MEASURING RESONANCES OF THE VOCAL TRACT USING FREQUENCY SWEEPS AT THE LIPS

F. Ahmadi and I. V. McLoughlin

School of Computer Engineering
Nanyang Technological University
Nanyang Avenue, Singapore 639798

ABSTRACT

Precise measurement of the resonances of the human vocal tract is important in the research of acoustical phonetics. It has also significant applications in speech therapy and language learning – providing feedback about the shape of the vocal tract and position of the tongue. This paper investigates a novel method of measuring these resonances using linear frequency sweeps at the lips. To investigate the effectiveness of the method, tests have been completed on constructed tube models of the vocal tract and also human subjects for six English vowels. The precision of the measurement in the current implementation is shown to be superior compared to traditional electro-larynx method.

Index Terms— Speech, vocal tract, Kelly-Lochbaum

1. INTRODUCTION

Formant frequencies are peaks of the spectrum of the vocal tract. This spectrum is being sampled with harmonics of the pitch frequency when a voiced phoneme of speech is generated. If this phoneme is used to estimate the formants, the precision of estimating cannot be very much better than the harmonic spacing of the pitch. The lack of precision becomes more significant, particularly when the pitch frequency is comparable to or greater than the resonance frequency of interest. Consequently, it is more challenging to determine the formants of high-pitched voices (such as children and some women). Techniques for measuring vocal tract resonances can be classified in four groups: i) Estimation from formants of normal speech (e.g. linear prediction [1]) which has the short comings mentioned above, ii) estimation from whispered speech [2] which is intrinsically noise excited and the resulting formants become noisy, iii) estimation using an external source at the glottis [3] which needs three to four times acoustic power of the speech signal and makes the subject uncomfortable and iv) using an external source at the lips which is more accessible and provides better precision [4]. Referring to the theorem by Epps et. al. [5], resonances of the vocal tract can be measured as the peaks of the vocal tracts impedance Z_{VT}, measured at the lips. Vocal tracts impedance is in parallel the external radiation impedance Z_ϵ which is defined as:

$$Z_\epsilon = az\frac{jkr}{1+jkr} \tag{1}$$

with k being the wavenumber $k = 2\pi f/c$, c the speed of the sound, f the frequency, r the radial distance, z the specific acoustic impedance and a being a geometrical constant [6]. In the current work $f \leq 3.5\,\text{kHz}$ and r is several millimetres (specifically, it is the radial distance between the lips and the microphone). This means that $kr << 1$ in eqn. 1 and consequently, $Z_\epsilon \approx jkraz$. The sound source at this research derives the external radiation impedance Z_ϵ and Z_{VT} in parallel.

$$Z_{\|} = \frac{1}{1/Z_{VT} + 1/Z_\epsilon} \tag{2}$$

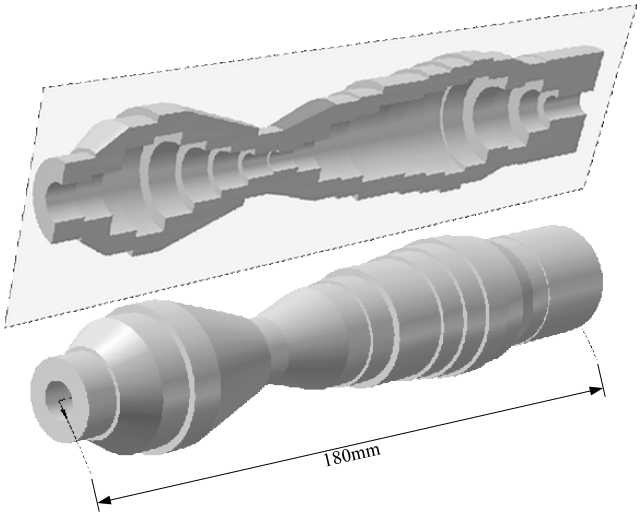

Fig. 1. A 22-tube Solidworks model and cutaway representation for vowel /u/, constructed to match the area functions of [7].

