Emic: Portable Hardware-Based TTS for Budget Sensitive Environments

Joe Grand

Grand Idea Studio, Inc.
joe@grandideastudio.com

Abstract

The speech technology industry as a whole is trending towards software-based speech synthesis and recognition solutions, which require high-powered processors and large amounts of memory. This translates to an increased cost of the target platform for which the speech solution is being used, such as a personal digital assistant (PDA) or desktop PC.

The average unit cost of a current PDA suitable for a software-based solution is $500. Dedicated text-to-speech hardware platforms are bulky and expensive and can cost upwards of $1000 each. These are prohibitive amounts for use in most limited budget environments, especially when multiple units are required. The answer to this problem is a low-cost, hardware-based solution designed to interface to a similarly low-cost, readily available PDA.

This paper introduces "*Emic*," a modular hardware design for text-to-speech (TTS) applications. The current form of the device is a removable cartridge that integrates with an off-the-shelf Handspring Visor organizer running the Palm operating system (OS), though the core circuitry of *Emic* can be ported to other platforms. The *Emic* hardware is controlled through a simple set of function calls, allowing for unlimited uses of TTS without requiring large computational overhead. This solution could potentially be sold for less than $100 in quantity.

1 The Problem

The speech technology industry's view of TTS is focused on niche yet mainstream markets, such as automotive or call-center navigation. The solutions are often expensive and are not highly customizable outside of their specific purpose.

A large problem is that there are no readily available, low-cost TTS solutions for use in mobile or portable environments that are generalized (requiring only the necessary computational power and functionality) and are user programmable (allowing for an infinite number of applications and uses). This limits the usefulness of employing an actual TTS solution for many vertical markets, such as educational and special needs environments, where budgets are limited but the technology is still desperately needed.

---

* Published by the Applied Voice Input/Output Society in the *Proceedings of the 2003 AVIOS Speech Developers Conference*, San Jose, California, March 31-April 3, 2003.

Currently available, state-of-the-art TTS products fall into two categories. The first category consists of products that are designed for a single, specific purpose, such as TTS synthesizers for visually impaired users. These devices are often bulky, expensive, and cumbersome to use. Typically, the devices will operate for only 4 hours and are housed in heavy, aluminum enclosures. Some current products, such as Access Solution's Triple Talk USB device, are smaller, but still require USB-enabled laptops or desktop PCs for operation.[1] Thus, that solution still remains too large to distribute to a group of people.

The second category consists of products that are too generalized with no defined target market, such as a software-based TTS application running on a PC with Microsoft Windows. The desktop PC is designed for more complicated, multipurpose functionality and contains an overkill of computational resources for the selected task. Additionally, using a desktop PC or laptop solution does not fit into the market area of a portable solution for mobile use.

In both categories, the use of TTS is either too narrow and restrictive, making the device limited in its potential functionality, or too generalized, making the device too expensive. Another category needs to be defined, as we attempt to do in this paper.

1.1 Lack of Solutions

Industry perception is that hardware TTS is not practical given such readily available high-end computers and mobile devices that can support TTS in software. The goal of *Emic* is to create a simple yet expandable solution that costs less, requires fewer computational resources (e.g., processor speed, RAM, ROM), and is arguably easier to use than other existing mobile TTS solutions.

Portable, software-based speech synthesis products, such as Elan Speech's PocketSpeech, currently exist on expensive Windows CE/Pocket PC platforms, which require a fast processor capable of performing speech synthesis.[2] For example, a Pocket PC-based iPaq H3950 handheld with an Intel PXA250 400MHz processor, 32MB of Flash ROM, 64MB of SDRAM, 3.8" active matrix LCD display with 65,536 colors, and internal speaker, costs slightly less than $600.[3]

Another disadvantage of software-based TTS is that licensing is often required in order to integrate it into an end-user product. Redistribution of the vendor's TTS engine is not allowed without arranging a licensing agreement or purchasing additional user licenses. *Emic*, the proposed hardware-based solution detailed in this paper, requires no outside licensing or other fees.

Table 1, consisting of a small sampling of the available TTS solutions aimed at the portable market, shows that the prices are too great and/or the functionality is too limited to be effectively used in group environments or those with budget constraints.[4] Devices in Table 1 that

---

[1] `http://www.axsol.com`

[2] `http://www.elantts.com/products/fp_pocketspeech.html`

[3] Price current as of February 2003 from `buy.com`.

