OPEN-SOURCE SOFTWARE FOR ESTIMATING VOCAL TRACT RESONANCES USING BROADBAND EXCITATION AT THE LIPS

Marie Jeanneteau, Noel Hanna*, André Almeida, John Smith, and Joe Wolfe

School of Physics, University of New South Wales
n.hanna@unswalumni.com

ABSTRACT

Broadband excitation of the vocal tract at the lips (Epps, Smith & Wolfe, 1997. Meas. Sci. Technol. 8, 1112–1121 [1]) has been used by the present authors and in other laboratories to estimate the resonances of the vocal tract in 'ecological' conditions of speaking and singing. A sound source and microphone are both held at the lower lip of the subject, then the ratio of microphone pressure with lips open to that with lips closed is measured in response to a broadband excitation signal. The peaks of this ratio provide an estimate of the resonances Ri of the vocal tract, without disturbing the subject's vocal gesture. Here we present an updated open-source implementation of this technique written in Matlab that allows real-time measurements with inexpensive equipment. This simplifies use of the technique and allows some new applications including measurements during respiration, whisper and breathy speech.

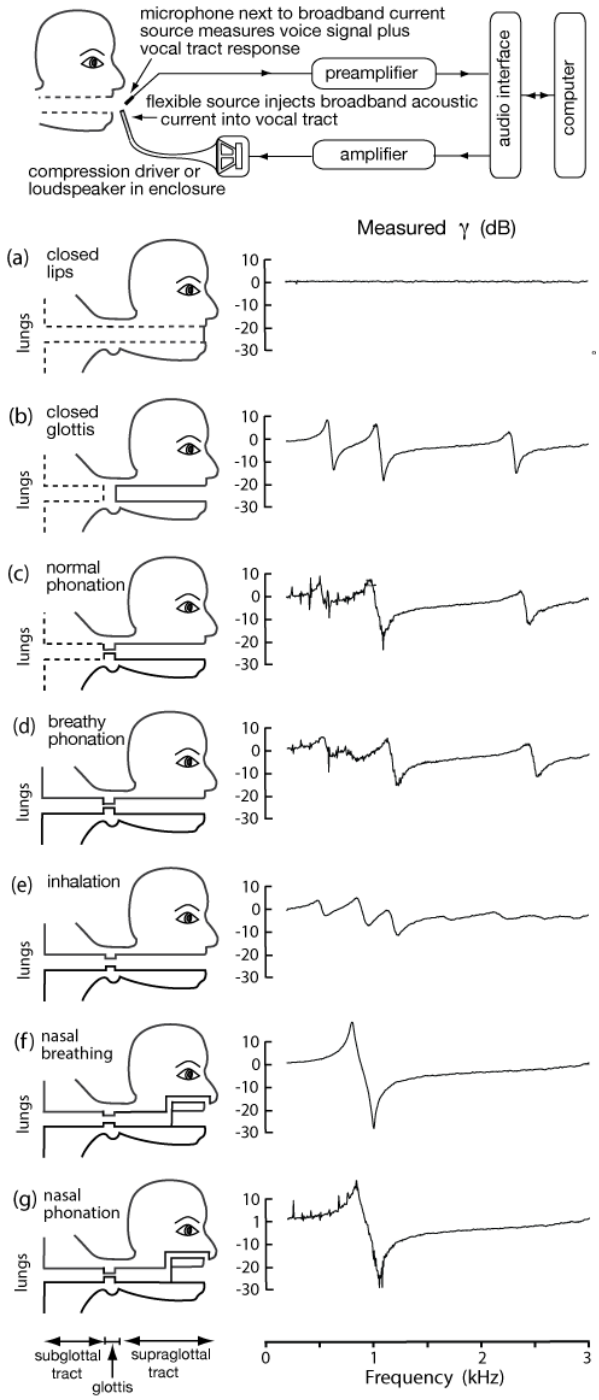
Keywords: Open-source Software, Resonance, Formant, Vocal Tract, Acoustic Impedance.

1. INTRODUCTION

Excitation of the vocal tract with a broadband source at the lips [1] has been used by several groups to estimate the resonances of the vocal tract in 'ecological' conditions of speaking and singing. Measuring the tract resonances, rather than the formants [2] they usually produce, has the advantage that resonances can be measured with greater precision in most situations.

In this technique, a small broadband source of acoustic current and adjacent small microphone are placed at the lips of the subject, as shown in Figure 1. Suitably calibrated, the combination comprises a simple impedance head. With the lips open, the vocal tract and the radiation field in parallel form the load for the acoustic current source.

