

Software considerations for adding a speech recording system to a Digital cellular telephone

1.1 Abstract

Explosive growth in the digital cellular market has provided an opportunity and a challenge to cellular phone manufacturers at the same time. Their customers are demanding smaller size, lower cost, longer talk times, data modem interface and other attractive features such as a built-in speech recording system. This article features software considerations for adding speech recording to GSM cellular telephone using traditional digital memory storage and an innovative voice recorder chip using multi-level analog storage. Conclusions are drawn in favor of the innovation for software simplicity, ease of implementation, time-to-market advantage, and relative power consumption.

2.1 Introduction

Product differentiation has become a necessity for GSM phone manufacturers in a very competitive market place. Time to market and design cycle time reduction is important to every phone manufacturer to remain profitable and gain market share. A brief description of the GSM cellular phone architecture and hardware design considerations for speech recording and playback in digital and analog memory have been presented in an earlier white paper¹. This paper will focus mainly on the software considerations for the recording and playback of speech in conventional digital memory and the ISD33000 family single-chip multi-level storage (MLS) analog memory. A simple analysis between the two approaches will be shown for memo recording, off-the-air recording, and playback modes. Conclusions will be drawn based on the simplicity and elegance of the analog recording scheme through the analysis of software requirements and power consumption.

2.2 The GSM Digital Cell Phone

The typical base-band signal processing for a GSM phone is shown in Figure 1.

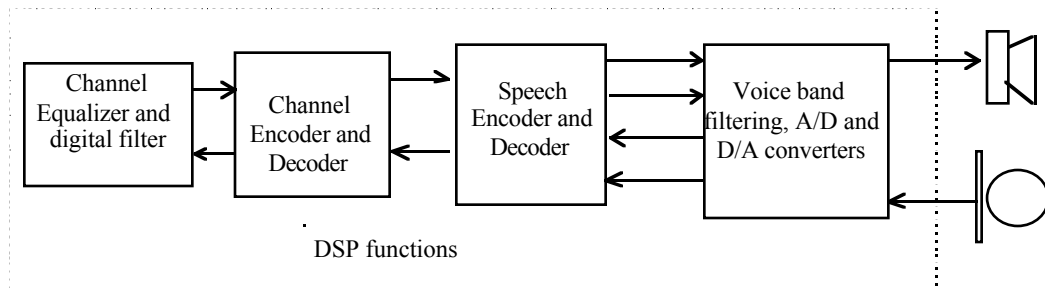


Figure 1

The DSP (Digital Signal Processor) in a GSM digital cell phone has numerous tasks to perform in a very short period of time. Its list of functions include digital filtering, Viterbi decoding, channel equalization, block coding and convolutional coding, interleaving, channel decoding, speech encoding, speech decoding and voice band filtering including voice activity detection (VAD), comfort noise insertion (CNI), lost speech frame substitution and muting. In addition to these tasks, data modulation, demodulation, burst building, tone generation, echo cancellation, etc. are also required which further tend to maximize the MIPS usage of the DSP. In order to perform these tasks within the required time frame requires a dedicated DSP with enough capability to conform to the GSM specifications.

There is an economic pressure on the phone manufacturers to squeeze the last ounce of performance out of a DSP, but at the same time there is pressure to reduce the MIPS requirements for longer battery life.

Hence, power consumption should be given special attention during the discussions here in this paper.

The GSM (Global System for Mobile Communications) is a Time-Division Multiple Access (TDMA) system for mobile communication of voice and data information. There are 124 transmit and receive channels in the system with 200 KHz bandwidth for each channel. The data rate on each channel is 270.83 kb/s. On each transmit or receive channel, there are frames of 4.615 ms duration, and each frame has eight time slots. Each time slot is assigned to a single mobile subscriber. Thus it is necessary to complete all transmit or receive signal processing and data handling requirements in this time slot.

GSM uses RPE-LTP (Regular Pulse Excitation with Long Term Prediction) as the speech coder. The full rate speech coder has a rate of 13 kb/s for the speech data. The speech coded operates on a 13-bit speech data sample with a sampling rate of 8 ksp/s. There are 160 samples in each 20 ms speech window, which are converted to coefficients of 260 bitsⁱⁱ.

