Lung and Larynx Coordination in a Composite Model of Speech Production

C. Scully and E. Allwood
Leeds, United Kingdom

1. Introduction

From different combinations of a few rather simple articulatory actions, a variety of quite complex aerodynamic conditions and acoustic outputs can be created. The most basic requirement of all for speech is the creation of voice; this is easily achieved by new-born babies. What is examined here is the building up of a repertoire of lung and larynx actions appropriate for *controlled* operation of the voice source. Even apparently simple speech sounds demand correct coordination. The auditory goal of the simulation described here was an [i] vowel quality with 'modal' as opposed to 'breathy' or 'pressed' ('laryngealised') phonation type and with falling pitch. The tasks of speech production are by no means clear, but one basic aim is to achieve a subglottal pressure suitable for the onset and maintenance of voice.

2. The model

A model of speech production processes implemented on a VAX 11/780 computer was used. The stages modelled are shown in Figure 1. Inputs to the model define speaker dimensions, initial conditions, larynx type for a functional model of voicing and articulatory transitions. Eight quasi-independent articulators are used, as controllers of the geometry rather than as anatomical structures. Most articulatory actions are represented by changes in cross-section area of a few constrictions of the vocal tract. Articulations of the lung walls are represented either by air pressure in the lungs Pl, or as in the study described here, by the rate of change of lung volume DVLU. Vocal fold articulations are represented by the slowly changing (d.c.) component of glottal area Ag and by a variable called Q, for the effective stiffness and mass of the vocal folds. Vertical movements of the vocal folds are not modelled at present. The bases for the modelling have been described (Scully, 1975; Allwood and Scully, 1982).

Timing and coordination in the articulatory block determine aerodynamic conditions throughout the respiratory tract. Articulatory states and aerodynamic conditions combine to determine the magnitude of turbulence noise sources for aspiration and frication. A pulse source, derived from rate of pressure change in the oral cavity, has been introduced recently, but was not

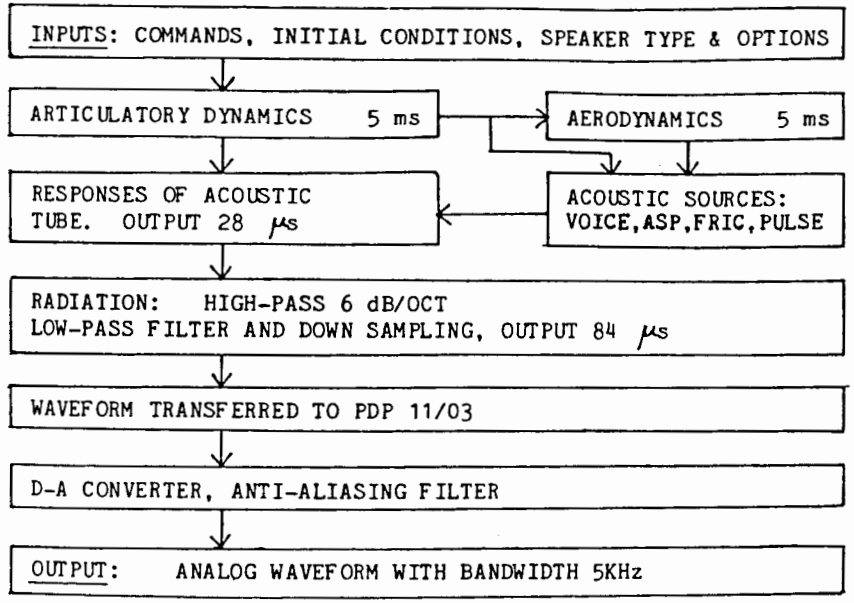

Figure 1. Block diagram of the model.

used in this study. A parametric description of the voice source is used as shown in Figure 2 (Fant, 1980). A minimum Δ P of 2 cm H_2O was assumed for the onset and offset of voicing. Fundamental frequency F_0 was derived from $F_0 = \Phi + 4.\Delta P$. A voicing 'plateau' region was defined between AG= 0.04 cm^2 and Ag = 0.08 cm^2. F_0 decreased for Ag less than 0.04 cm^2. K varied inversely with Ag. TCR was constant at 0.1. Aspiration and frication sources were weakened and modulated when voicing was present. In an alternative form of the voicing model the wave parameters VOIA, K and TCR can all be made to vary as linear functions of three controlling physiological variables: Ag, Δ P and Q. Using the model interdependence of vowel and consonant durations have been demonstrated for voiced and voiceless fricatives having constant supraglottal articulation and for open and close vowel contexts. The effects were similar to those of real speech and the model's outputs were intelligible and speech-like (Allwood and Scully, 1982).

3. Modelling of aerodynamic processes

The system in Figure 3. A set of first order differential equations expresses the assumptions made and the physical principles invoked in the model, which are as follows:

1. The compliance of the lung walls need not be included. It is assumed that the speaker takes the nett compliance (recoil) into account when adjusting muscle pressures at different lung volumes so as to give a pre-planned rate of lung volume decrement. Passive changes in rate of lung volume decrease are not modelled at present.

