Experimental evaluation of inverse filtering using physical systems with known glottal flow and tract characteristics

Derek Tze Wei Chu and Kaiwen Li
School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia
derek.chu@unswalumni.com, k.w.li@unswalumni.com

Julien Epps
School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, New South Wales 2052, Australia
j.epps@unsw.edu.au

John Smith[a] and Joe Wolfe
School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia
john.smith@unsw.edu.au, j.wolfe@unsw.edu.au

Abstract: The technique presented here uses an impedance head to measure the input impedance spectrum of a physical model of a vocal tract, and then to inject a known glottal flow waveform into the tract. The sound measured outside the mouth is used to evaluate inverse filtering techniques by comparison with the known glottal flow and measured acoustical properties of the tract. The normalized least square errors in the glottal flow were typically a percent or less in the time domain and several percent in the frequency domain. Accurate determination of resonance frequencies and bandwidths required a suitable order of inverse filter.

PACS numbers: 43.72.Ar [DO]
Date Received: February 19, 2013 **Date Accepted:** March 14, 2013

1. Introduction

The waveform of the acoustic flow through the glottis—the aperture between the vocal folds—is fundamental to our understanding of speech, and a fundamental input to models of speech. Unfortunately, it cannot usually be measured directly in normal speech. Instead, it is often estimated from the pressure signal measured outside the mouth via inverse filtering; a procedure whereby the filtering effects of the vocal tract and the radiation from the lips are removed by adjusting anti-resonances of inverse filters until the derived glottal flow has the expected characteristics. Such an approach, involving the separation of two unknowns with complex behavior (the glottal flow and the tract transfer function) from a single signal, cannot be unequivocal and will inevitably introduce some uncertainty (Drugman, 2012).

Numerical simulations are widely used to estimate the reliability of the parameters extracted by inverse filtering. Thus inverse filtering might be applied to a synthetic speech signal obtained by numerically filtering a synthesized glottal flow. However, such simulations will have their own limitations and uncertainties; often because the same numerical model used to synthesize the sound is later used to perform the inverse filtering. More realistic models can simulate some features of vocal fold vibration and/or use finite element models of the vocal tract that differ from the

[a]Author to whom correspondence should be addressed.

filter models used for the subsequent inverse filtering (Alku *et al.*, 2006a, 2006b). However, numerical methods inevitably involve some approximations associated with the mesh and the boundary conditions.

Measurements using a physical system in which the glottal flow is independently known can provide an experimental test of the accuracy and reliability of inverse filtering, including situations where inverse filtering techniques might be compromised; e.g., where large open quotients are involved or where fundamental frequencies are sufficiently high to approach the first formant frequency ($F1$) (Alku, 2011). Physical measurements also allow researchers to assess any experimental problems that might be present due to noise, interference, phase shifts, non-linearities, etc.

In this paper, an experimental system is presented that uses an impedance head to synthesize and to inject a known, periodic, glottal flow signal into a rigid, unvarying physical model of the vocal tract. The same impedance head is also used to measure precisely the acoustic properties of the tract. The measured output sound is then analyzed using techniques of inverse filtering with a range of parameters. Both the glottal flow and acoustic properties of the tract derived from the inverse filtering techniques are then directly compared with their known experimental values, and thus provide a rigorous test of the performance of the algorithms and assumptions used.

2. Materials and methods

2.1 Physical models of the tract

Motivated by piecewise constant area functions seen in the numerical simulation literature, these were constructed using stacked plexiglass discs 5 or 10 mm thick; each with a circular concentric hole with a diameter appropriate to the desired area function of the vocal tract—see Fig. 1. They were constrained inside an external cylinder 170 mm in length and 60 mm in diameter, and the joints between discs were sealed with a very thin layer of petroleum jelly. The area function used for the results reported herein was appropriate for the vowel /æ/ and based upon MRI images (Story *et al.*, 1996) with interpolation where necessary.

