CrossMark

An Experimentally Measured Source–Filter Model: Glottal Flow, Vocal Tract Gain and Output Sound from a Physical Model

Joe Wolfe[1] · Derek Tze Wei Chu[1] · Jer-Ming Chen[1] · John Smith[1]

Abstract A known acoustic flow is generated and input to a physical (hardware) model of a vocal tract at its 'glottis'. The output sound is measured, along with the gain of the tract (the transpedance or transimpedance: the ratio of output pressure to input flow). These allow all stages of the Source–Filter model to be illustrated by actual experimental measurements on the same physical model in the time and frequency domains, rather than by the usual qualitative, illustrative sketches.

1 Introduction

The Source–Filter model [1] treats the glottis as a source of harmonic frequency components for voiced speech or broadband sound for whispering. The vocal tract—the respiratory tract downstream from the glottis to the lips—is regarded as a filter, whose resonances produce broad spectral peaks or formants in the output sound. In the simplest version of the model, the source and filter are treated as being independent, although the glottal mechanism has been shown to affect the vocal tract resonance [3–5] and mathematical models of the vocal folds suggest that their motion is affected by the acoustical impedance of the tract [6,7]. The Source–Filter model is so fundamental to our understanding of speech that many books and didactic articles on speech science, and many in acoustics, include a qualitative, sketched, schematic representation (e.g. [6,8–10]). First, the flow signal at the glottis is represented in the frequency domain (and sometimes also the time domain). Then the gain spectrum of the vocal tract is shown. Finally the output sound spectrum is shown, showing broad peaks in the spectral envelope, or formants [2].[1]

There are good reasons why such figures in textbooks are qualitative, illustrative sketches, rather than experimental measurements. Because of the relative inaccessibility of the glottis and the difficulty of measuring small acoustic flows, the glottal flow cannot readily be measured in vivo. Instead, it is usually estimated from the pressure signal measured outside the mouth by adjusting the parameters of filters that simulate the inverse transfer function of an idealised vocal tract. For the same reason, the gain (usually displayed as the transpedance: output pressure to input flow) of the vocal tract cannot be directly measured, though it too can be estimated using inverse filtering (e.g. [11,12]).

The combination of a calibrated impedance head and a hardware model of the vocal tract can allow all the parameters of interest to be directly measured for the first time. The impedance head injects the desired glottal current [14]. The transpedance of the tract and the output sound can be

✉ Joe Wolfe
  j.wolfe@unsw.edu.au

[1] School of Physics, The University of New South Wales, Sydney, New South Wales 2052, Australia

---

[1] The American National Standard Acoustical Terminology [2] is followed here: the formant is a maximum in the spectral envelope of the sound. Some speech researchers use formant to mean the resonance that gives rise to the spectral maximum.
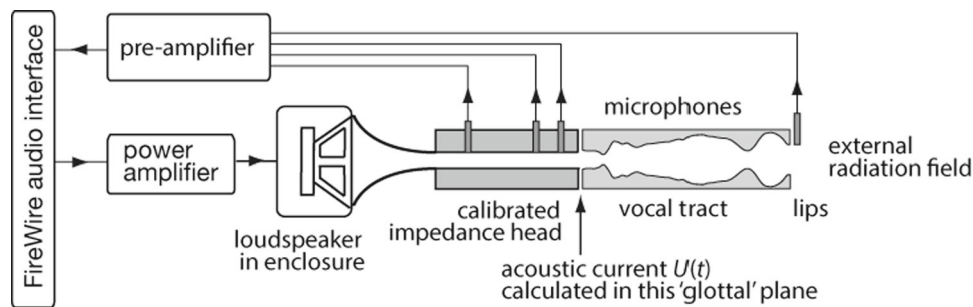
Fig. 1 Schematic diagram (not to scale) showing how the 3-microphone impedance head is used to generate the desired glottal flow. A fourth microphone, located at the edge of the 'lip' aperture, is used to measure the sound pressure at the lips and to measure the tract transpedance. For clarity, the impedance head, vocal tract and external microphone are shown slightly separated

measured using the same impedance head and an additional microphone at the output. The purpose of this tutorial paper is to present a set of experimental data that illustrate the simple Source–Filter model in the frequency and time domains, and which may be used as illustrations by acousticians and teachers.