In eqn. 1, Z_ϵ increases only linearly with frequency (as endorsed by $Z_\epsilon \approx jkraz$) but there are relatively strong resonances in Z_{VT} over the frequency range of application. Con-

sequently while the radiation load at the lips can still be considered small, $Z_{\|}$ will have the same resonances as Z_{VT}. Since the sound source is located near to the lips, it can be approximated as an ideal current source independent of the load with velocity v. Consequently, the measured pressure response p by the microphones will have the same resonances as $Z_{\|} = p/v$.

Using the above theorem, the present work uses frequency sweeps for measuring resonances of the VT at the lips of human subjects and also constructed plastic models of human vocal tract. When human vocal tract is excited by chirp pulse train, the exact value of its resonances is not known and is estimated by harmonics of the pitch pulse. Consequently, the advantage of using models of VT in the measurements is that, the resonances of the model are determined using simulations and this provides a better evaluation of the measurement results. Nevertheless, the method is also tested on real vocal tract measurements.

2. VT MODELLING

To verify the approach of measuring resonances of the VT at the lips the present work used simulations of models of human vocal tract. Later, the models were constructed using epoxy material and measurements of the resonances were compared with the simulation results.

Perhaps the most widely used model of vocal tract is the Kelly-Lochbaum model which is a cascade of uniform tubes having similar length and different cross sectional areas. The magentic resonance imaging (MRI) data used in this paper was derived from the work of Story et. al. [7] in terms of area functions, and then converted using MATLAB to a 22-order interconnected tube model. Six models were constructed, for six different phonemes. The vocal tract models were then simulated using the Comsol multi-physics simulator to determine resonances in their frequency response when excited at the lips (glottis closed), using eigenfrequency analysis.

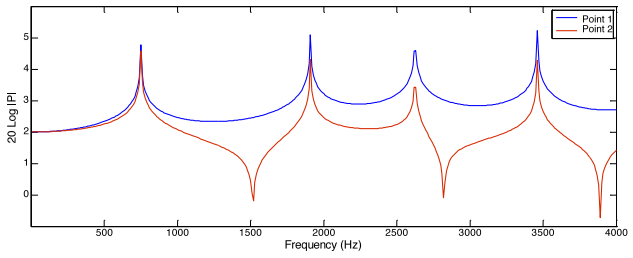

Fig. 2. Logarithmic values of pressure in a 22 tube model of vocal tract for vowel /æ/ at two different recording points located directly in front of the mouth.

One of the 22-tube models is shown in Fig. 1 as a three dimensional representation and a cut-away cross section. In these models, the impedance values of the VT walls have been set to match the mean acoustic impedance of soft tissues in the human body (approximately 1.7×10^6 Rayl). Table 1 shows the resulting formants obtained by an eigenfrequency analysis of the model, compared with previously published values [7] for six simulated English vowels /ɔ, æ, ɛ, u, i, o/ (as in the spoken RP words *paw, had, head, who, heed* and *hoe* respectively). Fig. 2 shows the spectral envelope at the output of the 22 tube model for the vowel /æ/, as measured at two alternative sites in front of the lips, showing clear resonances at approx. 750 Hz, 1.9 kHz, 2.6 kHz and 3.5 kHz corresponding to the first four formants. The results in table 1 only record F1, F2 and F3, but demonstrate the precision of the finite element model (FEM) implemented in this research. In all but two cases, the expected measurements are within 16% of the values reported by Story et. al. [7].

Table 1. Frequency of first three formants for six common vowels derived through Comsol eigenfrequency analysis of our 22 tube model, compared with the simulation and measured naturally phonated results of simulations [7], showing the minimum percentage correspondence between each.