[4] This information is adapted from a table provided at `http://www.tiresias.org/equipment/eb8table.htm` and combined with independent research. It is not a comprehensive listing of all known products.

are battery-powered also ship with an AC adapter for use in a fixed location. This leads us to believe that the total operating time before the need to replace or recharge batteries is small and therefore these devices would not be suitable for "in the field" use.

Company	Product	Price	Description
Access Solutions	Triple-Talk PCI	$395	PCI expansion card for desktop PC.
	Triple-Talk USB	$495	Serial port and USB interface. Internal rechargeable battery. 5.5" x 4.0" x 1.5" dimensions. Internal speaker.
Artic Technologies	DECTalk Express	$1195	Serial port interface. Internal rechargeable battery.
	SPIRIT	$299	Serial port interface. Large aluminum case. 9V battery. Internal speaker.
RC Systems	DoubleTalk PC	$289	ISA expansion card for desktop PC.
	DoubleTalk LT	$299	Serial port interface. 9V battery. 5.3" x 4.0" x 1.5" dimensions. Internal speaker.
Quantum Technology	Q-Talk	$400	Serial and parallel port interface. 5.0" x 3.3" x 1.5" dimensions. Internal rechargeable battery (14 hours for full charge). Internal speaker. Metal case.

Table 1: A small sampling of currently available hardware-based TTS solutions. Most are too large for true portable use and too expensive for group environments.

Many mobile devices in the marketplace do not have the computational power and resources to support a software-based TTS solution. However, they also cost much less. For example, a used Handspring Deluxe organizer can be purchased for approximately $50 on eBay. By off-loading the actual TTS functionality to a separate hardware component, a low-cost, portable TTS solution can be created.

1.2 Mobile Devices and Vertical Markets

TTS can be used for any number of general-purpose applications, such as browsing web pages, or reading text files, e-mail, or electronic books. The real benefit of *Emic* is its integration into educational and special needs vertical markets, both of which are often extremely limited in budgets and are often overlooked by manufacturers producing the newest technologies. Because of this, these markets end up with unsupported equipment and resources. Most TTS solutions employed in these small-budget environments run on desktop PCs, usually one or two per classroom, which limits the number of students who can benefit from the technology at one time.

Portable devices are ubiquitous in the consumer marketplace, and it is only natural that they will become a target system for speech technologies. It is estimated that handheld mobile phone and device shipments will reach nearly 250 million by 2006 [1]. Today's handheld devices usually allow third-party applications to be written for the platform, which allows the end-user to customize the device for a specific purpose.

One example of using a portable TTS solution like *Emic* in a classroom environment is as follows: Before class, a teacher can load each student's project plan to his or her PDA. Each plan could vary per student, allowing for specialized needs. The single-ear headphone in the *Emic* module can be used to keep the student in tune to the TTS application running on the device while not disturbing others, but still be aware of ambient activities in the classroom. Any number of students could be individually executing their project plans on their PDAs. This example is an extension of the typical "learning lab" commonly used for speech and language education in which each student wears a set of headphones to listen to a pre-configured project or exercise.

2 Palm OS Architecture

Palm OS, through PalmSource, Inc., is licensed to a number of major vendors in the portable device and cellular phone markets, such as Handspring, Sony, IBM, Kyocera, Samsung, Qualcomm, Franklin Covey, TRG, Symbol, Fossil, and Garmin.[5] The *Emic* prototype was developed for Palm OS because devices running Palm OS make up nearly 80 percent of the global handheld computing market [2].

2.1 Hardware

Hardware devices running Palm OS currently use either the Motorola DragonBall MC68328 microprocessor family[6] or the Texas Instruments OMAP-1510 processor[7] (available in new Palm, Inc. devices such as the Palm Tungsten T). Software-based speech synthesis and recognition applications are available for the OMAP platform, though not yet ported to Palm OS as of February 2003. Still, high-powered portable devices are still too expensive for many deployment scenarios.

Early and low-end Palm devices are powered by two standard alkaline AAA batteries, which are removable and replaceable. High-end and newer models contain a non-removable, internal Lithium-Ion battery (rechargeable through the HotSync cradle or external charging device). Alkaline batteries typically last longer than their rechargeable counterparts and add a level of convenience in emergency situations, as they are available in almost any retail store. The disadvantage to using non-rechargeable batteries is that they must be properly discarded after use.