Figure 1: The acoustic source and a microphone are placed at the speaker's lower lip. The microphone pressure p_{closed} is measured with the lips closed then open p_{open}. The ratio $\gamma = p_{open}/p_{closed}$ is shown for the same male subject with tract configurations (a-f).



Using a flexible tube connected to the loudspeaker allows the sound source to remain in the same position at the subject's lip. This allows speakers or singers to produce reasonably natural vocal gestures with little disturbance.

In the usual mode of operation, the source and microphone are first placed at the subject's lower lip with the lips closed. The 'impedance head' is thus loaded with the radiation impedance Z_{rad} at that position, baffled by the face. The spectrum of the source signal is then usually adjusted so that the measured acoustic pressure p_{closed} is approximately independent of frequency with this load. Applying the same output signal to the source, the acoustic pressure p_{open} is measured with the lips open, and the 'impedance head' is now loaded with the vocal tract impedance at the lips Z_{in} in parallel with the radiation impedance.

Usually, the quantity recorded and displayed is the magnitude of the ratio $\gamma = p_{open}/p_{closed}$. Making the approximation that the source has a very high output impedance (an ideal current source), the output current is the same in the open and closed conditions. It follows that

$$\gamma \cong (Z_{in}||\mathrm{Z_{rad}})/Z_{rad} = Z_{in}/(Z_{rad} + Z_{in}) \quad (1)$$

This ratio has maxima when $Z_{in} \cong -Z_{rad}$.

The limited capabilities of the 20th century computers originally used (16 MHz, 4 Mb RAM) required two computers to operate in 'real-time' [1,3] for language training. However, for many studies 'real-time' response is not necessary and delayed display from a single computer was adequate for studies of speech [*e.g.* 4] and singing [*e.g.* 5-8]. Others have implemented this technique, sometimes using software derived from that by this lab [e.g. 9-11], and/or different hardware [*e.g.* 12-16].

Here we provide a stand-alone, open-source graphical user interface (GUI) for the technique [17,18], which displays the impedance ratio γ in real-time with modest hardware requirements. It may be used and modified for the purposes of research into the acoustics of speech and singing.

2. MATERIALS AND METHODS

2.1. Hardware requirements

For the measurements made here, research equipment is used: both output and input pass via a Firewire interface (MOTU 828, Cambridge, MA). The microphone (DeltaTron Pressure-field 1/4" Type 4944A; Bruel & Kjaer, Denmark) is connected via a conditioning amplifier (Bruel & Kjaer, Denmark) to the interface. A second channel is available for inputs from an electroglottograph, accelerometer or other devices, but these were not used in the results shown here. The output signal is delivered to a 20 W power amplifier (constructed in-house) and a 60 W, 1" compression driver (BM-D450 MKII, P.Audio, Nakhon, Thailand). It is matched via a conical horn to a 1.0 m length of flexible hose with internal diameter 5 mm.

However, for many studies, inexpensive elements are quite adequate. We have also used a cheap, tie-clip microphone (Optimus 33-3013, Tandy, Melbourne, Australia) connected to the sound card of a computer via the microphone input and a USB powered loudspeaker, placed close to the lips, for sound output.

2.2. Software requirements

The source Matlab files are available for download at [17]. These require Matlab 2014b or higher and have been developed and tested on Mac OS 10.12 and Windows 7 and above. To use the software, copy the directory to your chosen location, then 'Add to path, folders and Subfolders' in Matlab. The software can be started by typing GUI_RAVE into the command window.

Alternatively, a precompiled version can be downloaded from [18]. This version can be installed as a Mac app or Windows program, and does not require Matlab to be installed on the computer. Once the file is downloaded, double click the installer icon to begin the guided installation process.

2.3. Graphical user interface (GUI)

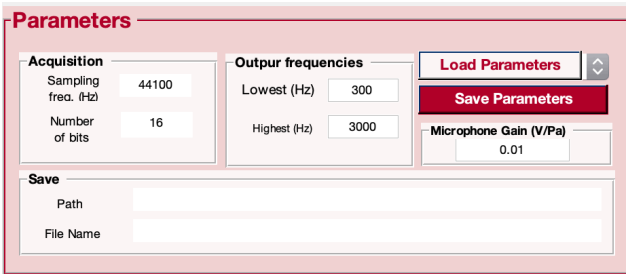
The GUI comprises several windows. Screen shots of some of these are displayed in the sections below as we list the protocol.

