameters. Fig. 4 shows the spectrum view, which plots the spectra of the glottal and vocal tract filter transfer functions given by IAIF together with the spectra of the original signal and the glottal flow. Since no resonant frequencies should appear in the glottal flow, the monotonicity of its spectrum can be used to inspect the quality of the vocal tract fitting of the IAIF algorithm. Fig. 5 displays the phase-plane view of the glottal waveform [15]. The existence of minor loops (or lack thereof) in the phase-plane plot give an indication of the amount of formant ripple present in the glottal waveform. Also the kurtosis measure of inverse filtering quality is given in the phase-plane view.

Parametrisation of the glottal flow illustrated in the lower pane of the signal view in Fig. 3 is performed automatically, once a portion of the glottal flow is selected by dragging with the mouse. These parameters may be further inspected using the glottal flow parameter window. An example of the window showing parametrisation of a male voice is given in Fig. 6. It shows the selected portion of the glottal flow together with its derivative. The time and amplitude instants from which the pa-

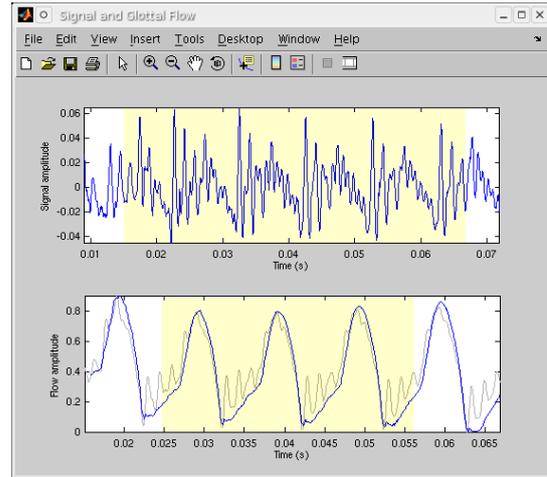
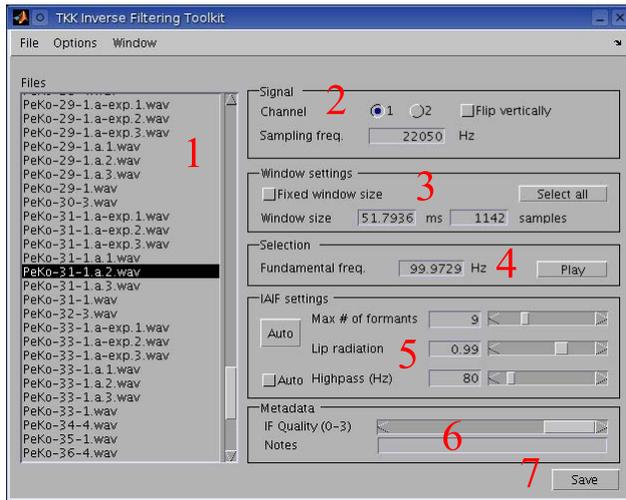


Figure 3: The main dialogue and the signal view of Aparat. Please refer to the text for number descriptions.

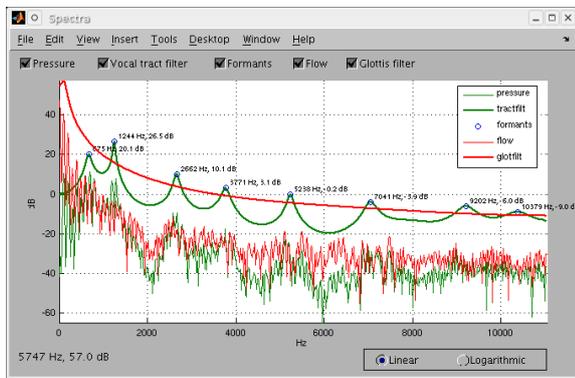


Figure 4: A view showing the spectra of the speech signal, calculated glottal flow and used vocal tract filter.

