

A model of jet modulation in voiced fricatives

A. Barney¹, P.J.B. Jackson²

¹ ISVR, University of Southampton, UK, Email: ab3@soton.ac.uk

² CVSSP, University of Surrey, UK, Email: p.jackson@surrey.ac.uk

Introduction

Voiced fricatives such as /v/ and /z/ are characterised by simultaneous generation of sound from acoustic sources at two different locations within the vocal tract: the voicing source associated with glottal vibrations and a noise source, related to a constriction of the oral cavity, with a precise location that depends on the target phoneme. The output sound is characterised by both a (quasi-) periodic element related to the glottal vibrations and a random element associated with pressure fluctuations of the jet flow exiting the constriction.

The non-linear interaction between the sources leads to amplitude modulation of the noise at the fundamental frequency of the glottal vibration. Although the occurrence of an amplitude modulation of the noise source in voiced fricatives was noted by Fant [1], Flanagan and Cherry [2] were the first to incorporate the effect into a speech production model. They multiplied band-passed Gaussian noise by the square of the instantaneous volume flow rate. Stevens [3], motivated by the result of perceptual tests, included an amplitude modulation in his fricative models with a modulation depth related to the glottal area while Klatt [4] used a square wave with a 50% duty cycle to modulate the frication amplitude in his synthesiser.

More recently, Pincas and Jackson [5] have studied modulation effects in sustained fricatives and fluent speech, showing that the depth of modulation varies little between speakers but does vary with place of articulation, usually being strongest for /z/. Ganapathy et al. [6] and Jepsen et al. [7] have found modulation effects to be important in speech perception.

The aim of our research is to provide a simple parametric model of the interaction between the sources based on empirical data obtained in a dynamic mechanical model of the larynx and vocal tract [8] configured with a tract geometry representative of a rudimentary voiced fricative.

We are predominantly interested in the strength of the interaction between the two sources rather than in the absolute strength of either of the sources individually and therefore our analysis concentrates on measurement of the magnitude and phase of the modulation.

The Modulation Index

In an amplitude modulated signal a carrier signal w_n is multiplied by a modulating signal a_n to produce an amplitude modulated signal x_n where, if a_n is a sinusoid of frequency f_0 :

$$x_n = w_n [1 + m_0 \cos(2\pi f_0 nT + \phi_0)] \quad (1)$$

In equation (1), m_0 is the modulation index at frequency f_0 , T is the sampling period and ϕ_0 is an arbitrary phase shift, assumed constant.

To estimate the modulation index, m_0 , the Fourier transform of equation (1) is first computed [5]:

$$X(k) = F\{w_n\} \otimes \dots \left[\delta(k) + \frac{m_0}{2} (\delta(k - k_0)e^{-j\phi_0} + \delta(k + k_0)e^{+j\phi_0}) \right] \quad (2)$$

where \otimes denotes convolution and $k_0 = f_0 nT$ is the frequency bin that contains f_0 . We can estimate a value for m_0 by comparing the spectral coefficients at f_0 and d.c.:

$$m_0 = 2 \frac{|\bar{X}(k_0)|}{|\bar{X}(0)|} \quad (3)$$

The modulation index, m_0 , can be thought of as the fraction of the carrier signal by which the modulated signal varies, e.g., if $m_0 = 0.4$, then the carrier signal fluctuates by 40% above and below its original, un-modulated value. It is most often given in standard index form (in the range 0–1).

The Mechanical Model

The mechanical model used in this study is a life-size replica of the human, male vocal folds and tract. It is based on a previously described mechanical model, [8], with the main differences being the use of a square cross-sectioned duct for the vocal tract and a more realistic shape at the glottal inlet and outlet sections. A schematic diagram of the model is shown in cross-section in Figure 1.

