A T1-Based Speech Data Collection Platform

Kornman	Paul	Institute for Signal and Info. Processing	601-325-8335/3149	kornman@isip.msstate.edu
Duncan	Richard	Institute for Signal and Info. Processing	601-325-8335/3149	duncan@isip.msstate.edu
Picone	Joseph	Institute for Signal and Info. Processing	601-325-3149/3149	picone@isip.msstate.edu

Dr. Joseph Picone
Institute for Signal and Information Processing
Mississippi State University

413 Simrall, Hardy Rd.
Box 9571
Mississippi State, Mississippi 39762

3.11 (Speech Processing Hardware) 4.7

Digital speech data collection, that is, the capability to collect speech data digitally from the telephone network has long been an elusive goal in the speech research community. Yet, this is the preferred method of data collection now, and is the method used for several years by a number of leading speech research organizations. Existing platforms are based on PCs or ISDN-based Unix workstations. Each approach has its drawbacks and has proven not to be extensible.

In this work, we present the first low cost accessible platform based on a Sun Sparcstation 5 and a third-party T1 interface board. In addition, we provide a software environment that facilitates the development of data collection-oriented applications. This system is being developed for the Linguistic Data Consortium for use in future speech data collection efforts.

N. Deshmukh, J. Ngan, J. Hamaker and J. Picone, “An Advanced System to Generate Multiple Pronunciations of Proper Nouns”

1. ABSTRACT

In the last five years, a direct digital interface to the telephone network has become the standard method of speech data collection for large speech corpora involving telecommunications applications. In particular, the T1 interface is popular because it is a cost-effective way to deliver large numbers of voice channels. Unfortunately, such systems, most often based on PC hardware, use closed architectures consisting of proprietary software and hardware designs, resulting in a strong dependence on custom software from a single vendor. Vendors have repeatedly demonstrated an inability to deliver timely and cost-effective solutions for speech research, resulting in a great deal of wasted time and money, with no industry-standard solution in sight.

The Institute for Signal and Information Processing (ISIP) has developed a Unix-based platform for speech data collection based on an open-architecture design. The platform uses a very inexpensive and popular workstation, a Sun Sparcstation 5, as its host. The digital interface is provided in two flavors. First, the Sun is interfaced to a standard T1 link using a two board system. This is attractive because it provides a look-alike replacement for the current system, and uses a widely-available telephony protocol. Second, a system based on a new telecommunications standard — Asynchronous Transfer Mode (ATM) will be developed. This approach is attractive because it provides a much higher bandwidth solution at a reduced hardware and software cost and complexity (only a single SBus slot is required).

2. HISTORICAL BACKGROUND

The first system featuring a digital interface that was deployed for full scale data collection in speech research was the T1-based system described in [1]. This system was based on the Intervoice Robot Operator hardware platform. This environment featured an IBM PC platform that used an interface consisting of a two proprietary boards, an OS/2 operating system (in its most recent generation), and a 4GL programming language for rapid application prototyping. It was used to collect the SWITCHBOARD corpus, and is currently in use on the CALL HOME and Voice Across Hispanic America (VAHA) projects.

What is wrong with this system? First and foremost, it is a closed-architecture based on a platform that is incompatible with those currently in use in speech research. Hence, its acceptance by the community as a general purpose platform has been slow. Second, for most projects, extensive firmware modifications have been required by the vendor to perform a particular style of data collection. Such modifications have historically been very expensive, and have caused numerous delays to the projects requiring these modifications. In short, the vendor has traditionally been very unresponsive to the need for such modifications. Third, the current platform requires a steep learning curve involving a nontrivial custom environment. Operators often have to be educated directly by the vendor, and require several months to come up to speed. The cost of maintaining and operating the system is extremely high.

At the time the previous system was developed, there were no alternatives. Now, there are several vendors that have announced similar capabilities on general purpose Unix workstations. For the first time, the functionality of the current system can be duplicated in a much less expensive Unix environment, and built from standard programming tools well-known within the speech research community.

3. TOWARDS DIGITAL INTERFACES

T1 interfaces have only recently become available for the Sun Sparcstation product line. The majority of these simply support data transmission using the 1.5 Mbps bandwidth of the T1 (e.g. an Internet connection). Virtually all of these systems have no native support for voice communications, which requires splitting the signal into 24 voice lines, and supporting some basic call processing capabilities. One exception is Linkon Corporation, a vendor previously used in previous LDC-funded data collection platform projects. Linkon has teamed up with Newbridge Microsystems to offer a two-board solution for voice interfaces using a T1.

