First European Conference on
Speech Communication and Technology
EUROSPEECH '89
Paris, France, September 27-29, 1989

Multi-DSP and VQ-ASIC Based Acoustic Front-End for Real-Time Speech Processing Tasks

Abdulmesih Aktas and Harald Höge

Siemens AG, München
Otto-Hahn-Ring 6, D-8000 München 83, West Germany

Abstract

This paper describes the architecture of a multi-DSP based acoustic front-end (AkuFE) developed within the speaker adaptive continuous speech understanding and dialogue system SPICOS[1]. The AkuFE is designed as a configurable high performance signal processing VMEbus system employing up to five Texas Instruments TMS320C25 signal processors and an ASIC for vector quantization (VQ) developed within the project. The system is realized on three boards and achieves a total computational power of more than 100 MIPs. In this case two VQ processors can work in parallel. A 68020 based workstation serves as host computer. The AkuFE is employed for the real-time acoustic-phonetic decoding task in the SPICOS system. Due to its flexibility, it can be used for a wide range of real-time speech processing tasks.

1. Introduction

The multi-DSP and ASIC based acoustic front-end (AkuFE) has been developed in the framework of the SPICOS system /1/ and is build up of three boards: A so called master board and two slave boards. The system is designed for the analysis and synthesis of telephone quality speech (8 KHz sampling frequency and 3.2 KHz bandwidth) and high quality speech (16 KHz sampling frequency and 6.4 KHz bandwidth). *Figure 1* shows an general overview of the AkuFE system.

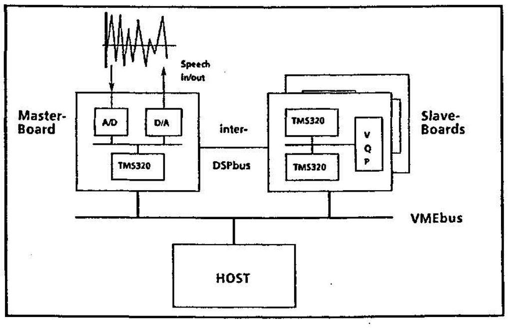

Figure 1: The multi-DSP acoustic front-end system (AkuFE)

The master board is the basic control system of an extendable multi-DSP architecture. The complete system is based on Texas Instruments TMS320C25 digital signal processor VLSI chips and built as a VMEbus system. Within the distributed architecture of SPICOS /2/ it is plugged in a SUN workstation and employed for the real-time acoustic-phonetic decoding task. The master board of the AkuFE consists of an analog (A/D and D/A converter) and digital subsystem, while the slave board employs two TMS320 DSPs and a VLSI chip for VQ /3/. A local data and program memory is dedicated to each signal processor. Two slave boards can be connected to a master board, achieving a computational power of more than 100 MIPs. But the slave boards can be used also without a master board.

The inter-DSPbus allows fast communication between the master and the slave processors without interfering with the VMEbus. A global memory scheme is realized in order to perform data transfers between the processors. Control is handled by interrupts. In order to test, download and control the system, the host machine has read and write access to the whole memory of the AkuFE system. Because of its configurability and free programmability, the AkuFE system can be used in a wide range of signal processing applications, such as speech analysis and synthesis, speech coding and speech recognition. Used as a fast array processor for algorithms with extensive computational requirements, it can highly improve the performance of a standard workstation. The codebook generation /4/ particularly can be accelerated by employing the VQ Chip on the slave board in connection with a host processor. Further applications are spectral analysis, instrumentation and image processing.

A more detailed description of the master board is given in section 2. In section 3 the description of the slave DSP subsystems and the VQP follows. Finally an application of AkuFE system for real-time acoustic-phonetic decoding is given in Chapter 4. Performance results for the specific application are given.

It should be mentioned that the notation master and slave are not used here in the sense of VMEbus definitions.

----

[1] SPICOS (Siemens, Philips, IPO Continuous Speech Recognition and Understanding) is carried out as a joint project and sponsored by the German Federal Ministry for Research and Technology (BMFT) under the grant No. ITM 8801 B9.

