[LOGO]

[Publication metadata: please leave blank]
[Publication metadata: please leave blank]
[Publication metadata: please leave blank]
[Publication metadata: please leave blank]
[Publication metadata: please leave blank]

Aalto Aparat – A Freely Available Tool for Glottal Inverse Filtering and Voice Source Parameterization

Paavo Alku, Hilla Pohjalainen, Manu Airaksinen

Aalto University, Department of Signal Processing and Acoustics, Aalto University, Finland
E-mail: paavo.alku@aalto.fi

Accepted: [date].

How to cite this publication: Paavo Alku, Hilla Pohjalainen, Manu Airaksinen: Aalto Aparat – A freely available tool for glottal inverse filtering and voice source parameterization. Proc. Subsidia: Tools and Resources for Speech Sciences, Malaga, Spain, June 21-23, 2017.

ABSTRACT: A software tool, Aalto Aparat, is introduced for glottal inverse filtering analysis of human voice production. The tool enables using two inverse filtering methods (Iterative adaptive inverse filtering, Quasi closed phase analysis) to estimate the glottal flow from speech. The inverse filtering analysis can be conducted using a graphical interface either automatically or in a semiautomatic manner by allowing the user to select the best glottal flow estimate from a group of candidates. The resulting glottal flow is parameterized with a multitude of know parameterization methods. Aalto Aparat is easy to use and it calls for no programming skills by the user. This new software tool can be downloaded as a stand-alone package free of charge to be run on two operating systems (Windows and Mac OS).

Keywords: glottal inverse filtering; voice source; speech research tool.

1. INTRODUCTION

Voiced speech is excited by a quasiperiodic airflow pulse form which is generated at the vocal folds. This excitation waveform, referred to as the glottal volume velocity waveform (shortly glottal flow), is the source of some of the most important acoustical cues embedded in speech. The fluctuation speed of the vocal folds determines the cycle length of the glottal flow which in turn affects the sensation of pitch from speech signals. The human speech production mechanism is capable of varying not only the fluctuation *speed* of the vocal folds but also their fluctuation *mode* thereby generating glottal flow pulses whose shape varies from smooth (i.e. large spectral tilt) to more abruptly changing (i.e. smaller spectral tilt). The shape of the glottal pulse is known to signal acoustical cues which are used, for example, in vocal communication of emotions (Gobl & Ni Chasaide, 2003).

Direct non-invasive recording of the glottal flow is, unfortunately, not possible due to the position of the vocal folds in the larynx behind cartilages. Non-invasive analysis of the glottal flow is, however, enabled by using an alternative to direct acoustical measurements, the technique known as *glottal inverse filtering* (GIF) (Alku, 2011; Drugman et al., 2014). This corresponds to using the idea of mathematical inversion: by recording the output of the speech production system, the pressure signal captured by microphone, a computational model is first built for those processes (i.e. vocal tract, lip radiation) that filter the glottal excitation. By feeding the recorded speech signal through the inverse models of the filtering processes, an estimate for the glottal flow is obtained. Analysis of speech production with GIF consists typically of two phases: (1) the estimation phase in which glottal flow signals are estimated from speech utterances with a selected GIF method, and (2) the parameterization phase in which the obtained waveforms are expressed in a compressed form with selected glottal parameters.

Given the fact that digital GIF methods have been developed since the 1970's, there are plenty of known algorithms available today both for glottal flow estimation and parameterization. (For further details of GIF history, see recent

reviews by Alku (2011) and Drugman et al., (2014)). It is delighting to observe that there is currently a growing interest among developers of GIF algorithms in open source practices and open repositories (Kane, 2012; Kane 2013; Degottex et al., 2014; Drugman, n.d.,). Inverse filtering and parameterization methods developed so far are, however, almost exclusively published in a manner which unfortunately hinders the utilization of these techniques by researchers who do not have programming skills. Therefore, the corresponding speech research methods can be fruitfully utilized only by those researchers who have engineering or computer science background while these open source tools (which are mostly made available today as MATLAB scripts) remain of limited practical value for individuals with non-technical background. While providing openly available MATLAB implementations in GIF helps, for example, in evaluating different GIF methods by the algorithm developers, we argue that it would be desirable to have GIF analysis available also for a wider speech research community. In other words, estimation and parameterization of the glottal flow should be made as easy as the Praat system (Boersma & Weenink, 2013) to researchers such as linguists, phoneticians, and physicians who typically do not have skills in programming languages such as MATLAB.