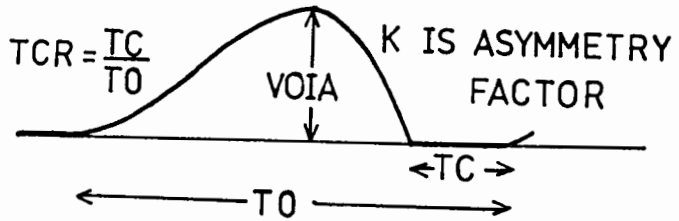

Figure 2. The parametric description for the voice source.

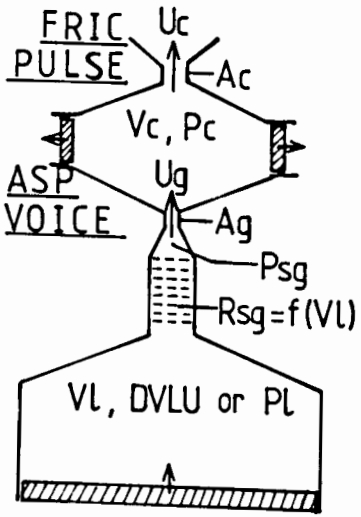

Figure 3. The aerodynamic system.

2. The walls of the subglottal airways are taken as rigid, with flow rates in speech well below limiting flow rate.
3. The supraglottal cavity has an active component of volume change due to articulatory actions, added to a passive component associated with wall compliance (Rothenberg, 1968).
4. All but 4% of the subglottal volume is located in the respiratory zone of small airways, with generations higher than 16. Subglottal flow resistance is almost totally confined, on the contrary, to the large tubes of generation less than 10. This striking separation of subglottal volume and flow resistance justifies a model with one lumped lung volume and a separate single flow resistance linking it to the glottal orifice. This constrasts with the more complex representation in the model of Rothenberg (1968)
5. Subglottal flow resistance is an 'ohmic' conductance which increases linearly with lung volume, up to a maximum value of about 2 L/cm H_2O.
6. Inertance of air and tissues may be neglected.
7. The air in the respiratory tract is assumed to be an ideal gas and to be compressible. Departures from atmospheric pressure are small. Isothermal conditions are assumed. The flow is taken as one-dimensional. There is continuity of mass flow for each of the two cavities.
8. For each of the two orifices (constrictions) there is conservation of energy at the inlet (the Bernoulli effect), but energy is lost in the turbulent mixing region at the outlet. This gives a turbulent, flow-dependent component of pressure drop. A laminar 'ohmic' component of pressure drop is added to this. The same empirical constants are used for both orifices.
(Space does not permit reference to the relevant respiratory literature).

Parameter values are chosen to define cavity wall compliance, subglottal properties and initial conditions, Lung volume Vl, lung and supraglottal air pressures Pl and Pc are integrated at each time step to obtain values for the next sample. Merson's method (NAG library, 1981) was used here. There were problems with numerial instabilities in the aerodynamic variables, especially when oral pressure Pc was very low, in vowel-like segments. Other methods for the integration, including Gear's method for dealing with 'stiff' equations (NAG library, 1981), have recently given improved stability and much reduced computation time for the aerodynamics.

4. The modelling of lung and larynx coordination

Some articulatory plans yielded inappropriate pitches, voice qualities or vowel lengths. Two of a series of articulatory plans are shown in Figure 4

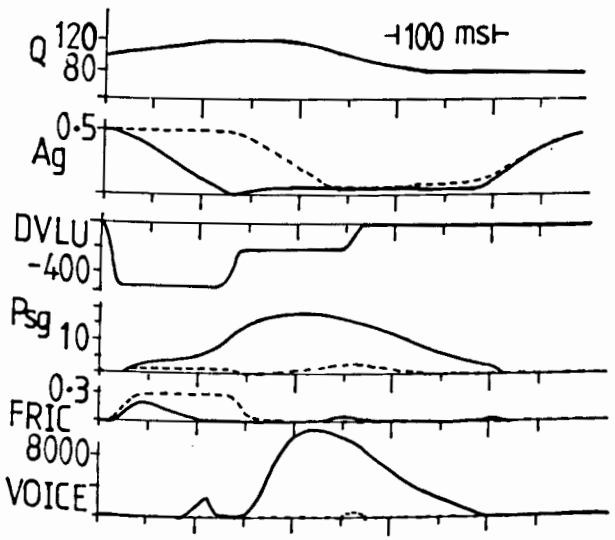

Figure 4. Two contrasting coordinations for the lung walls (DVLU in cm^3/s) and the larynx (Ag in cm^2 and Q in Hz). Also shown: a computed aerodynamic variable Psg in cm H_2O and the envelopes of acoustic sources voice and frication noise (FRIC) in arbitrary units. (a) --- (b) ——.