2 Materials and Methods

The materials and methods are similar to those reported previously [14], except in this case the physical models of the tract were constructed out of PLA (a thermoplastic aliphatic polyester) using a 3D printer with an axially symmetric profile based on the open area functions $A(x)$ for the vowels /æ/ and /ɜ/ taken from a study based on MRI images [15]. It is worth noting that the resonances of this physical model are expected to have higher Q-factors than would be present in a human vocal tract, where the magnitude of the (complex) visco-thermal loss factor α is typically increased by a factor of five [16].

The three-microphone impedance head was calibrated with three non-resonant loads; an acoustically infinite waveguide, an open circuit and a large baffle. This technique allows high precision and dynamic range over a wide frequency range [17]. The microphones (4944-A, Brüel and Kjær, Denmark) in the impedance head were located 10, 50 and 250 mm from the reference plane, which in this case is the 'glottis' of the model tract. This gives a smallest microphone spacing of 40 mm, which imposes an upper limit around 4 kHz to the useful frequency range. The microphone at the lips was located immediately outside the vocal tract model at the edge of the 'lip' aperture. Fig. 1 gives a schematic of the apparatus. Their signals are passed to the computer via a conditioning preamplifier (Nexus 2690, Brüel & Kjær, Denmark) and a FireWire audio interface (MOTU 828, Cambridge, USA).

The flow $U(t)$ at the reference plane of the impedance head is, by continuity, the flow into the tract at the 'glot-

tis'. Pressure measurements at the three microphones in the impedance head are used to calculate the (complex) amplitudes of the travelling pressure waves in the head: $p_{left}(x)$ and $p_{right}(x)$. The acoustic flow at the reference plane ($x = 0$) is thus given by $U(t) = (p_{right}(0) - p_{left}(0))/Z_0$, where Z_0 is the characteristic impedance of the cylindrical duct calculated from its cross-sectional area.

The $U(t)$ chosen closely approximates a typical glottal waveform calculated by [12] from inverse filtering with a glottal contact quotient (CQ) of 0.5. (CQ = the ratio of the glottal contact duration to the period of vibration.) It has a fundamental frequency of 172 Hz (equal to the sampling frequency divided by an integral power of 2; 172 Hz = 44.1 kHz/2^8). This value is intermediate to the typical values of male and female speakers; it allows harmonics up to the 22nd harmonic (3.8 kHz). An inversion technique reported earlier [18] is used to produce the desired $U(t)$ with Fourier transform $U(f)$. Thus a voltage signal proportional to $U(t)$ is synthesised and output to the amplifier and loudspeaker. The resultant flow $U_0(t)$ at $x = 0$ is then determined as described above and the Fourier transform $U_0(f)$ is calculated. A new waveform, with Fourier components $U(f)/U_0(f)$, is then synthesised and output to the amplifier and loudspeaker. In a completely linear system, this would produce $U(t)$ at the 'glottis'. In practice, the loudspeaker is not quite linear, so this procedure is iterated to produce the desired $U(t)$ [18,19].

The transpedance $T(f)(= p_{lips}(f)/U(f))$ of the tract was then calculated from the ratio of the pressure measured at the 'lips' at the far end of the tract to the flow injected at the glottis.