To the best of our knowledge, there are currently only two freely available GIF tools that do not call for any programming by the user to run the analysis. DeCap (Granqvist et al., 2003; Tolvan Data, n.d.) is a tool that enables voice source analysis in which the user adjusts each antiresonance of the vocal tract using the computer mouse by simultaneously monitoring the waveform of the GIF output on the computer screen. DeCap users typically define the optimal antiresonance setting as the one that results in the glottal flow pulse with the longest horizontal closed phase thereby utilizing a prevalent subjective inverse filtering criterion (Gauffin-Lindqvist, 1965; Rothenberg, 1973; Lehto et al., 2007). DeCap enables parameterizing the obtained glottal flow with, for example, H1-H2 (Titze & Sundberg, 1992) and NAQ (Alku, Bäckström, & Vilkman, 2002). TKK Aparat (Airas, 2008) is another user-friendly tool for glottal flow estimation and parameterization. (TKK stands for Teknillinen korkeakoulu, the former name of Aalto University.) Differently from DeCap, the user of TKK Aparat is given an option to select the best glottal flow signal from a set of candidates that have been computed from the input speech by varying two inverse filtering parameters (order of the vocal tract model, coefficient of the lip radiation). After the user has selected the best glottal flow candidate, the selected waveform can be parameterized in TKK Aparat by a rich set of parameterization methods. It is also worth noting that in addition to DeCap and TKK Aparat there are tools, such as VoiceSauce (Shue et al., 2011; VoiceSauce, 2016), which have been developed for the parameterization of voice production based on quantifying the speech pressure signal or its spectrum with measures such as H1*-H2* (Kreiman et al., 2012). These tools, however, do not estimate the glottal flow as a time-domain signal and therefore they cannot be regarded as (true) GIF tools.

The current study introduces a new, updated version of TKK Aparat, named Aalto Aparat. Similarly to its predecessor described by Airas (2008), Aalto Aparat is a speech inverse filtering and parameterization software that enables analyzing the voice source using a user-friendly graphical interface. The interface enables the user to conduct GIF analysis and parameterization with no need to use a specific programming language or environment. The tool has been originally programmed in MATLAB but, importantly, it can be downloaded freely as a stand-alone package which can be used without access to MATLAB. Compared to its predecessor published by Airas (2008), Aalto Aparat includes three major improvements. First, the tool now supports a new GIF algorithm, Quasi closed phase analysis (QCP), which has been shown to be one of the most accurate, if not the most accurate, GIF method (Airaksinen et al., 2014). Second, the user interface of Aalto Aparat has been improved, for example, by allowing the user to save the estimated flow waveforms as digital signals, not just their parameters. Third, the tool is now available (Aalto Aparat, 2016) as a stand-alone package that can be run in two operating systems (Microsoft's Windows, and Apple's Mac OS).

2. FEATURES OF AALTO APARAT IN A NUTSHELL

Aalto Aparat is a MATLAB-based tool designed for glottal inverse filtering studies of speech production. It supports the two phases (estimation and parameterization) that are typically needed in inverse filtering research. Given its user-friendly interface, the tool is well-suited particularly for studies in which large amounts speech signals need to be inverse filtered and parameterized. Inverse filtering in Aalto Aparat has been implemented in such a form that the user can fine-tune certain GIF

settings thereby affecting the estimated glottal flow estimate if desired. The user is given a possibility to select the best glottal flow estimate from a group of candidates, hence enabling running GIF analysis that is not completely automatic (and therefore maybe more prone to errors) but allows feedback from the user.