Many Palm OS-based devices are designed to be modular in that they have an external expansion slot of some sort to support additional memory or other custom circuitry. For example, some Palm, Inc. devices contain a slot for MultiMediaCard and SecureDigital expansion cards. The Sony Clié family supports the Memory Stick. Handspring Visor organizers support the Springboard module, on which the *Emic* prototype is based.

[5] `http://www.palmsource.com`
[6] `http://e-www.motorola.com`
[7] `http://focus.ti.com/omap/docs/omaphomepage.tsp`

2.1.1 Handspring Springboard Module

Certain Handspring devices, such as the Handspring Visor Deluxe, Neo, Platinum, Pro, Prism, and Edge, contain a Springboard expansion slot to accept physical cartridges inserted into the back of the device. The Springboard module allows for practically unlimited expansion of such Handspring organizers by providing a 68-pin interface directly to the address and data buses, microphone, and other control signals of the device. There are niche markets that make use of the Handspring Springboard module to transform the device for such applications as a digital camera, mobile phone, wireless modem, MP3 player, blood glucose meter, or GPS receiver.

Although the first revision of *Emic* is designed to work with the Handspring Visor through its custom Springboard module interface to show an example of a low-cost, mobile TTS solution, the *Emic* circuitry and driver routines can be customized to interface with a wide range of other portable devices. It is likely that future versions of *Emic* will be in a CompactFlash expansion card format.

Handspring is transitioning away from the Visor organizer and Springboard module and focusing their efforts on the Treo family of mobile phone communicators. However, the Springboard module version of *Emic* is still feasible and attractive in limited deployment scenarios (up to 1,000 units), given the new and used availability and low-cost of the Visor organizer product line.

The Handspring Developer Solutions web site has archived all publicly available information related to Springboard module development, including complete physical and electrical interface specifications, mechanical housing drawings, application notes, sample code, and manufacturing information.[8]

2.2 Application Development

Use of the Palm OS Application Programming Interface (API) provides the application developer with a notion of hardware independence and provides a layer of abstraction. Palm OS was designed to be open and modular to support application development by third parties.

Applications are commonly written in C using the Metrowerks CodeWarrior for Palm OS development suite, which comes with a graphical-based integrated development environment including compiler, resource editor, and debugger.[9] Palm, Inc. offers complete technical documentation, support, application notes, and sample source code, altogether providing a solid framework for third-party application development.[10] For more details, refer to the Palm OS Programming Development Tools Guide [3] and the Palm OS SDK Reference [4].

[8] `http://www.handspring.com/developers`
[9] `http://www.metrowerks.com/products/palm`
[10] `http://www.palmsource.com/developers`

3 Emic

Emic is the premiere hardware-based TTS solution for portable devices. The prototype form exists as a removable module that integrates with any Palm OS-based Handspring Visor organizer that supports the Springboard expansion interface. Any application running on the Visor can communicate with the *Emic* hardware through a simple set of function calls, allowing for unlimited uses of TTS without requiring large computational overhead.

3.1 System Details

The system block diagram of *Emic* (Figure 1) shows the elegant simplicity of the design. By requiring *Emic* to be controlled by a PDA, much of the design overhead common in a custom embedded system is avoided.

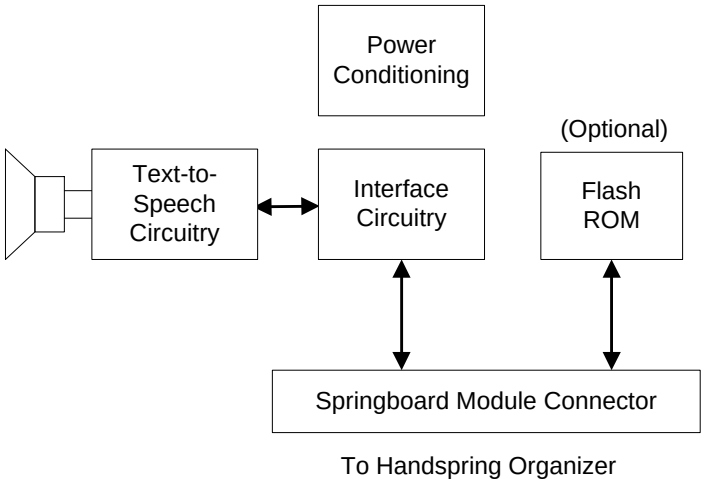

Figure 1: *Emic* TTS Springboard Module System Block Diagram

Emic requires only 3 integrated circuits and some supporting discrete components. The reduced parts count keeps the device cost down and simplifies manufacturing and troubleshooting. Because all the parts are in small footprint integrated circuits, the overall dimensions of the *Emic* hardware are extremely small, requiring just 1-2 square inches. The prototype *Emic* circuitry fits into a standard form-factor Springboard module and contains a 3.5mm headphone jack. Figure 2 shows an early development picture of *Emic*.

