MODELLING OF THE INTERACTION OF THE WALLS OF THE VOCAL TRACT WITH THE ACOUSTIC PRESSURE WAVE.

G.C.F. Clark & C. Scully

Dept. of Linguistics and Phonetics, University of Leeds, Leeds, U.K.

INTRODUCTION

Although the basic principles of speech production are well understood it is still very difficult to obtain accurate measurements on the vocal tract during speech or to predict the effect of a change in the conditions. One method which is of general applicability is to develop computer simulations and to test theories and refine parametric estimates on these.

As reported previously [1, 2] Scully and her co-workers have built up a large computer model of human speech production. It operates in three stages, the specification of how the geometry of the system changes with time, the derivation of sound sources and the filtering of the sources through the vocal tract. This paper describes enhancements which have been introduced into the filter stage of the model in order to model the mechanical properties of the walls of the vocal tract more closely.

The propagation of the acoustic pressure wave through the vocal tract can be modelled by treating it as a transmission line [3] consisting of a series of concatennated cylindrical sections. A pair of pressure waves travels forwards and backwards in this, exchanging energy at each boundary. If the sections are treated as having rigid walls the system has resonances which are sharper than those observed in speech. It is known that these can be broadened by taking into account acoustic losses at the walls of the tube [4] as well as elsewhere. The losses can be ascribed to boundary effects which will be present even in a rigid walled system and those due to an interaction of the pressures wave with a yielding wall.

Flanagan et al. [5] produced a composite model of speech production in which they were able to estimate the contributions made by various loss processes to the overall formant bandwidths. The losses due to air viscosity and heat conduction were small enough to be disregarded. The portion due to the glottal losses was relatively independent of frequency. The contribution from wall losses fell with increasing frequency whilst that for lip radiation increased. Below 500Hz only glottal losses and wall effects made a significant contribution. Maeda [6] showed in a similar study that modelling the wall losses in the time domain gave a fairly good match to the observed formant bandwidths as did frequncy domain synthesis as employed by Mryati [7].

METHODS

Transmission line modelling

As mentioned above the vocal tract is described as a series of concatennated tubes with pressure waves travelling in both directions along it. Since the calculations must be repeated whenever a wave reaches the end of a section the interval between calculations is governed by the length of the section and the speed of sound in the vocal tract. Using 0.5cm sections the calculations are repeated at 14µs intervals.

The two pressure waves are treated as occupying alternate sections. If the sections are numbered from the glottis then in the first time interval the forward wave is in the odd-numbered sections, and the backward in the even ones. At the end of the interval the waves interact at the boundaries and then travel onwards, so that the forward wave now occupies the even-numbered sections. The various sound sources are added into the pressure wave at the appropriate points, the voice and aspiration sources at the glottis and the main frication noise and transient sources in front of the major constriction in the tube.

In the original form of the model the interaction of the two waves was treated by calculating an instantaneous total pressure at the boundary which was then decomposed into two new pressures, as shown in eqn. (1) - (3).

$$Sum = 2 * (Pf_{old} * Area1 - Pb_{old} * Area2) / (Area1 + Area2) \quad (1)$$

$$Pf_{new} = Sum - Pf_{old} \quad (2)$$

$$Pb_{new} = Sum - Pb_{old} \quad (3)$$

Pf - Forward pressure wave
Pb - Backward pressure wave
Area1, Area2 - Areas of adjacent sections

The geometry is specified at 5ms intervals and interpolated to provide values for the vocal tract filter. Originally this was done every 14µs. In the new formulation described here the reflection coefficients are calculated at 1ms intervals so simplifying the calculations of the pressure which are still conducted at 14µs. The intermediate value in the calculation is here a pressure shift which is applied to both waves.

$$Refl.\ Coeff. = (Area1 - Area2) / (Area1 + Area2) \quad (4)$$

MODELLING OF THE INTERACTION OF THE WALLS OF THE VOCAL TRACT WITH THE ACOUSTIC PRESSURE WAVE.

$$\text{Shift} = \text{Refl. Coeff.} * (Pf_{old} - Pb_{old}) \tag{5}$$
$$Pf_{new} = Pf_{old} - \text{Shift} \tag{6}$$
$$Pb_{new} = Pb_{old} - \text{Shift} \tag{7}$$

These two procedures are mathematically equivalent. At the naso-pharyngeal junction the original method has been retained as this leads to a simpler calculation.

Modelling of wall losses

The basic transmission line model is of a hard-walled tube. The sharp resonances of this were originally broadened by modelling the glottal and mouth radiation losses [8] and by introducing an empirical 'loss coefficient' which was used to delay a portion of each pressure wave and add it back in at the next time interval.

The vocal tract walls show a very great variation in composition, from the soft yielding material of the cheeks to the rigid material of the hard palate. An exact model of this would be impossibly complex and we have approximated it by treating each section as being made of an isotropic material describable by three parameters covering the viscous, compliant and inertial properties.