The input to Aalto Aparat is a speech pressure signal in the wav format. In the estimation phase, Aalto Aparat enables using two glottal inverse filtering algorithms, Iterative adaptive inverse filtering (IAIF) (Alku, 1992) or Quasi closed phase analysis (QCP) (Airaksinen et al., 2014), to estimate the glottal flow from the input speech. In IAIF, the user can select either conventional linear prediction (LP) (Makhoul, 1975), discrete all-pole modeling (DAP) (El-Jaroudi & Makhoul, 1991) or minimum variance distortionless response (MVDR) (Wölfel & McDonough, 2005) as a vocal tract all-pole modelling method. In QCP, the user can fine-tune the parameters of the attenuated main excitation (AME) (Alku et al., 2013; Airaksinen et al., 2014) weighting window. Once the user has selected the best estimate (see section 3.2), the obtained glottal flow is parameterized with several parameters both in the time domain using, for example, ClQ (Timcke, von Leden, & Moore, 1958) and NAQ (Alku, Bäckström, & Vilkman, 2002), and in the frequency domain using, for example, H1-H2 (Titze & Sundberg, 1992) and PSP (Alku, Strik, & Vilkman, 1997). In addition, it is possible to fit the Liljencrants-Fant (LF) waveform (Fant, Liljencrants & Lin, 1985) into the obtained glottal flow derivative. The parameterization procedures are equal to those in (Airas, 2008) where more details can be found.

3. DEMONSTRATION OF AALTO APARAT

The best way to describe Aalto Aparat is to study an example demonstrating the major parts that are needed in order to inverse filter and parameterize an input speech signal by this new tool. Given the space restriction in the current article, interested readers are referred to the manual of Aalto Aparat (Aalto Aparat, 2016) to get a more in-depth view on the system.

3.1. Step 1: Importing speech

When the Aalto Aparat tool is opened, the system displays two windows (Figure 1): control window (left) and signal view window (right). The former lists all the pre-recorded wav files (i.e. speech pressure signals) that the user wants to analyze. As a pre-processing step, the system enables removing ambient noise from the recorded signals with a liner phase high-pass filter whose cut-off frequency can be set automatically (according to the fundamental frequency of the input speech) or manually. In addition, the speech signal's sampling frequency can be changed and its polarity can be swapped if desired.

3.2. Step 2: GIF analysis

After the speech signal has been imported to the system, an analysis frame in which the GIF analysis is to be computed is set to a default duration (50 ms) and position (in the middle of the input signal). If desired, the user can, however, adjust both of these values. Next, the user selects the GIF method (either IAIF or QCP) after which the system automatically depicts the obtained glottal flow (Figure 1, right window, second pane from top) and its derivative (Figure 1, right window, bottom pane) on the computer screen. By pressing the corresponding buttons (Figure 1, left window, two red circles) the user can vary the value of two parameters of the selected GIF algorithm: the vocal tract filter order (Figure 1, upper red circle) or the lip radiation coefficient (Figure 1, lower red circle). After this, the system opens a new window which depicts a group of candidate glottal flow estimates that have been computed by varying the corresponding parameter (Figure 2 shows an example where the vocal tract order is varied). Once the user has screened the depicted waveforms, he/she can select the one that he/she considers best by clicking the waveform with the mouse. Finally, the selected glottal flow and its derivative appear into signal view window (Figure 3).

The procedure described above is flexible because it enables running the inverse filtering analysis either in an automatic or a semi-automatic mode. In the former, no user feedback is required by Aalto Aparat (i.e. default parameter values are used for the corresponding GIF algorithm). In the latter, the tool allows utilizing subjective criteria in letting the user to take advantage of his/her expertise to select the waveform that is he/she considers to be the best estimate of the unknown true glottal flow.

3.3. Step 3: Parameterization

After inverse filtering, the obtained glottal flow is parameterized in a completely automatic manner using a multitude of parameters (for further details, see Airas (2008)). Parameterization is activated from the corresponding menu, after which a new window pops up indicating the obtained parameter values (Figure 4). By pressing the

corresponding button (Figure 4, "LF-model, Evaluate"), the system matches the obtained glottal flow derivative with the LF pulse form, and shows the obtained LF parameter values (Figure 4, right bottom corner). In addition, Aalto Aparat also depicts the output of the LF fitting by depicting both the synthetic flow and its derivative as time-domain waveforms (Figure 5).