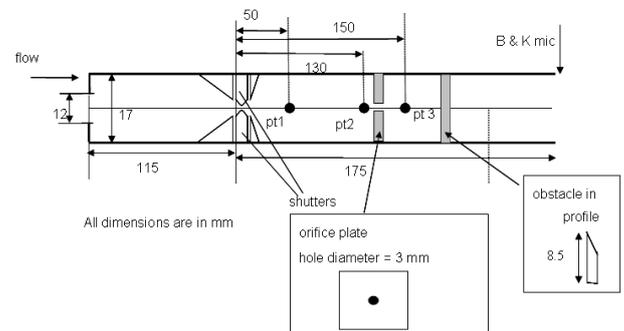


Figure 1: Schematic diagram of the mechanical model showing (inset) details of the orifice plate and obstacle.

The Perspex duct, which has a cross-sectional area of 17x17 mm², is 290 mm in total length. It is intersected at 115 mm from the air inlet by a pair of brass shutters that represent the

vocal folds. In the direction of the duct axis the shutters have a length of 3 mm. The glottis is rectangular with a height of 17 mm and variable width with a maximum value of approximately 6 mm. On the inlet side, a gentle constriction over a distance of about 7 mm channels the air flow through the glottis. On the outlet side there is a more abrupt expansion into the wider duct.

Mechanical shakers (Ling Dynamics model 202) drive the shutters in the direction perpendicular to the duct axis. In this respect, the model differs from the self-oscillating vocal folds found *in vivo*. The phonation neutral position of the shutters can be adjusted and complete glottal closure can be achieved in each glottal cycle if desired. The shutter excitation waveform is generated by a Hameg HM813C waveform generator and amplified by two Ling Dynamics power amplifiers (PA25E), one for each shutter.

The friction 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.

A steady flow of air into the model is provided by a compressor and the volume flow rate can be monitored at the duct inlet by a rotameter and controlled by a valve. Upstream of the duct, a compliant model of the lungs with a volume of 8 litres, lined with foam, settles the air. Just downstream of the duct inlet, a flow straightener helps to remove any remaining turbulent fluctuations from the air stream to give a laminar flow in the duct upstream of the glottis and at the glottal inlet. The mean static pressure upstream of the glottis is measured using a manometer.

Miniature pressure transducers (Entran EPE-541-2P) are mounted in machined pressure taps with their measurement plane flush with the duct walls. These permit wall-pressure measurements to be obtained at the three locations indicated by pt# 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 1/2" microphone with pre-amp and amplifier is located at the plane of the duct exit to measure the acoustic pressure due to the combination of friction and voicing at the outlet. The signal from this microphone is digitised at the same sampling frequency as used for the pressure transducers.

Measurements and Analysis

This paper reports the outcome of two sets of measurements: one to investigate the variation of the modulation index estimated from the radiated sound as a range of different flow parameters were varied within the model and one to investigate the effect of the constriction-to-obstacle distance on the estimated phase of the modulation.

For the first set of measurements (to be referred to as set A) two locations of the orifice plate and obstacle within the duct were considered:

- upstream face of orifice plate 32 mm from duct open end and upstream face of obstacle 12 mm from the open end, giving constriction-to-obstacle distance of 20 mm.
- upstream face of orifice plate 38 mm from duct open end and upstream face of obstacle 9 mm from the open end, giving constriction-to-obstacle distance of 29 mm.

For each of these geometric configurations a sample of possible permutations for the model parameters was investigated with fundamental frequency of the shutter vibration, f_0 , between 10 and 150 Hz, glottal amplitude, A , (per shutter) from 0 mm to 1.5 mm, inlet volume flow rate, U , from 167 to 400 cm³ s⁻¹, minimum glottal width, w , from 0 to 1 mm, where $w = 0$ indicates that the shutters just touch, and two choices of driver waveform, s : sine wave or square wave.