When a time-varying pressure impinges on a yielding wall it will be attenuated by three processes: moving the wall against a viscous load, stretching it against an elastic load and accelerating its mass. The last two effects lead to the storage of energy in the wall so that there can be an exchange of energy between the pressure waves at different points in time. The three effects are additive and lead to an equation which gives the shift in the pressure due to the wall interaction. We have taken the displacements to be small relative to the total area so that their effect on the reflection coefficients can be ignored. Likewise we are obliged by the interrelationship of geometry and time in our model to ignore the variation of the velocity of sound due to yielding walls. [4].

We are modelling the attenuation of the pressure by the wall interaction in terms of a second order differential equation in the pressure.

$$\Delta P = \text{alpha} * P + \text{beta} * dP/dt + \text{gamma} * d^2P/dt^2 \tag{8}$$

where alpha, beta and gamma are constants which are functions of the mechanical properties of the wall and ΔP is the shift in the pressure due to the wall interaction. The model works in discrete time steps so that eqn. (8) must be replaced by the difference equation

$$\Delta P = alpha * P_0 + beta * (P_0 - P_1) / \Delta t + gamma * (P_0 - 2P_1 + P_2) / \Delta t^2 \quad (9)$$

P_0 - Pressure at current time
$P_{1,2}$ - Pressure in previous two time intervals.
Δt - Time interval between calculations.

This can be rearranged to give three terms in P_0, P_1 and P_2 by introducing three new constants

$$A = alpha + beta / \Delta t + gamma / \Delta t^2 \quad (10)$$

$$B = -beta / \Delta t - 2 * gamma / \Delta t^2 \quad (11)$$

$$C = gamma / \Delta t^2 \quad (12)$$

$$\Delta P = A * P_0 + B * P_1 + C * P_2 \quad (13)$$

We now have an expression for the shift in the pressure due to the wall interaction, ΔP, in terms of the current pressure and that in the two previous time intervals. This is used to modify the pressure in the forward wave, Pf_{old} in eqn. (5) - (7). After this the boundary interaction is calculated using these equations and the new pressure waves Pf_{new} and Pb_{new} are used to update the stored pressures used in the wall calculations.

<u>Parameter estimation</u>
The alpha, beta and gamma terms above are functions of the viscous, compliant and inertial properties of the wall. There are two approaches which may be made when assigning values to them. In the first a rigorous derivation of equation (8) in terms of the basic mechanical properties can be made and experimental values for these substituted in. This approach was followed by Maeda [6] using the results of Ishizaka *et al.* [9], who measured the mechanical properties of the vocal tract wall by mechanically exciting it from the outside. Maeda's results showed a close fit between the measured and calculated bandwidths so that an analytical approach using these figures is likely to give a close match to actual utterances. The second approach is to use an analysis by synthesis approach as adopted elsewhere in our model, in which an initial rough estimate of a parameter is refined by comparison of the calculated and measured acoustic waveforms.

The choice of approach is in part dictated by the use to which our model is to be put. Our main interest is in modelling specific utterances by known speakers, for whom measurements made on another person may be quite inappropriate. An utterance is recorded with simultaneous measurements of the air flows and pressures. The geometric and aerodynamic stages of the model are iterated until there is a good match between the observed and calculated air flows and pressures. We then are able to use a consistent set of geometric and aerodynamic data as an input when refining the parameters of the acoustic sources and the vocal tract filter. In these the comparison is between the observed and calculated acoustic output. Thus whilst we may find that initial estimates derived from experimental data shorten the analysis time we shall always be involved in an iterative refinement of these estimates. At present we are investigating the effects on the output as the parameters are systematically varied.

RESULTS

Transmission line enhancements

The deterioration in the computation time for the model introduced by the fuller description of the wall losses has been largely offset by the gain in reformulating the reflection processes in the transmission line. With the vocal tract calculations repeated at 71.4KHz (the frequency imposed by using 0.5cm sections) the computation:real time ratio is roughly 200, roughly 3 minutes per second of speech.

Wall losses

Until we have completed the installation of the radiation loading stages in the new model we are not able to present the results of using the formulation described here in simulating individual speakers.

Figs. 1 and 2 show the results obtained by varying the parameters for a given geometry and input signal. The utterance was created using a female vocal tract 14.5 cm. in length. It represents the second [ ] in [p z p]. The vertical lines on the graphs indicate 5.95KHz, which is the bandwidth of the output from the vocal tract filter. In fig. 1 the coefficients are both set to zero, i.e. a hard-walled tube, whilst in fig. 2 they are set to 0.03. The major effect is on the broadening of the formant peaks but there is also some shift in their positions.

CONCLUDING REMARKS

We have produced a model of the interaction of the vocal tract wall with the acoustic pressure wave incorporating second order effects.

MODELLING OF THE INTERACTION OF THE WALLS OF THE VOCAL TRACT WITH THE ACOUSTIC PRESSURE WAVE.

