Dynamic mechanical modelling of speech

A. Barney

University of Southampton, Highfield, SO171BJ Southampton, UK
ab3@soton.ac.uk

In voiced speech, the source of sound is the vibrations of the vocal folds within the larynx. The steady flow of air passing out from the lungs is modulated by a complex fluid-tissue interaction in the larynx and excites acoustically the vocal tract downstream of the glottal exit. The system is small and inaccessible and, being situated inside a living person, is unsuitable for many kinds of experimentation. In vitro models offer the opportunity for systematic empirical studies in a controllable environment. This paper describes experimental measurements of the flow and pressure regimes made on a driven mechanical model of the vocal folds and vocal tract. The model has been used in a range of studies to explore jet formation and development downstream from the glottal exit, the effect of small changes in the duct geometry and, latterly, the interrelationship between the vibrating glottal source and a noise source further downstream as found in voiced fricative sounds.

1 Introduction

The source of sound in vowels is the vibration of the vocal folds within the larynx. This turns the steady flow of air from the lungs into a time-varying flow in the supra-glottal region, exciting the vocal tract acoustically. The sound is then shaped by the vocal tract, acting as an acoustic filter, to produce the radiated vowel sound. A more complex situation exists for voiced fricatives and plosives where a second source of sound, due to a stationary or time-varying constriction in the vocal tract, leads to an output sound based on the non-linear interaction between the multiple sound generation mechanisms as well as the filtering effect of the vocal tract.

In order to understand the detail of the processes involved in the production of speech sounds, and to synthesize realistically the output, numerical models are often developed, More recently these have combined sophisticated computational fluid dynamics (CFD) and finite element analysis (FEA) techniques to model the flow, the tissues and their interaction [1,2]. These approaches have undergone a rapid increase in complexity and processing speed in recent years, but there is still a fundamental question about how to validate such models against the true laryngeal and vocal tract behavior. This is an instance where *in vitro* models can offer at least a partial solution through the possibility of making direct measurements of experimental data..

The anatomical structures involved in speech production are small, delicate and inaccessible which makes direct measurement of the flow and pressure regimes within the larynx and vocal tract problematic. Although some direct measurements have been attempted [e.g. 3,4], subject numbers are usually small in such studies and validity and reliability across a wider population are therefore impossible to confirm. An often more profitable approach is the use of mechanical models, which offer control, flexibility of design, ease of access for instrumentation and repeatability of flow conditions in a way *in vivo* experiments rarely can.

In this paper we present three case studies of measurements in a driven dynamic mechanical model of the larynx and vocal tract which illustrate the advantages of mechanical models but also that some caution must be exercised in using them as a reference for numerical models.

2 The model and its instrumentation

The dynamic mechanical model considered here and shown schematically in Figure 1 consists of a square duct, 295 mm in length and 17 x 17 mm^2 in cross-sectional area, intersected by a pair of shutters at 175 mm from the outlet. A linear-profile converging section 11 mm in length and positioned just upstream of the shutters reduces the duct cross section to 17 x 3 mm^2. On the downstream side of the shutters there is an abrupt expansion over 2 mm to the full duct width. In the axial direction, the shutters are 3 mm in depth.

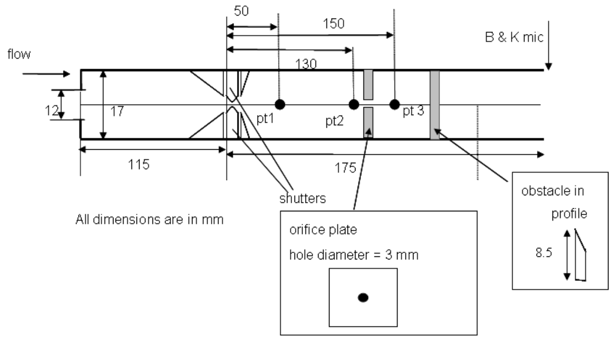

Figure 1 : Schematic diagram of the DMM showing dimensions and including the orifice place and obstacle used for dual source measurements

Each shutter is driven by a Ling Dynamics LD202 vibration generator and can be used to periodically widen and close the model glottis mimicking the vibrations of the vocal folds. The glottis is rectangular in cross-section with a height of 17 mm.