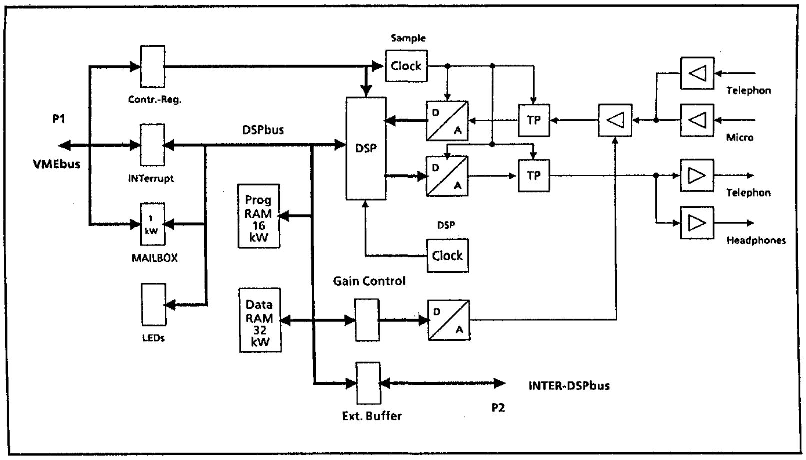

Figure 2: Block diagram of the AkuFE master board

2. The AkuFE Master System

The master board /5/ performs the initial signal preprocessing steps, like amplification, sampling, A/D and D/A conversion. Computationally intensive operations required for a spectral or time analysis can be executed on the DSP subsystem. Further task of this board is the control of data flow to the host machine over a so-called Mailbox. *Figure 2* shows the general block diagram of the AkuFE master board.

2.1. Analog Subsystem

Two different audio channels, one for telephone and one for a high quality microphone input, are provided. The board allows the signal acquisition, filtering and sampling of telephone quality speech (8 KHz sampling frequency and 3.4 KHz bandwidth) and high quality speech (16 KHz sampling frequency and 6.8 KHz bandwidth). The corner frequency of the anti-aliasing filters for both sampling rates are software selectable. In order to increase the dynamic range of the 12 Bit A/D converter, a programmable gain preamplifier (0-24dB) is implemented.

The D/A subsystems consists of a 12 bit D/A converter and the associated interpolation filter with two different software selectable corner frequencies.

The master board provides further outputs for a telephone and a headset.

2.2. Digital Subsystem

The second generation chip TMS320C25 from Texas Instruments provides a high performance 16 bit digital subsystem for the AkuFE. The modified Harvard architecture of the processor allows a fast memory access and high data through-put with a 100 ns instruction cycle. The CMOS version of the DSP is capable of executing ten million instructions per second, where most of the instructions require one cycle.

A RAM based local data and program memory is dedicated to the DSP. The host processor has read and write access to the whole memory. A local data memory of 32 KWords and program memory of 16 KWords are provided on the master board. The remaining 32 KWords of data memory are reserved for global memory communication areas of four further DSP subsystems located on two slave boards. The memory expansion and the fast communication to the slave board subsystems is performed via inter-DSP bus. Data exchange can be done via VMEbus interface as well.

A data memory region of 1 KWord called MAILBOX is provided as dual-ported memory enabling a parallel access from both the DSP and the host processor without holding the DSP during a block data transfer routine. This allows, e.g.in case of signal acquisition, a real-time transfer of the preprocessed and analyzed data.

For more flexible VMEbus communication an interrupt control register is provided. Hold and reset control is performed by a further register which additionally can be used to toggle between the sampling rates. A third register allows the digital control of the gain amplifier. Several LEDs provide visual control of the communication.

3. The AkuFE Slave System

The AkuFE slave is a powerful optional system that can also be used without the AkuFE master board /6/. The two TMS320C25 based subsystems and the vector quantization processor (VQP) provides a computational power of about 50 MIPs and is appropriate for use as an array processor in many speech applications with high computational requirements. Two slave boards can be used in conjunction with the AkuFE master board.