3.4. Step 4: Exporting data

Aalto Aparat enables saving both the obtained parameter values as well as two signals (estimated glottal flow and input speech, both as time-domain signals spanning the frame that was selected in the GIF analysis). In a typical inverse filtering session, the user has many input signals to be analyzed. Once all of these have been processed, one by one, the system enables combing the corresponding parameter data in a single array which can be later imported to, for example, Excel to be further processed (e.g. for statistical analysis and visualization).

Figure 1: Two windows of Aalto Aparat: control window (left) and signal view window (right). In control window, red circles show two settings (vocal tract filter order, lip radiation coefficient) that the user can vary if desired. In signal view window, the three panes show the input speech signal (top), the estimated glottal flow (middle), and the derivative of the estimated flow (bottom).

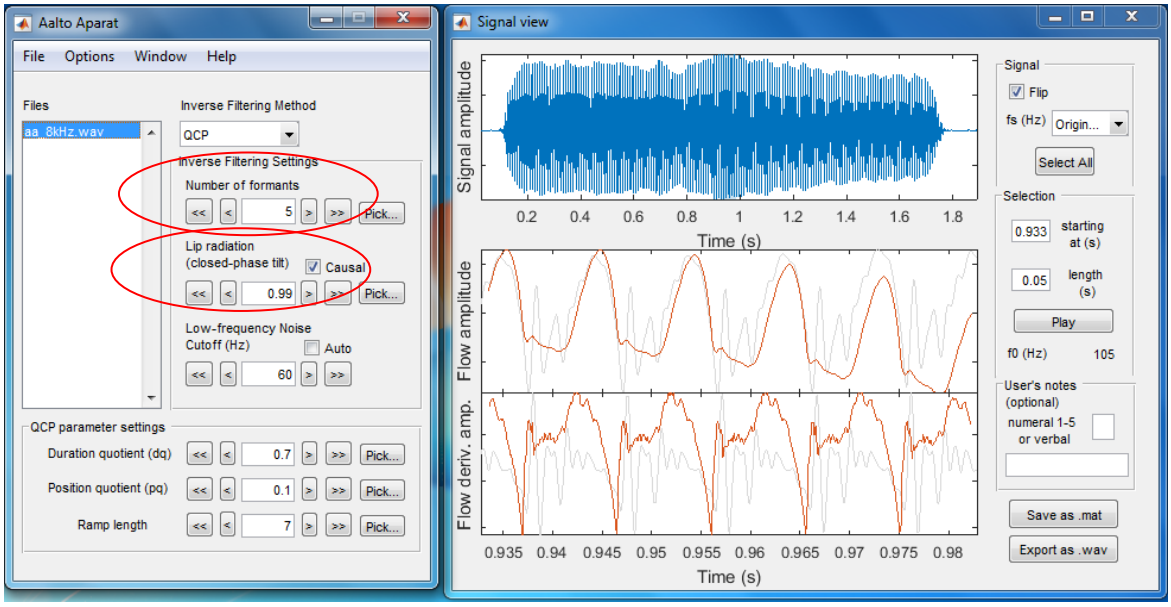


Figure 2: A group of candidate flow signals which have been obtained by varying the vocal tract filter order from 4 (top signal) to 16 (bottom signal).

