

# A DSP's HARDWARE ARCHITECTURE FOR VOCAL TRACT MODELLING

A Benkrid\* M Koudil \*\* M Atmani\* and TE Cross\*\*\*

\*Electronic Engineering Dpt, College of technology at Dammam, PO box 7650, Saudi Arabia.

E-mail : Benkridah@hotmail.com

\*\* Institut National de formation en Informatique INI Oued smar BP 68M Algiers, Algeria.

\*\*\*Electrical and Electronic Eng. Dpt, University of Nottingham, NG7 2RD,England

## Abstract:

An architecture of a multiprocessor platform based on the Texas Instrument TMS 320's digital signal processor and its application to the problem of real-time modelling of the human vocal tract is discussed. The need for a multiprocessor system is justified and various methods of inter-processor communication are explored. Solutions are also given for the problems associated with the division of the modelling task amongst the processors in the system. Finally the design of a system which has been used to model the human vocal tract in real time will be described.

## 1 Introduction

The introduction over the last years of fast digital signal processors has made it possible to contemplate the construction of computing platforms capable of modelling problems in real time. This presents a challenge in two fields, firstly in the design of the hardware necessary to build such a platform and, secondly, in the formulation of the problem to be modelled in such a way that it can take full advantage of the processing speed available. This paper examines approaches, which may be taken to solve the first problem and considers one in some detail. Solutions to the second problem are also outlined. The relative computational simplicity of the TLM<sup>[1]</sup>. modelling technique opens up the possibility of using the latest generation of digital signal processors to build a multiprocessor platform. In order to guarantee the performance of this modelling, it is essential to calculate the number of sections, which are required to model the system with an acceptable degree of accuracy. It is necessary to establish precision at which the TLM model has to be used to minimize the error arising only from the ability of the transmission line model to represent the system and not from the numerical solution of the model<sup>[2]</sup>. Conventionally, for acceptable performance of a TLM model, the node spacing should be about one tenth of the smallest wavelength to be modelled. With a maximum frequency of 3.4 kHz ( $\lambda_{\min}=10\text{cm}$ ) the section lengths should be about 10mm. A typical vocal tract is 175mm long so using sixteen sections would give a model of acceptable resolution. Using sixteen sections to model the tract gives a time-step (T) of 32 $\mu\text{s}$ . Including the nasal tract, which is typically 120mm long, produces a model with a total of 27 scattering nodes (fig 1.). For a signal bandwidth of 3.4kHz, working at the Nyquist limit, about 147 $\mu\text{s}$  would be available for the calculations at each model time-step. Therefore the task cannot be achieved in real-time by a single

processor<sup>[3]</sup>. Consequently a multiprocessor system is required in order to perform the modelling in real time.

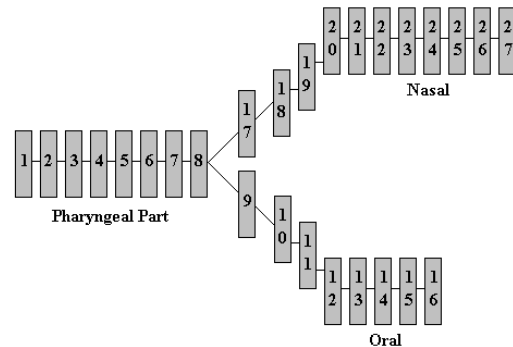


Figure 1.

## 2 System hardware

Although the TMS320 family<sup>[4]</sup> was not explicitly designed to be used in a multi-processor environment it does have some features, which make this possible. It supports various types of inter-processor communication such as direct memory access (dma), global memory, serial and parallel input-output ports.

The vocal tract model can be split amongst a number of processors providing a mechanism exists to pass the output from one section of the model to the input of the adjacent section. This exchange of information may be achieved using any of the techniques mentioned described later in this paper. However, the overheads associated with global memory would be significant as all processors would, in principle, wish to access the global memory simultaneously at the end of each time-step to exchange model values. A similar problem would be encountered using dma techniques. A serial system of communication would be slow and was therefore not considered suitable for the current system. If the model is split amongst a number of

processors then a parallel port may be used to pass the output from the last section on one processor to the first section on an adjacent processor. This technique has a number of advantages, it allows the processor architecture to closely resemble the model and hence the physical situation. The software and hardware overheads involved in passing data from one processor to another are small and so the time delay due to this exchange of information will be negligible. This technique was therefore adopted for this work a schematic of the configuration is shown in Fig. 2.

