# OBJECTIVE QUALITY MEASURES FOR GLOTTAL INVERSE FILTERING OF SPEECH PRESSURE SIGNALS

*Tom Bäckström, Matti Airas, Laura Lehto and Paavo Alku*

Helsinki University of Technology (HUT), Laboratory of Acoustics and Audio Signal Processing
P.O.Box 3000, FIN-02015 HUT, email: Tom.Backstrom@hut.fi

## ABSTRACT

Glottal inverse filtering is a process where the effects of the vocal tract are cancelled from the speech signal in order to estimate the voice source. Traditionally, inverse filtering methods have involved a high level of manual tuning of parameters, such as the vocal tract model order. In this article, we present objective heuristics for the measurement of the quality of the resulting glottal flow estimate. In addition, we propose an automatic method for determining the order of the vocal tract all-pole model in inverse filtering based on phaseplane analysis and estimation of the glottal flow kurtosis.

## 1. INTRODUCTION

Analysis of the source of voiced speech, the glottal volume velocity waveform, is of fundamental importance in speech science [1]. However, direct observations of the glottis are difficult and they require invasive methods such as insertion of instruments through the vocal tract. An alternative, indirect, but noninvasive method for studying the behaviour of the glottal source is inverse filtering, which corresponds to cancelling the acoustic effects of the vocal tract in order to estimate the glottal source. Inverse filtering methods take as an input either the oral flow signal obtained by using a pneumographic mask [2] or the acoustic pressure waveform recorded by a microphone in a free field outside the mouth [3].

Inverse filtering methods developed typically require manual tuning of certain parameters of the underlying inverse filtering algorithm. In some techniques, these user adjustments concern determining the positions of the vocal tract resonances [4] while in other approaches the manual parameter tuning concerns selection the order of the all-pole filter to model the vocal tract [3]. In all the cases, the selection of the best parameter selection relies on visual evaluation of the output of the inverse filtering algorithm, the obtained estimate of the glottal flow. Since the parameter tuning process is subjective, it is possible that it introduces bias to the results. In addition, manual tuning of parameters is labourintensive and automatising of the process would thus be highly desirable.

In this article, we will study objective measures, both heuristic and statistical, that quantify the quality of the inverse filtering result, the estimate of the glottal flow. These measures are then used to construct an algorithm that automatically determines the parameters of the inverse filtering algorithm. As an inverse filtering method, we use a semi-automatic technique that estimates the glottal flow directly from the acoustic speech pressure waveform recorded in a free field, i.e., no flow mask is required. The method, Iterative Adaptive Inverse Filtering (IAIF) [3], is based on the separated speech production model by Fant [5] and it is described in detail in [6]. In the present implementation of the IAIF algorithm, the estimation of the vocal tract transfer function is based on an allpole modelling technique, called Discrete All-Pole Modelling (DAP) [7] instead of the conventional Linear Predictive Coding (LPC). Due to the application of the Itakura-Saito distortion measure, the estimation of the vocal tract model given by DAP is more accurate than that obtained by LPC especially for high-pitched voices.

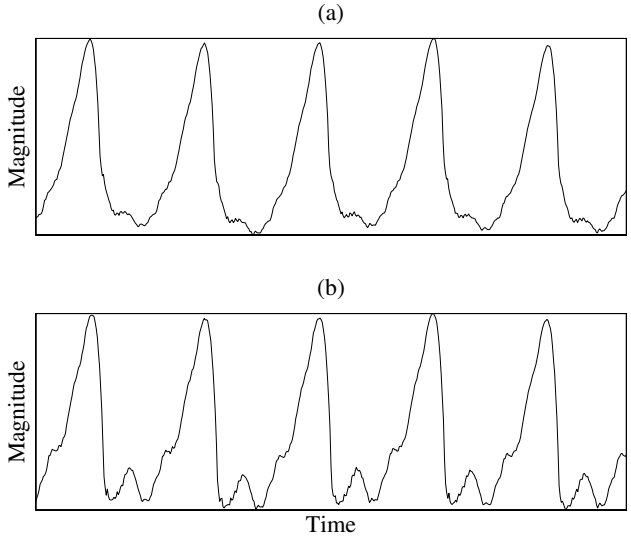


**Fig. 1**. Two glottal flow estimates of a female vowel /a/. Estimate (a) without formant ripple, (b) corrupted by formant ripple.

## 2. QUALITY MEASURES

Inverse filtering methods attempt to estimate the glottal source, but, with noninvasive methods, it is impossible to exactly capture the true flow generated by the vocal folds. Hence, it is also impossible to compare how accurately the estimated glottal flows given by inverse filtering correspond to the true glottal flows. Therefore, assessment of the performance of inverse filtering is based, more or less, on an ambiguous concept, the *quality* of the glottal flow estimate. The major artifacts affecting the quality of obtained glottal flow estimate is formant ripple, that is, an undesirable component present in the output of inverse filtering due to unsuccessful cancellation of the formants of the vocal tract model. This type of errors are often visible as ripple in the closed phase of the glottal cycle (See Fig. 1b).