Since we know the time slot duration, channel data rate, speech rate and speech coder requirements for GSM, we can estimate the MIPS requirements for various functions. DSP performs channel encoding/decoding (4.5 MIPS), GMSK modem and data filtering (15 MIPS), Viterbi decoding and channel equalization (6 MIPS), speech encoding/decoding (5 MIPS), voice activity detection (0.5 MIPS), etc.ⁱⁱⁱ Thus a minimum of 35-40 MIPS is required to implement the GSM. Additionally, the system micro-controller is active with the GSM protocol implementation, data handling, timing, radio resource management and man-machine interface.

Now that we understand the critical timing, speech processing, MIPS requirements and the micro-controller loading, we will analyze digital and analog speech storage schemes in the following sections, giving special attention to power consumption, software overhead, and implementation time.

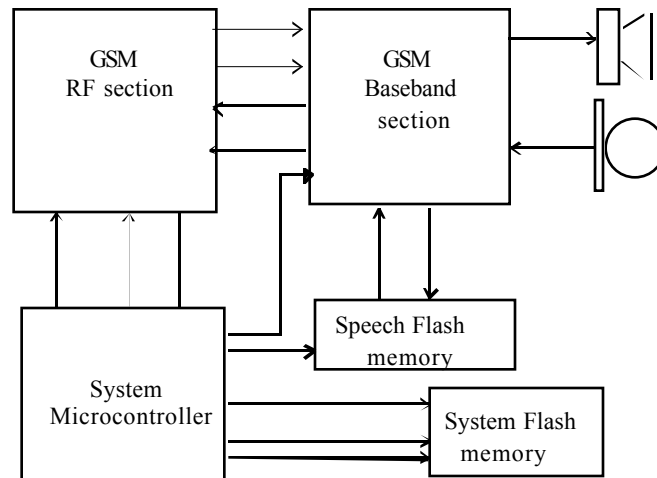


Figure 2

3.1 Digital implementation of speech recording and playback

Figure 2 shows a typical block diagram for voice recording using a traditional digital approach with the DSP that is being used for all the GSM signal processing requirements. It is assumed that the speech will be recorded in a serial FLASH memory dedicated to digital speech recording and storage. Since the DSP may have two serial ports available and one is used for the coded interface^{iv}, the other serial port can be used for

interfacing with the serial FLASH. If a parallel FLASH is used for speech storage, a serial-to-parallel converter would be needed to interface between the DSP and the FLASH memory, or the speech data will have to be routed from the DSP into the system micro-controller and then into the memory. Such processing is not covered in this paper but additional loading on the DSP, the system micro-controller and

the system buss will occur. A separate speech FLASH memory is shown in the figure, but it is also possible to use the system FLASH memory for speech storage as long as care is taken to make sure that the system code cannot be accidentally destroyed during audio recording. Other memory technologies may be used as well, but nonvolatility may be a design requirement.

3.2 ISD33000 Implementation of speech recording and playback

Figure 3 shows a typical block diagram for voice recording using an MLS analog voice recorder chip such as the 4 minute storage ISD33240, a member of the ISD33000 Family, in a GSM cellular telephone. Since the inputs and outputs of ISD33000 are analog, the interface to the microphone and earpiece is very simple. It is not necessary to burden the DSP or its ports for any interface needs. The interface between the system micro-controller and ISD33000 is via an SPI bus.

