Forum Acusticum 2005

The Effect of Glottal Opening on the Acoustic Response of the Vocal Tract

Anna Barney, Antonio De Stefano
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, UK, **ab@isvr.soton.ac.uk**

Nathalie Henrich
Laboratoire d'Acoustique Musicale (LAM),11, rue de Lourmel , 75015 Paris, France, **henrich@lam.jussieu.fr**

In the source-filter model of speech production it is assumed that the acoustic source mechanism is independent of the vocal tract acoustics. In fact, the changing acoustic boundary condition at the glottal end of the vocal tract may be expected to affect the formant frequency and the formant bandwidth, with a corresponding effect on voice quality.

In this study, a geometrically simplified physical model of the vocal tract, larynx and sub-glottal tract is used in the presence of an external applied sound field to quantify the changes to the acoustic response of the vocal tract with changing glottal width. The applied sound field is tonal, generated by a loudspeaker, at frequencies ranging from 300 Hz to 2 kHz. The acoustic response of the vocal tract is measured with a pair of pressure transducers, spaced 8 cm apart, and mounted flush with the vocal tract wall. The acoustic transfer function from glottis to lips of the vocal tract is calculated from the measurements using the method described by Holland and Davies (JSV, 230(4), 915-932, 2000) and shows an increase in the first formant frequency of the order of 13% at F1 when the glottis is open to a width of 3 mm compared to the closed-glottis case.

The experimental measurements will be compared to a theoretical model that takes the change in glottal impedance with changing glottal width into account.

1 Introduction

This paper describes an investigation of the effect of the glottal open phase on the vocal tract formant frequencies. The study is carried out using a mechanical model of the larynx and vocal tract in order to allow independent variation of glottal and vocal tract parameters which might not be possible in real human speech. The model is used in two modes: i) with the glottis static, no flow and an imposed external sound field and ii) with a forced glottal oscillation and an applied, low-speed mean flow.

In the classical source-filter theory of speech production [1] it is assumed that the glottal acoustic source mechanism and the vocal tract filter do not interact. This model has formed the basis of many speech synthesis systems. The output speech from such systems is intelligible, but is frequently found to sound unnatural [2]

Childers and Wong [3] define several possible types of source-tract interaction: loading of the source by the vocal tract impedance, dissipation of vocal tract acoustic energy, especially at F1, due to glottal opening, and in cases of low glottal damping, carry-over of vocal tract energy from one glottal period to the next. Historically, much research into source-tract interaction has tended to concentrate on the first of these effects [4, 5], while [6] considered the third. It is the second of these interaction types that we will consider here.

Childers and Wong [3] noted the tendency of increased glottal damping during the open phase of the glottal cycle to truncate the F1 waveform, to shift the formant frequencies and to cause an increase in the formant bandwidths. Flanagan [7] showed using transmission line models that for a fixed vocal tract geometry the presence of a finite glottal impedance could be expected to increase glottal damping and to raise the formant frequencies in comparison to those found for the case of infinite glottal impedance. For his steady state models he predicted an increase in the frequency of F1 of about 1.4% and about 1% for F2 for a glottal area of 5 mm^2.

Ananthapadmanabha and Fant [8] also considered the effect of the glottal damping, but were concerned principally with a way to incorporate its effects into a glottal waveform model. They found the effect of glottal inertance to be small. Badin and Fant [9] investigated the effect of the glottal impedance on the formant frequencies of the vocal tract. For a glottal area of 0.027 mm^2 they found that with a short-circuited sub-glottal system and a uniform vocal tract, and with the glottis modelled by an inductance only, there was an upward shift in F1 of approximately 10%. A glottal resistance alone had no effect on F1 while a resistance and inductance together reduced the upward shift to around 0.2%. They concluded that the effect was insignificant during most instances of vowel production. However, despite the hypothesis that in speech the formant shifts may generally be small, they are likely to be an acoustic correlate of changes in glottal open quotient or glottal vibration amplitude and hence may be of interest in voice quality considerations and in enhancing naturalness in synthesised speech. Furthermore, the glottal areas considered by Badin and Fant [9] and by Flanagan [7] correspond to glottal widths rather smaller than the peak widths generally reported in the literature for vowel production.