For the second set of measurements (referred to as set B) $f_0 = 80$ Hz, $U = 300$ cm³ s⁻¹, $w \approx 0.1$ mm (shutters just not touching). The driver waveform, s , was sinusoidal and adjusted in amplitude, A , to give a fluctuating pressure at pt2 (see Figure 1) 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 locations 3, 9, 13, and 25 mm from the open end to give constriction-to-obstacle distances of 10, 22, 26, 32 mm respectively with two time-histories obtained for each position of the obstacle.

Analysis

To extract the modulation index from the signal recorded by the microphone, the time-series was filtered using an order 20, zero-phase bandpass filter with a pass band of 2.25-7.75 kHz. The filter bandwidth was chosen by observation of the measured signals to correspond to the portion of the spectrum where the harmonic content, related to the acoustic excitation of the duct by the shutter vibrations, had negligible influence. The Fourier transform of the filtered signal was then calculated. For the measurements of set A, the modulation index, m_0 , was of primary interest and was calculated using equation (3). For the measurements of set B the phase of the modulated signal relative to the phase of the the pressure signal upstream of the orifice plate (measured by pt2) was the focus. To calculate this the phase of the filtered, Fourier transformed microphone signal at f_0 was subtracted from the phase estimate at f_0 for the Fourier transform of the pressure at pt2.

Results

The signal recorded by the microphone at the duct exit for a constriction-to-obstacle distance of 20 mm, an inlet volume flow rate of 333 cm³ s⁻¹ and a sinusoidal driving frequency with a minimum glottal width of 0.3 mm can be seen in the top graph of Figure 2.

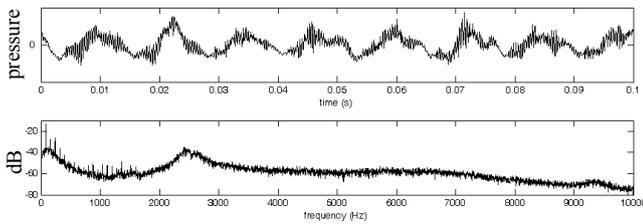


Figure 2: (Top) time history and (bottom) power spectrum of the of the signal from the microphone.

The slowly varying, periodic part of the signal corresponds to the oscillation frequency of the shutters; the higher frequency, noise-related part of the signal appears to have the time varying amplitude expected of a modulated signal.

The spectrum of the microphone signal can be seen in the bottom graph of Figure 2. At frequencies of less than approximately 2 kHz the shutter harmonics can clearly be seen as regularly spaced peaks. The broader peak centred at approximately 2.5 kHz is associated with the quarter-wavelength resonance of the duct section from the orifice plate to the exit.

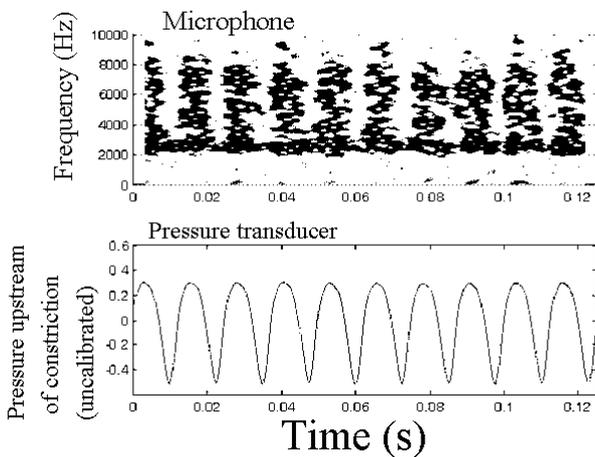


Figure 3: Spectrogram of the microphone signal (top) and pressure measured upstream of the orifice plate by pt2 (bottom).

The top graph of Figure 3 shows the spectrogram, for the same conditions as Figure 2, of the band-pass filtered microphone signal where the amplitude fluctuations at the fundamental frequency are evident. The driving pressure measured upstream of the orifice plate is shown in the bottom graph.