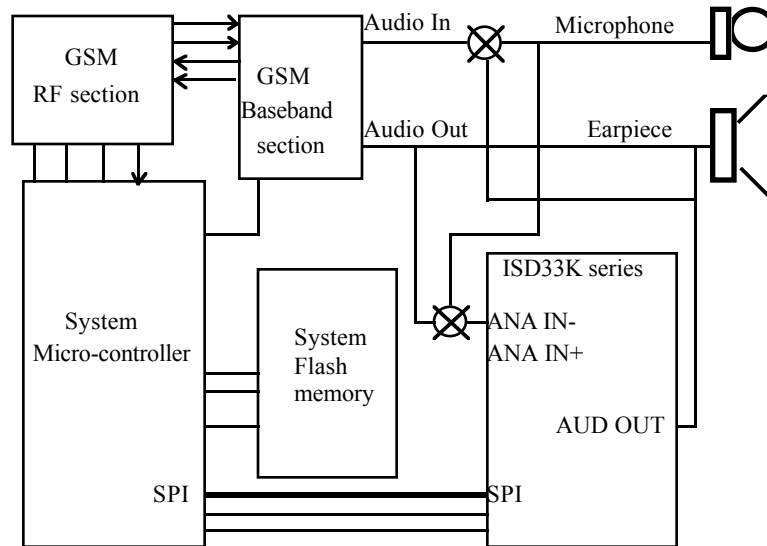


Figure 3

In the paragraphs that follow, we will investigate the functions required for voice storage in a cellular phone: (1) recording during telephone standby time, as required for a personal voice memo recorder, (2) recording off the air during a phone call, and (3) playback of the recorded voice during standby time so that the user can listen to the recorded messages. Additionally, we will look at (4) playback during a phone call; this function is necessary to play an outgoing message or OGM to emulate the operation of an answering machine. And finally, we will look at what is required to do message management in each technology to maximize use of the available memory.

4.1 Voice memo recording - digital:

In the digital speech recording setup, the voice memo recording function is only feasible when the GSM phone is not handling a cellular call. It should be noted that in this mode the DSP and system micro-controller are mostly in the stand-by or power down state. They are only active to the extent necessary to detect an incoming call. In order to initiate the recording function, the DSP and the system micro-controller will have to be brought up to the full operating state. Figure 4 shows the data flow path through the DSP while in the voice memo recording mode. The coded provides digital data at 8k samples/s that are digitally filtered by the DSP.

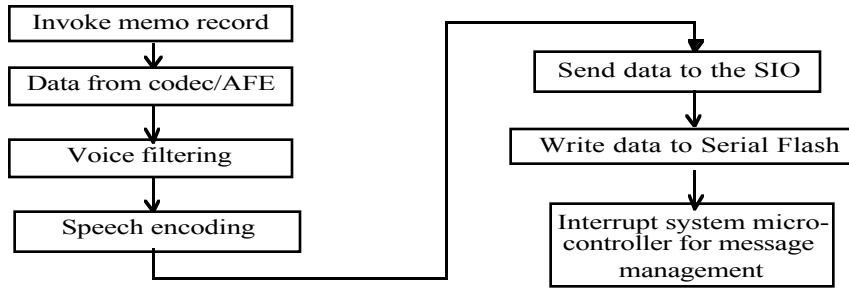


Figure 4

The DSP will form speech frames of 260 bits based on the RPE-LTP coding that can be stored in the Flash memory. The serial I/O port on the DSP interfaces with the memory, and message management task is handled by the system micro-controller. The rest of the operation is very simple and straightforward. This process will use about 3.5 MIPS in the DSP for speech coding, filtering, etc. and about 300 lines of software overhead in the system micro-controller for message management. One drawback of this digital solution is the additional power consumption while invoking the voice memo recording function. The DSP, Coded, memory and the controller together will consume more than 300 mA of current while in the record mode. If the system Flash memory is shared with the speech data, then additional code would be required and a need for increased MIPS in the DSP, resulting in additional power consumption. Depending on how often memo recording is performed, the power consumption could reduce the overall stand-by and talk time for the GSM phone. In a market where every manufacturer is trying to increase the stand-by and talk times, the excess power consumption using this method of implementation could have a negative impact.

4.2 Voice memo recording, ISD33000:

Voice memo record is very easy and simple to implement with the ISD33000 with much less increase in power consumption. Since the DSP is not required for this function it will remain in stand-by or idle mode. The system controller software will also be minimal with liberal timing requirements. Figure 5 shows the flow diagram for voice memo record mode.