The heart of the *Emic* design is the Winbond Electronics WTS701, the industry's first single-chip TTS processor.[11] The chip accepts ASCII input via a simple 4-wire serial interface and converts it to spoken audio via an analog output. As detailed in the WTS701 technical data sheet [5], the device uses a concatenation algorithm and "integrates a text processor, smoothing filter, and multi-level memory storage array. Text-to-speech conversion is achieved by processing the incoming text into a phonetic representation that is then mapped to a corpus of naturally spoken word parts." The device can be reprogrammed in-circuit to provide firmware

[11] `http://www.winbond-usa.com`

revisions or change the corpus (e.g., languages or gender). This is a major advantage to *Emic*, which can be easily updated through its connector interface without any special programming equipment, simplifying customer service and technical support issues. Playback speed, voice pitch, and volume can all be adjusted through simple commands to the WTS701.

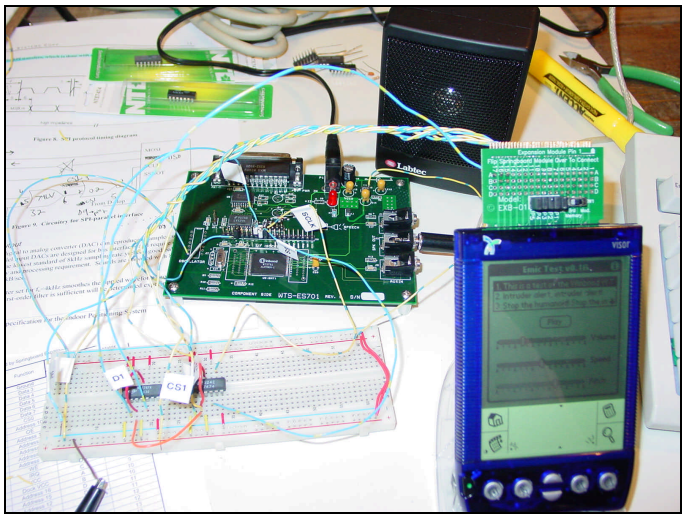

Figure 2: *Emic* while in development, right: Handspring Visor Neo with Springboard module prototype board, left: Winbond WTS-701 evaluation kit and interface circuitry.

Power consumption for the WTS701 is very low, with an operating current typically of 35mA and standby of less than 1uA. *Emic* requires no external batteries and gets its power from the Springboard port (which can provide 5VDC up to 100mA). A typical iPaq Pocket PC device, depending on usage, will last 1 to 6 hours before requiring a recharge [6]. A Handspring Visor Neo will last 1 to 2 months on average before the batteries require replacing [7]. Initial current measurements show that the *Emic* hardware itself draws between 22mA and 31mA while audio is being played and 10uA in idle mode. Thus, given the typical current capacity of 1175mAh for an alkaline AAA battery, the complete prototype *Emic* solution (Visor Neo with Springboard module) can last for approximately 10 hours of continuous use (longer if the device is not constantly active) before battery replacement is necessary.

Customized user applications can be loaded and stored onto the optional Flash ROM component. This way, even when battery power is removed (e.g., the Springboard module is extracted from the Handspring device), the contents of the application will remain in the non-volatile Flash ROM.

3.2 Application Programming Interface (API)

The *Emic* API consists of a number of functions for core TTS functionality and abstracts the actual low-level hardware of *Emic* from the end-user Palm OS application (Figure 3).

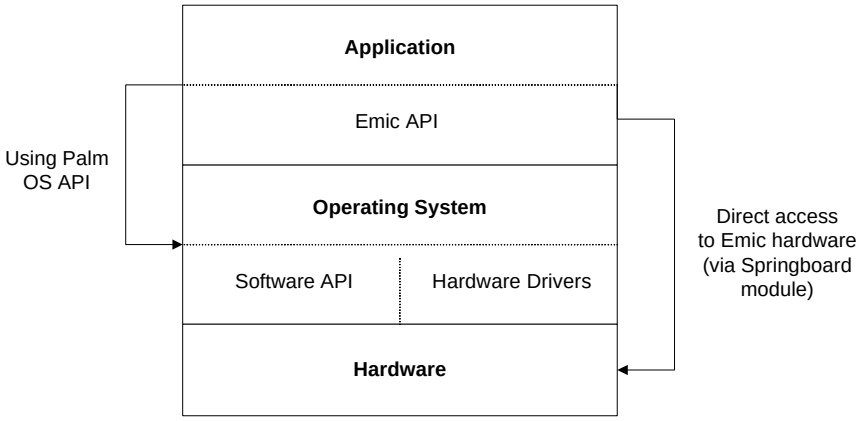

Figure 3: Typical architecture for controlling the *Emic* hardware with a PDA.