Local memory is attached to each DSP subsystem. In order to test, download and control the system, the host machine has read and write access to the whole memory of the AkuFE slave system. Four memory banks each with 1 Mbit are available, allowing storage of codebooks up to the size 8,000 in our application. The mapping of the VQP control register into the global memory area permits the control of codebook search from either one of the DSP subsystem of the AkuFE or from host computer. *Figure 3* shows the basic VQP-DSP interfacing.

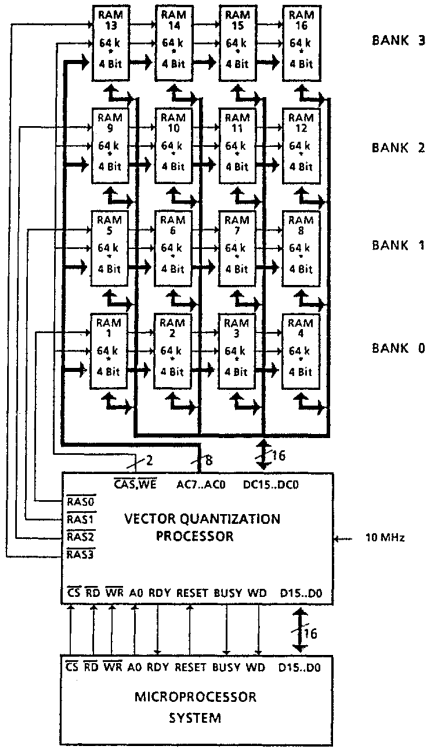

Figure 3: Vector Quantization Processor and DSP interfacing

3.1. DSP Subsystems

Two TMS320C25 based subsystems can be employed for parallel processing. Each subsystem has a dedicated data/program memory of 56 KWords/64 KWords. A data area of 8 KWords is reserved for global communication. Each global RAM area is accessible from all processors, including the VQPs. A priority controlled access is implemented. The master DSP has the highest priority. The serial input of each DSP can be used to build up a second communication path to other processors.

The very highly flexible interrupt handling allows the implementation of various software architectures and communication schemes. A bus arbiter controls the access to the internal DSP bus.

3.2. Vector Quantization Processor (VQP)

As codebook search is one of the most computationally extensive operations in speech processing, a VLSI Chip using high concurrency was developed to perform this task for real-time applications /3/. The VQP implements a full search algorithm. It supports two software selectable distance measures for the search : Euclidian and city-block. The VQP is able to handle vectors of variable dimensions, codebooks of programmable size. Each codebook vector can be extended to up to 64 dimensions of 8 Bit components, without noticeable loss of performance. The chip delivers the codebook entry and the distance of the best codebook candidate.

Finally, some important features of this ASIC should be mentioned:

- on-chip cache RAM for input vector
- vector dimensions up to 64 with word-length of 8 bit
- 4 software selectable codebook banks, each of 1Mbit
- throughput rate of 10^7 vector components/s
- on-chip watchdog
- 16 bit parallel interface
- easy communication with the host processor
- needs two addresses in the host addressing space

4. Real-time Acoustic Phonetic Decoding

Within the SPICOS system, a man-machine dialogue system allowing data queries about office activities using a vocabulary of about 1,000 words, the AkuFE is used for real-time acoustic phonetic decoding and diphone synthesis /8/. The acoustic-phonetic decoding is based on articulatory features which describe the place (labial, dental, alveolar, etc.) and the manner (consonant, liquid, nasal, plosive, etc.) of articulation /7/ The feature vector generated from these categories is extracted every 10 ms and used by subsequent Hidden Markov Models for the recognition of subword units. It has been shown /7/ that the mapping of the speech signal onto articulatory categories has several advantages, such as the reduction of the feature set and speaker invariability. For the place of articulation, formants are required. Instead of implementing a computationally expensive root solving algorithm to calculate the LPC poles, a faster alternative approach using vector quantization was chosen. The spectral peaks in precomputed LPC spectra are stored in a codebook and used as formant candidates.