In the following sections we consider the theoretical formant shift in the case of a static glottal configuration and compare it to measurements in an imposed sound field in our mechanical model. We then consider the case of formant shifts for an oscillating glottis in terms of their dependence on glottal open quotient and glottal vibration amplitude.

2 The Mechanical Model

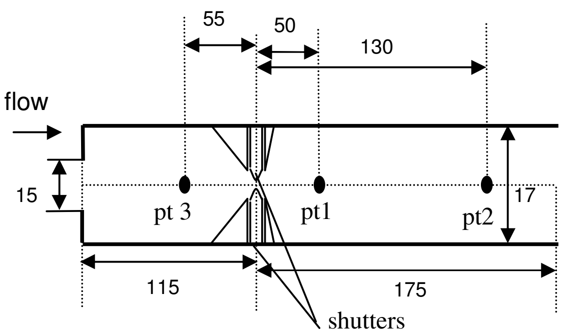

Figure 1: Cross-section through the vocal tract model showing the location of the shutters and the pressure transducers (pt). All dimensions are in millimetres. Figure not to scale.

The mechanical model used in this study is shown in cross-section in Figure 1 and consists of a square duct with sides of length 17 mm, intersected at 115 mm from the inlet by a pair of shutters that represent the vocal folds. Mechanical shakers can drive the shutters in the direction perpendicular to the duct axis and complete glottal closure can be achieved. In the direction of the duct axis the shutters have a thickness of 3 mm. The glottis is rectangular with a height of 17 mm and variable width with a maximum value of 6 mm. On the inlet side, a gentle constriction over a distance of about 7 mm channels the air flow through the glottis, while on the outlet side there is a more abrupt expansion into the wider duct.

A flow of air into the model can be provided by a compressor and controlled by a rotameter. 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. Pressure transducers (Entran EPE-541-2P) mounted flush with the duct walls permit measurement of static or time-varying pressure at the three locations marked (pt) in Figure 1. The output of these transducers is amplified and digitised at a sampling frequency of 8928 Hz.

3 Theoretical Modelling of the Formant Shift

A lumped element transmission line [7] may be used to investigate the expected formant frequencies for different values of the glottal width. This corresponds to a static glottal model. In Figure 2 is the output from such a model for a rigid-walled duct with a cross-sectional area of 289 mm^2 and a length of 175 mm; the same dimensions as for the vocal tract of our mechanical model. The radiation impedance is modelled as that for an unbaffled duct and the sub-glottal system is short circuited. The vocal tract is modelled by four T-sections with the elements defined as proposed by Flanagan [7].

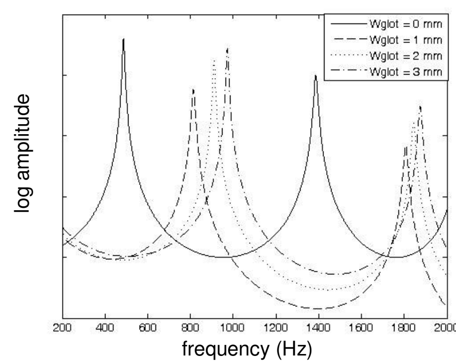

Figure 2: Theoretical model of transfer function from glottis to lips for a duct, open at the lips, of length 17.5 cm with various glottal widths.

For the solid line, the glottal impedance is infinite while for the cases with finite glottal impedance a value of :

$$Z_g = \frac{12\mu l_g}{w_g{}^3 h_g} + j\omega \frac{6\rho l_g}{5h_g w_g} \qquad (1)$$

is used [10] where μ is the coefficient viscosity of air, ρ is the density of air, ω is the angular frequency l_g is the length of the glottis parallel to the axis of the vocal tract, h_g is the height of the glottis perpendicular to the vocal tract axis and w_g is the variable glottal width.

The exact values of the formants in the cases with finite glottal width will depend on the expressions chosen for the glottal and sub-glottal impedance, thus while Figure 2 may be representative of the behaviour for a static glottis as the glottal width is increased, it should not be considered as a predictive model for the formant frequencies that may be observed in the mechanical model. Nevertheless, one can observe the increase in F1, and the rather smaller increase in F2, as the glottis is widened. The curves have been shifted for ease of comparison so that the first minimum of each is at the same level and thus no conclusions should be drawn about the relative levels of the formants from this figure.