Vowel	Model Hz	Story (sim) Hz [7]	Story (nat) Hz [7]	Min. diff %
/ɔ/	646	618	654	1
	1062	958	944	11
	2477	2195	2739	10
/æ/	750	732	692	2
	1908	1689	1873	2
	2625	2370	2463	7
/ɛ/	645	621	624	3
	2077	1795	1853	12
	2751	2436	2475	11
/u/	257	356	389	28
	1194	1108	987	8
	2393	2334	2299	3
/i/	225	337	333	32
	2447	2340	2332	5
	3463	3158	2986	10
/o/	388	461	540	16
	874	861	922	2
	2435	2217	2584	6

3. VT MEASUREMENT AT THE LIPS

3.1. The hardware

Fig. 3 shows the acoustic hardware setup of the system: a linear chirp is generated using a computer-driven signal generator board (DT Translations 9836). The signal is transferred to a 32Ω 0.1 W flat speaker to act as an ideal acoustic current source located close to the tube model opening (or the mouth

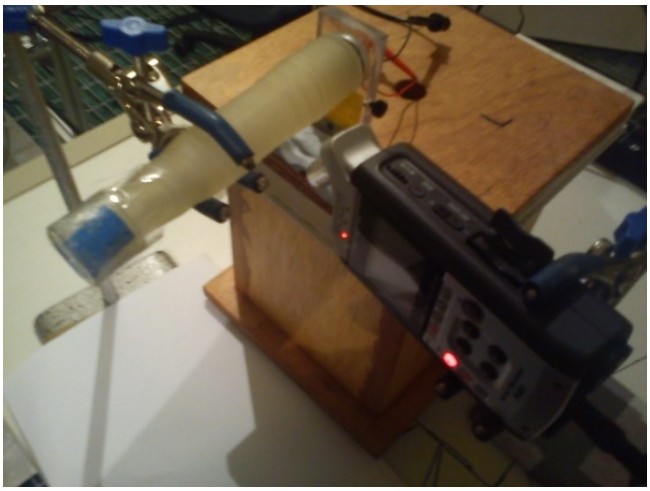

Fig. 3. A photograph of the experimental setup showing a fabricated 22-tube vocal tract model incident with a driving loudspeaker and audio recorder.

in the human experiments). The sound enters the 'mouth', resonates throughout the cavity, and is then captured using a recorder (Zoom H4n) located at a distance of less than 1 cm from the 'mouth'. All experiments were conducted in an anechoic chamber.

The excitation signal is a linear sweep-frequency cosine chirp pulse train. Each chirp sweeps the frequency range of 100-3500 Hz and is repeated at 2 Hz intervals (the repetition rate could be faster, but in practice was reduced to aid in automatic signal capture and analysis. To measure resonances of the mouth under practical conditions, the repetition rate should be at least 40-100 Hz). The chirp pulse train recorded in front of the speaker has a flat frequency response, and the speaker can be considered as an acoustic current source essentially independent of the load for ease of analysis.

3.2. Experimental design

Fabricated tube models were tested in the anechoic chamber. Each tube was, in turn, placed in front of the loudspeaker at a distance of 1 mm from the active surface of the loudspeaker. Following the tube model evaluation, a female speaker was analysed in the anechoic chamber, generating the same vowels as modelled (i.e. /ɔ, æ, ɛ, u, i, o/) using exactly the same setup (i.e. a linear chirp generated at the mouth), and the formant frequencies determined through LPC and power-spectrum estimation. The volunteer was then requested to generate the same vowels with glottal excitation (i.e. spoken), while in the same positon.

During the resonance test, chirp trains repeating twice a second were produced at the mouth of the test subject while she mimed the action of speaking the vowels in question, without phonation, sustained without breathing for at least 10 s (glottis closed). Multiple recordings were made during two three-hour sittings, and assessed for repeatability.

When the subject posed her mouth in front of the loudspeaker, the impedance of the vocal tract appeared acoustically in parallel with the free field impedance baffled by the face, and the resonances of the tract appeared as strong variations of the measured frequency spectrum. Visible variation in F1 and F2 spectral peak locations within the duration of a single test recording were grounds for discarding that particular recording - this could happen when the subject moved excessively, breathed during a test or did not maintain the same vowel shape. Eventually, all chirp results were time-aligned and a cumulative distribution function obtained. Very clear resonances are obvious as peaks in the resulting spectrum.