A method based on the phaseplane analysis was proposed for the assessment of the quality of the glottal flows computed by inverse filtering [8]. This technique is based on the assumption that the vocal tract can be modelled as a cascade of second order resonators [9], whereby the glottal waveform can be consider by the second order harmonic equation:

$$\frac{d^2x}{dt^2} + x = 0. \tag{1}$$

This system can be analysed in the phase-plane $(x, y)$ by:

$$\frac{dx}{dt} = y \quad \text{and} \quad \frac{dy}{dt} = -x \tag{2}$$

Based on our assumption, a glottal waveform without formant ripple should be cyclic in the phase-plane, corresponding to the fundamental frequency (see Fig. 2). The resonances of the vocal tract yield different solutions that are also periodic. Therefore, if the glottal waveform estimate is corrupted by uncancelled vocal tract resonances, these should be visible in a phase-plane plot as minor loops (see Fig. 3). Conversely, by checking if the phase-plane has loops inside the fundamental frequency periods, we can asses the quality of the inverse filtering output.

The phaseplane analysis, proposed in [8], included only a visual assessment of the phaseplane plot. In order to implement this idea with no subjective evaluations by the user, one needs an algorithm to automatically determine if the phaseplane plot has loops other than the those caused by the fundamental frequency. Before this, however, the high-frequency noise components present in the phase-plane have to be removed by lowpass filtering both the flow and differentiated flow. (We used low-pass filtering with a cut off frequency at 3.5 kHz.) The number of loops can then readily be calculated by taking the slope angle at each time step. Furthermore, the angles must be unwrapped (by compensating for jumps larger than $\pi$ by $-2\pi$).

Comparing the initial and final angle with the number of loops corresponding to the fundamental frequency gives

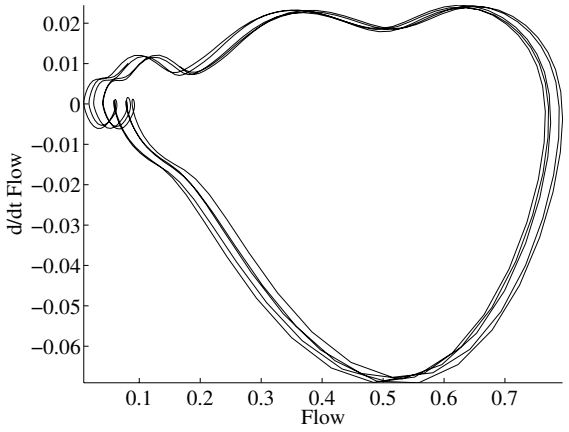


**Fig. 2**. Phase-plane plot of the glottal waveform, without formant ripple, from Fig. 1a (low-pass filtered at 3500Hz).

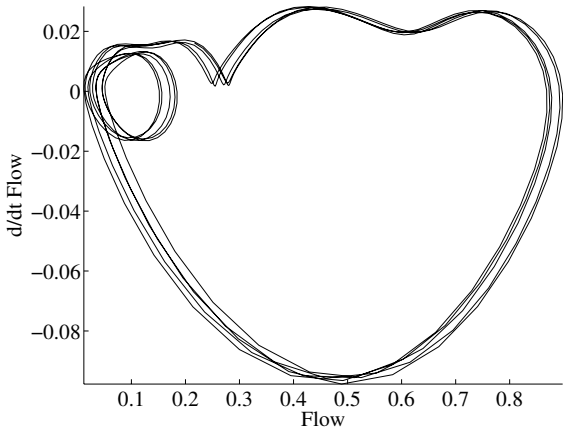


**Fig. 3**. Phase-plane plot of a glottal waveform corrupted by formants ripple from Fig. 1b (low-pass filtered at 3500Hz).

us an estimate of the number of loops to fundamental frequency period ratio. In other words, if the loops per period ratio is 1, then there is no formant ripple present in the signal. Correspondingly, if the ratio is 2 or higher, then there exists some minor loops. In practise, however, we must threshold the estimate somewhere in the range $[1, 2]$, for example, at 1.3, to make the estimate robust.

The size of the sub-cycles matters as well. That is, the size of sub-cycles directly corresponds to the magnitude of formant ripple. By estimating the size of sub-cycles, it is possible to give a measure of the quality of inverse filtering. In order to do this, we must first find the points in the phase-plane, where the trajectory crosses itself. From the trajectory intersections, it is then possible to easily estimate the subcycle area.