2.2 Measurement of input impedance of model tract

Acoustic input impedance spectra of the model tract were measured using a three microphone impedance head and calibrated with three non-resonant loads (Dickens *et al.*, 2007). This technique allows high precision and dynamic range over a wide frequency range. The microphones (Brüel and Kjær 4944-A) were located 10, 50, and 250 mm from the reference (i.e., glottal) plane. The smallest microphone spacing was thus 40 mm, which produces a singularity (Jang and Ih, 1998) at 4.3 kHz when it corresponds to a half wavelength. Spectra were measured over the frequency range 100–4000 Hz with a resolution of 2.69 Hz using a synthesized signal that distributed measurement errors equally with frequency. The non-resonant loads were an

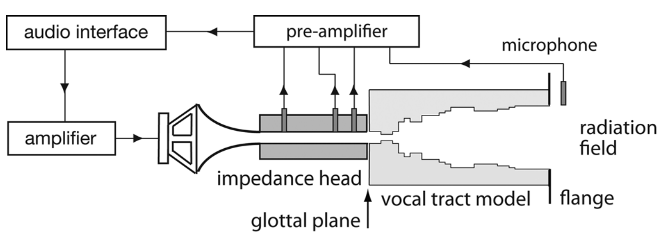

Fig. 1. Schematic diagram (not to scale) showing the 3-microphone impedance head connected to a physical vocal tract model in the plane of the glottis. The impedance head can first measure precisely the input impedance of the tract seen at the glottis, and then inject a known periodic glottal flow at the glottis. The pressure signal measured by the microphone in the radiation field can then be used to test various inverse filtering techniques. The loudspeaker and impedance head are well isolated from each other and the radiation field.

acoustically infinite waveguide, an open circuit, and a large baffle. Performance was checked by measuring cylindrical metal pipes of known geometry (Dickens *et al.*, 2007).

2.3 Generation of a known glottal flow waveform

The flow $U(t)$ at the reference (i.e., glottal) plane can be determined from the measured signals of the microphones in the impedance head. This involves determining the complex amplitudes of the traveling pressure waves in two opposite directions within the head [$p_{\text{left}}(x)$ and $p_{\text{right}}(x)$]; the acoustic flow to the right at the reference plane ($x = 0$) is then simply given by $[p_{\text{right}}(0) - p_{\text{left}}(0)]/Z_0$, where Z_0 is the characteristic impedance of the cylindrical duct.

The production of the desired flow signal at the reference plane involved an iterative, inversion technique similar to that used previously to generate a known acoustic flow (Smith *et al.*, 1997). The first step was to digitize the non-zero flow portion of a representative glottal waveform from Alku (1992) so that the open quotient could be varied by changing the length of the zero flow portion. Once a particular set of parameter values was chosen, the complex Fourier components $U(f)$ of this desired periodic glottal flow $U(t)$ were determined. Initially, a signal was synthesized with spectral components of magnitudes $U^*(f)$ proportional to $U(f)$. This was connected to the amplifier and the acoustically insulated loudspeaker, and the Fourier components $U'(f)$ of the flow $U'(t)$ at the reference plane were determined. A new signal was then synthesized with complex components proportional to $U^*(f)/U'(f)$ that should, in a completely linear system, produce the desired $U(t)$ at the reference plane. Sometimes additional iterations were required, due to small nonlinearities in the loudspeaker. The glottal waveform used had a fundamental frequency of 172 Hz ($= 44.1$ kHz/2^8) with harmonic components extending to 3.8 kHz (i.e., the 22nd harmonic).

2.4 Estimation of glottal flow and the resonances and bandwidths of the tract

An identical fourth microphone was placed near the output of the model tract to allow measurement of the sound in the radiation field—see Fig. 1. Acoustic measurements were made in a room insulated to reduce the influence of outside sound and treated to reduce reverberation.

Inverse filtering was performed using the program *Aparat* (Airas, 2008) to implement direct inverse filtering (DIF) and iterative adaptive inverse filtering (IAIF) (Alku, 1992). The sound, originally recorded at 44.1 kHz, was re-sampled at 8 kHz with the high-pass filter set to 150 Hz. Window lengths of 50 ms were used.

3. Results and discussion

To demonstrate this experimental approach to assessing the performance of inverse filtering, several different glottal flows with OQ ranging from 0.004 to 0.996 were synthesized and injected into a physical model of the vocal tract. The resultant pressure signal was then analyzed using DIF and IAIF, with the inverse filter model order ranging from 6 to 12.