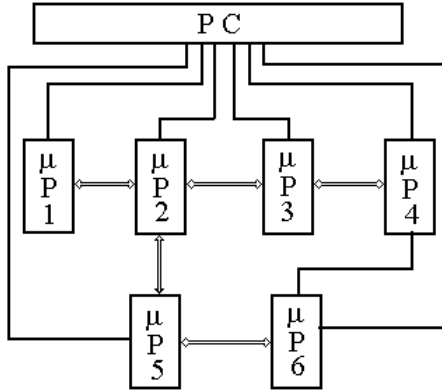


Figure 2.

During the articulation of a sound, the shape of the vocal tract will change and the model parameters must also be changed to reflect this. The host PC is used to calculate the reflection coefficients, which will be used by the models for the subsequent time frame. These numbers are passed to the chip memory using DMA techniques. In order not to interfere with the model calculation, the transfer of reflection coefficients is performed during idle time at the end of a scattering calculation. The whole transfer will take several time-steps to complete. At the beginning of a new time frame, the new reflection coefficients will be transferred from off-chip memory to on-chip memory for use in subsequent calculations.

### 2.1 Inter-processor communication

For the majority of inter-processor links, the hardware required is straightforward and is shown in Fig. 3. However, for the part of the model which includes the six port junction between the nasal, oral and pharyngeal parts of the tract, a more complex link is required, this is illustrated in Fig. 4. The first technique uses the BIO pin<sup>[5]</sup> of each processor to arbitrate the handshake. When a processor has completed its calculation and wishes to exchange information with its adjacent processor, it writes its current output to the common latch and then examines the status of its BIO input. If this is

high, it indicates that the other processor has already written its data. The data can then be read and processing continues. If the BIO is low, the processor will wait for it to become high. The act of

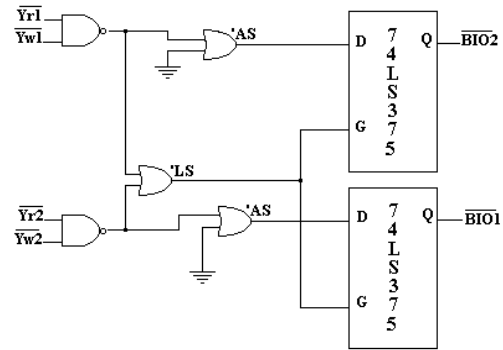


Figure 3.

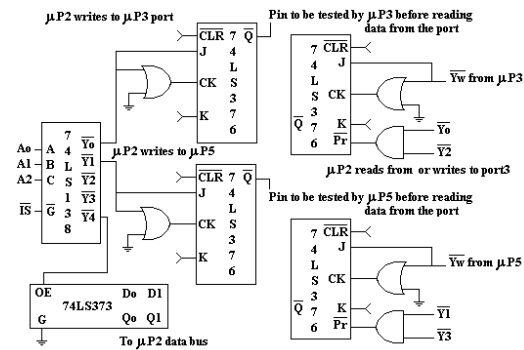


Figure 4.

reading the data from the common latch will reset the BIO input of the processor. This type of handshake is used between μP1 and μP2, μP3 and μP4, μP5 and μP6 (Fig. 2.). The second technique uses two bistables to arbitrate the transfer of data from one processor to another. It is used between μP2 and μP3, μP2 and μP5, (Fig. 2.).

### 2.2 Host PC to processor communication

The handshake between the host PC and the model processors (figure 5.), responsible for simulating the different type of excitation sources and updating vocal tract configuration by sending the corresponding reflection coefficients to the processors uses the XF, HOLD and HOLDA pins of the TMS320 (figure 6.). When XF is low, the corresponding board is performing no computation. During this time, the PC sends a low pulse to the HOLD pin and waits for it to be acknowledged before sending data. Due to short time available at the end of every calculation the transfer is restricted to one byte per time-step. In order to speed up the communication, the software performing this task on the PC is written directly in machine language. Prior to the first data item being sent to the

processor the host loads an on-board counter with the base address, within the TMS32020 off-chip memory space, at which the data will be stored. This counter is automatically incremented after each two bytes have been transferred from the host.