Parameter screen (Figure 2). The initial parameters, including sampling frequency, desired frequency range and file location, are selected here and may be saved for subsequent re-use.

Calibration screen (Figure 3). First, the audio input and output must be selected from the available devices. Then the speed and frequency resolution of the broadband probe signal is chosen. The signal is synthesised as a sum of sine waves. Their frequencies are high harmonic

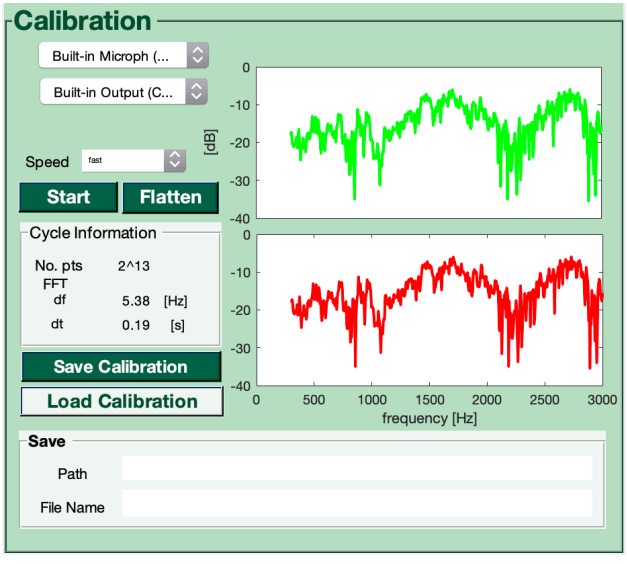
multiples of a missing fundamental at $f_s/2^n$, where f_s is the sampling frequency (44.1 kHz) and n is an integer in the range 12 to 16, so that the measured frequency resolution is sampled at spacings between 0.67 and 11 Hz. Each cycle, which is also a sample window, therefore has a length $2^n/f_s$ *i.e.* 1.5 to 0.093 s, respectively.

Figure 2: Parameter selection window.



The calibration process is then started by holding the microphone and source against the subject's lower lip with the lips closed. Pressing 'Start' (Figure 3) outputs the initial, flat, uncorrected broadband signal and measures the frequency response of the whole system, including the radiation impedance.

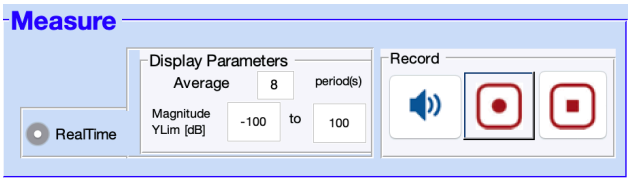
Figure 3: Calibration window. Shows the initial microphone signal (red) and the second channel (green) (a second microphone in this example).



The 'Flatten' process adjusts the output signal iteratively until the microphone signal for the calibration condition (usually lips closed) is independent of frequency to a desired tolerance. This optional process has the advantage of putting more power into frequency bands having low system gain. In each iteration, the phases of the sine waves are adjusted to improve signal-to-noise ratio [19]. Figure 3 shows the initial gain of the system in red, which in this case is weak at low frequencies (due to the radiation impedance and compression driver), and has resonances spaced at 170 Hz (due to the pipe).

Measure screen (Figure 4). Microphone recording is started by pressing the (red) 'Record' button. The broadband probe signal may be turned on or off at any time by toggling the 'Loudspeaker' button. 'Stop' ends both the recording and source signals.

Figure 4: Measure window.



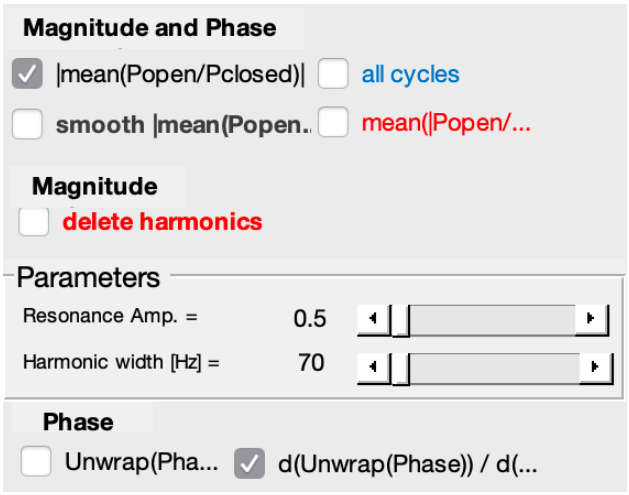
During recording, the time domain waveform of the two recording channels, microphone and optional additional sensor, and the frequency spectrum of the microphone are displayed in real time. The spectrum may optionally be averaged over a chosen number of cycles—this improves the signal:noise ratio, at the expense of time resolution and response time. However, this only affects the displayed spectrum and does not change the recorded signal.

