Distributed Processing in the Home using a PC with a Wireless Speech Interface

Debashis Chowdhury and Ser Chia

Intel Corporation
Computing Enhancement Architecture Lab
5000 W. Chandler Blvd
Chandler, Arizona 85226, USA

ABSTRACT

The Personal Computer is evolving from a purely personal device to one that can support multiple applications in different locations simultaneously. This paper describes how the connectivity and processing capability of the PC can be used in a distributed manner in the home to provide a variety of services like speech activated environmental command and control functions, digital video decoding, Internet telephony and entertainment control. As they are architected today, current PC's are challenged when trying to perform intensive signal processing tasks while managing several external connections (e.g. dial-up internet) and multiple internal connections (e.g. cordless phone interface) at the same time. We will describe some of these challenges, and what remains to be done to make the PC more capable in undertaking such a multifunctional challenge.

1. USAGE AND APPLICATIONS

Currently PCs only provides value to the user when he/she is within arm's length of the keyboard and mouse. This usage model limits the functionality of the PC to the vicinity of the PC. By using a distributed architecture with a wireless connection, the power of the PC can be accessed from every corner of the home through speech I/O.

Command and Control:

Traditionally, speech driven command and control applications have been PC application centric, meaning that speech could be used to control ordinary PC applications. You could use voice to control functions normally controlled through the keyboard or mouse. However, with this approach to wireless enabled distributed computing, command and control can be extended to our everyday life - and for multiple members of the family.

There are two approaches to PC aided command and control that have been experimented with. One is the always listen mode, where the computer is listening for a specific set of commands or key words to initiate an action. A conceptual example of this is the *Star Trek* computer. The second approach is the 'Push to Speak' mode - which requires one to be in close proximity of a portable listening device. With the current state of the technology, much better results have been obtained with the second approach - the primary advantages coming from a much better signal strength and the suppression of spurious commands picked up by the microphone. A second usability issue - privacy - will probably still bias users towards Push to Speak, even when the technology exists to make the *Star Trek* model work.

PC Enhanced Entertainment Center:

The days where an entertainment system is a collection of CD players, VCRs, audio-cassette players, and DVD players could be coming to a close. With an Intel® Pentium® II processor-based PC equipped with a TV card, all of these capabilities are built into a PC, including soft DVD playback, digital recording/playback of TV shows, and music recording/playback. These features can now be performed with the PC located away from the entertainment center through the use of wireless transmitter/receivers for A/V. The need for multiple remote controls are also eliminated since all of the functions can be voice controlled.

The usability issue we run into here is: while speech input is great for choosing what you want, conversational systems have trouble conveying possible options to the user with speech output. The latter shortcoming is readily apparent to all of us who have tried to navigate complex voice-menu options on the telephone. In many cases a small display that complements the Speech I/O would go a long way towards enhancing ease of use.

Home Automation:

A powerline based home control system can be integrated in with your home PC to provide a wide range of computer automation of everyday appliances. Using control devices such as X-10*, for example, we can provide lighting control, HVAC automation, security functions, and much more. The PC provides voice-based control through voice recognition and feedback. The usability issue of being able to view status and options through a display carries over to this area as well.

Internet Telephony / Telephony assistant:

With voice interface the PC can manage all of your telephony contacts for you. You additionally get the option of placing a call through the Internet. Based on caller ID and an internal directory lookup, your PC can announce for you who is calling. Depending on filters that you can set up, one possible usage is to forward all phone solicitations to the answering machine - which can also be hosted in the PC.

There are several usability issues when it comes to using the PC in this manner. First, the wirelessly connected handheld devices (e.g. a 900MHz cordless phone) today does not have the physical characteristics of the devices they might functionally replace. For example, when using the PC as the answering machine, where do I locate the 'You have new Messages' indicator - usually a red blinking lamp? Do I have to pick up a PC connected phone and ask how many messages I have? This is an easy problem to fix in the next generation of handheld devices - where the ergonomics and usability can be designed in with the functionality in mind. Also, if I have several messages, I would like to know who they were from in order to decide which ones to check out first. This too points towards the integration of some sort of display into the handheld device.

Personal Assistant:

With speech recognition, the PC can become an interactive speech based assistant. Applications such as lists, notes, calendar, alarm clocks, and messaging are a few applications that can be speech enabled. From anywhere in the home, you can make lists, take notes, set your alarm clock, and take care of all of your messaging needs.