The value of glottal resistance used here is the same as that used by Flanagan [7] while the glottal inductance term has an additional factor of 6/5 as given by Beranek [10]. In experimental studies Laine and Karjalainen [11] and Rösler and Strube [12] have measured the impedance for a static model glottis.

Laine and Karjalainen [11] found a fairly good match between measurement and theory for the zero flow case over a frequency range of 0 to 1.5 kHz in a rectangular glottis using the same glottal resistance model as in Equation (1) but an inductance term that included an end correction and excluded the factor of 6/5. They added a term to the theoretical glottal resistance to account for an absorbing boundary condition for the sub-glottal termination which was measured as part of the glottal impedance in their set-up. Using an absorbing sub-glottal boundary condition has the effect of adding additional damping mainly at the first formant frequency and would result in a reduction both in the amplitude and in the upward frequency-shift of F1. These reductions would be greater for a wider glottis.

Rösler and Strube [12] used a more realistically shaped rigid glottal model with an inlet contraction and an abrupt expansion at the outlet. They too included an end correction in their theoretical expression for the inductance to account for the effective length of the oscillating air mass in the glottis. They subtracted a value for the measured sub-glottal impedance from their measured impedance values and compared the theoretical and measured glottal impedance in isolation from the sub-glottal system. They were unable to obtain valid measurements for frequencies below 800 Hz, but found that for frequencies above, the inductance was generally smaller than predicted by the theoretical expression.

While both these experimental studies provide valuable information about the glottal impedance, the details of the resistance and inductance that were observed are likely to be very dependent on the precise geometry of the model used in each case and therefore not directly transferable to conditions in our model. We therefore adopted Equation (1) for this study as our theoretical models are expected only to be representative of the gross behaviour of the formants.

4 Experimental Measurement of the Formant Shift for the Static Glottis

4.1 Measurements

The mechanical model was used to obtain a measurement of the vocal tract acoustic response for a static glottis with widths of 0, 1, 2, & 3 mm. A loudspeaker at the vocal tract exit applied the external sound field. In order to get a sufficiently good coupling between the applied sound field and the duct, the speaker was placed close to the end of the duct such that the boundary condition there was that of a closed end and the effective length of the duct from that end to the downstream face of the vocal folds was 205 mm. While this raises the overall values of the resonances, it will not alter the direction of the frequency shift due to changes in glottal width. Furthermore, to avoid large variations in the in-duct sound level arising from zeros in the speaker-duct coupling, the analysis was carried out at a series of tonal frequencies and the amplitude of the speaker excitation was adjusted to give a strong signal at each microphone. Since we will only be interested in relative levels this strategy does not affect the validity of our findings. For frequencies up to 900 Hz the frequency resolution was 50 Hz and above 900 Hz a resolution of 100 Hz was used.

4.2 Analysis

The pressure transducers were located at positions pt1 and pt2 of Figure 1 with a distance between them of l = 80 mm. The microphones signals were used to make a decomposition of the standing wave field into its forward- and backward-travelling component waves. The predicted components become unreliable at frequencies where the microphone spacing is close to a half wavelength. This gives an upper frequency limit to this analysis of approximately 2 kHz. At frequencies below 300 Hz the speaker did not produce enough sound power in the duct to give a sufficient signal to noise ratio for the analysis to be carried out. A plane wave assumption is inherent in the analysis.

For each frequency the measurement protocol was as specified by Holland and Davies [13]. The pressure signals were measured twice, once with transducer #1 at pt1 and transducer #2 at pt2 and again with the locations of the transducers and their entire signal paths swapped. From these pairs of measurements, two estimates of the pressure transfer function between the measurement locations were obtained, H_A and H_B.