The sub-cycles are located with a brute force algorithm, since more intelligent approaches would be very complex. Namely, at each point, we test for trajectory crossings at all points ahead in a window of length $T/2$, where $T$ is the length of the fundamental period.

The area of the sub-cycle can now be estimated in the following way. Let $l$ be the mean circumference length of all sub-cycles. If we assume that the sub-cycle is a circle, then its area $a$ is proportional to the square of the circumference length, that is, $a \propto l^2$. A suitable measure for the quality of the flow estimate is the total area of sub-cycles $A = Na = Nl^2$, where $N$ is the number of sub-cycles per fundamental frequency cycle.

In addition to the phase-plane analysis, we propose another quality measure for evaluation of the glottal flow estimate. This quality measure, the kurtosis, is well known in, for example, the optimisation criterion of independent component analysis [10]. The kurtosis measures the similarity of a distribution to the Gaussian distribution. According to the central limit theorem, mixtures of equally distributed signals converges to the Gaussian distribution as the number of signals grows. Convolution by the vocal tract transfer function can be considered as mixing of the glottal waveform at different time delays and consequently, the output should be more similar to Gaussian than the glottal waveform. In other words, we can use kurtosis as a performance measure for the inverse filtering.

The range of kurtosis is $[-3, +\infty]$, where positive and negative values correspond to supergaussian and subgaussian, respectively, that is, sharper or flatter peak of the distribution. The kurtosis of a Gaussian distribution is zero. A glottal waveform is subgaussian by nature, since it has two distinct peaks (see Fig. 4). Therefore, generally speaking, the glottal waveform with the lowest kurtosis is the best waveform.

We have thus obtained three measures for inverse filtering quality assessment, namely, number of cycles per fundamental frequency cycle, total area of sub-cycles and the kurtosis. A fourth measure, which can be readily obtained, is the Itakura-Saito distance of the DAP model in the inverse filtering algorithm.

## 3. AUTOMATIC INVERSE FILTERING

Let us recall our objective: We are supposed to find such model parameters for the inverse filtering task that the resulting flow signal has the best possible quality, i.e., effects of the uncancelled vocal tract resonances are minimised. The sub-cycle analysis provides a robust first quality measure; if the phase-plane has sub-cycles, then the flow signal is corrupted. We can therefore begin by searching for such parameters using which the flow signal has no sub-cycles.

However, observe that the measures obtained from sub-cycle analysis are useful only when there are sub-cycles present in the phase-plane. If the cycles corresponding to the fundamental frequency are simple, then the sub-cycle analysis will not provide any new information. In such cases, we must use the kurtosis and the Itakura-Saito distance measure. However, the Itakura-Saito distance criterion was used

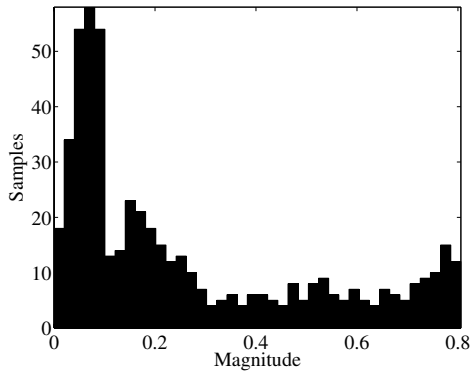


**Fig. 4**. Histogram of the glottal waveform from Fig. 1a.

in the optimisation of the DAP model and as a quality criterion for the inverse filtering process it can be biased. In other words, the Itakura-Saito measure relies on the assumptions made in the IAIF process, whereas the kurtosis is free from these assumptions.

In some cases, we may not be able to find any such parameters using which the flow signal would be free from sub-cycles. In such a case, the best we can do is to minimise the number and size of sub-cycles.

Computational complexity of the different methods varies greatly. Estimation of the kurtosis and Itakura-Saito distance are deterministic and inexpensive. Calculation of the number of sub-cycles is also deterministic but slightly more complex. The estimation of sub-cycle size is an order of magnitude more complex and it depends greatly on the length and complexity of the flow signal. The processing of a signal with many sub-cycles will be far more expensive than for a flow signal without formant ripples.

In our experiments, where the sampling frequency was 22.05 kHz, we used even model orders of the vocal tract from $m = 8$ upward. Each increment of the model order then brings one additional formant to the model. The upper limit of the model order was set at $m = 50$. The range of the lip radiation coefficient $c$ (the zero of the first order FIR modelling the lip radiation effect [9]) was set to $[0.98,\ 0.9999]$.

Based on informal experiments with IAIF, we have found that the model order should be large enough so that the phase-plane representation does not contain sub-cycles, but as small as possible to avoid over learning. By visual assessment, the best solution was usually found among the three smallest model orders that showed no sub-cycles.