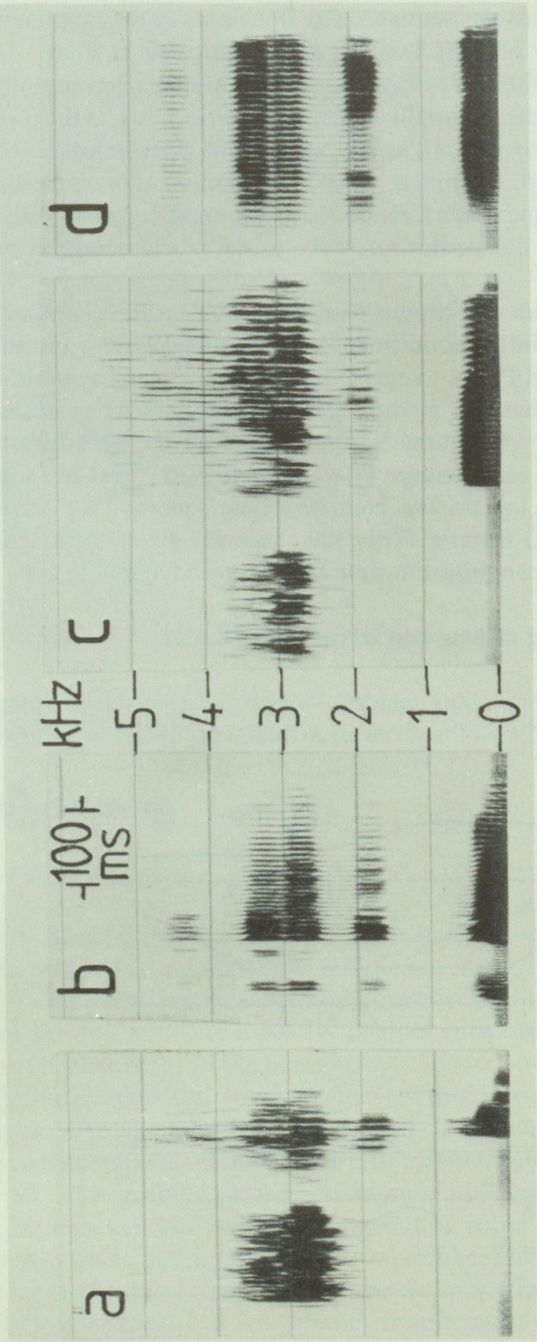

Figure 5. Spectrograms for (a) and (b) in Figure 4 and for additional runs (c) and (d).

together with some of the aerodynamic and acoustic results. Unwanted sounds were generated in both cases. (a) was an attempt at 'braethy' attack. It was transcribed auditorily as [ç^{h}i̥ʔi] with falling pitch. (b) was an attempt at 'hard' (or 'glottalised') attack and was transcribed as [hiʔi·] with 'gulp' effect, sudden onset and falling pitch. Spectrograms for (a) and (b) are shown in Figure 5. Two other unsuccessful attempts at the auditory goal are shown as (c) and (d) in Figure 5. (c) gave [breath drawn in sharply] then [i] falling pitch. (d) gave a 'strong' [i] sound with no audible noise, but not a falling pitch. In another set of syntheses for target words 'purse' and 'purrs', unwanted vowel-like segments were often generated at the speech offset. By trial and error, combinations of lung and larynx actions could be found which avoided unwanted onset and offsets. It is suggested that auditory feedback must be of overwhelming importance for the acquisition of speech, as in our modelling. The onset and offset of speech present speakers with specific problems. The options selected by a particular speaker for the achievement of rather broadly defined auditory goals will be reflected in the details of acoustic structure. Modelling of the kind outlined here may be able to assist in defining the probable acoustic variations within one accent, with potential applications in automatic recognition of speech.

Acknowledgement

This work is supported by the Science and Engineering Research Council, Grant Gr/B/34874.

References

Allwood, E. and Scully, C. (1982). A composite model of speech production. In: *Proceedings of the IEEE International Congress on Acoustics, Speech and Signal Processing, ICASSP 82,* Paris, 932-5.

Fant, G. (1980). Voice source dynamics. STL-QPSR 2-3/80. Stockholm: Department of Speech Communication and Music Acoustics, RIT. 27-37.

Numerical Algorithms Group (1981). Numerical Algorithms Group FORTRAN Library Manual, Mark 8, Vol. 1.

Rothenberg, M. (1968). *The Breath-Stream Dynamics of Simple-Released-Plosive Production.* Bibliotheca Phonetica, No. 6. Basel: Karger.

Scully, C. (1975). A synthesizer study of speech segment durations. In: *Speech Communication* (G. Fant, ed.), Vol. 2, pp. 227-234. Stockholm: Almqvist and Wiksell.