Each routine beginning with '`WTS701_`' is a high-level user function. The others are low-level, device specific routines. Table 2 shows the functions, which are currently written in C targeted for Palm OS, but could easily be ported to other platforms. Additional routines could be written to support the extra functionality of the WTS701 device, such as user customization for special characters and application specific abbreviations.

Function Declaration	Description
`void WTS701_Init (void);`	Initialize WTS701 and set defaults
`void WTS701_SendCmd (UInt8 cmd, UInt8 data);`	Send command and data
`void WTS701_SetVol (UInt8 vol);`	Set speech volume
`void WTS701_SetSpeed (UInt8 speed);`	Set speech speed
`void WTS701_SetPitch (UInt8 pitch);`	Set speech pitch
`void WTS701_PowerDown (void);`	Put WTS701 into power-saving mode
`void WTS701_InitSpeaker (int init_length, char* ptr);`	Begin TTS conversion, text string pointed to by 'ptr'
`UInt8 WTS701_ReadInt (void);`	Read interrupt bytes
`UInt8 WTS701_ReadStatus (void);`	Read status bytes
`void hardware_setup (void);`	Initialize physical hardware and enable WTS701
`UInt8 spi_wbyte (UInt8 wbyte);`	Sends one byte via bit-banged SPI, received byte is returned
`void spi_begin (void);`	Starts SPI transaction
`void spi_end (void);`	Ends SPI transaction

Table 2: *Emic* API functions. Routines beginning with '`WTS701_`' are high-level user functions; others are low-level, device specific routines.

A simple Palm OS TTS application, with fully commented source code, was created to read Memo Pad memos (Figure 4). It is a reference design to aid developers in understanding the process of controlling *Emic* and in integrating the *Emic* hardware with their particular application. To exercise the full capacity of *Emic*, an application will need to be created by the end-user for his or her specific purpose.

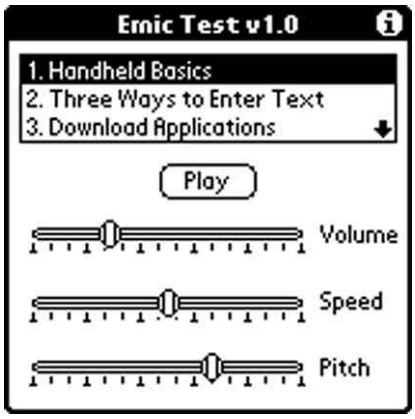

Figure 4: Simple Palm OS application written to evaluate the *Emic* TTS hardware.

4 Summary

As the cost of mobile devices decreases and the computational power and functionality increases, there will eventually come a time when hardware-based TTS may not be required. Until that time comes, *Emic*, the hardware-based product detailed in this paper, provides a simple and affordable TTS solution. It is fully customizable for a wide range of hardware platforms and software applications and is able to be deployed in relatively large numbers, especially within budget-constrained environments.

5 Availability

As of February 2003, the core *Emic* TTS design is complete. Information on upcoming production and availability of the technology can be obtained by contacting the author or at `http://www.grandideastudio.com`.

References

1. Datamonitor, *US Mobile Devices to 2006*, July 2001.
2. NPD Intellect, February 2002.
3. Palm, Inc., *Palm OS Programming Development Tools Guide*, DN 3011-002.
4. Palm, Inc., *Palm OS SDK Reference*, DN 3003-003.
5. Winbond Electronics, *WTS701 Single-Chip Text-to-Speech Processor Data Sheet*.
6. Dale Coffing's Pocket PC Passion, *Battery Life Run Times*, `http://www.pocketpcpassion.com/General/BatteryMonitor/BatteryRunTimes.htm`.
7. Handspring Customer Support FAQ, `http://support.handspring.com/esupport/forms/hsResolutionView.jsp?ResolutionId=6638`.