To obtain the complex pressure components for each frequency at location pt1, p_1^+ in the positive *x*-direction and p_1^- in the negative *x*-direction, the following expressions are used:

$$\left|p_1^+\right| = \frac{(G_{11})^{1/2}}{|1+R|} \text{ and } p_1^- = R\,p_1^+ \qquad (2a,b)$$

$$R = -\left\{\frac{H_A^{0.5}H_B^{0.5} - e^{-jk^+l}}{H_A^{0.5}H_B^{0.5} - e^{jk^-l}}\right\} \qquad (3)$$

where R is the acoustic reflection coefficient and k^+ and k^- are the complex wave numbers in the positive and negative *x*-directions which take into account visco-thermal losses. From the complex pressure amplitudes at pt1 it is straightforward to calculate the estimated transfer function from the glottis to the lips.

4.3 Results

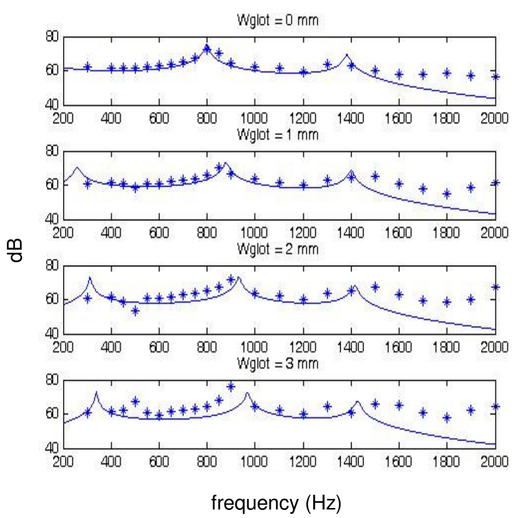

Figure 3: Measured (*) and theoretical (solid line) transfer function from glottis to lips for various glottal widths. Theoretical transfer function based on a duct of length 20.5cm closed at lips.

The transfer functions between volume velocity at the glottis and pressure at the speaker, derived as described above from the measurements on the mechanical model, are shown in Figure 3 as the data points (*). The measured values for F1, taken as the maximum of the data are 800, 850 900 and 900 Hz for glottal widths of 0, 1, 2 and 3 mm respectively. These should be taken as a guide only since the frequency resolution of the measurements is insufficient to determine the precise location of the formant peak. A significantly more damped peak in the region 1.3 to 1.5 kHz defines F2 but while there appears again to be an overall increase in centre frequency with increasing glottal width it is difficult to feel confident in associating specific peak frequency values with this formant.

Also shown in Figure 3 (solid line) is a predicted transfer function between glottal volume velocity and closed-end pressure, derived for a 205 mm duct with the glottal impedance infinite (top plot) or given by Equation (1) (lower 3 plots) The vocal tract is modelled as a lumped element transmission line with 3 T-sections and, as for Figure 2, the sub-glottal system has been modelled as a short circuit.

For a glottal width of 0 mm the match between measurement and prediction is excellent up to a frequency of 1.2 kHz. For a widening glottis the prediction follows the data fairly well between 400 Hz and 1.2 kHz although for the 3mm glottal width case, the formant amplitude around 900 Hz is under predicted and the frequency of this peak is somewhat over predicted. At the peak around 1.4 kHz the theory appears to predict a smaller frequency increase with increasing glottal width than is measured and above 1.4 kHz the measured amplitude is significantly above the theoretical level.

At low frequencies for finite glottal width, there is a predicted peak in the region of 300 Hz that is not found in the measured data. Adoption of an absorbing model for the sub-glottal system damps this peak significantly. We can also improve the fit between the measured and predicted location of the peak around 800 Hz in this way, but the amplitude of the prediction in this frequency region is then reduced significantly, giving a bigger discrepancy between the measured and predicted levels. Additionally, while the match in amplitude between 900 Hz and 1.4 kHz is improved, the peak around 1.4 kHz is predicted at a lower frequency than for the short-circuited sub-glottal system giving an even worse match to the measurements. A small peak in the measured amplitude at 500 Hz for a glottal width of 3 mm is assumed to be due to experimental error as there is no evidence of a similar peak developing at other glottal widths.

4.4 Discussion

The measurements for the static glottis show an increase in the formant frequency with increasing glottal width. Although the frequency resolution is insufficient to assign precise values to the formant frequencies, the increase compared to the formant frequency for the closed glottis is of the order of 6% for a glottal width of 1 mm and close to 13% for a glottal width of 3 mm. This is much greater than predicted by Flanagan [7] or Badin and Fant [9] due in large part to the much greater glottal widths considered in this study.