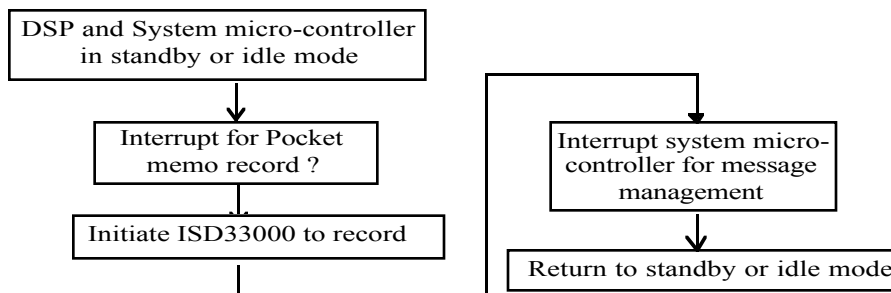


Figure 5

During voice memo recording, the ISD33000 will require about 40 mA and will need about 45 lines of code. This is very easy and quick implementation compared to the digital storage approach and there are very minimal risks in affecting any GSM system critical timing requirements.

4.3 Voice Memo Recording, summary:

Voice Memo Recording is an easy function to perform for both digital and the MLS analog methods. The ISD33000 solution has significant advantages, however, since it requires very much less total power and much less code to be written for the micro-controller. Table 1 summarizes the most important data.

	Digital Memo Record	ISD33000 Memo Record
DSP Mips Required	3.5	0
Power	300 mA	40 mA
Lines of code for micro-controller	300 lines	45 lines

Table 1

5.1 Off-the-air recording - Digital:

There are two Off-the-air recording methods in use today in cellular telephones. Some manufacturers chose to record only the incoming call while others desire to record both sides of the conversation, i.e. full duplex recording. In the digital storage implementation these two methods are very different. In either case, several points have to be considered carefully. The DSP is running full speed to perform the data demodulation, channel equalization, Viterbi decoding, de-interleaving, channel decoding, speech decoding and other essential tasks within each specific time slot. The system micro-controller is busy with the protocol timing loops and other housekeeping tasks. We will look at simplex, incoming call only recording first.

The incoming speech data frame of 260 bits will have to be intercepted from the DSP while all the processing is taking place. If we assume that a serial FLASH memory is used for speech storage and is interfaced through a serial port on the DSP, the software for the DSP will need to be modified and additional instructions added to make the speech data available on the serial port. Figure 6 shows data flow diagram through the DSP when off-the-air command is invoked.

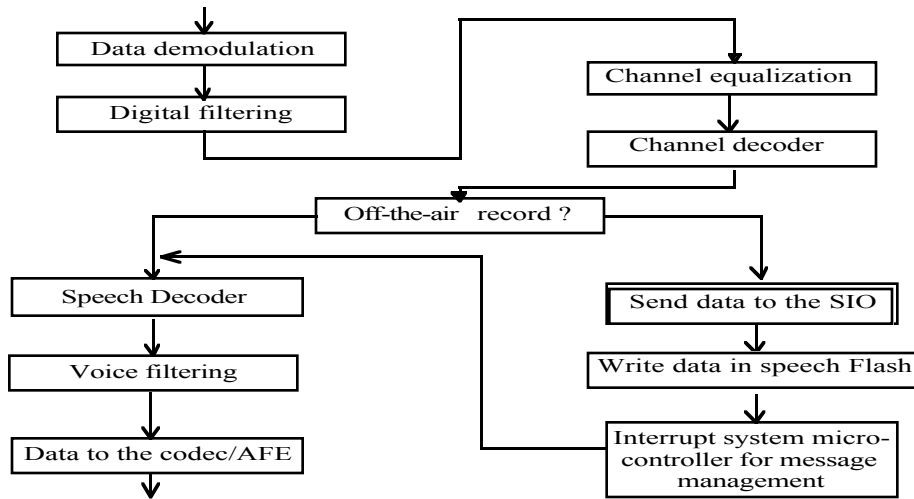


Figure 6

Either the DSP or the system controller will have to handshake with the speech FLASH memory for storage of the speech data. As it can be seen from the data flow diagram, every byte from the DSP will have to be sent out through the serial port to the speech flash memory. Depending on how the GSM protocol is implemented and resource availability, this could pose some problem with the system timing since all the processing will have to be done within the confines of 4.615 milliseconds. The system controller software will have to be modified to add message management functions. About 0.3 MIPS of additional DSP processing will be required to implement the off-the-air recording and approximately 200-300 lines of code will be required in the system micro-controller. Depending on the access time and programming time for the serial Flash, this could pose timing problems in the protocol implementation.