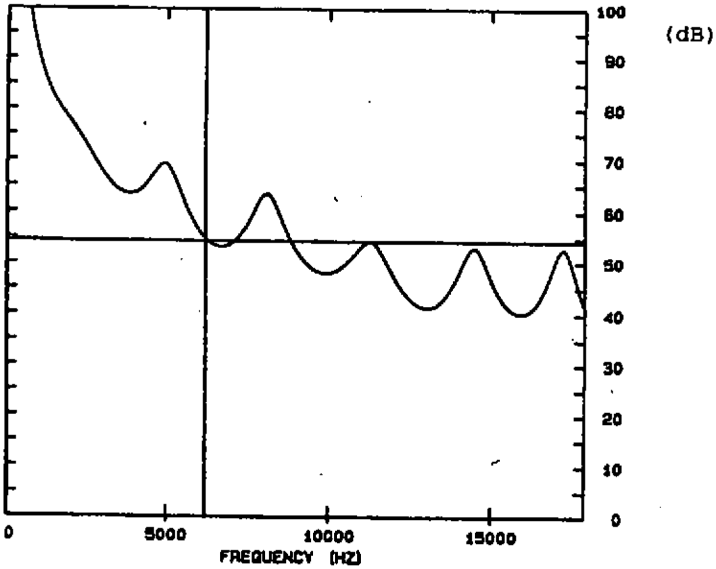

Fig 1 Spectrum of output with all coefficients set to zero

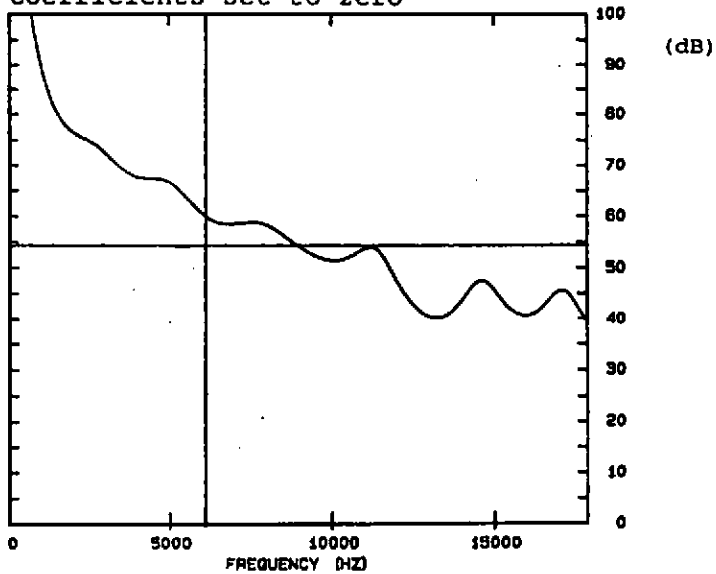

Fig 2 Spectrum of output with all coefficients set to 0.03

MODELLING OF THE INTERACTION OF THE WALLS OF THE VOCAL TRACT WITH THE ACOUSTIC PRESSURE WAVE.

This has been implemented in a Kelly-Lochbaum representation of the vocal tract in such a way that we can model these effects without incurring heavy computational penalties.

Acknowledgements
This work was supported by SERC grant GR/D/40791 - MMI/056 as part of the SPAR consortium of the Alvey Project, Speech Interfaces and Phonetic Algorithms, MMI 009

REFERENCES

[1] E. Allwood & C.Scully, 'A composite model of speech production', Conf. Rec. ICASSP 82, Paris, Vol 2, 932-935, (1982)

[2] C. Scully & G.C.F. Clark, 'Analysis of speech signal variation by articulatory synthesis', IEE Conf. Publ. 258, 83-87, (1986)

[3] J.L. Kelly & C.C. Lochbaum, 'Speech synthesis', Proc. 4th Intl. Cong. Acoust., Copenhagen, Paper G42, 1-4, (1962)

[4] J.L. Flanagan, 'Speech analysis, synthesis and perception', Springer-Verlag, Berlin, 2nd Edtn., pp 65-69, (1972)

[5] J.L. Flanagan, K. Ishizaka & K.L. Shipley, 'Synthesis of speech from a dynamic model of the vocal cords and vocal tract', Bell System Tech.J., Vol. 54, 485-506, (1975)

[6] S. Maeda, 'A digital simulation method of the vocal tract system',Speech Communication, Vol. 1, 199-229, (1982)

[7] M. Mryati & R. Carré, 'Relations entre la forme du conduit vocal et les caractéristiques acoustiques des voyelles françaises', Phonetica, Vol. 33, 285-306, (1976)

[8] C.Scully & E. Allwood, 'Glottal losses in a composite model of speech production', Proc I.O.A., A1.1 - A1.5, (1982)

[9] K. Ishizaka, J.C. French & J.L. Flanagan, 'Direct determination of vocal tract wall impedance", IEEE Trans. ASSP-23, 370-373, (1975)