Transmission line modelling of the vocal tract

John Loasby (J.M.Loasby@derby.ac.uk)
School of Engineering
University of Derby

Speech synthesisers typically fall into two classes; the so- called terminal analogue, or parallel, approach which employs a variety of oscillators and filters to mimic the speech waveform directly and the source/filter, or series, approach which sets out, perhaps only in a crude way, to model the human articulatory system. The use of an electrical transmission line analogue to implement the filter for the latter approach was pioneered by Kelly and Lochbaum in the early 1960s. The relationship between linear predictive coding of speech, which has been widely adopted, and the transmission line model has been discussed by a number of authors.

An arrangement by Benkrid and Cross at the University of Nottingham is depicted in Figures 1 and 2. The system is excited by a current waveform representing the sound volume velocity glottal flow and each section is assigned a characteristic impedance representing the cross-sectional area of the vocal tract at that position. The model may be described as a quasi- two-dimensional system and can only represent longitudinal sound flow patterns. The generated output is simply a combination of the outputs from the nasal and oral ports which is fed to a further TLM representing the radiation load. The potential for further sophistication of the model is clear. Never-the-less speechlike sounds are readiliy generated by models of this form.

Since it is highly desirable to operate the model in real time the computational rate demanded is high and multiple digital signal processing systems based on the Texas Instruments 32020 family have been designed and constructed. One system incorporating a total of six processors is depicted in Figure 3.

An essential characteristic of the model is the temporal evolution of the tract profile, and hence elemental characteristic impedances, during articulation. This has led to a discussion about what, if anything, should be done to the travelling waves (forward and backward) within an element when its characteristic impedance is changed. An analysis considering each transmission line element as a parallel conductor pair, assuming conservation of elemental energy during the change and allowing for the work gained or lost by adjusting the conductor spacing (to change the characteristic impedance) leads us to the view that the travelling current waves should be adjusted by the factor

$$\sqrt{\frac{A'}{A}}$$

where A is the original elemental section area and A' is the new elemental section area.

It is important close to closure of an element. Upon closure the area becomes zero and any travelling waves within the element are thereby closed down. Without this correction any travelling waves are totally internally reflected within the closed element to be released on opening - a situation which is clearly not representative of real tract behaviour.

Further work has related to the possibility of employing the model in a search strategy aimed at identifying possible tract configurations which generate a given speech waveform - the so- called inverse problem. Associating one parameter, rate of change of area, with each element and using a systematic search process the author has been able to demonstrate an ability of the system to determine the necessary evolutionary pattern given the glottal waveform and in very simple cases. The error used to guide the search process is derived as a spectral magnitude difference between the required and actual time waveforms.

Figures denote TL14 section identification number.