Further options (see Figure 5) are to smooth the spectrum and/or to remove periodic signals, such as the voice fundamental frequency f_0 and its harmonics (Figure 6). The 'Harmonic width' parameter determines the bandwidth (width on the spectrum) of peaks to be removed. 'Resonance Amp' is a threshold of relative change in magnitude (height on the spectrum) to consider a spectral feature as a resonance.

2.4. Playback and analysis

For retrospective use, the waveforms are displayed in a selectable window (not shown). Users can scroll through the waveform to choose individual cycles, or the start and end points of several cycles to display an average γ magnitude spectrum. After selecting a start and end point, clicking 'Save' will save the selected data in a .mat file with .wav files for each recording channel. By clicking on a γ curve, it is saved in memory and may be toggled on or off, to compare it with other parts of the recording.

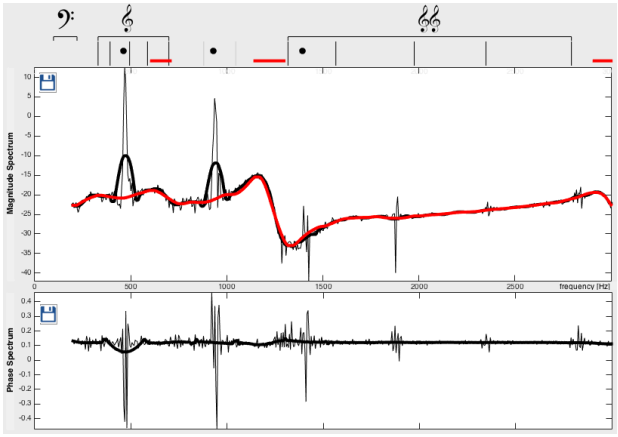
Figure 5: Display adjustments.



3. EXAMPLE RESULTS

The example measurements in Figure 1 illustrate a range of respiratory and phonatory gestures. After an earlier calibration with closed lips, a subsequent measurement with lips closed (a) shows the ratio $\gamma = p_{closed}/p_{closed} \sim 1$. (b) shows the magnitude of γ of the vocal tract, alone with the glottis closed, as seen in mimed vowel gestures. During normal phonation (c), the glottis is small enough that the inertance of air in the opening seals the tract at high frequencies. So, the γ curve for normal phonation is largely similar to glottis closed (b) but with the addition of the periodic voiced signal, which can be removed (as shown in Figure 6). (d) shows a very breathy voice with incomplete glottal closure, in which an additional resonance due to the combined length of the subglottal and supraglottal tracts is visible at approximately 600 Hz. In (e), the subject has opened the glottis widely during inhalation, so that the vocal tract and trachea are acoustically connected, making a duct almost twice as long as the vocal tract. In this case, γ shows about twice as many resonances, though these are weaker due to greater losses (details in [20]). In (f), the velum is lowered, so the impedance head at the lips 'sees' only a short buccal tract, with behaviour between that of a short waveguide and a Helmholtz resonator: it shows a single strong resonance in the frequency range measured. The addition of the periodic voiced signal is evident in (g).

Figure 6: The measured magnitude and phase of the impedance ratio γ when a woman singer produces f_0 at A#4 (466 Hz). The raw γ signal (pale) is shown along with a smoothed version (black) and with harmonics removed (red). Black circles above the plot indicate detected harmonics, and horizontal red bars show resonances that meet the 'Resonance Amp' Criteria. (The bass, treble and super-treble clefs show their pitches, which is useful in singing applications.) In this case, the singer is not tuning either tract resonance.



4. ADVANTAGES AND LIMITATIONS

This technique has the advantage that the frequencies of resonances of the vocal tract can be estimated with greater precision than can usually be achieved from formants using microphone recordings. It allows reasonably normal speech and singing gestures. However, the vocal tract is measured in parallel with the (low impedance) radiation field; consequently the magnitudes and bandwidths of the resonances cannot be determined. Further, the approximations made in deriving (1) are more severe at high frequencies, so caution should be used interpreting results above about 2 kHz.