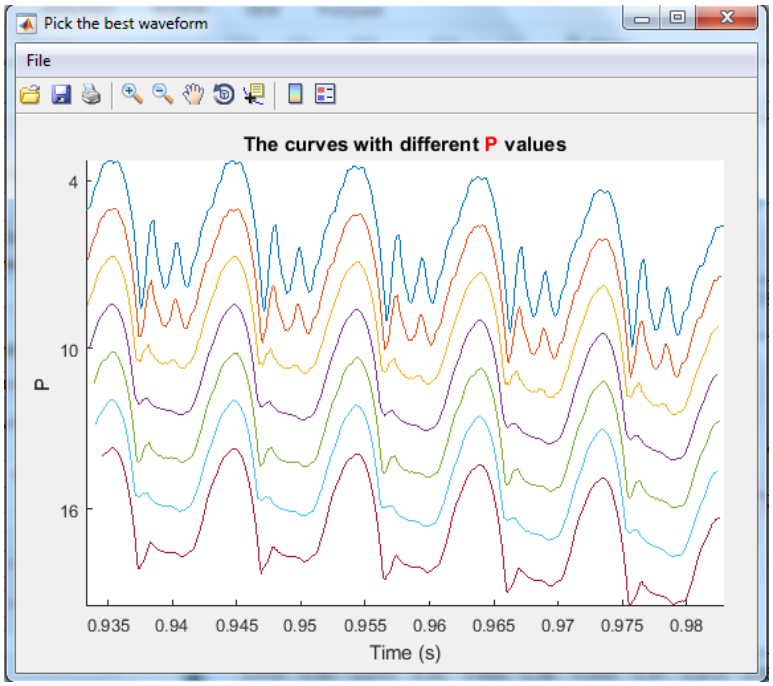


Figure 3: Signal view window after the user has made his/her selection for the best glottal flow estimate

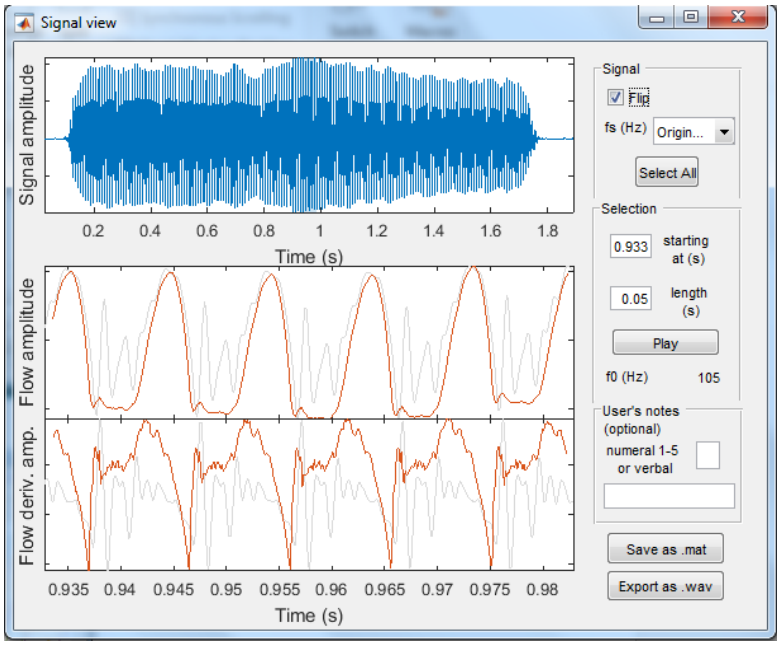


Figure 4: Results of parameterizing the glottal flow shown in Figure 3. Parameters are organized into time-based, frequency-based and LF model -based.