This function will cause some increase in the power consumption in the telephone, but other DSP and RF processes will dominate power usage. It is of utmost importance to notice that the additional message management code and the programming time for the 260 bits of serial data into the serial Flash could

impact the GSM timing requirements.

In our second off-the air recording case, we have a very difficult situation because now we have the outgoing audio stream to also capture and we are still limited by the critical 4.615 mS window dictated by the GSM channel requirements. The Coded digitizes the analog speech from the microphone and the DSP performs the speech encoding. This encoded 260 bits represents the 20mS speech window, which also has to be stored into the speech memory prior to channel encoding. The two data streams will have to be mixed, and possibly attenuated. The data from the DSP will now have to be ported out to the speech memory via the SIO port. This total process now requires a total of 1.2 Mips, 0.9 Mips additional to record the microphone input and mix the two bit streams. This will increase current consumption still further, though the RF power amplifier and other DSP functions will still dominate this specification. About 300 lines of code will be required to implement both the modifications to the DSP code and the system micro-controller code. It should be noted that some DSPs may reach capacity before the implementation of this function rendering it impossible to accomplish without further changes to the GSM platform.

In some cases, a third party software GSM protocol stack is used in the design of the phone. This implementation is usually already optimized for specific hardware. In this case, it would be very difficult to modify the code and still maintain the system timing.

5.2 Off-the-air recording - ISD33000:

Off-the-air speech recording using ISD33000 is very simple compared to the digital approach. The interface between ISD33000 and the system is made through a simple 3-wire serial bus. The analog interface is also simple between the earpiece output of the coded and the analog input on ISD33000. While the system is active and in voice conversation mode, and an off-the-air record command is invoked, initiate the chip into record mode. This operation only requires the serial transfer of 3 bytes of data. In the simplest case, no further intervention is required until recording is commanded to end, or the chip's memory is exhausted. If full message management is performed using the ISD33000's built in control architecture, the micro-controller needs to be interrupted only once in every 150 mS to 300 mS⁴. And then, only a few lines of code need be executed to update the Message Address Table (MAT) used in the chip's message management operation. This is very easy since the GSM time slots are 4.615 mS. This will not affect any critical timing requirements for the GSM protocol. The total additional software overhead required to service the ISD33000 is estimated to be less than 45 lines of code. The current consumption of the ISD33000 in the record mode will be only 40 mA. Again, the DSP and RF power usage dominate during this operation.

If a third party software GSM protocol stack is used in the design of the phone, it would be relatively simple to add the few lines required for implementing the off-the-air recording function and still maintain the system timing.

It is very easy to record full-duplex conversation off-the-air by simple analog mixing of the two audio streams. In addition the ISD33000's ANA-IN input configuration allows the combining of two unequal level signals with little additional circuitry. In fact, there is no extra code or current consumption required to implement full-duplex record. This is a significant advantage over the digital speech storage in the full duplex off-the-air recording mode. Figure 7 shows the steps necessary to perform duplex and simplex recording off-the-air.