2. ARCHITECTURE

Figure 1 shows how a typical high-performance PC could be connected for distributed computing in the home. It has

access to the outside world through a high speed Internet connection. The Home PC also has access to within the home using several wireless connections. These include a PC connected 900MHz cordless telephone (one commercially available product was recently announced by Microsoft*), a 2.4GHz wireless NTSC output (e.g. WAVECOM*Sr. by RF-Link), a simple IR transmitter (e.g. PC to IR Linc* by SmartLinc), and a 2-way X-10 controller (e.g. CM11A by ActiveHome*). The 900MHz cordless phone system provides audio I/O capabilities using the sound system on the PC. The wireless video signal transmitter also acts as an IR repeater to different parts of the house. The powerline based X-10 network provides control and monitor capability for appliances, lighting and other AC powered devices.
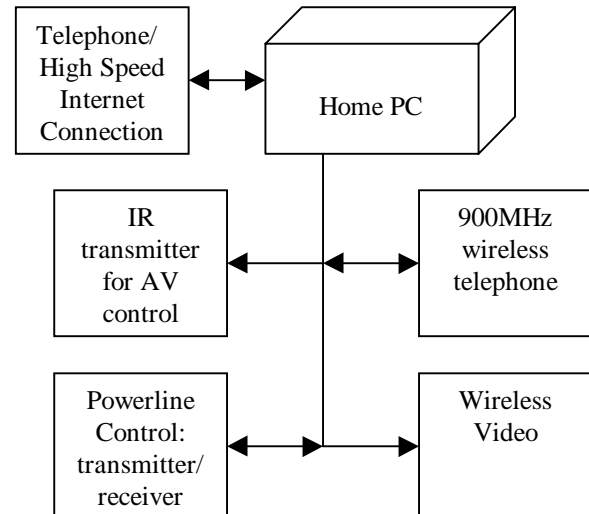
Figure 1. Typical architecture of a PC connected for distributed computing.

For now, let us assume that the PC remains powered all the time, and that it has a low latency Internet connection. Let us also assume that we have sufficient processing capability in the host CPU to run all of these interfaces in parallel (which is not a big challenge for today's high-end PC processors).

The primary command and control interface described here is a speech interface. User access using the speech interface can be delivered to the PC from anywhere in the home, by the use of the cordless telephony handset for command and control (or a higher quality version, for continuous speech recognition). The PC has the

capability to notify the user about actions taken, recognized or unrecognized commands, or more information needed by the use of short beeps, or synthesized audio output. We can set our preferences for high audio feedback (for when the action taken is not visible to the user), or for zero feedback in cases for the action is clearly visible (example, when you're changing the settings of your entertainment system).

We performed many experiments comparing different commercially available speech recognition products. These products fell into two broad categories: Discrete recognition systems and continuous recognition systems. The discrete recognition systems performed well in command and control tasks. They tended to be user-independent, required little training, and could use a lower quality cordless connection. However, they were capable of only recognizing from a limited set of commands, and commands had to be separated by distinct pauses. Continuous speech recognition systems allowed the user to dictate more complex phrases (for example, grocery lists) and provided a much larger vocabulary that was recognized. These systems tended to be user dependent, required a significant amount of training, and required a higher quality microphone connection (e.g. Sony* WCS-990 wireless transmitter-receiver pair).

Continuous speech recognition systems in general required two times the processing MIPS compared to discrete limited vocabulary versions – which roughly translates to about 200MIPS or more just to do large vocabulary continuous speech recognition. Such a system requirement is not a big deal for today's high-powered desktop CPUs, but remains a significant challenge for any small handheld device. Hence, we have an advantage running the application in a distributed manner with the handheld providing the remote human interface and the desktop providing the CPU horsepower.

The Internet connection shown in Figure 1 works best when it is always connected, but isn't too bad when there is a responsive ISP at the other end of a wire. You can then set up the PC to continuously collect information that is pertinent to the user like weather, stock market information, road conditions, television listings and the like. Users can then query for specific information; one example would be to ask for specific television show information that might not be cached in your local machine - which the PC can then get for you. In many cases, the requested information is available in audio format (e.g. in the local news report), or can be synthesized for you using text to speech.

The X-10 power line based control network is connected to the PC using a serial port with a powerline transmitter-receiver device. One example of such a device is the X-10 ActiveHome computer interface (CM11A). By using X-10 modules that have two way communication capabilities, the PC can give you the status of any piece of appliance in the home such as the living room light, garage door (open/closed), bedroom TV, AV equipment, etc. For homes already outfitted with one-way devices, this means changing out the one-way modules with ones that are capable of two-way communication. CeBus based [2] systems will have this capability to begin with, and also promises to provide more reliable and secure in-home control.