These limitations do not affect measurements made with the lips sealed on an impedance head [*e.g.* 20,21]. Such measurements can also be used to estimate vocal tract shape [22] but these require that the lip geometry is constrained.

The software described here is now freely available for non-commercial use [17,18].

5. ACKNOWLDEGEMENTS

We thank the Australian Research Council for its support and our volunteer subjects.

6. REFERENCES

[1] Epps, J., Smith, J., Wolfe, J. 1997. A novel instrument to measure acoustic resonances of the vocal tract during speech. *Meas Sci Technol* 8, 1112–1121.

[2] ANSI. 2004. ANSI S1.1-1994, Am Nat Standard Acoust. Term. Acoust. Soc.Am., Melville, NY.

[3] Dowd, A., Smith, J.R., Wolfe, J. 1998. Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time. *Language and Speech* 41, 1-20.

[4] Swerdlin, Y., Smith, J., Wolfe, J. 2010. The effect of whisper and creak vocal mechanisms on vocal tract resonances. *J. Acoust. Soc. Am*. 127, 2590-2598.

[5] Joliveau, E., Smith, J., Wolfe, J. 2004. Tuning of vocal tract resonances by sopranos. *Nature*, 427, 116.

[6] Garnier, M., Henrich, N., Smith, J., Wolfe, J. 2010. Vocal tract adjustments in the high soprano range. *J. Acoust. Soc. Am*. 127, 3771-3780.

[7] Henrich, N., Smith, J., Wolfe, J. 2011. Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. *J. Acoust. Soc. Am*. 129, 1024-1035.

[8] Henrich Bernadoni, N., Smith, J. and Wolfe, J. 2014. Vocal tract resonances in singing: variation with laryngeal mechanism for male operatic singers. *J. Acoust. Soc. Am*. 135, 491-501.

[9] Bourne, T., Garnier, M. 2012. Physiological and acoustic characteristics of the female theatre voice. *J. Acoust. Soc. Am*. 131, 1586-1594.

[10] Vos, R.R., Daffern, H., Howard, D.M. 2017. Resonance tuning in three girl choristers. *J.Voice* 31, 122.e1-122.e7

[11] Balamurali, B T, Jer-Ming Chen, J–M. 2018. Automated Classification of Vowel-Gesture. Interspeech 2018, Hyderabad.

[12] Kob, M., Neuschaefer-Rube, C. 2002. A method for measurement of the vocal tract impedance at the mouth. *Med. Eng. & Physics.* 24, 467-471.

[13] Agerkvist, F. T., Selamtzis, A. 2011. The effect of vocal tract impedance on the vocal folds. Proc. 9th Pan European Voice Conference.

[14] Ahmadi, F., Mcloughlin, I. 2012. Measuring resonances of the vocal tract using frequency sweeps at the lips. *Communications Control and Signal Processing (ISCCSP), 2012 5th Intl. Symp.,* 1-5.

[15] Uezu, Y., Kaburagi, T. 2016. A measurement study on voice instabilities during modal-falsetto register transition. *Acoust. Sci. and Tech.* 37, 267-276.

[16] Graf, S., Schwiebacher, J., Richter, L., Buchberger, M., Adachi, S., Mastnak, W., Hoyer, P. 2018. Adjustment of Vocal Tract Shape via Biofeedback: Influence on Vowels. *J. Voice* 2018.

[17] Jeanneteau, M., Hanna, N., Almeida, A., Smith, J., Wolfe, J. 2019. noelhanna/real-time-resonance-estimation: International Congress of Phonetic Sciences Version. doi:10.5281/zenodo.2616871

[18] Vocal resonances and broad band excitation http://newt.phys.unsw.edu.au/jw/broadband.html

[19] Smith, J. R. 1995. Phasing of harmonic components to optimize measured signal-to-noise ratios of transfer functions. *Meas. Sci. Technol*. 6, 1343– 1348.

[20] Hanna, N., Smith J., Wolfe, J. 2018. How the acoustic resonances of the subglottal tract affect the impedance spectrum measured through the lips. *J. Acoust. Soc. Am*. 143, 2639-2650.

[21] Hanna, N., Smith J., Wolfe, J. 2016. Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation. *J. Acoust. Soc. Am*. 139, 2924–2936.

[22] Rodriguez, A., Hanna, N., Almeida, A., Smith, J., Wolfe, J. 2018. Estimation of vocal tract and trachea area functions from impedance spectra measured through the lips. *Australasian Intl. Conf. Speech Science and Technology*, 77-80.