ELECTRONIC SPEECH SYNTHESIS

Techniques, Technology and Applications

Geoff Bristow

CHAPTER 9

CHIPS USING TIME-DOMAIN SYNTHESIS

Joe Costello and Forrest Mozer (9.1, 9.2), Toshio Oura (9.3), Tom Durgavichs (9.4)

In this chapter we explore two techniques of time-domain synthesis which were developed specifically with silicon implementation in mind

9.3 *Speech segment synthesis*

In this section an alternative method of time-domain synthesis will be described in which idealised speech segments are stored in memory and read out every pitch period. This technique has much in common with the Mozer technique and is representative of the methods used by several manufacturers. However, as an example a device manufactured by the Nippon Electric Company will be described here.

The device chosen is a single-chip 8-bit *Sound Synthesis Microcomputer* (SSM), the μPD1776C, containing 64 kbits of on-chip

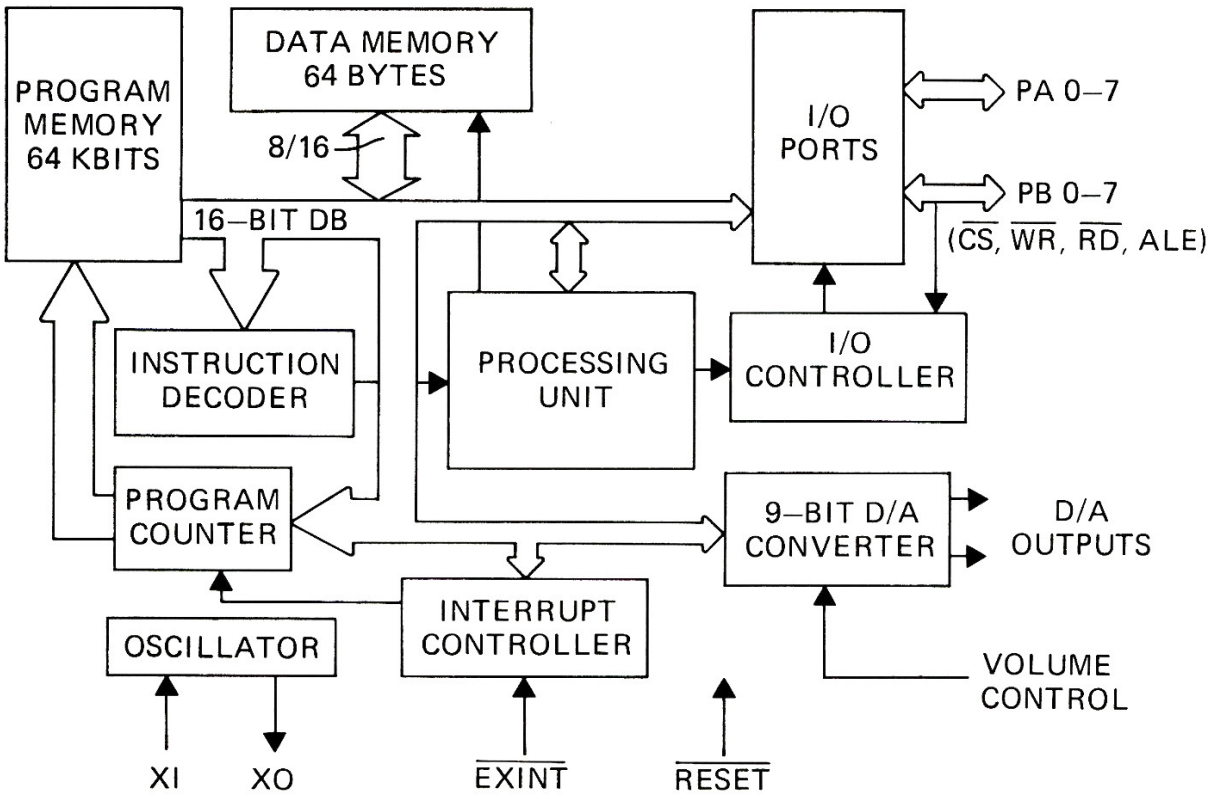

Figure 9.7 Block diagram of sound synthesis microcomputer

program memory (ROM) and 64 bytes of data memory (RAM). The SSM can synthesise high quality speech, musical sounds and imitation sounds with a low bit rate using a method known as *Variable Sampling Rate Speech Segment Synthesis* (VSRSSS). The μPD1776C is, in fact, a member of a family of SSM devices, known as the μPD1770 series.