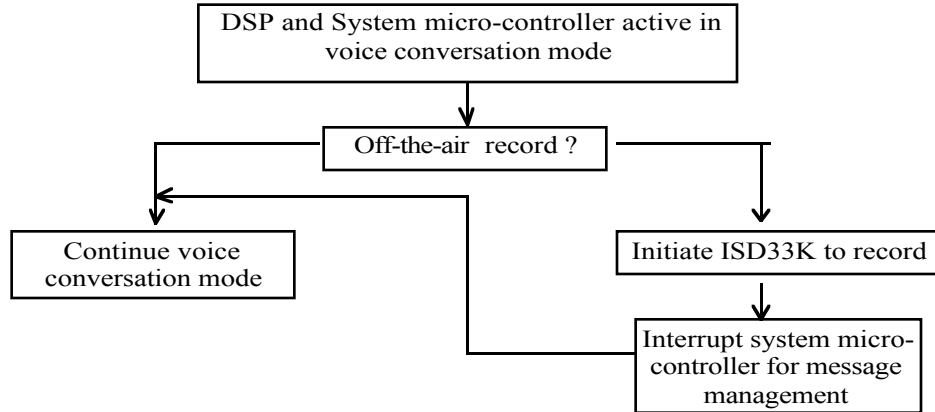


Figure 7

5.3 Off-the-air Recording, summary:

Off the air recording is a task for the digital solution. In fact, the bandwidth required for DSP processing may be a limiting factor. Additionally, timing constraints must be carefully watched to insure that no data is lost or required call processing compromised. Existing or third party software may already be optimized for most efficient operation, and the required modifications to add this feature may be difficult to fit in. The ISD33000 solution, however, requires minimal microprocessor overhead and no modifications to already existing DSP code. In both cases, the additional power requirement needed to add this feature is minimal compared to that required by the call processing and RF components of the phone. Table 2 summarizes the most important data.

	Digital Off-the-air record	ISD33000 Off-the-air record
DSP Mips required simplex	0.3	0
DSP Mips required duplex	1.2	0
micro-controller code duplex	200 lines	45 lines
micro-controller code duplex	300 lines	45 lines

Table 2

6.1 Playback of recorded speech during standby - Digital:

Since full access to the DSP is essential in playback of the recorded speech, it is only available during the standby mode. The system controller and the DSP need to be put in the normal operating mode and the data from the speech memory needs to be fed into the DSP for speech processing. Figure 8 shows the data flow through the DSP and system micro-controller. Upon an interrupt to playback, the micro-controller will enable the DSP to read data from the serial Flash, which will then be decoded and digitally filtered. The coded converts digital to analog signals for playback through the earpiece. This process will require about 1.8 MIPS in the DSP and about 200 lines of code overhead in the system controller for message management.

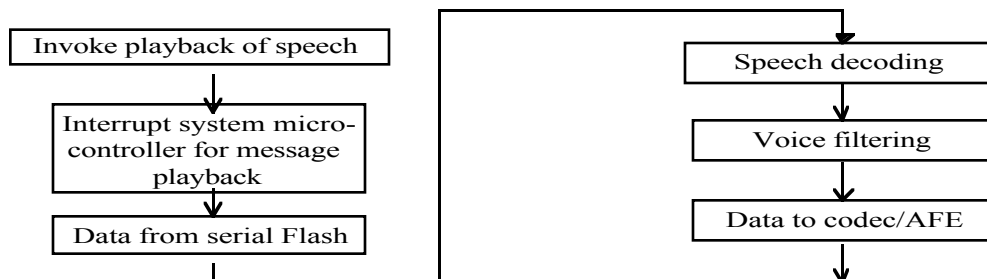


Figure 8

In the playback mode about 200 mA of current consumption is estimated for the DSP, Coded, speech memory, and system controller with its memory. Once again, if the system memory is utilized for speech storage, the data will have to be routed through the micro-controller into the DSP and would increase both the current consumption and code requirements.

6.2 Playback of recorded speech during standby - ISD33000:

Since the DSP is not needed for the playback of recorded speech when using the ISD33000, it can remain in the idle or stand-by mode. The system micro-controller will be used only to initiate the ISD33000 to playback and provide the addresses from the MAT. Figure 9 shows the simple flow diagram for speech playback.

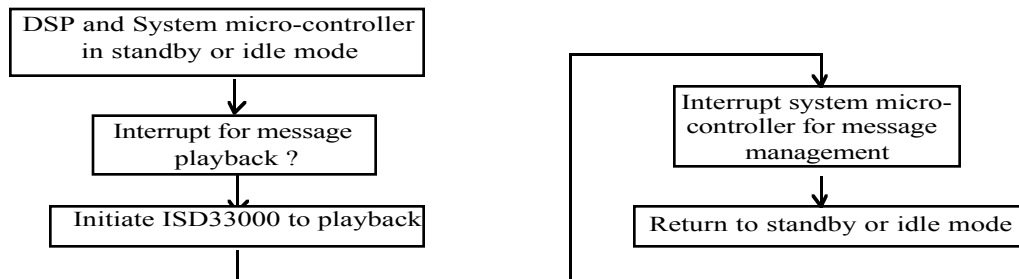


Figure 9

During the playback mode, the ISD33000 will consume about 30 mA and the controller will need 35 lines of code. The timing requirement is also very liberal due to the intelligent message management features of the ISD33000.