In the final test, an electro-larynx (Servox TM) was used to excite the vocal tract at the lips. The vibrating tip of the electro-larynx was connected to a narrow pipe to act as an impedance matching unit. The pipe was placed in front of the mouth of the subject for six vowel configurations.

4. ANALYSIS OF THE RESULTS

Previous authors have demonstrated a reasonable degree of correspondence between formant frequencies expected from MRI measurements of vocal tract dimensions and the actual recorded formants. This paper has validated the results using a 22 tube model multiphysics simulation. We have also fabricated models of the tube systems from the MRI area function data, and determined formant positions empirically – the actual measurements, excited at the output end of the tube, are found to correspond closely to the other data.

4.1. Measuring resonant frequencies of the tube models

The area functions of the models used in the simulations of section 2 were used by the authors to construct Solidworks computer aided design (CAD) representations of 22-tube model vocal tracts for the same vowels. The CAD models were fabricated through a high-resolution three-dimensional printing process using epoxy deposition to create a set of six physical tubes. The tube model for vowel /u/ plus a cutaway cross-section was shown in Fig. 1 where the 22 equally-spaced tube sections are visible between input flange and output flange. A constructed tube is also visible in Fig. 3

Each tube model was tested, in turn, within an anechoic chamber. A computer-generated linearly swept frequency chirp was generated from 100 to 3500 Hz and input into one end of the tube using the calibrated loudspeaker. A directional pair of microphones captures the reflected sound from the same end of the tube, while the opposite end is obstructed.

Based on the discussions of section 1, when the glottis is closed, measuring the impedance at the lip end of the model using an ideal current source is equivalent to measuring the resonances of the model when excited at the glottis. Table 2

Table 2. Frequency of first two formants for vocal tract models of six common vowels as measured using chirp pulse trains when excited at the lips (glottis closed) compared to Comsol simulation eigenfrequencies and simulated results from simulations [7]. Again, the degree of correspondence (%) between each is highlighted.

Vowel	Measured, Hz	Std. Dev., Hz	Simulated, Hz	Story_sim	Min. diff%
/ɔ/	572	3	633	618	7
	1035	2	1041	958	1
/æ/	685	4	735	732	6
	1760	19	1870	1689	4
/ɛ/	570	9	632	621	8
	1795	21	2037	1795	0
/u/	257	9	252	356	2
	1045	23	1170	1108	6
/i/	280	6	221	337	17
	2434	3	2398	2340	2
/o/	357	7	381	461	6
	837	2	857	861	2

summarizes the measurement results, and the standard deviation of each measurement is also reported in Herz. The measurement results are compared with simulations of the 22-tube models of the present work and to the simulations reported by Story et al [7][1]. Minimum difference of the values of the measurements and the values of simulations are also listed in this table.

4.2. Excitations using electro-larynx

As discussed in section 3.2, an electro-larynx (EL) was also used as the excitation source of the human vocal tract, using a flexible tube to convey the excitation pulse train into the mouth cavity, and employing the same measurement and analysis setup. It should be stressed that the EL oral connector and tube attachment mentioned is standard equipment with such devices, preferred when the user's neck is sore, or for user convenience (for example hands-free operation). The EL resonance results are shown in table 3 for a human test subject, compared with chirp-excited resonances of the human subject and her natural phonation. In all but one case, the EL resonance results are more extreme than either the chirp resonances or the naturally phonated resonances: it is observed that the chirp-excited signal generated at the mouth provides a more accurate representation of F1 and F2 compared to mouth-generated EL resonance.

[1]The difference between simulation values of Table 1 and 2 are partly the result of the difference in the sound speed in simulating vocal tract models at body temperature (37°C) and simulations at room temperature (25°C) where the measurements are done.