The shutters insert into closely machined slots in the side of the duct. The slots around the shutters are filled with a film of graphite grease to reduce frictional vibrations arising from the shutter oscillations.

Air from a compressor passes through the duct, and the volume flow rate can be set by a rotameter with a control valve. Just inside the duct inlet there is a flow straightener to reduce turbulence in the flow.

The following measurements can be recorded from the model:

- Inlet volume flow rate using a rotameter.
- Static pressure on the upstream side of the shutters using a manometer.
- Pressure at the duct wall for up to 4 locations chosen from a set of 7 pressure taps drilled in the duct wall Pressure transducer positions are chosen to meet the required experimental conditions in each study and are available both upstream and downstream of the shutters. Entran EPE-54 miniature pressure transducers with a diameter of 2.36 mm and a range of 0 to 14kPa are used for these measurements.

- Shutter position measured from a linear variable differential transformer attached to the shutters.
- Flow velocity at points within the duct downstream of the shutters using an ISVR Constant Temperature Hotwire Anemometer.
- Radiated pressure at the duct exit using a B & K ½" microphone and measuring amplifier.

All time-histories are captured by a simultaneous-sampling ADC connected to a PC.

The duct geometry can be varied by the insertion of various geometric elements to model different speech conditions.

3 Case Studies

The case studies chosen here are representative of the kind of measurements that may be obtained with a model which could be uncontrollable *in vivo*. The first two are chosen to illustrate different parameters that can be recorded, in somewhat inaccessible locations and the final case demonstrates the need for caution when trying to obtain reference measurements for numerical models.

3.1 Case study 1: Glottal jet measurements

A particular interest in CFD modeling is the structure and temporal evolution of the glottal jet. Understanding how the jet develops through a glottal cycle is critical to knowing how the vocal fold vibrations are driven and how the vocal tract is excited acoustically.

For these measurements the model was driven with an oscillation frequency of 80 Hz and the glottal width varied approximately sinusoidally from 0 to 3 mm over each cycle. Fluid velocity data was measured downstream of the glottal exit with a hotwire anemometer that was tracked across the horizontal midline of the vocal tract (x = 0 was the midpoint of the glottis). The sampling frequency was 8928 Hz with a sampling duration of 1 second.

Using the shutter timing signal, the hotwire time history for each location was segmented into glottal cycles and a coherent average was performed to obtain the average flow behavior over a cycle in each location.

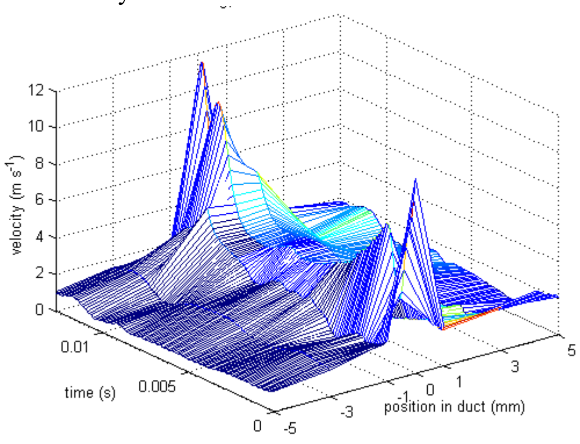

Figure 2 : Glottal jet velocity measured with a hotwire anemometer across the horizontal mid-line of the duct. The time t = 0 corresponds to the start of glottal closure and the location x = 0 is the mid-point of the horizontal mid-line of the duct.