THE SOUND SYNTHESIS MICROCOMPUTER

Figure 9.7 shows a block diagram of the SSM. It has a set of 182 instructions which mostly execute in one machine cycle (1.33 μs). To facilitate the high speed processing required of a microcomputer approach to synthesis the word length of the ROM is 16 bit. In the ROM, the program for the sound synthesis, and all the speech data are stored.

Thirty-two bytes of the 64-byte RAM are accessible directly and can be used to store table addresses as 16-bit Pair Registers, or used as temporary memory for speech parameters. The rest of the on-chip RAM can be used as an eight-level stack register and can be addressed in either 16-bit or 8-bit units. The processing unit contains an 8-bit special ALU (Arithmetic/Logic Unit), as well as multiplication registers, a quasi-random noise generator and other logic. In addition to the general functions of an ordinary ALU, the processing unit can execute a high speed multiply between a speech segment and suitable envelope data.

To enable the SSM to be used in the read-out of speech segments, there is an on-board interrupt controller which determines the sampling rate. The interrupt controller contains an 8-bit timer and 9-bit binary counter and the timer periodically generates a *tone interrupt* signal determined by a divide ratio. The minimum period of the tone interrupt is identical to sixty-four times one machine cycle. Two kinds of interrupt signals are generated from the binary counter; one is a *noise interrupt*, selected as one of four periods (of 32, 64, 256 or 512 times the machine cycle), while the other is a *timed interrupt* of 512 machine cycles. An external interrupt may also be generated by the *EXINT* line. These interrupts are enabled by the *Mode Register* and only one of them may be accepted at one time.

The on-chip 9-bit D/A converter can drive directly a 32–64 ohm speaker at 200 mW output power, and the maximum output power can be varied by the volume control line. This D/A converter can supply sound signals with up to 23.4 kHz sampling rate, as generated by the VSRSSS method.

There are also two multi-purpose 9-bit I/O ports, which have four modes determined by the mode register. The *PB* port can be set to receive control signals from other microcomputers, or can be set to supply signals to peripherals by software, so that the SSM can control them. The SSM can receive and transmit byte-wide data through the PA port connected to an external data bus. It can also interface to a key-matrix directly and/or drive display devices as a stand alone microcomputer. The PB port is TTL and key-matrix compatible, and the PA port is TTL compatible, allowing the SSM to be applied to various systems. Figure 9.8 shows an example of a single-chip system and fig. 9.9 shows an example of a multi-chip system.