The following processing steps are performed on the AkuFE: The master board performs the primarily signal preprocessing steps, like amplification, sampling, A/D and D/A conversion with a linear resolution of 12 bits. The analogue speech signal is low-pass filtered at a corner frequency of 6.4 kHz and digitized with a sampling rate of 16 KHz. Further tasks of this board are the control of data flow over a so-called Mailbox and the computation of some basic parameters like signal energy and zero crossings.

The acoustic analysis and synthesis algorithms are implemented on the second board. In order to reach high spectral resolution and detect reliable formants in the 0-6.5 KHz region of the spectrum, a 16th order LPC analysis is performed. The LeRoux and Guegen algorithm is implemented in mixed 16/32-bit arithmetic /9/.

The VLSI VQ-Chip performs a full search over 4,000 LPC codebook entries represented by 16 coded reflection coefficients each coded with 8 bits. The codebook entries deliver formant candidates and corresponding sub-band energies for the succeeding formant tracking algorithm. Both formant tracking and articulatory analysis with the composition of the AFV consisting of the occurrence probabilities of articulatory categories /7/ are performed on the host machine.

The homogeneous articulatory feature vector which is composed of probabilities describing the manner and place of articulation, serves as an input parameter for a phoneme recognition scheme based on HMMs. Finally the phoneme models are concatenated in order to model 400 subword units of consonant clusters and syllabic nuclei. The Viterbi algorithm to recognize these clusters is now going to be implemented on an accelerator board with an ASIC. This board will be built up of a high performance 32 bit floating-point DSP and and a RISC processor.

5. Performance

The processing steps, from the sampled speech data through the calculation of the reflection coefficients, require about 8 ms on a TMS320C25, leading to a frame rate of 10 ms. Besides the weighting and calculation of the autocorrelation coefficients, which require about 1.5 ms, about 0.5 ms is necessary for the double-buffered processing and real-time signal acquisition. As the VQ processor can perform a full search over the total codebook in less then 9 ms, the processing can be performed in real-time.

6. References

/1/ M. Brenner, H. Höge, E. Marschall and J. Romano: "Word Recognition in Continuous Speech Using a Phonological-Based Two Network Matching Parser and a Synthesis Based Prediction", Proc. of the ICASSP '89, Glasgow/U.K. (1989), pp. 457-460

/2/ H. Höge : "Verteilte Prozessarchitektur für ein sprachverstehendes Echtzeitsystem", Proc. of the DAGM-Symposium, Braunschweig/W. Germany (1987), pp. 106-110

/3/ E. Preiss, A. Stölzle, W. Drews and J. Pandel: "Architecture of a CMOS VLSI Vector Quantization Processor", Proc. of the EUSIPCO 88, Grenoble/ France (1988), Vol. 3, pp.1241-1244

/4/ T.M. Liu and H. Höge: "Phonetically Based LPC Vector Quantization of High Quality Speech", Proc. of this Conference

/5/ G. Gohn and A. Peham: "Pflichtenheft, Akustisches-Frontend", Internal Report, Siemens Wien (1986)

/6/ S. Szikora: "AkuFE-Slave, Spezifikationen", Internal Report, Siemens Wien (1988)

/7/ O. Schmidbauer : " Robust Modelling of Systematic Variabilities in Continuous Speech Incorporating Acoustic Articulatory Relations", Proc. of te ICASSP '89, Glasgow/U.K. (1989), pp 616-619

/8/ A. Aktas and H. Höge: "Real-Time Recognition of Subword Units on a Hybrid Multi-DSP/ASIC Based Acoustic Front-End", Proc. of the ICASSP '89, Glasgow/U.K. (1989), pp. 100-104

/9/ J. Sokat : "Untersuchung der Autokorrelation-Ladder-Algorithmen zur Berechnung von Prädik-tionsfilter-Koeffizienten in Integer-arithmetik", Diplom-Thesis, Uni Duisburg/W. Germany (1987)