6.3 Playback of recorded speech during standby, summary:

Playback of recorded speech during standby is an easy function to perform for both digital and the MLS analog methods. The ISD33000 solution again has significant advantages, however, since it requires much less total power and quite a bit less code to be written for the micro-controller. Table 3 summarizes the most important data.

	Digital Playback - Standby	ISD33000 Playback - Standby
DSP Mips Required	1.8	0
Power	200 mA	30 mA
Lines of code for micro-controller	150 lines	30 lines

Table 3.

7.1 Playback of Speech during an active phone call - digital:

This function is required to enable the answering machine function in a cell phone. The telephone is put into a mode such that it automatically answers and plays an outgoing message (OGM) back to the calling party. As demonstrated in the section on Off-the-air recording, the DSP and system micro-controller are then heavily involved with call processing, data demodulation, channel equalization etc. The stored speech data in the Flash memory will be read into it's associated serial port and injected into the digital data stream by the DSP for transmission by the phone. The system micro-controller will have to keep up with message management functions, and have to be routed to the output data stream. About 1.8 Mips will be required from the DSP and approximately 150 lines of code added to the micro-controller.

7.2 Playback of speech during an active phone call - ISD33000:

Playback of speech during an active phone call is very simple compared to the digital approach. The system micro-controller uses the SPI control to initiate the ISD33000 to playback and provide the addresses from the MAT. Analog gates are added to the hardware to direct the playback to the telephone's audio input. This function requires no additional Mips from the DSP and requires only 30 lines of additional code for the micro-controller.

7.3 Playback of speech during an active phone call, summary:

As in the previous Off-the-air recording example, the digital approach adds a significant burden to the DSP and system micro-controller. In fact, the bandwidth required for this function is 0.5 Mips higher. The ISD33000 example again adds no additional Mips requirements and only a small amount of system micro-controller code. Table 4 summarizes the most important data.

	Digital Playback - Active	ISD33000 Playback - Active
DSP Mips Required	1.8	0
Lines of code for micro-controller	150 lines	30 lines

Table 4.

8.1 Message management software - digital:

An additional task is required by the digital implementation of speech storage when full message management is incorporated into the phone. As recording and erasure of random length messages proceeds in the phone, occasionally, the Flash memory must be cleaned up and reorganized. This is a background task that would have to be carried out at a very low priority. The system timers for GSM protocol would be at the highest priority level. The access to the speech memory is via the DSP SIO port. The message management tasks become very cumbersome and involve powering up the system controller and the DSP. Figure 10 shows a simple flow diagram for message management and speech memory optimization.

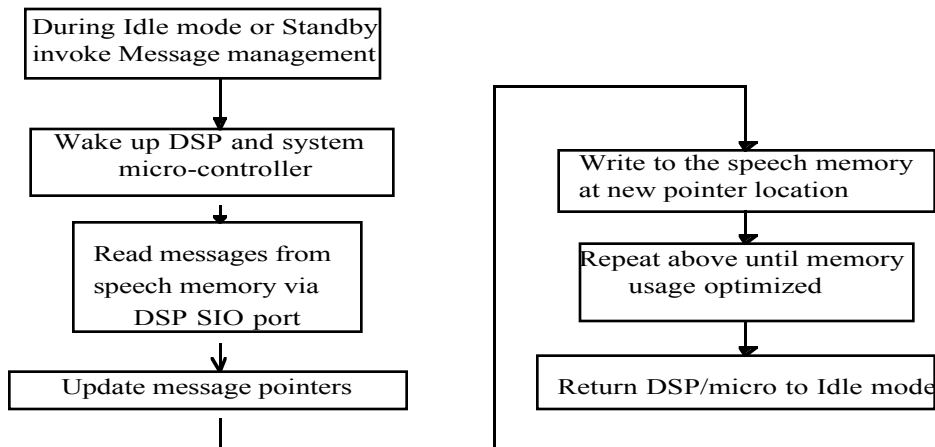


Figure 10

It is necessary to read message blocks from the speech memory into the DSP and update the message pointers to new locations, and write back the message blocks in the speech memory.