The PC can source high-quality audio, video, and family album snapshots through the S-Video output and transmitted to a TV through a 2.4GHz transmitter - as described earlier. A relatively recent addition to the PC peripheral set is the DVD-ROM player. DVD-ROM players can also play DVD movies, with the MPEG 2 video stream being decoded directly on the host CPU, or optionally with a hardware assist.

Incidentally, the distributed use of the PC in the home is about to receive a major enhancement with the arrival of the HomeRF* wireless connectivity [1]. This is expected to provide a much tighter connection to the telephony network, and to also enable other wireless connected data devices (e.g. another PC), all of which can use the Internet connection of the primary PC as a resource.

3. Signal Processing and Storage Required

Command and Control
The amount of processing required for recognition of commands depends on the type of recognition system you are running. In the case of running a continuous speech recognition system, roughly 200 MIPS are required. In a discrete speech recognition system about half that much is usually sufficient.

PC Based Entertainment System
A mid-range system based on an Intel Pentium II Processor performed the functions described in the previous section, including software DVD decode.

Home Automation
The requirements for Home Automation are similar to command and control. The commands recognized by the

recognition systems are executed by the home automation system. The back end powerline based control network runs at a much slower speed and does not use any significant processing power.

Internet Telephony
For Internet telephony, significant processing is required for the audio compression and decompression. For users of 28.8kbps connections, the audio is generally compressed to 8kbps [3] which can be easily sent over a modem.

Data Storage
The PC can act as a centralized storage area for digital data. The type of data stored can be items which can help us organize our daily activities such as shopping lists, grocery list, appointments, family schedules or daily reminders. Or the PC can be used to hold and replay A/V content as varied as home videos, family album snapshots, or pre-recorded TV shows. The data rate for digitally compressed (e.g. MPEG2-NTSC) video averages around 5Mbps, so you could easily record multiple hours of programming on today's 10+ Gigabyte hard drives.

4. NEXT STEPS AND FUTURE DIRECTIONS

The distributed model computing using the PC in the home is about to receive a major capability improvement with the ability of HomeRF to support both voice and data devices simultaneously. For now, the video output still remains analog, but that is likely to change with the more widespread adoption of IEEE 1394 standard in PCs and Consumer Electronics devices. A wireless version of the IEEE 1394 for a high-speed version of HomeRF will be needed before the wireless video stream can be distributed digitally around the home.

Other infrastructure improvements are required for the PC as well. One significant effort here is the "Instantly Available" initiative. When implemented in PCs, this will allow the desktop PC to go to a low-power state when not in use, but snap back to a full-power 'On' state as soon as activity is detected upstream (from the Internet or telephone connection) or downstream (e.g. from the portable voice I/O device). Another improvement area is the ability of the PC to handle multiple applications simultaneously, especially those that require low latency or isochronous support. These will have to come from the OS and applications stack. Intel is working actively with the Personal Computing industry to implement these kinds of improvements to the PC platform.

There is also the promise that some all the radio functions could be done by the host CPU. This has the significant advantage of being highly reconfigurable and of not requiring specialized peripherals outside the PC.

5. SUMMARY

Our work has demonstrated the ability of the home PC to interface by voice with the user for command and control of a host of home information, entertainment and control applications. We expect to see significant improvement in the ability of multiple users to access the PC simultaneously for voice and data with the advent of HomeRF. Distribution of multimedia streams will remain primarily in the analog domain until a high-speed wireless home connectivity standard is proposed and implemented. Key areas in the PC itself (e.g. instant availability and multitasking) has been identified as areas for improvement.

Looking forward to the day when all of the new technologies identified in the previous paragraph are in place, we may find that the home PC undertaking all these tasks may be left alone to manage these tasks - and not used as the current PC usage scenario, i.e. with a monitor and keyboard/mouse in front of it. Add to that some significant multimedia storage capabilities and we have basically what we can call a Home Server. On the other hand, the PC might still retain its current two-foot user paradigm, and additionally support the new usage as an extended function.

Disclaimer:
Unless specifically noted, the products mentioned in this paper are not Intel products. Intel does not take a position towards recommending or not recommending these products, and makes no claims as to their usability.

*All brands and names are the property of their respective owners.

6. REFERENCES

[1] HomeRF Working Group Web Site: http://www.homerf.org
[2] CeBus Industry Council Web Site: http://www.cebus.org
[3] "Microsoft Netmeeting 2.1 Resource Kit", #098-80106, page 168