Conditions in a mechanical model are rather different to those in the human glottis and care must be taken in interpreting these results for speech production modelling. The choice of sub-glottal impedance had a strong influence on the exact value of the predicted formant frequencies and amplitudes but while a model with absorbing boundary conditions gives a better match than a short-circuited one to the frequency of the peak around 800 Hz, it is correspondingly worse at predicting the amplitude of the transfer function in this region. At higher frequencies the short-circuit model gives an improved amplitude match while both under-predict the frequency of the upper formant peak. Further work on the choice of glottal and sub-glottal impedance appears to be needed for a satisfactory theoretical description.

5 Experimental Measurement of the Formant Shift for the Time-Varying Glottis

5.1 Measurements

The mechanical model was used dynamically to obtain a measurement of the formant shift arising when the glottal width varied with time. The acoustic excitation of the vocal tract was due to the driven oscillation of the vocal fold shutters with a steady airflow of 200 cm^3s^{-1} delivered from the compressor. Pressure transducers were located as for the static case.

The shutter vibration was monitored by an electro-mechanical system and the output from this was digitised at the same sampling frequency as the signals from the pressure transducers. We wished to control both the amplitude of the shutter vibration and the open quotient of the glottal cycle. Accordingly the shutters were driven with a square wave signal for which the mark-space ratio and the amplitude could be independently controlled. At fundamental frequencies approaching those of speech, compliance in the driver system and phase lag in the location monitoring system combine to make the instant of opening and closing somewhat difficult to define precisely from the electro-mechanical monitor signal. The measurements reported here were therefore obtained at the rather lower fundamental frequency range of 10 to 40 Hz where the behaviour of the shutters could be reliably determined. For fundamental frequencies that do not approach F1 for the vocal tract, we do not expect the influence of the glottal impedance on the frequency of F1 to be related to the value chosen for f0.

Measurements were obtained for open quotients of 20, 40, 60 and 80% and fundamental frequencies of 10, 20 and 40 Hz. The peak glottal width ranged from 0.25 mm to 4 mm. The range of peak glottal widths for each open quotient is shown in Figure 4 where it can be seen that while there is some correlation between the minimum peak glottal width and the open quotient, there is generally a good overlap in the range of glottal widths investigated for the open quotients of 20 to 60%.

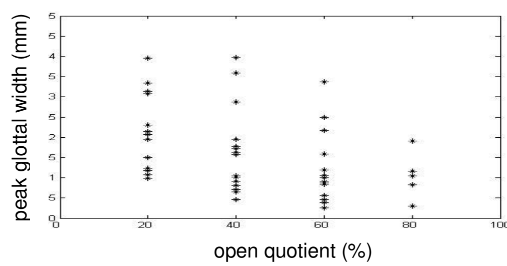

Figure 4: Peak glottal width vs open quotient for all f0

5.2 Analysis

The first formant peak was found from the pressure time-history measured at pt1 using an AR spectral estimation method of order 11 and a 2048-point Fourier transform. The AR estimation is based on the autocorrelation function of the signal and operates on the entire signal, not just the glottal closed phase, in order to capture the effect of the change in glottal impedance during the open quotient. The F1 peak was defined as the maximum value of the spectral estimate between 200 Hz and 1 kHz. Each spectrum was inspected visually and where no identifiable peak was found, the F1 estimate for that data set was rejected. Hence the data set for an open quotient of 80% in particular is of a rather smaller size than for the other three open quotients tested.

5.3 Results and discussion

Figure 5 shows the relationship between glottal width and F1 estimate for each open quotient. A further division of the data by fundamental frequency showed no dependence of F1 on f0 for the range investigated. It is immediately evident that the F1 estimates in all cases are higher than the expected value of about 500 Hz for a 175 mm duct open at one end.