5. CONCLUSION

This paper describes a number of experiments used to assess the effectiveness of chirp excitation at the lips at determining vocal tract resonances. In the first part of the paper, area function data from MRI scans [7] is converted to a Kelly-Lochbaum style 22-tube uniform model. This model was created in Solidworks, then evaluated in the Comsol multiphysics simulation with flesh-like wall characteristics. Eigenfrequency analysis yielded close matches, in most cases, between the first three formant frequencies for six common vowels and the published simulation and measurement data, shown in table 1.

Next, the Solidworks models were fabricated with a high resolution three dimensional printer using epoxy deposition. These models were tested for 'lip' excitation in an anechoic chamber using calibrated audio test equipment. Post-processing with 16-pole LPC and spectral estimation techniques revealed resonant frequencies for F1 and F2 that were good matches to the above mentioned data, with very low standard deviation.

Following this, a human subject was tested in three ways using the same experimental equipment, location and setup. Firstly to determine resonances from chirp pulse train lip excitation while maintaining the vowel mouth and throat shape, secondly by measuring the subjects naturally vocalized resonances for the same vowels and finally by using electro-larynx resonance under the same conditions. Again, close corresponance was found for the natural and lip-excited results, while the electro-larynx resonances were generally more extreme.

In general, it has been shown that lip excitation with a chirp pulse train is a viable and accurate method for determining vocal tract resonances. This leads to the possibility of applications such as speech communication enhancement,

Table 3. Long-term average frequencies of the first two formants for six common vowels as measured with a human volunteer, using chirp pulse trains excited at the lips, compared with electrolarynx excitation (at the lips) and natural phonation (glottal excitation).

Vowel	Chirp, Hz	EL, Hz	Natural, Hz
/ɔ/	664	877	685
	1011	1044	1054
/æ/	761	1001	730
	1096	1382	1032
/ɛ/	750	982	517
	2145	1488	2132
/u/	396	387	385
	-	1500	1074
/i/	316	265	380
	2216	2139	2550
/o/	639	705	527
	947	1481	886

speech recognition, prosthetic devices for speech therapy and so on.

However, it must be noted that impedance matching between the loudspeaker face and the open mouth is very important for all resonances to be observed in the resulting spectrum. The alignment of both source and recording equipment with the mouth opening is also critical, as is the physical distance between the devices. Finally, this present study has been conducted on vowels: it may not work as well for for all phonemes, and we have evidence to suggest that nasals will require further analysis, as may some stops, bilabials (i.e. lips pressed together), and perhaps some plosives. Despite this, as mentioned, results to date indicate that the technique works well for voiced sounds, which are some of the most important phonemes for quality and intelligible communciations.

6. REFERENCES

[1] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[2] H. R. Sharifzadeh, I. V. McLoughlin, and F. Ahmadi, "Reconstruction of normal sounding speech for laryngectomy patients through a modified CELP codec," *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 2448–2458, Oct. 2010.

[3] A. Djeradi, B. Guerin, P. Badin, and P. Perrier, "Measurement of the acoustic transfer function of the vocal tract: a fast and accurate method," *J. Phonetics*, vol. 19, p. 38795, 1991.

[4] A. Dowd, J. Smith, and J. Wolfe, "Real time, non-invasive measurements of vocal tract resonances: Application to speech training," *Acoust. Aust.*, vol. 24, pp. 53–60, 1996.

[5] J. Epps, J. R. Smith, and J. Wolfe, "A novel instrument to measure acoustic resonances of the vocal tract during phonation," *Meas. Sci. Tech.*, vol. 8, pp. 1112–1121, July 1997.

[6] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*. Springer, 1991.

[7] B. H. Story, I. R. Titze, and E. A. Hoffman, "Vocal tract area functions from magnetic resonance imaging," *J. Acoust. Soc. Am.*, vol. 100, pp. 537–554, July 1996.