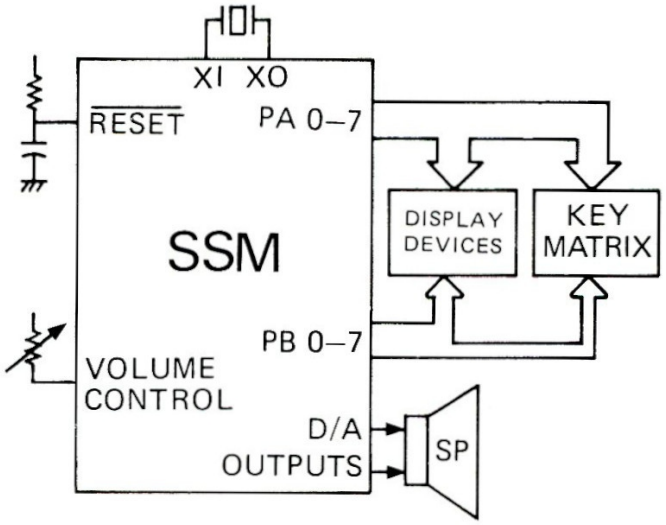

Figure 9.8 Example of single-chip system

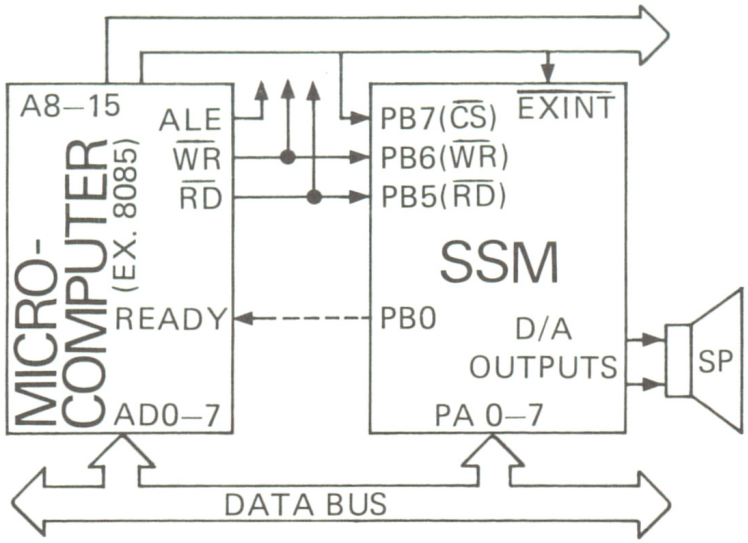

Figure 9.9 Example of multi-chip system

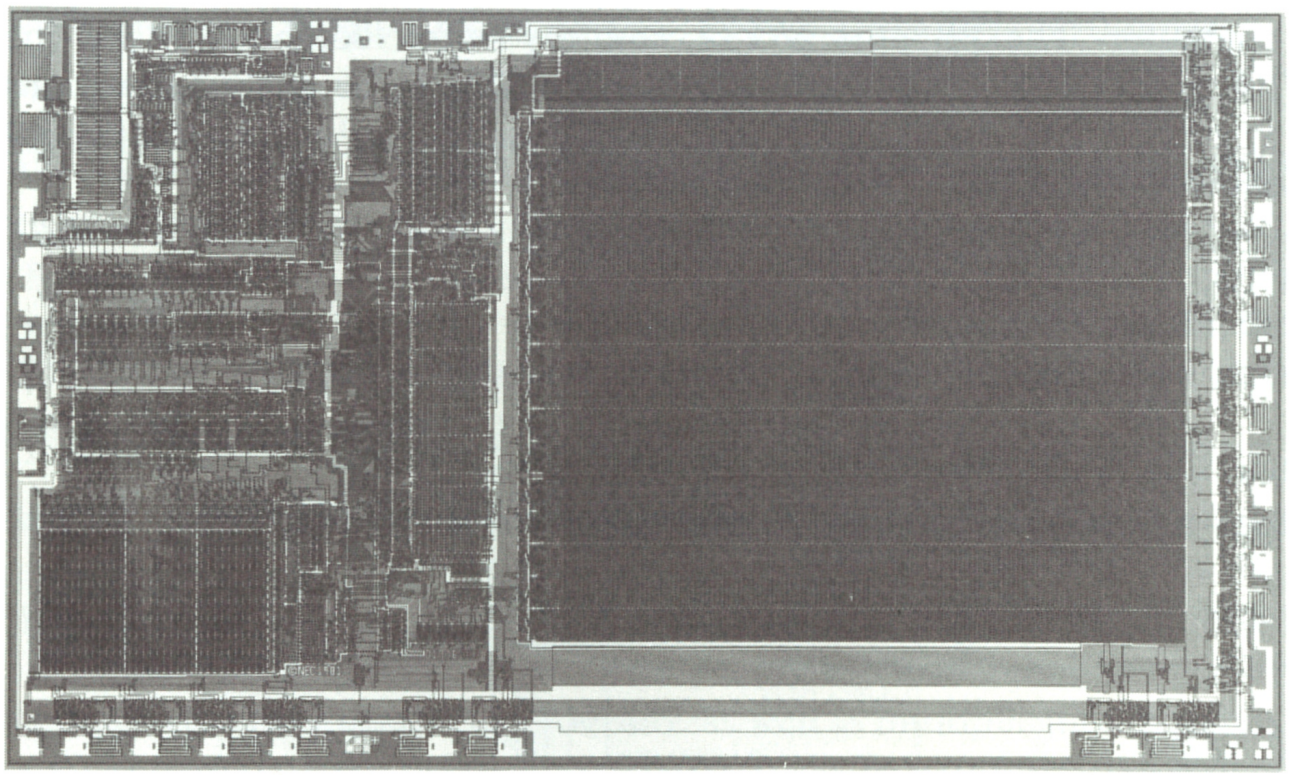

Figure 9.10 Photomicrograph of the SSM μPD1776C. (Reproduced by permission of the Nippon Electric Co. Ltd)

Figure 9.10 shows a photomicrograph of the SSM chip. It is fabricated by an N-channel MOS process, contains 47 000 transistors and has a die size of 4.83 mm × 8.09 mm (246 mil square). It can operate with a single power supply of +4.5 to +6.6 *V* with a 6 MHz clock. In operation, the power dissipation is 450 mW, but on stand-by, it is down to 50 μW. It is housed in 28-pin plastic DIP.

THE VSRSS METHOD

Figure 9.11 shows the Variable Sampling Rate Speech Segment Synthesis method. The ROM contains the software for speech synthesis, speech segments normalised for amplitude and time, envelope data, pitch data and sequence data. The *Normalised Speech Segments* (NSS), envelope data and pitch data are read from the ROM in order, in accordance with the sequence data. The sampling rate of the NSS is varied by the pitch data, namely the divide ratio. The processing unit multiplies the NSS data and the envelope data every sample, and supplies multiplied data to the D/A converter. The shape of the envelope and the pitch can be freely varied along the time axis by changing the appropriate parameters, and the amplitude of the envelope can also be varied, as can the number of times that the NSS data are repeated. The sequence data table can contain commands of *subroutine call*, *jump* and *return* for the current table, as well as *initial set*, *change* and *rest*.