Figure 2 shows a typical sequence of average velocity profiles across the duct horizontal axis with the duration of a single shutter cycle and with the hotwire placed 5 mm downstream of the glottal exit. The volume flow rate at the rotameter was 200cm^{3}s^{-1}. Measurements were made at locations –5, -3, -1, 0, 1, 3, and 5 mm along the horizontal midline of the duct parallel to the plane of the shutter motion. In examining Figure 2 it should be remembered that the velocities at different x-positions were not obtained simultaneously. The time t = 0 corresponds to the start of glottal closure. The glottis remains closed for a very short period (~ 1.5 ms) compared to the closure time that might be expected *in vivo*.

There is a large peak in the jet flow just prior to glottal closure with an amplitude, at x = 0 of 11.5 m s^{-1}, followed by a smaller peak coinciding with glottal opening with an amplitude, at x = 0, of 8.5 m s^{-1}. The peak at closure and the peak at opening each last approximately 1/3 of a shutter cycle. In the remaining 1/3 cycle the flow directly in line with the glottis is relatively constant at around 3.6 m s^{-1} or 28% below the mean flow.

Although the glottal geometry in the model is simplified and the shutters are driven, not self-oscillating, we observe here flow data that could not be obtained *in vivo*, and which has many of the same qualitative behaviors as that observed in CFD models [e.g. 1,2].

3.2 Case study 2: Effect of a downstream noise source

In this case study we consider the interaction between the periodic glottal source and a frication source downstream of the glottis [5]. The frication source arises from the combination of the constriction, formed by an orifice plate with a circular hole of 3 mm diameter, and a sharp edged obstacle further downstream. The obstacle is half the height of the duct and extends across the full width. Both the orifice plate and the obstacle have a thickness of 3.5 mm and are sealed into the duct to allow no leakage of air around the perimeter (see Figure 1).

The wall pressure is measured by pressure transducers mounted in taps at the locations indicated by pt2 and pt 3 in Figure 1. The output from these transducers is amplified and digitised with 16 bit resolution at a sampling frequency of 20 kHz for a duration of 2 seconds using a simultaneous-sample-and-hold system to retain an accurate estimate of the phase relationship between pressure signals at different tap locations. A B&K ½" microphone, located at the plane of the duct exit, measures the acoustic pressure due to the combination of frication and voicing at the outlet. The signal from this microphone is digitised at the same sampling frequency as used for the pressure transducers.

Figure 3 shows the effect of changing the glottal volume flow rate, measured at the rotameter, on the characteristics of the sound measured at the duct exit. Two orifice plate locations were used: with the upstream face of the plate 32 mm and 38 mm from the duct exit respectively.

For the first configuration the obstacle was placed with the upstream face 12 mm from the open end of the duct giving an constriction to obstacle distance of 20 mm and for the second the obstacle was placed with the upstream face 9 mm from the open end of the duct giving an constriction to obstacle distance of 29 mm.

For each of these geometric configurations a fundamental shutter frequency of 80 Hz, a minimum glottal width of approximately 0.1 mm (shutters just not touching) with a sinusoidal driver waveform and a glottal vibration amplitude of 1mm were held constant while the volume flow rate was varied from 10 to 24 l/min.

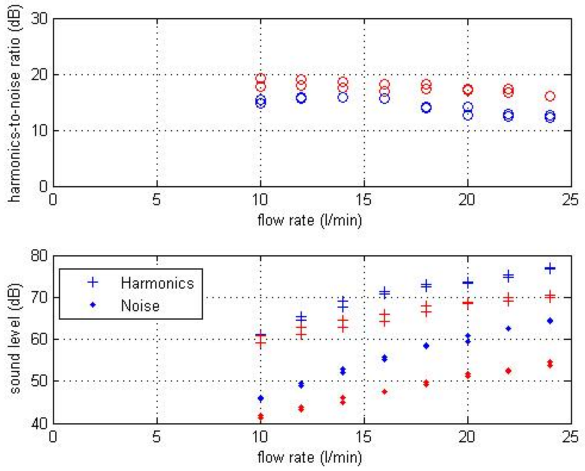

Figure 3 : Effect of changing glottal volume flow rate on the characteristics of the sound measured downstream of an orifice plate-obstacle combination

Figure 3 (top) shows the harmonics to noise ratio calculated from the pressure at the duct exit. Figure 3 (bottom) shows the absolute levels of the energy in the harmonic and noise parts of the signal.