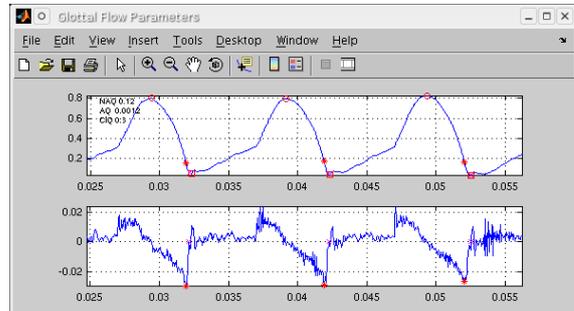


Figure 6: A view illustrating the time-based parameters of the glottal flow together with automatically obtained time instants for parameter calculation. In this example, NAQ, AQ, and CIQ are shown for a male voice with a fundamental frequency of 100 Hz [11].

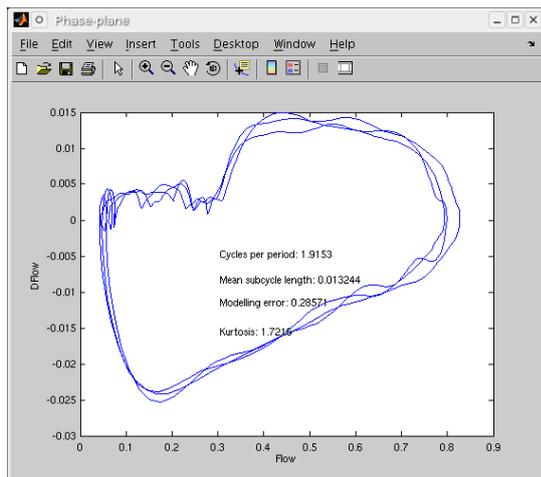


Figure 5: A view displaying the phase-plane graph of the glottal flow. The phase-plane plot may be used to inspect the quality of the IF process. [15]

rameters are calculated without the need of user intervention and marked in the graphs. Although Fig. 6 only shows calculation of the normalized amplitude quotient (NAQ), the amplitude quotient (AQ), and the closing quotient (CIQ), automatic deduction of for example the open quotient (OQ), the speed quotient (SQ), and the difference of first two harmonics (H1-H2) have already been implemented and their inclusion in Aparat will be trivial [16]. Another example of the parameter window is given in Fig. 7, which illustrates parametrisation of a traditionally rather difficult voice, a female voice with a fundamental frequency of 440 Hz. Although the glottal flow waveform already exhibits some formant ripple due to imperfect inverse filtering, the parameters can still be accurately calculated.

All data acquired by Aparat are automatically exported to the main MATLAB workspace for further interactive processing, making data exploration quite straightforward. Furthermore, by pressing the Save button in the main dialogue, the user may save the data to a standard MATLAB mat file, the name of which is automatically deduced from the wave file name. These files may be read back easily into the MATLAB environment for

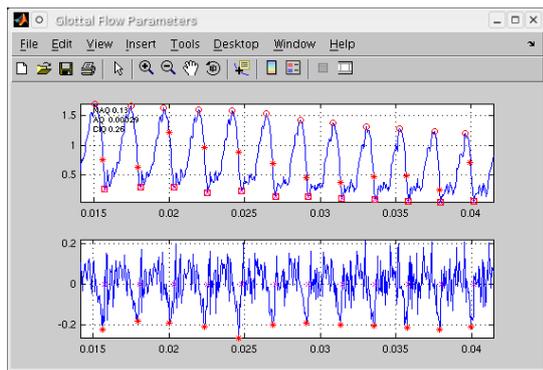


Figure 7: Parametrisation of a female voice sample with a fundamental frequency of 440 Hz. While the waveform shows some formant ripple due to imperfect inverse filtering, the parameters are still calculated accurately.

further processing or exporting to statistical software packages.

5. Discussion

This paper has described HUT Voice Source Analysis and Parametrisation Toolkit (Aparat), a MATLAB-based software for voice inverse filtering and analysis of the glottal flow. Aparat is a streamlined user interface for the semi-automatic IAIF inverse filtering algorithm and several time-based glottal flow parameters. According to the authors' knowledge, no other open source software for voice inverse filtering currently exists.