3.1 Glottal flow derived using inverse filtering techniques

Figure 2 compares two examples of the glottal flow as estimated via inverse filtering from the output sound with the actual known flow. To make these comparisons, the scale of the estimated glottal flow was adjusted and shifted along the time axis to minimize the mean-square error from the injected glottal flow. As might be expected, the harmonic content of the sound cannot provide sufficient information to estimate the glottal flow exactly, the main difference being the presence of "ripples" at higher frequencies. Table 1 shows the mean-square error in the inverse filtered glottal flow in both the time and frequency domains. This process was repeated for different values of OQ, and different orders of the DIF and IAIF models. For DIF the errors were smallest when the order of the inverse filter model was 8 or 10. However for IAIF, the

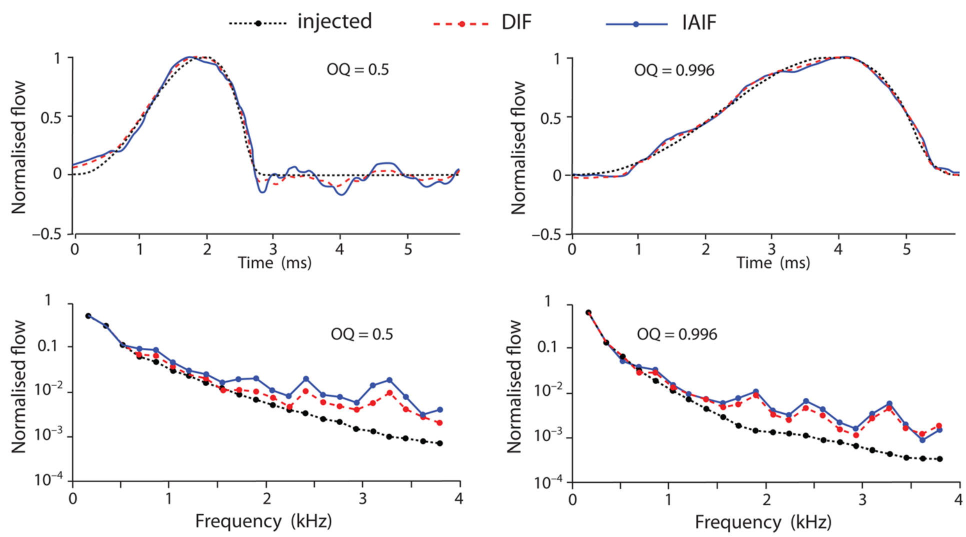

Fig. 2. (Color online) A comparison of glottal flow estimates made using direct inverse filtering (DIF) and iterative adaptive inverse filtering (IAIF) with the injected synthesized glottal flow. Results of the experiments are shown in the time domain (upper) and frequency domain (lower), for two flow waveforms with OQ = 0.5 (left) and OQ = 0.996 (right). Both used models of order 8.

errors were considerably smaller for a model of order 8 rather than 10. For both DIF and IAIF, the error generally increased when the model order was increased to 12. From the above results, it appears that, in this particular situation, DIF produced distinctly smaller mean-square errors than IAIF.

3.2 Resonance characteristics of the tract derived using inverse filtering techniques

Figure 3 shows the resonance characteristics of a physical model of the tract as determined by inverse filtering (IAIF) for a wide range of OQ values. The values calculated from precise measurements of the minima in the acoustic impedance are also included for comparison. When a model of order 8 was used, some resonances were completely missed or estimated poorly. Furthermore, the lowest resonance frequencies corresponding to $F1$

Table 1. Mean-square error, expressed as a percentage in the time and frequency domains, calculated between the synthesized glottal flow waveform and the estimates made via inverse filtering using DIF and IAIF. Synthesized flow waveforms with different open quotient values were used, with the model order ranging from 6 to 12.