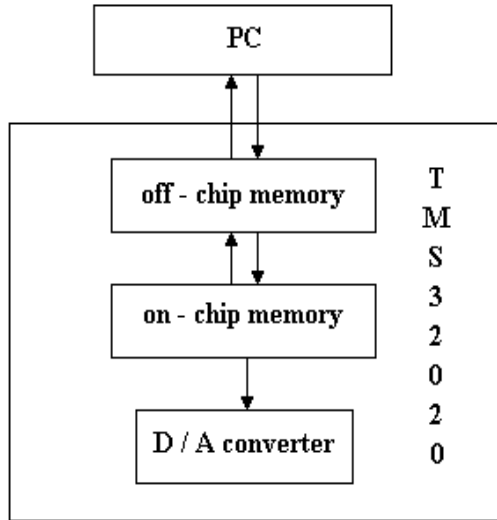


Figure 5.

### 2.3 TMS320's processor boards

In order to validate the design described previously, the part of the system corresponding to  $\mu P1$  and  $\mu P4$  was implemented first, with the other processors being added subsequently. All processor boards are of a similar design except that  $\mu P1$  only requires a single inter-processor communication port[3]. The program held in the EPROM is transferred to on-chip program memory after reset and the model uses entirely on-chip data memory. The off-chip data memory is used to receive coefficients from the host these being transferred by the TMS32020 to on-chip data memory at the appropriate time.

To test the two-processor system, the model was reduced to seventeen sections producing a time-step of  $51\mu s$ . This model was then split over processors  $\mu P1$  and  $\mu P4$ . In this reduced model the five sections simulating the pharyngeal part of the tract, a software implementation of the six port, one section of the nasal part and three sections of the oral part were handled by  $\mu P1$ . The eight remaining sections of the oral and nasal parts of the tract together with the free-space were modelled on  $\mu P4$ .

### 2.4 Partitioning of the tasks on 6 DSPs

In the previous section, the case of two processors was examined. This section described how the task could be divided amongst six processors to obtain a

high performance platform. The four remaining boards are identical to  $\mu P1$  with some additional logic for the second inter-processor port. The system using six processors models the vocal tract using twenty-seven sections as described previously. The corresponding time-step being  $32\mu s$ . About  $7\mu s$  are used for housekeeping and about  $5\mu s$  for host-processor communication.

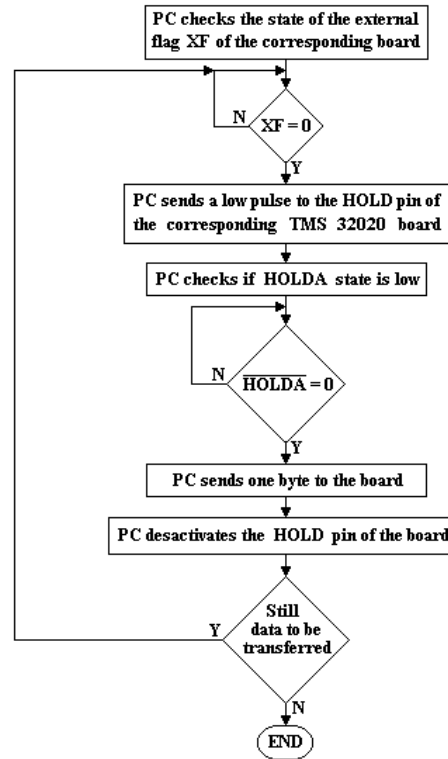


Figure 6.

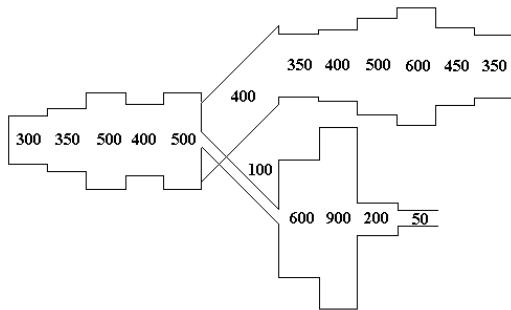
Consequently each processor can handle up to five sections of the model. Processor one is used to compute the first five section (1-5) of the tract and also generates the glottal wave. Simulation of the next four sections (6-8,17) the six-port junction between the pharyngeal, nasal and oral parts of the tract are modelled by  $\mu P2$ . Processor three simulates four sections of the oral tract (9-12) and also acts as a noise source for fricatives or steady state source for stops. this can be injected in any one of the four sections. This last four sections of the oral tract (13-16) and the modelling of the oral free space impedance is performed by  $\mu P4$ . It also can act as a noise or steady source generator. Processor four receives the nasal output from  $\mu P6$ . Processors five and six handle the remaining ten sections of the nasal tract (18-27) and the nasal free space impedance.