Newbridge supplies a Sun-compatible SBus board that plugs directly into a T1 line, and performs all of the packetization, synchronization, and formatting functions for a T1. Linkon resells this product, and adds to it an SBus board containing four DSPs that can process 4 to 8 voice channels depending on the nature of the application. Linkon also supplies an API for the entire system that supports all of the functions required for speech data collection, including the conference bridge connection required for SWITCHBOARD-style corpora. Since Linkon is a direct competitor of Intervoice, this system was designed as a lower-cost alternative to Intervoice's system.

3.1. An ATM-Based Platform

The major drawback to the T1 system is its cost. The T1 connection is expensive, typically costing approximately $1K/month plus a $1K connection fee. In addition, the Linkon solution is expensive both in dollars and real estate. The two-board system costs approximately $4K. It requires two SBus slots. A Sun Sparcstation 5, our preferred platform because of its low price and good performance, has typically three SBus slots available for use (the fourth slot is used by the monitor). Adding the T1 interface to this system does not leave much room for expansion (an additional ethernet card and or I/O channel are often desired), though it provides a reasonable configuration for a dedicated data collection system.

Fortunately, a better solution is on the horizon. Telephone companies have started pushing a low cost alternative to T1 and ISDN: the Asynchronous Transfer Mode (ATM) interface. This is a digital communications link that operates at 155 Mbps, and can be subdivided into fractional channels such as a bank of 64 kbps voice channels. We have selected an ATM card from Efficient Networks, Inc. as the preferred ATM-interface vendor. These cards occupy a single slot on a Sparcstation, cost in the neighborhood of $1.3K, and offer a much simpler and more flexible software interface. Though ATM is an experimental service offered by most telephone companies, and as yet not tariffed for general public use, its cost is projected to be close to the cost of a single ISDN line, and will be much less expensive than a T1 interface (which has 1/100 of the bandwidth). We believe that the ATM technology offers a system more likely to be upwardly compatible with the future direction of digital telephony.

3.2. Software to Support Rapid Prototyping

Both hardware platforms in the data collection system are supported from a common base of software. This software has been developed as an extension of ISIP's object-oriented public domain software environment [3]. The software uses GNU's g++ compiler, a highly portable freeware C++ compiler. The software development effort consists of two main components.

The low-level programming interface, referred to as the API, consists of a collection of C++ classes that provide a hardware-independent interface to the digital telephony hardware. Beneath this veneer, there is independent support for the T1 and ATM boards. Users are able to perform low-level call processing functions on individual telephone lines within the T1 or ATM span. In addition, support is provided to interface two lines together for the conference bridge application, yet retain access to the transmit sides of both telephone lines. Audio software is provided to support playout of prompts, real-time segmentation of the incoming speech data, and recording of the incoming speech data to disk.

A syntactically simple object oriented scripting language with dynamic flow control provides a powerful interface to the low-level functionality of the hardware. This is used in the second component to the software — a high-level graphical user interface (GUI). This collection of tools supports creation of applications by graphically flow charting the application using an application builder. All attempts are being made to follow the conventions of several such tools currently on the market (including the Intervoice system), yet preserve the low-level control required for speech research. This software is being written using the C++ extensions to the Tcl programming language (a public domain scripting language for X windows applications).

4. CORPORA CERTIFICATION SOFTWARE

Historically, one of the problems with corpus collection has been certification of the corpus. This often comes after the corpus has been collected, which means if serious problems are discovered, the corpus is more or less trashed. It is much more desirable to do data certification on-line, so that problems are discovered as they are encountered. While this has been discussed for many years within the speech research community, no one has produced a viable implementation.

Following closely the work described in [4], which recommends a number of bit-level analyses of the data, we are developing software to automatically monitor the incoming data and report any problems. In addition, we support some of the more traditional measures of the quality of the data, including energy level histograms, SNR computations, long-term frequency response analysis for silence and active-speech intervals, etc. These statistics can be compared to some reference statistics developed from an analysis of existing corpora representative of the modern-day digital telephony environment.

5. REFERENCES

1. J.J. Godfrey, E.C. Holliman, and J. McDaniel, “SWITCHBOARD: Telephone Speech Corpus for Research and Development,” in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. I-517-I-520, San Francisco, California, USA, March 1992.
2. B. Wheatley and J. Picone, “Voice Across America: Toward Robust Speaker Independent Speech Recognition For Telecommunications Applications,” *Digital Signal Processing: A Review Journal*, vol. 1, no. 2, pp. 45-64, April 1991.
3. J. Picone, “Managing Software Complexity in Signal Processing Research,” in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. III-41-III-44, Minneapolis, Minnesota, USA, April 1993.
4. J.E. Porter, “Features of T1 Line CODEC Code Distributions,” ITT Aerospace / Communications Division, 10060 Carroll Canyon Road, San Diego, CA 92131, July 1991.