It is evident that increasing the volume flow rate increases the level of the radiated harmonic signal and the level of broadband noise. Above 20 l/min the noise level increases faster than the harmonic level reducing the harmonic to noise ratio accordingly.

For the measurements shown in Figure 4 the fundamental shutter frequency was 80 Hz, the volume flow rate was 300 $cm^3 s^{-1}$, the minimum glottal width was approximately 0.1mm (shutters just not touching). The driver waveform was sinusoidal and adjusted in amplitude to give a fluctuating pressure at pt2 of 55 Pa rms. The orifice plate was located 35 mm from the open end of the duct and the obstacle was placed at a series of positions 3, 9, 13, 18 and 25 mm from the open end to give constriction-to-obstacle distances of 10, 17, 22, 26, 32 mm respectively with two pressure time-histories obtained for each position of the obstacle.

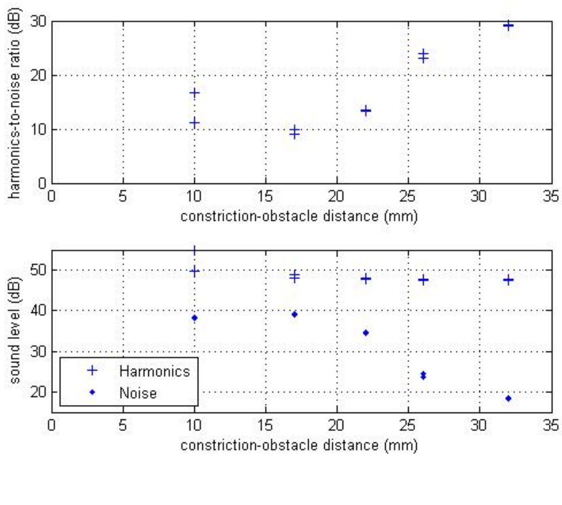

Figure 4 : Effect of changing orifice plate to obstacle distance on the characteristics of the sound measured downstream of an orifice plate-obstacle combination

For this set-up the harmonics to noise ratio increased with increasing plate to obstacle distance due to a reduction in the noise level relative to an approximately constant harmonic level.

As with the glottal jet measurements in the first case study, we see here a set of measurements that could not easily be obtained *in vivo*. The degree of control over the plate to obstacle distance permits a systematic study that cannot be achieved by human subject.

3.3 Case study 3: Vocal tract pressure

Considering again the set-up used to derive the data in Figure 4 we can look at the power spectrum of the wall pressures just upstream of the orifice plate, down-stream of the orifice plate between the plate and the obstacle and at the duct exit. These are shown in Figure 5.