	DIF				IAIF			
OQ	model order of 6	model order of 8	model order of 10	model order of 12	model order of 6	model order of 8	model order of 10	model order of 12
	Percent of error in time domain							
0.5	0.76	0.18	0.19	0.35	0.90	0.51	0.70	0.90
0.75	0.11	0.035	0.038	0.078	0.23	0.11	0.25	0.78
0.9	0.068	0.036	0.034	0.075	0.18	0.1	0.24	0.62
0.996	0.12	0.076	0.067	0.060	0.15	0.11	1.49	1.73
	Percent of error in frequency domain							
0.5	6.0	0.64	1.0	2.65	6.0	2.18	5.15	6.9
0.75	0.82	0.15	0.23	0.54	1.66	0.64	2.27	6.13
0.9	0.65	0.38	0.27	0.54	1.60	0.81	2.81	5.57
0.996	0.86	0.18	0.36	0.36	0.73	0.34	9.61	11.7

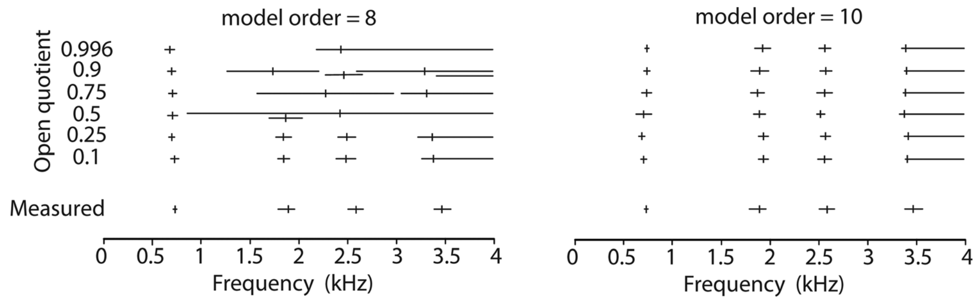

Fig. 3. The resonance characteristics of a physical tract as estimated by inverse filtering (IAIF), compared with those determined from precise measurements of the acoustic impedance. Results are shown for glottal flows with different open quotients. Horizontal lines are not error bars: they indicate the bandwidths of each resonance, while vertical lines indicate the resonance frequency. Some have been slightly displaced vertically for clarity.

are systematically underestimated (by an average of 28 Hz for $OQ \geq 0.1$), and their average bandwidths were approximately twice the correct values. The frequencies of the higher order resonances were also systematically underestimated. These errors are not surprising as the fundamental frequency was 172 Hz, and so errors of the order of $172/2 = 86$ Hz might be expected. When models of order 10 or 12 were used, there was much better agreement with the measured values. The wide estimates of the upper bandwidth for F4 in the results presumably occur because only frequencies below 3.8 kHz were analyzed. Not unexpectedly, there were very significant errors when a model of order 6 was used.

The presented technique of using an impedance head to inject a known periodic glottal flow has many possible applications, including the rigorous testing of different approaches to inverse filtering. It could be useful in studies employing excised larynxes. It could also be used to evaluate glottal flow masks with a wide range of known glottal waveforms.

Acknowledgment

The support of the Australian Research Council is gratefully acknowledged.

References and links

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

Alku, P. (**1992**). "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," Speech Commun. **11**, 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 **36**, 623–650.

Alku, P., Horácek, J., Airas, M., Griffond-Boitier, F., and Laukkanen, A.-M. (**2006a**). "Performance of glottal filtering as tested by aeroelastic modelling of phonation and FE modelling of vocal tract," Acta Acust. Acust. **92**, 717–724.

Alku, P., Story, B., and Airas, M. (**2006b**). "Estimation of the voice source from speech pressure signals: Evaluation of an inverse filtering technique using physical modelling of voice production," Folia Phoniatr. Logo. **58**, 102–113.

Dickens, P., Smith, J., and Wolfe, J. (**2007**). "High precision measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution," J. Acoust. Soc. Am. **121**, 1471–1481.

Drugman, T., Bozkurt, B., and Dutoit, T. (**2012**). "A comparative study of glottal source estimation techniques," Comput. Speech Language **26**, 120–134.

Jang, S.-H. and Ih, J.-G. (**1998**). "On the multiple microphone method for measuring in-duct acoustic properties in the presence of mean flow," J. Acoust. Soc. Am. **103**, 1520–1526.

Smith, J. R., Henrich, N., and Wolfe, J. (**1997**). "The acoustic impedance of the Bœhm flute: Standard and some non-standard fingerings," Proc. Inst. Acoust. **19**, 315–320.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**, 537–554.