Magnitude of the modulation index

In Figure 4, the modulation index, m_0 , is plotted against the rms pressure, p_{us} , at pt2 just upstream of the orifice plate. The modulation index appears to increase with increasing p_{us} for low values of this driving pressure, but tends to saturate at a modulation index of approximately 0.6 for larger p_{us} values with only a very small number of measurements estimated to have a modulation index exceeding that. Although the modulation index as defined by equation (3) is biased by noise, a second, unbiased estimator [9] has been shown to give values in good agreement with these estimates.

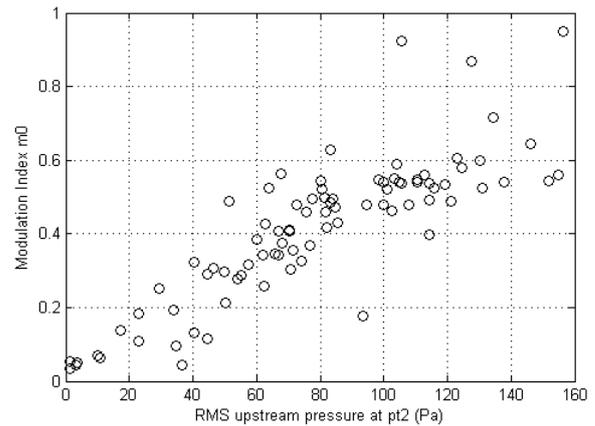


Figure 4: Modulation index m_0 plotted against upstream pressure at pt2

Phase of the modulation index

The phase of the modulation envelope relative to the phase of the pressure upstream of the orifice plate at pt2 is shown in Figure 5. In the figure the error bars represent the standard error in the mean for the relative phase estimates and for each constriction to obstacle distance there are two estimates. The straight line is the line of best fit to the data, which has a correlation coefficient $R = 0.84$ ($P < 0.01$).

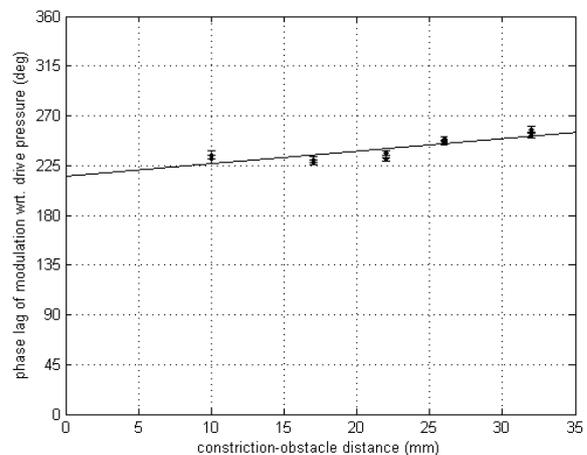


Figure 5: Phase of the modulation envelope relative to the phase of the upstream pressure measured at pt2

Discussion and Conclusions

The interruption of the airflow through the model, by the oscillations of the shutters, creates pressure fluctuations in the section of the duct between the shutter exit and the upstream face of the orifice plate. These fluctuations have a harmonic structure related to the fundamental frequency of the shutter motion, f_0 , and generate a pressure excitation just upstream of the orifice plate that can also be seen (Figure 3) to have a significant component at f_0 .

From Figure 4 it is apparent that, when the pressure fluctuation is small, the index of the amplitude modulation at

f_0 varies directly with the rms amplitude of the f_0 component of the pressure. This pressure, p_{us} , is a function of shutter frequency f_0 , inlet volume flow rate, U , shutter amplitude A , and driver waveform shape (sin or square), s , such that:

$$m_0 \propto p_{us}(f_0 U, A, s). \quad (4)$$

Thus in this region of the graph, if the size of the upstream pressure fluctuation is known, the depth of modulation is predictable. Predicting the upstream pressure from the parameters f_0 , U , A , and s is, however, not trivial since the mapping of these variables to p_{us} exhibits redundancy.