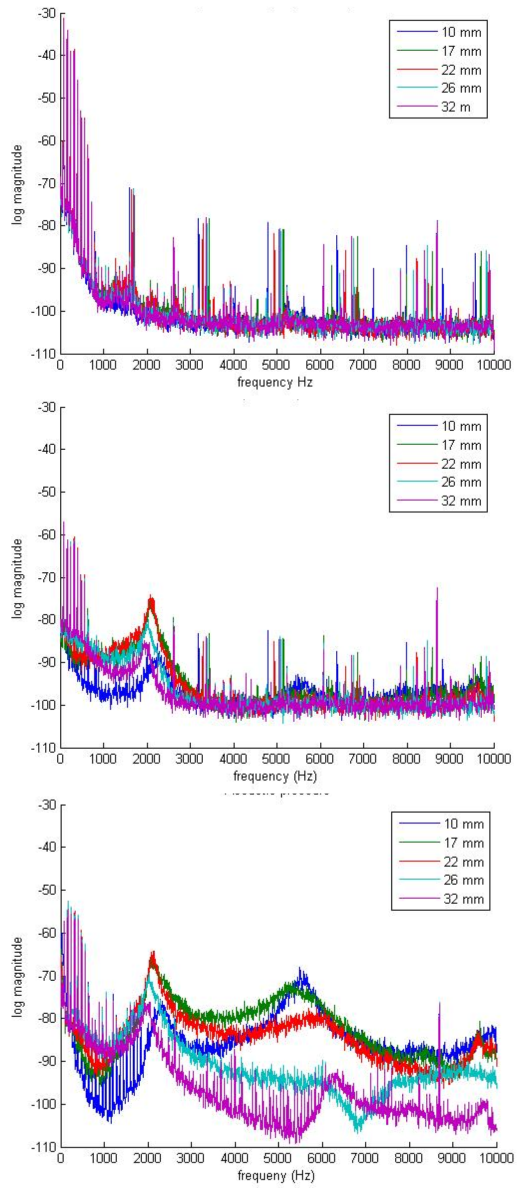

Figure 5: (top) Duct wall pressure just upstream of the orifice place (pt2), (middle) duct wall pressure just downstream of the orifice plate (pt3) and (bottom) radiated acoustic pressure outside the duct for orifice plate to obstacle distances of between 10 and 32 mm

The power spectra were calculated by the Welch method using a 4096 point Hanning window with 50% overlap and a 4096 point FFT.

In Figure 5 (top) we can see that moving the obstacle on the downstream side of the orifice plate affects in particular the higher resonance frequencies measured on the upstream side.

Figure 5 (middle) shows a similar pattern in the high frequency resonances but with the addition of a broad peak around 2 kHz with a peak frequency and bandwidth that is somewhat geometry dependent. Maximum excitation of this resonance is for plate to obstacle distances of 17 and 22 mm with lower levels of excitation at both shorter and longer distances. This resonance frequency is of the order of the quarter-wavelength resonance of the orifice plate to duct exit distance, allowing for an end correction at the open end of the duct.

Figure 5 (bottom) shows that this resonance is radiated beyond the duct exit. A further resonance is revealed by the more sensitive microphone located here, which was buried in the noise floor of the in-duct transducers. This is of the order of the ¾ wavelength resonance of the orifice plate to duct exit distance

Once again we see a set of measurements that could not have been obtained *in vivo* but we should also note that both the in-duct and the radiated measurements are sensitive to very small changes in the duct geometry. This must be carefully accounted for if we are to use them as a reference for validation of numerical models.

4 Discussion

Mechanical models are clearly a useful and flexible tool for systematic investigation of flow and pressure events in a duct with a time-varying constriction and a range of more complex geometric configurations further downstream. They allow experimentation that is systematic and controllable in a way that their anatomical counterparts can never be. Instrumentation can be introduced to regions completely inaccessible in a live subject and measurements can be made over long sampling durations in a repeatable manner. Care is required however in matching such data to the output of CFD simulations. As demonstrated, the flow and pressure behaviours are highly sensitive to local model conditions. It is therefore imperative that CFD simulations and their validating models are designed together in order that they be clearly analogous to each other. Mechanical-CFD model pairs are likely to achieve more if designed to answer specific questions about flow and pressure events rather than to try to obtain full simulations of all aspects of laryngeal behaviour.

References

[1] J. Horáček, S. Gráf, S, "Mathematical modeling of airflow in the glottal region and its comparison with experimental data", *Proceedings of MAVEBA*, Firenze, Italy, 203-206, (2009)

[2] S. Zörner, M. Kaltenbacher, M. Döllinger, "Finite element model of the human phonation process", Proceedings of MAVEBA, Firenza, Italy, 207 – 209 (2009).

[3] B. Cranen, L. Boves, "Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production", J. Acoust. Soc. Am., 77(4), 1543-1551, (1985).

[4] E Finnegan, E. Luschei, H. Hoffman, "Estimation of Alveolar Pressure During Speech Using Direct Measures of Tracheal Pressure", J. Speech, Language, Hearing Res, 42 1136-1147 (1999).

[5] A. Barney, P Jackson, "A model of jet modulation in voiced fricatives", Proc Int Conf Acoustics NAG-DAGA, Rotterdam, March 2009, 1733-1736, (2009).