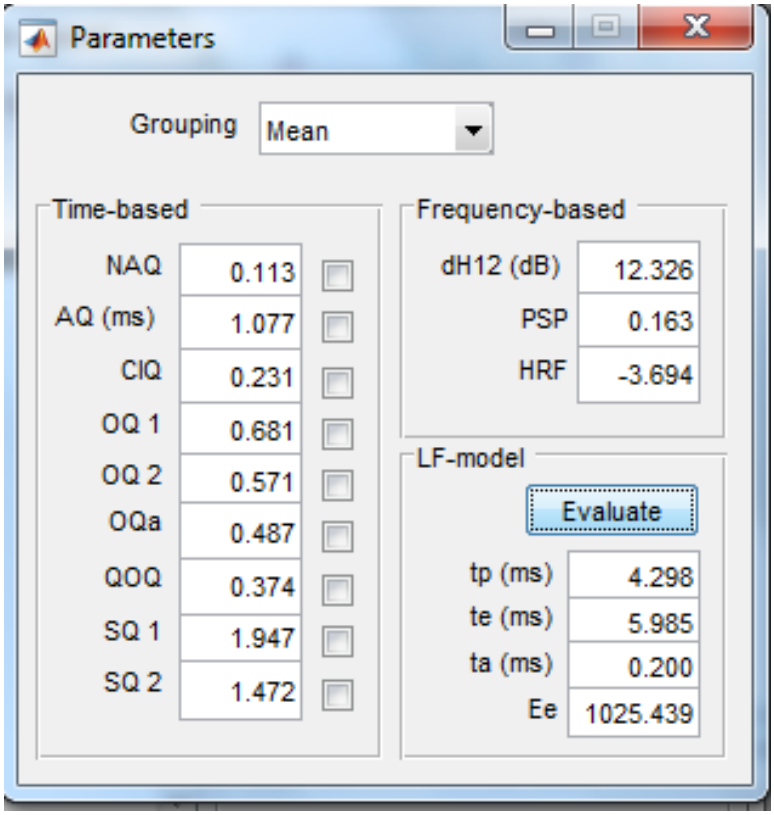
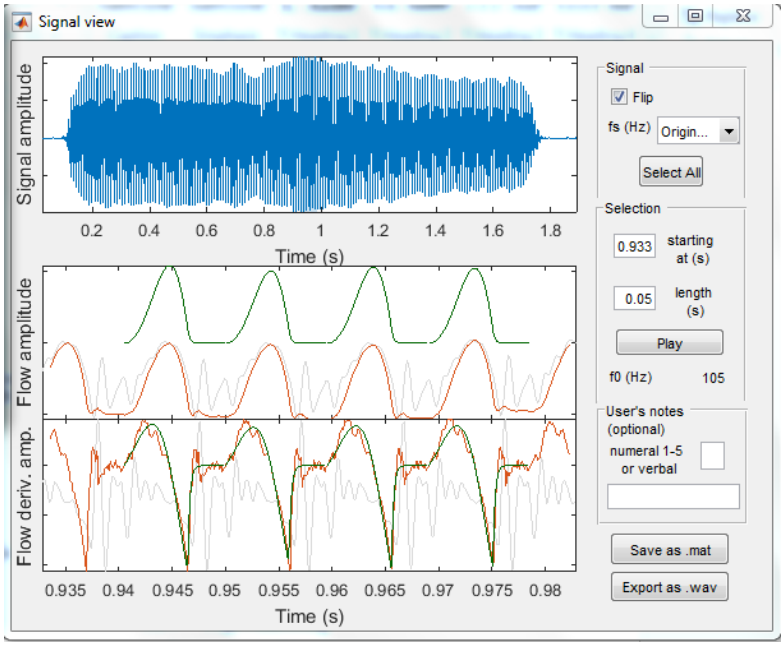


Figure 5: Signal view window after the user has selected the LF model based parameterization. Top pane shows the input speech signal. Middle pane shows the LF-synthesized flow (upper) and the estimated flow (lower). Bottom pane depicts two flow derivatives on top of each other: the one computed from the estimated flow (red) and the LF-modelled one (green).



4. CONCLUSIONS

A new glottal inverse filtering and voice source parameterization tool, Aalto Aparat, has been described in this article. Aalto Aparat is based on its predecessor, TKK Aparat, both offering a graphical interface using which a user with no programming skills can conduct glottal inverse filtering analysis and parameterization of the estimated flow signals. The tool has been programmed in MATLAB but it can be downloaded as a stand-alone package which can be run without having access to MATLAB. In comparison to its predecessor, Aalto Aparat involves a few major changes, the most important one being an opportunity to use a recently proposed potential GIF method, QCP. In addition, the Aalto Aparat stand-alone package can be installed into two operating systems (Windows and Mac OS).

Usability of Aalto Aparat has not been formally evaluated. However, the tool's predecessor, TKK Aparat, went through a formal evaluation process in which the interface was developed into its current form by collecting user feedback in a usability test (Airas, 2008). As a conclusion, the usability test of TKK Aparat indicated that the system can be easily taken advantage of by anyone who have basic knowledge in glottal inverse filtering. Since the user interface of Aalto Aparat has been changed only slightly from that of TKK Aparat (e.g. by correcting minor bugs), we argue that also the Aalto Aparat software is easy to use by anyone who knows the basics of glottal inverse filtering.

Researchers interested in glottal inverse filtering and voice source parameterization are welcome to download the Aalto Aparat software free of charge from Aalto Aparat (2016).

5. REFERENCES

Aalto Aparat. (2016). Retrieved from http://research.spa.aalto.fi/projects/aparat/.

Airaksinen, M., Raitio, T., Story, B., & Alku, P. (2014). Quasi closed phase glottal inverse filtering analysis with weighted linear prediction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(3), 596–607.