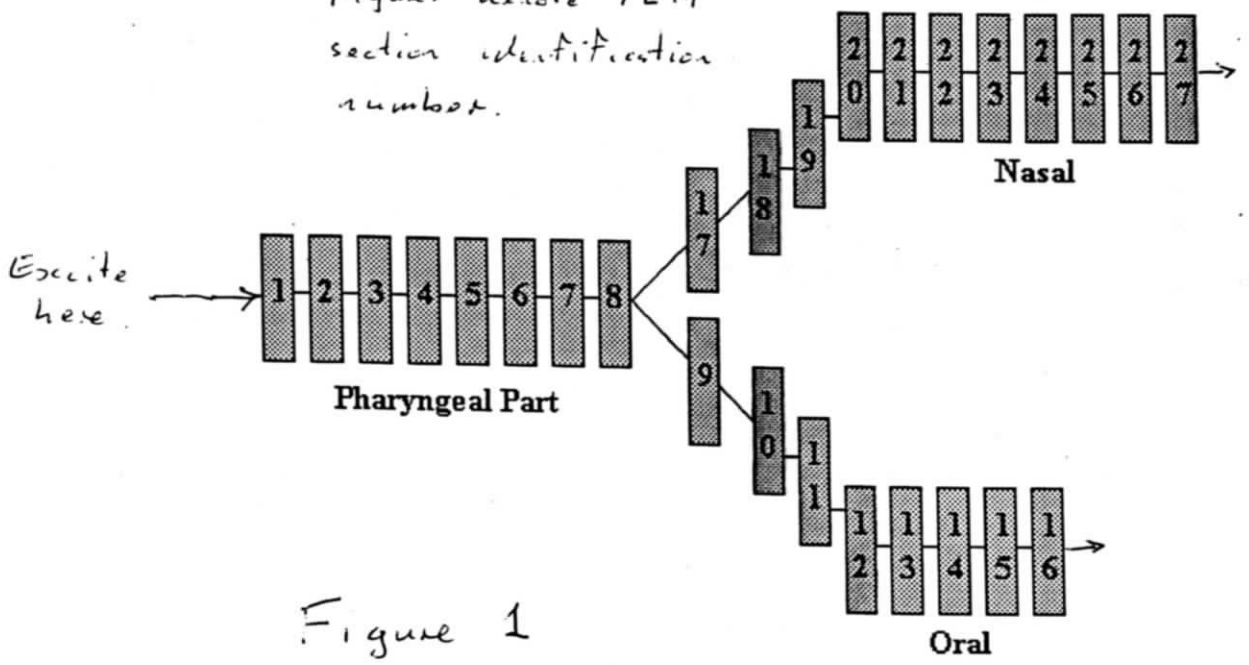


Figure 1

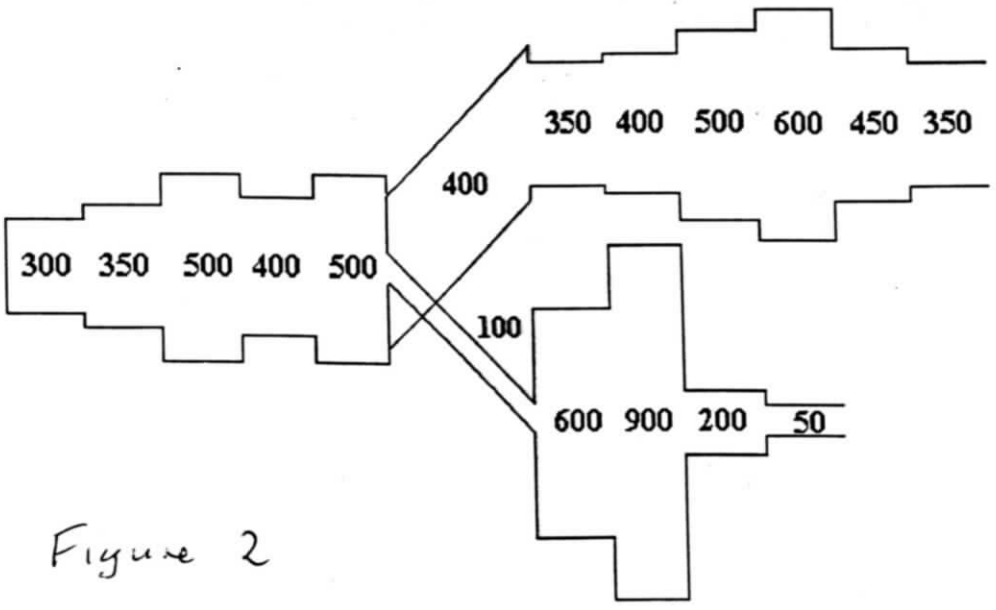
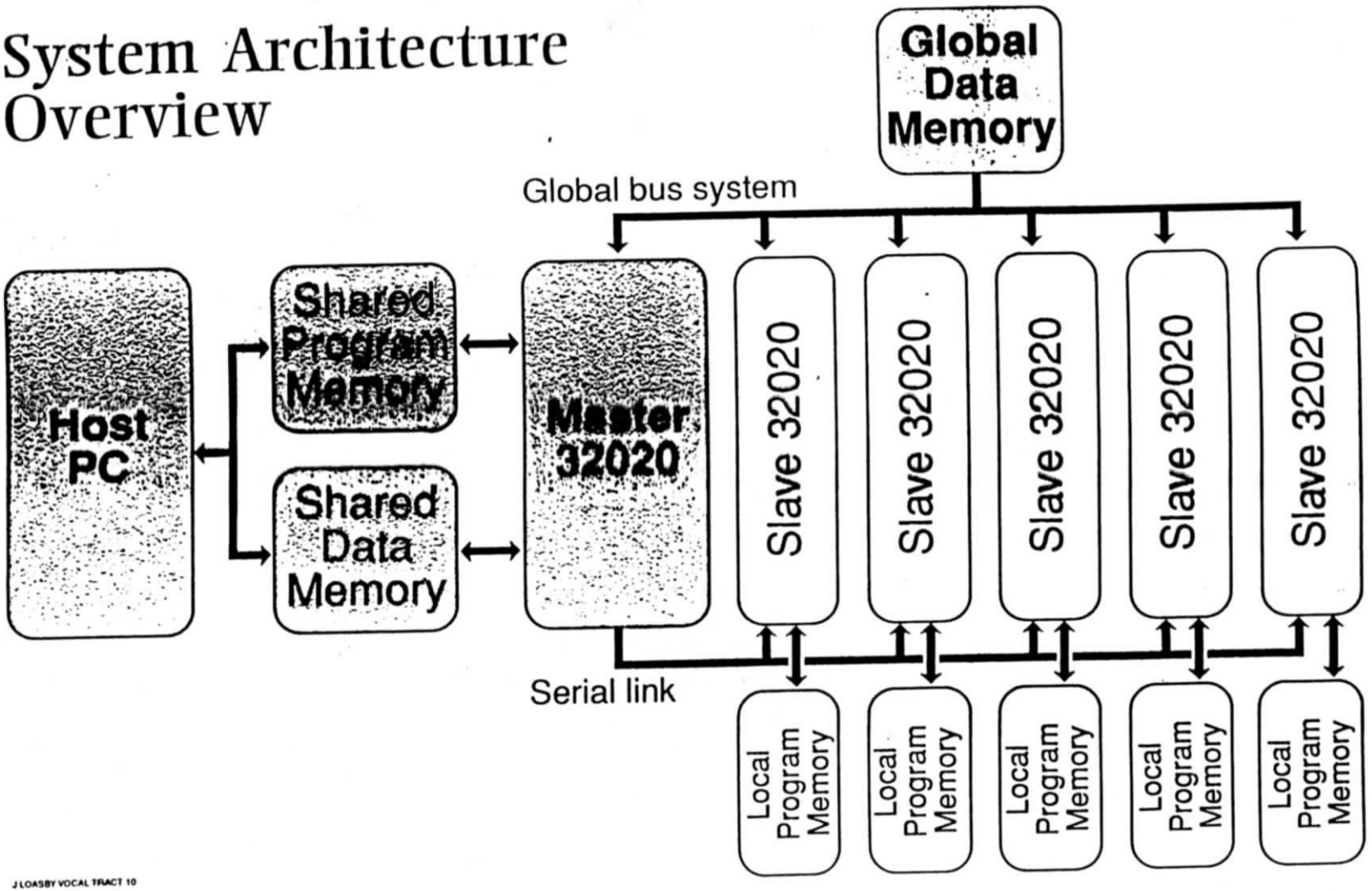


Figure 2

Figures denote section areas. Only relative values significant.

System Architecture Overview



J LOASBY VOCAL TRACT 10

Figure 3