For larger values of p_{us} the modulation index appears to saturate at approximately $m_0 = 0.6$ with few observations exceeding this value. A similar saturating behaviour has been observed for *in vivo* measurements of voiced fricatives [5] and also in a free jet [10]. No investigation has been made to date regarding modulation by higher harmonics of f_0 although both [5] and [10] report such an effect.

The source strength of sound generated by the interaction of an unsteady flow with a sharp-edged obstacle is known to depend on the flow velocity and on a relevant length scale [11], [12]. The flow velocity through the constriction depends on the pressure difference across the orifice plate. In data set B the driving pressure, and hence the flow velocity, was held constant, while changing the distance between the orifice plate where the separated flow regime is initiated and the obstacle where the sound is generated. We observe a phase relationship where the phase-lag between the driving pressure and the modulation envelope of the radiated sound increases with increasing constriction-to-obstacle distance.

Modelling the phase relationship as linear, gives rise to a moderate fit to the small number of measurements presented. This approach provides a pragmatic way to predict the relative phase between the fundamental frequency component of the harmonic part of the radiated spectrum and the envelope of amplitude modulation of the noise source. This may be useful as a first approximation for use in articulatory synthesis of voiced fricatives. A model that represents the underlying aeroacoustic processes realistically is likely to need rather greater sophistication: The linear model presented here is clearly unlikely to be valid for very short constriction-to-obstacle distances; in the limit, for a coincident constriction and obstacle, the phase lag is expected to tend to zero. Further, the linear model has an underlying assumption that the flow velocity on the duct axis between the orifice plate and the obstacle is independent of distance from the orifice plate. This may be a reasonable first approximation, but it is well known that a jet expands in area as it travels away from its origin and associated with the expansion is a reduction in axial velocity.

Empirical and theoretical studies of the interaction of a vortex or of turbulence with a sharp edge, that develop scaling laws for the intensity of the generated sound [11], [12], do not generally consider the case of a jet confined within a duct and are therefore not directly applicable to the circumstances of voiced fricative generation.

In order to progress with deriving an improved model for predicting the modulation phase, the next step will be to acquire a larger set of experimental measurements to explore the joint effect of variations in constriction-to-obstacle distance and flow velocity on the radiated sound field. A further study will explore models for predicting the intensity of the noise source.

References

- [1] Acoustic theory of speech production. Mouton, Den Hague, 1960.
- [2] Excitation of vocal-tract synthesizers. *J. Acoust. Soc. Am.* **45**, (1969), 764-769.
- [3] Airflow and turbulence noise for fricative and stop consonants: Static considerations. *J. Acoust. Soc. Am.* **50**, (1971), 1180-1192.
- [4] Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.* **67**, (1980), 971-995.
- [5] Amplitude modulation of turbulence noise by voicing in fricatives. *J. Acoust. Soc. Am.* **120**, (2006), 3966-3977.
- [6] Modulation frequency features for phoneme recognition in noisy speech. *J. Acoust. Soc. Am.* **125**, (2009), EL8-EL12.
- [7] A computational model of human auditory signal processing and perception. *J. Acoust. Soc. Am.* **124**, (2008), 422-438.
- [8] Fluid flow in a dynamic mechanical model of the vocal folds and tract I: Measurements and theory. *J. Acoust. Soc. Am.* **105**, (1999), 444-445.
- [9] Time-frequency modulation representation of stochastic signals. *Proc. 15th Intl. Conf. on Digital Signal Processing*, (2007), 639 – 642.
- [10] Orderly structure in jet turbulence. *J. of Fluid Mech.* **48**, (1971), 547-591.
- [11] Aerodynamic sound generation by turbulent flows in the vicinity of a scattering half plane. *J. Fluid Mech.* **40**, (1970), 657-670.
- [12] Acoustic wave emitted by a vortex ring passing near the edge of a half-plane. *J. Fluid Mech.* **155**, (1985), 77-103.