Airas, M. (2008). TKK Aparat: An environment for voice inverse filtering and parameterization. *Logopedics, Phoniatrics and Vocology*, 33(1), 49–64.

Alku, P. (1992). Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. *Speech Communication*, 11(2–3), 109–118.

Alku, P. (2011). Glottal inverse filtering analysis of human voice production – A review of estimation and parameterization methods of the glottal excitation and their applications. *Sadhana – Academy Proceedings in Engineering Sciences*, 36(5), 623–650.

Alku, P., Bäckström, T., & Vilkman, E. (2002). Normalized amplitude quotient for parameterization of the glottal flow. *Journal of the Acoustical Society of America*, 112(2), 701–710.

Alku, P., Pohjalainen, J., Vainio, M., Laukkanen, A-M., & Story, B. (2013). Formant frequency estimation of high-pitched vowels using weighted linear prediction. *Journal of the Acoustical Society of America*, 134(2), 1295–1313.

Alku, P., Strik, H., & Vilkman, E. (1997). Parabolic spectral parameter - A new method for quantification of the glottal flow. *Speech Communication*, 22, 67–79.

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer. Retrieved from http://www.praat.org/.

Degottex, G., Kane, J., Drugman, T., Raitio, T., & Scherer, A. (2014). Covarep – A collaborative voice analysis repository for speech technologies. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 960–964.

Drugman, T. (n.d). Retrieved from http://tcts.fpms.ac.be/~drugman/Toolbox/.

Drugman, T., Alku, P., Alwan, A., & Yegnanarayana, B. (2014). Glottal source processing: from analysis to applications. *Computer, Speech and Language*, 28(5), 1117–1138.

El-Jaroudi, A., & Makhoul, J. (1991). Discrete all-pole modeling. *IEEE Transactions on Signal Processing*, 39, 411–423.

Fant, G., Liljencrants, J., & Lin, Q. (1985). A four-parameter model of glottal flow. *Speech Transmission Laboratory – Quarterly Progress and Status Report*, 26(4), 1–13.

Gauffin-Lindqvist, J. (1965). Studies of the voice source by means of inverse filtering. *Speech Transmission Laboratory – Quarterly Progress and Status Report*, 6(2), 8–13.

Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, 40, 189–212.

Granqvist, S., Hertegård, S., Larsson, H., & Sundberg, J. (2003). Simultaneous analysis of vocal fold vibration and transglottal airflow: exploring a new experimental setup. *Journal of Voice*, 17, 312–330.

Kane, J. (2012). Tools for analysing the voice - Developments in glottal source and voice quality analysis (Doctoral dissertation). Trinity College Dublin.

Kane, J. (2013). Retrieved from https://github.com/jckane/Voice_Analysis_Toolkit.

Kreiman, J., Shue, Y-L., Chen, G., Iseli, M., Gerratt, B., Neubauer, J., & Alwan, A. (2012). Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation. *Journal of the Acoustical Society of America*, 132(4), 2625–2632.

Lehto, L., Airas, M., Björkner, E., Sundberg, J., & Alku, P. (2007). Comparison of two inverse filtering methods in parameterization of the glottal closing phase characteristics in different phonation types. *Journal of Voice*, 21(2), 138–150.

Makhoul, J. (1975). Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(3), 561–580.

Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *Journal of the Acoustical Society of America*, 53(6), 1632–1645.

Shue, Y-L., Keating, P., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. In *Proceedings of the 17th International Congress on Phonetic Sciences*, pp. 1846–1849.

Timcke, R., von Leden, H., & Moore, P. (1958). Laryngeal vibrations: measurements of the glottic wave. *Archives of Otolaryngology*, 68, 1–19.

Titze, I., & Sundberg, J. (1992). Vocal intensity in speakers and singers. *Journal of the Acoustical Society of America*, 91(5), 2936–2946.

Tolvan Data. (n.d). Retrieved from http://www.tolvan.com/.

VoiceSauce. (2016). VoiceSauce - A program for voice analysis. Retrieved from http://www.seas.ucla.edu/spapl/voicesauce/.

Wölfel, M., & McDonough, J. (2005). Minimum variance distortionless response spectral estimation. *IEEE Signal Processing Magazine*, 22(5), 117–126.