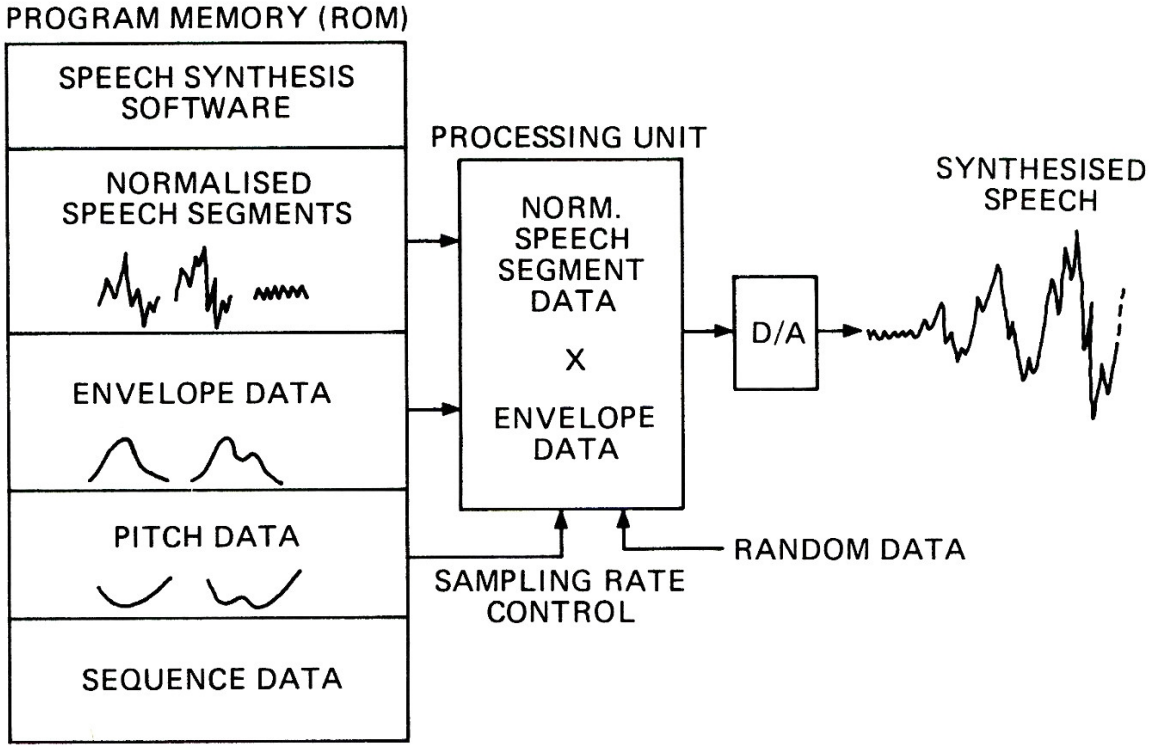

Figure 9.11 Variable sampling rate speech segment synthesis

As the SSM can synthesise speech by varying the amplitude and pitch of the NSS, and changing the NSS, and as it can repeat speech data sequences using subroutines, the data compression is higher than for many other speech synthesisers. It is capable of synthesising 90 seconds of high quality female speech with a bit rate of about 700 bits/s, or about 50 seconds of male speech. The SSM can also synthesise music and imitation sounds, even during the execution of a main program for control or numerical computation. Also, although the device has only been shown here operating with the VSRSSS method (for which it was designed), it is also capable of implementing other speech synthesis techniques, such as *Fixed Sampling Rate Speech Segment Synthesis* (FSRSSS), *ADPCM*, *DPCM*, *ADM*, *DM* and *PCM*. The only changes required to alter the synthesis method, or the speech to be uttered, will be in the software ROM.

ACKNOWLEDGEMENT

The author of this section wishes to thank M. Sakai, M. Endo, H. Aoyama, I. Fujitaka, T. Mukawa, E. Sugimoto, F. Tsukuda and N. Miyake for their helpful support.

Oura, Toshio

Currently supervisor of the special-purpose microcomputer department of the Nippon Electric Co. Ltd, Japan, Toshio has been engaged in MOS LSI design at NEC since 1970. He holds five patents and has worked with designs for calculators, video games and speech synthesis. He received a BS degree from Toyama University and is a member of the Institute of Electronics and Communications Engineers of Japan. (*section 9.3*)