With these considerations in mind, we can state our algorithm as:

1. Starting from a model with $m = 8$ and $c = 0.99$, iterate through even model orders to find three first models without sub-cycles.
2. If no models free of sub-cycles were found, go to Step 6.

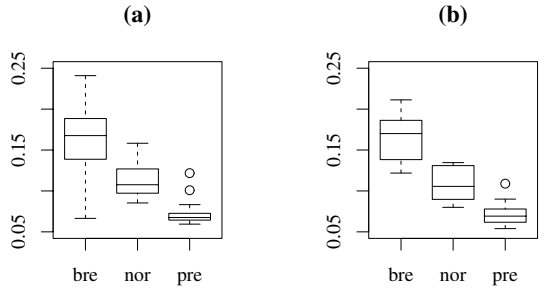


**Fig. 5**. Box-plots of NAQ values for breathy, normal and pressed phonations in analysis with (a) the automatic and (b) the manual approach.

3. Of the three models, choose the model order which yields the smallest kurtosis.
4. With an adequately fine grid, search for the lip radiation coefficient $c$ that minimises the kurtosis.
5. If the model with optimal $c$ is free from sub-cycles, then Stop. Otherwise, exclude this model order from further calculations and return to Step 1.
6. Iterate through even order models to find the model with the smallest total area of sub-cycles.
7. With an adequately fine grid, search for the lip radiation coefficient $c$ that minimises the total area of sub-cycles.

## 4. RESULTS

Six female and seven male participated in a recording session, where they pronounced sustained vowels /a/ in breathy, normal and pressed phonation. The speech samples were recorded in an anechoic chamber on a DAT tape, with the microphone placed at a constant distance of 40 cm from the mouth. The signals were digitally transferred from DAT tapes to a computer and down-sampled to 22.05 kHz. The analysis window was 10 glottal cycles or at maximum 60 ms.

The samples were manually inverse filtered with IAIF by three persons, each with prior experience of inverse filtering. The estimated glottal flows were parametrised with the Normalised Amplitude Quotient (NAQ), a voice source parameter that measures the characteristics of the glottal closing phase [11]. NAQ was selected because it has been shown to be robust against noise in glottal flows and its automatic implementation is straightforward. The results of these measurements were averaged to provide a comparison with the automatic procedure.

The automatic analysis was performed in the same way as the manual measurements, with the exception that the manual parameter tuning was replaced by the automatic procedure described in Section 3.

The NAQ values for different phonations, breathy, normal and pressed, are in both methods clearly separated (see Fig. 5). The variance of the automatic method is, however, larger in all phonation types. This was to be expected for two reasons. Firstly, the manual measurements are averaged over three manual measurements and are therefore bound to have smaller variance. Secondly, the visual quality criterion used for manual inverse filtering aims at obtaining consistent results. The automatic optimisation, on the other hand, operates in a different domain and does not take into account the consistency of NAQ values.

## 5. CONCLUSIONS

We have presented objective quality measures for inverse filtered flow signals of speech pressure wave forms and an algorithm that uses these to determine the IAIF model parameters. We have showed that the algorithm, based on phase-plane heuristics, provides consistent results when compared to manual parameter tuning and the NAQ parameter.

## 6. REFERENCES

[1] I. R. Titze, *Principles of Voice Production*, Prentice-Hall, Englewood Cliffs, NJ, 1994.

[2] M. Rothenberg, "A new inverse filtering technique for deriving the glottal airflow waveform during voicing," *J. Acoust. Soc. Amer.*, vol. 53, pp. 1632–1645, 1973.

[3] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Commun.*, vol. 11, pp. 109–118, 1992.

[4] M. Södersten, A. Håkansson, and B. Hammarberg, "Comparison between automatic and manual inverse filtering procedures for healthy female voices," *Log Phon Vocol*, vol. 24, pp. 26–38, 1999.

[5] G. Fant, *Acoustic Theory of Speech Production*, Mouton, The Hague, 1960.

[6] P. Alku, B. Story, and M. Airas, "Evaluation of an inverse filtering technique using physical modeling of voice production," in *Proc 8th Int Conf Spoken Language Proc (INTERSPEECH 2004 – ICSLP)*, Jeju Island, South Korea, October 4–8 2004.

[7] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal. Proc.*, vol. 39, pp. 411–423, 1991.

[8] J. A. Edwards and J. A. S. Angus, "Using phase-plane plots to assess glottal inverse filtering," *Electronics Letters*, vol. 32, no. 3, pp. 192–193, Feb. 1996.

[9] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[10] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Inc., New York, 2001.

[11] P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parameterization of the glottal flow," *J. Acoust. Soc. Amer.*, vol. 112, no. 2, pp. 701–710, August 2002.