## 3 Results

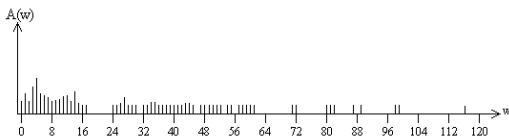
In order to test the performance of the modelling platform, simulations were run so the results of

which could be compared with previous non real-time results<sup>[3]</sup>. This was in agreement with results obtained from non real-time simulations<sup>[3]</sup>.

The performance of the system was established by simulating several vowel sounds Figs 7a and 8a show the model profiles for the sounds /m/ and /s/ respectively. The nasal sound /m/ is generated by a closing of the lips, a lowering of the velum to inject energy into the nasal tract and a glottal wave excitation. The fricative /s/ is produced by noise injection into the oral tract with isolation of the nasal part. the spectrum of the sound /m/ shown in Fig. 7.b shows formants at 270, 1090 and 2030 Hz which is close to those found in typical speech<sup>[3]</sup>. The output for the sound /s/ (fig 8b.) sound is also in agreement with previous work<sup>[3]</sup>.

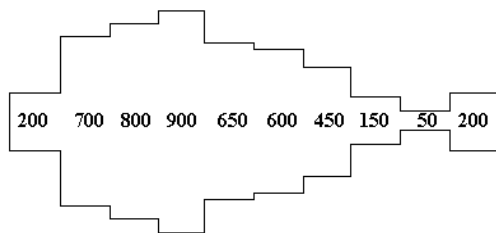


**Figure 7a.**

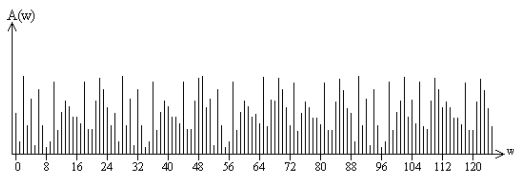


256 pts FFT from the I/O model above. The resolution of the FFT is equal to 78 Hz.

**Figure 7b**



**Figure 8a.**



256 pts FFT from the I/O model above. The resolution of the FFT is equal to 78 Hz.

**Figure 8b.**

## 4 Conclusion

The work described here has shown the feasibility of the development of a multi-processor TMS320's based platform for real-time modelling. This has enabled a physiologically realistic model of the vocal tract using the well-established technique of transmission line modelling to be developed. The appearance of DSP chips<sup>[6-10]</sup> with increased word size, floating point capability and increased support for parallelism would permits the construction of yet more powerful systems with higher performance and the architecture of the different boards are similar to those explained elsewhere<sup>[11]</sup>.

## REFERENCES.

- 1 P B Johns and R L Beurle, «Numerical Solution of 2-dimensional Scattering Problems Using TLM », Proc IEE Vol118, n°9, pp1203-1214, Sept 1971.
- 2 J W Bandler, P B Johns and M R M Rizk « Transmission Line Modelling and Sensitivity Evaluation for Lumped Network Simulation and Design in the Time Domain », Journal of the Franklin Institute, vol 304, N° 1, pp15-31, July 1977.
- 3 A.Benkrid, « Real-time TLM Vocal Tract Modelling », PhD Thesis, University of Nottingham, 1989.
- 4 « TMS32020 User's Guide » Texas instruments.
- 5 « TMS32020 board » User manual Loughborough Sound Images Ltd. June 1986.
- 6 TMS320C25 User's Guide. TEXAS INSTRUMENTS.
- 7 TMS320C30 User's Guide. Preliminary TEXAS INSTRUMENTS.
- 8 TMS320C40 User's Guide. Preliminary TEXAS INSTRUMENTS.
- 9 32-Bit Third Generation RISC Microprocessor. MOTOROLA INC., 1988.
- 10 H J Mitchell « 32-Bit Microprocessor » Williams Collins sons & Co Ltd. 1985.
- 11 A Benkrid TE Cross M Atmani and M Koudil "A multiprocessor TMS320 platform for a real-time modelling" IEEE, TEM, No6, pp174-178, 1999.