In this project, a detailed spectrum of the tract transpedance was required for illustrative purposes. However, the signal that generated the acoustic current is periodic with a fundamental frequency of 172 Hz; thus only 22 frequency components with a resolution of 172 Hz were below the 3.8 kHz limit of these measurements. Consequently, an additional separate measurement was made with greater frequency resolution using the impedance head, in conjunction with the microphone at the 'lips'; this used a broadband sig-
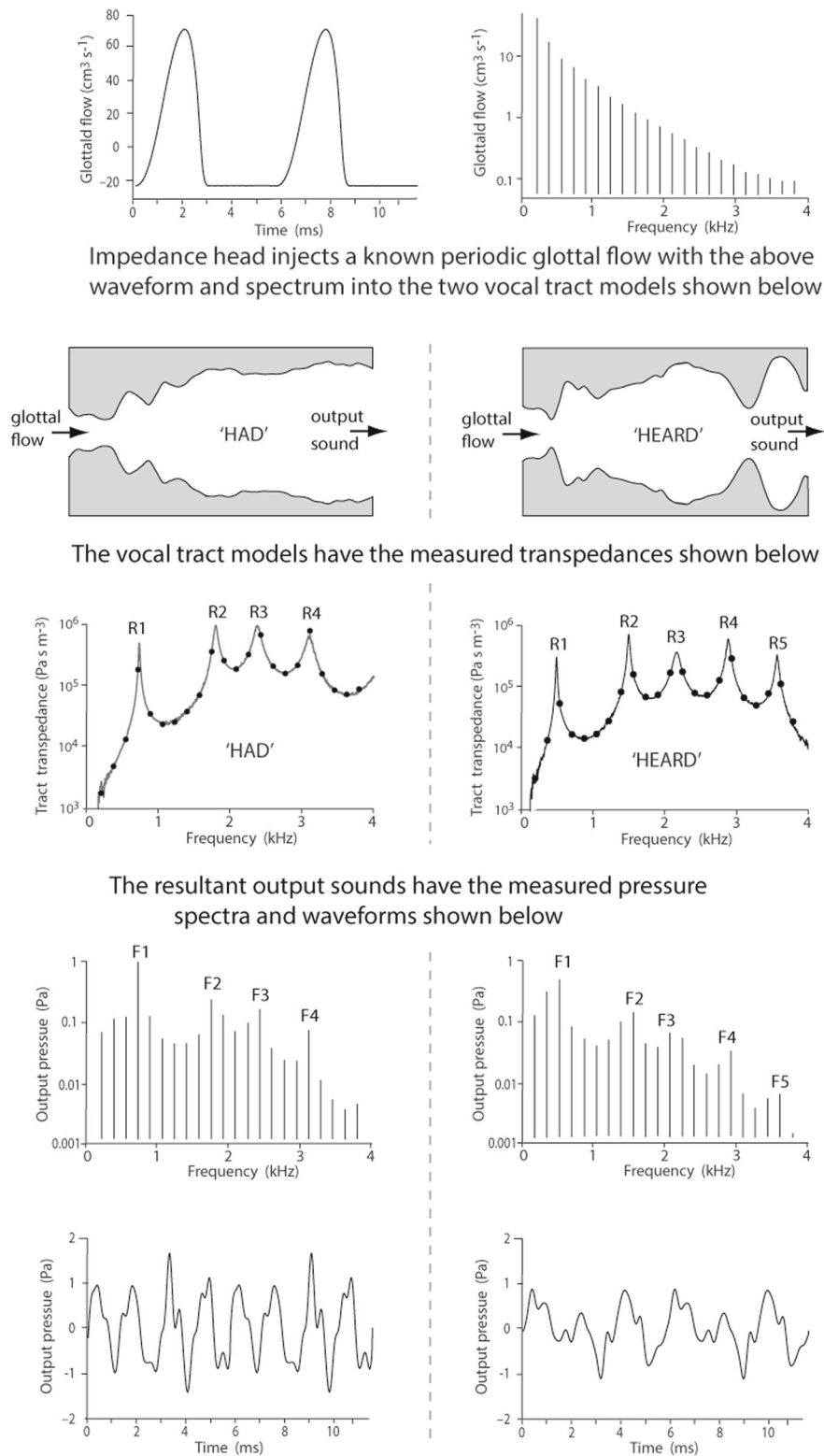
Fig. 2 The effect of different vocal tract shapes and transpedances on the measured output sound for two physical models of the tract with area functions corresponding to the vowels in 'had' and 'heard'. The vocal tract models on the second line show how their radius varied with distance along the tract [15]. The continuous curve for the tract transpedance displays the measurements made with 2.69 Hz resolution, whilst the black dots show the measurements made with the 172 Hz signal used to inject the glottal flow. The maxima in the transpedance that are associated with the vocal tract resonances are labelled R1–R5. The resultant maxima in the envelope of the measured output sound, i.e. the speech formants, are similarly labelled F1–F5. Two complete cycles are shown for the glottal flow and the output sounds

nal covering 100 to 3800 Hz with a resolution of 2.69 Hz ($=44.1$ kHz/2^{14}). This separate measurement also allows the injected current for the measurement of $T(f)$ to have a waveform that is different from the glottal flow and that can then be optimised to improve the distribution of measurement errors over the frequency range [17].

All measurements were made in a room insulated to reduce the influence of outside sound and treated to reduce reverberation.

3 Results and Discussion

The results of measuring the Source–Filter model are shown in Fig. 2: the glottal flow in both time and frequency domains, transpedance of the tract and sound output in time and frequency.[2] The waveforms and spectra are all experimental measurements made on a physical hardware model of the tract that is an axially symmetric representation of the open area function $A(x)$ measured using MRI [15].

The top row shows the time and frequency domain representations of the glottal flow synthesised (as described above) to match that calculated by [12]. Because of the rapid closure of the glottis, the flow signal is clipped at its lowest level, which produces the spectrum rich in harmonics.