Aparat has been used successfully in several research projects at TKK Laboratory of Acoustics and Audio Signal Processing. In the largest tasks so far, several hundreds of vowel segments were processed at a rate not feasible with traditional or computer-based manual inverse filtering approaches. Furthermore, the initial support for completely automatic deduction of inverse filtering parameters holds even greater promise for automatic inverse filtering of large speech materials.

The application of automatic inverse filtering has been problematic in high pitched signals. The sparse harmonic structure of the source signal causes LPC-based methods to bias on the spectral peaks instead of the formant structure of the signal. The use of DAP in Aparat reduces this problem. Furthermore, the majority of inverse filtering studies restrict themselves to /a/ and /ae/ vowels due to their high first formants, which eases the separation of the voice source and the vocal tract. In our experience, both high-pitched female voices and multiple vowels can be processed using Aparat with reasonably good results (although male /a:/ vowels still produce the best outcome).

To gather widest possible audience and to encourage adoption of algorithms developed at TKK, a decision was made to publish Aparat using a liberal open source licence, the GNU Library General Public License, which allows both modification and redistribution of the software. The software is available at <http://www.acoustics.hut.fi/software/aparat/>.

6. Acknowledgements

This research was supported by the Academy of Finland (project number 200 859).

7. References

- [1] G. Fant, *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations*, 2nd ed. The Hague, Netherlands: Mouton, 1970.
- [2] R. L. Miller, "Nature of the vocal cord wave," *J Acoust Soc Am*, vol. 31, no. 6, pp. 667–677, June 1959.
- [3] M. Rothenberg, "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing," *J Acoust Soc Am*, vol. 53, no. 6, pp. 1632–1645, 1973.
- [4] J. B. Allen and T. H. Curtis, "Automatic extraction of glottal pulses by linear estimation," *J Acoust Soc Am*, vol. 55, no. 2, p. 396, February 1973.
- [5] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Commun*, vol. 11, pp. 109–118, 1992.
- [6] M. Södersten, A. Håkansson, and B. Hammarberg, "Comparison between automatic and manual inverse filtering procedures for healthy female voices," *Logoped Phoniatr Vocol*, vol. 24, pp. 26–38, 1999.
- [7] E. Vilkman, L. Lehto, T. Bäckström, and P. Alku, "Vocal loading of call centre personnel," in *Proc. of the 6th International Workshop, Advances in Quantitative Laryngoscopy, Voice and Speech Research*, Hamburg, Germany, April 2003, pp. 1–7.
- [8] S. Granqvist, S. Hertegård, H. Larsson, and J. Sundberg, "Simultaneous analysis of vocal fold vibration and transglottal airflow: exploring a new experimental setup," *JVOICE*, vol. 17, no. 3, pp. 312–330, September 2003.
- [9] P. Milenkovic, "Glottal inverse filtering by joint estimation of an AR system with a linear input model," *IEEE T Acoust Speech*, vol. 34, no. 1, pp. 28–42, February 1986.
- [10] M. Lee and D. G. Childers, "Manual glottal inverse filtering algorithm," in *Proceedings of the IASTED International Conference on Signal and Image Processing*, Orlando, Florida, USA, November 1996, pp. 34–37.
- [11] P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parametrization of the glottal flow," *J Acoust Soc Am*, vol. 112, no. 2, pp. 701–710, August 2002.
- [12] P. Alku, B. Story, and M. Airas, "Evaluation of an inverse filtering technique using physical modeling of voice production," in *Proceedings of ICSLP 2004*, Jeju Island, South-Korea, October 2004.
- [13] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal Processing*, vol. 39, pp. 411–423, February 1991.
- [14] A. de Cheveigne and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *J Acoust Soc Am*, vol. 111, no. 4, pp. 1917–1930, 2002.
- [15] T. Bäckström, M. Airas, L. Lehto, and P. Alku, "Objective quality measures for glottal inverse filtering of speech pressure signals," in *Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, USA, Mar. 2005.
- [16] P. Alku, "Parameterisation methods of the glottal flow estimated by inverse filtering," in *Proc. ISCA Workshop on Voice Quality: Functions, analysis and synthesis (VO-QUAL03)*, Geneva, Switzerland, August 2003, pp. 81–87.