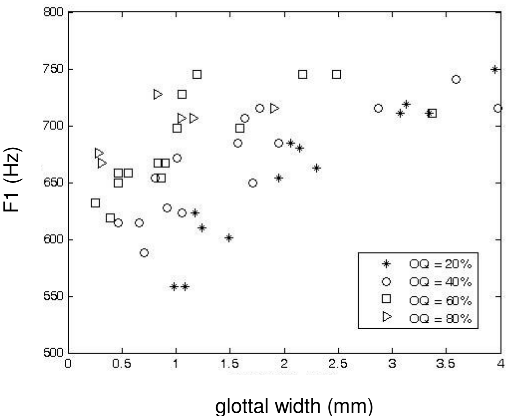

Figure 5: Frequency of F1 for changing glottal width and open quotient (OQ). For all f0. OQ = 20% (*) , OQ = 40% (o), OQ = 60% ( ), OQ = 80% (∇).

The general trend of increasing F1 with increasing glottal width is clear for each open quotient value. The trend of increasing F1 with increasing open quotient is more pronounced for smaller glottal widths. The apparent convergence of F1 values at higher glottal widths may be due to the paucity of data in this region arising from the difficulty in identifying a spectral peak associate with F1 for longer open quotients.

6 Conclusions and Future work

This paper has detailed an investigation of the variations in F1 with changes in glottal impedance in both static and dynamic glottal configurations. In both cases the frequency of F1 increases with increasing glottal width and in the dynamic case, increasing the glottal open quotient also caused an increase in the frequency of F1. For the dynamic glottis, the frequency of F1 was always significantly higher than the quarter-wavelength resonance of an open-closed tube with the same duct length. The exact values found for the F1 frequencies in this study are likely to be strongly dependent on the precise geometry of the model and thus the findings may be thought of only as indicative of the a likely F1 values to be found in real speech.

The next stage of our work will focus on developing a theoretical model that can predict behaviour in the dynamic case and on analysis of the first formant of voiced speech as the open quotient and glottal amplitude are varied.

References

[1] G. Fant, 'The Acoustic Theory of Speech Production', Mouton, The Hague, ISBN-9027916004 (1960)

[2] N.B. Pinto and D.G. Childers, 'Formant Speech Synthesis: Improving Production Quality', *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. 37. No. 12. pp. 1870-1887 (1989)

[3] D.G. Childers and C.-F. Wong, 'Measuring and Modeling Vocal Source-Tract Interaction', *IEEE Transactions on Biomedical Engineering*, Vol. 41. No. 7. pp. 663-671 (1994)

[4] M. Rothenberg, 'Acoustic Interaction between the Glottal Source and the Vocal Tract', in Vocal Fold Physiology, K.N. Stevens and M. Hirano, (Eds), University of Tokyo Press,, pp. 305-328 (1980)

[5] B. Guerin, M. Mrayati and R. Carre, 'A Voice Source Taking Account of Coupling with the Supraglottal Cavities', *Proc ICA 1976*, Vol. 1. pp. 47 - 50 (1976)

[6] G. Fant and T.V. Ananthapadmanabha, 'Truncation and Superposition', *STL-QPSR*, Vol. 2-3/1982. pp. 1-17 (1982)

[7] J.L. Flanagan, 'Speech Analysis Synthesis and Perception', Springer Verlag, New York, ISBN-0387055614 (1965)

[8] T.V. Ananthapadmanabha and G. Fant, 'Calculation of the True Glottal Volume Velocity and Its Components', *Speech Communication*, Vol. 1. pp. 167-184 (1982)

[9] P. Badin and G. Fant, 'Notes on Vocal Tract Computation', *STL-QPSR*, Vol. 2-3/1984. pp. 53-108 (1984)

[10] L.L. Beranek, 'Acoustics', AIP, New York, ISBN-088318494X (1986)

[11] U. Laine and M. Karjalainen, 'Measurements on the Effects of Glottal Opening and Flow on the Glottal Impedance', *ICASSP '86*, pp. 31.36.31 - 31.36.34 (1986)

[12] S. Rösler and H.W. Strube, 'Measurement of the Glottal Impedance with a Mechanical Model', *Journal of the Acoustical Society of America*, Vol. 86. No. 5. pp. 1708-1716 (1989)

[13] K.R. Holland and P.O.A.L. Davies, 'The Measurement of Sound Power Flux in Flow Ducts', *Journal of Sound and Vibration*, Vol. 230. No. 4. pp. 915 - 932 (2000)