The second row shows a cross section through the two vocal tract models for the vowels in 'had' and 'heard' based on [15].

The third row shows the magnitude of the measured transpedance $T(f)$ of the model tract: the ratio of the output sound pressure to the input acoustic flow. The continuous curve displays the magnitude of the transpedance measured with a resolution of 2.69 Hz; the black dots indicate the measurements of transpedance made with the 172 Hz signal. Four and five resonances occur within the experimental frequency range for the vowels in 'had' and 'heard', respectively.

The fourth row shows the frequency spectra of the sound output measured at the lips from the physical models. They show four or five strong formants with frequencies approximately equal to those of the resonances shown in the respective transpedance in the third row.

The fifth row shows two cycles of the measured sound output with the same period as the glottal flow shown above.

In these experiments, the impedance head was used first to inject a glottal current, and subsequently, several hours later, to measure the transpedance at higher resolution. The consistency of these two separate measurements was examined by comparing the measured $p_{lips}(f)$ in response to the injected $U(f)$ with that predicted from $U(f)/T(f)$ using the values of $T(f)$ with 2.69 Hz resolution. The average RMS differ-

ence between the two was 1.7 % and 0.9 % for *had* and *heard*, respectively (-35 and -41 dB, respectively); this difference was probably a consequence of small shifts in resonance frequencies caused by changes in the room temperature.

4 Conclusions

Experimental measurements on a hardware models illustrate the Source–Filter model by showing the glottal flow, the tract transpedance and the sound at the lips, in the frequency and time domain (Fig. 2).

References

1. Fant, G.: Acoustic Theory of Speech Production, p. 15. Mouton, The Hague (1970)
2. Acoustical Society of America: ANSI S1.1-1994, American National Standard Acoustical Terminology. Acoust. Soc. Am., Melville (2004)
3. Klatt, D.H., Klatt, L.C.: Analysis, synthesis, and perception of voice quality variations among female and male talkers. J. Acoust. Soc. Am. **87**, 820–857 (1990)
4. Barney, A., De Stefano, A., Henrich, N.: The effect of glottal opening on the acoustic response of the vocal tract. Acta Acust. United Acust. **93**, 1046–1056 (2007)
5. Swerdlin, Y., Smith, J., Wolfe, J.: The effect of whisper and creak vocal mechanisms on vocal tract resonances. J. Acoust. Soc. Am. **127**, 2590–2598 (2010)
6. Titze, I.R.: Principles of Voice Production. National Center for Voice and Speech, Iowa (2000)
7. Titze, I.: Nonlinear source-filter coupling in phonation: theory. J. Acoust. Soc. Am. **123**, 2733–2749 (2008)
8. Rossing, T.D.: The Science of Sound. Addison-Wesley, Reading (1983)
9. Lindblom, B., Sundberg, J.: The human voice in speech and singing. In: Rossing, T.D. (ed.) Springer Handbook of Acoustics, p. 682. Springer, New York (2007)
10. Clark, J., Yallop, C., Fletcher, J.: An Introduction to Phonetics and Phonology. Blackwell, Malden (2007)
11. Makhoul, J.: Linear prediction: a tutorial review. Proc. IEEE **63**, 561–580 (1975)
12. Alku, P.: Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Speech Commun. **11**, 109–118 (1992)
13. Music Acoustics: An experimentally measured source filter model. http://newt.phys.unsw.edu.au/jw/measured-source-filter.html. Accessed 3 Dec 2015
14. Chu, D.T.W., Li, K.-W., Epps, J., Smith, J., Wolfe, J.: Experimental evaluation of inverse filtering using physical systems with known glottal flow and tract characteristics. J. Acoust. Soc. Am **133**, EL358–EL362 (2013)
15. Story, B.H., Titze, I.R., Hoffman, E.A.: Vocal tract area functions from magnetic resonance imaging. J. Acoust. Soc. Am. **100**, 537–554 (1996)
16. Hanna, N., Smith, J., Wolfe, J.: Low frequency response of the vocal tract: acoustic and mechanical resonances and their losses. In: McGinn, T. (ed.) Acoustics 2012 Fremantle: Proceedings of

---

[2] For the use of authors and teachers, various versions of Fig. 2 are freely available via the UNSW Acoustics website [13].

Annual Conference of the Australian Acoustical Society, Hobart (AAS2012) (2012)

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

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

19. Smith, J., Chu, D., Wolfe, J.: You can't measure it but you can know it: precisely synthesising acoustic flow. Acoustics 2015, Hunter Valley, paper 72 (2015)