It is estimated that the message management and memory optimization could take more than 0.3 MIPS of the DSP and more than 400 lines of code to implement. The current consumption during this process could be as much as 300 mA while active and this would be too much drain during Idle or Standby modes. Considering the importance of talk times and standby times for a given battery capacity, it may be

desirable not to implement the memory optimization process and just have a larger capacity memory for the messages.

8.2 Message management software - ISD33000:

The hooks for message management functionality are built into the into the ISD33000. When message management is used, it is part of the micro-controller system software used to control the device. The code sizes quoted in the other sections of this paper assume message management is implemented.

8. Message management software, summary:

Table 5 summarizes the most important data.

	Digital Playback - Standby	ISD33000 Playback - Standby
DSP Mips Required	0.3	0
Power	200 mA	n/a
Lines of code for micro-controller	400 lines	n/a

Table 5.

9.0 Comparison of digital speech storage verses the ISD33000.

In the previous few sections, digital and analog approaches to off-the-air recording, memo recording and speech playback have been discussed. The digital implementation required more attention to the GSM system timing issues, more usage of the DSP and system micro-controller and more memory for storage. Tables in each section show a comparison between the two approaches. They show the additional incremental MIPS required to implement various record and playback modes for both the digital approach and the ISD33000. In the off-the-air record and OGM modes, the DSP is already in the active mode and so the additional MIPS or the current consumption is insignificant compared to the MIPS used for GSM and the current for RF power amplifier. In the ISD33000 implementation, there is no additional MIPS requirement and the current consumption is insignificant as well. It is most important to notice the increased current consumption and MIPS requirements in the Memo record and playback modes because the DSP and system micro-controller have to be activated from the standby or idle mode and there is no consumption from the RF power amplifier. Since both the micro-controller and DSP code have to be modified in the digital implementation, about 200 – 300 lines of additional code would be required to implement these various modes. The ISD33000 interfaces with the micro-controller via an SPI bus and a couple of hardware lines for the interrupt, RAC, etc., it requires very minor modification to the system code and has no impact on the DSP code.

Message management can be cumbersome with the digital approach depending on the block write/erase or page write/erase requirements for the Flash memories. If every 20 mS speech frame (260 bits of information) is individually addressed in the memory, the MAT could need as many as 500 bytes for the pointers, which is not practical. On the other hand, if the block sizes are fixed to 64 bytes, fewer pointers may be required for MAT but the total memory requirement for 120 seconds of speech storage could be more than 320 kbytes! There does not seem to be significant difference in the power consumption between the two implementations during the off-the-air record mode because it will dominated by the DSP and RF power amplifier. But there is a significant difference in the consumption during the voice memo function and speech playback indicating that the digital approach will need more power for the DSP, system micro-controller, and the flash memory compared to the consumption by ISD33000. The software requirement for digital implementation indicates more than a 1000 lines of additional code whereas the ISD33000 needs less than 200 lines. The timing requirement for implementing digital storage is significantly critical as compared to the ISD33000. And the overall time to implement the digital speech storage could be significantly longer than a simpler ISD33000 based approach.

10.0 Conclusion :

Speech storage and playback in a GSM cellular phone with digital and ISD33000 approach were discussed. Based on the above analysis it can be concluded that the ISD33000 approach is much better than the digital approach. It is a lower power consumption solution, less software overhead, less timing critical, easier and quicker to implement, and therefore the right choice for speech storage and playback in GSM cellular phones.

ⁱ Prasanna Shah and Phillip Pyo, "Designing the ISD33000 series into Digital Cellular Phones," ISD Corporation, 1997

ⁱⁱ GSM Specification series 03, 04, 05, 06

ⁱⁱⁱ Wireless Communications by T. S. Rappaport.

^{iv} Data Sheet – TMS320C54X family Texas Instruments Inc.

^v Data Sheet - ISD33K family Information Storage Devices, Inc

^{vi} ISD Application Note # 2 – Joe Jarrett, Field Applications.