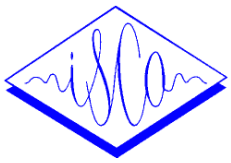

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

Applications of Speech Recognition Technology in Telecommunications

Jay G. Wilpon and *David Roe*
AT&T Bell Laboratories, Murray Hill, New Jersey 07974

Abstract

As the telecommunications industry evolves over the next decade to provide the products and services that people will desire, several key technologies will become commonplace. Two of these, automatic speech recognition and text-to-speech synthesis, will provide users more freedom on when, where and how they access information. While these technologies are currently in their infancy, their capabilities are rapidly increasing and their deployment in today's telephone network is expanding. The economic impact of just one application, automation of operator services, is well over $100 million per year. Yet, there still are many technical challenges that must be resolved before these technologies can be deployed ubiquitously in products and services throughout the wordwide telephone network.

The current and future applications of speech recognition in the telecommunications industry will be examined in terms of their strengths, limitations and the degree to which user needs have been or have yet to be met.

1 Introduction

A quarter century ago the influential John Pierce wrote an article questioning the prospects of one technology, speech recognition, and criticizing the "mad inventors and unreliable engineers" working in the field [8]. Pierce went on to describe the motivation for speech recognition research: "The attraction [of speech recognition] is perhaps similar to the attraction of schemes for turning water into gasoline, extracting gold from the sea, curing cancer, or going to the moon." His influential article was successful in curtailing, but not stopping, speech recognition research.

What Pierce's article failed to foretell was that even limited success in speech recognition—simple, small vocabulary speech recognizers—would have interesting and important applications, especially within the telecommunications industry. In the decades since Pierce's article, tens of thousands of these "limited" speech recognition systems have been put into use, and we now see the beginnings of a telecommunications based speech recognition industry [1, 3, 5, 6]. The total size of the speech recognition market is $200 million, in 1992, according to the Wall St. Journal. The economic impact of just one application, automation of operator services, is well over $100 million per year.

As the telecommunications industry evolves over the next decade to provide the products and services that people will desire, voice processing technologies will become commonplace.

Current applications using speech recognition technology center around two areas: those that provide cost reduction (e.g., AT&T and Bell Northern Research's [BNR] automation of some operator functions and Nynex and BNR's attempt to automate portions of directory assistance) and those that provide for new revenue opportunities (e.g., Nynex's directory assistance call completion service, BNR's stock quotation service and Nippon Telegraph & Telephone's [NTT] banking by phone service).

2 The Art of Speech Recognition

Current speech recognition practices encompass engineering art as well as scientific knowledge. Fundamental knowledge of speech and basic principles of pattern matching have been essential to the success of speech recognition over the past twenty-five years [10]. That being said, the *art* of successful engineering is critically important for applications using these technologies. There is an important element of craftsmanship in designing a successful speech recognition. Knowledge about the task also helps ASR based applications to be tuned to the user's requirements. Often, this engineering art has been developed through trial and error. It should be emphasized that improving the engineering art is a proper and necessary topic for applied research.

The art of speech recognition has improved significantly in the past few years, further opening up the range of possible applications [11]. For speech recognition some of these advances are:

- Wordspotting. The ability to spot key sounds in a phrase is the first step toward understanding the essence of a sentence even if some words are not or cannot be recognized [12, 13].
- Barge-in. When talking to a machine, it is desirable to be able to interrupt the conversation. Barge-in provides a necessary, easy to use capability for customers and, as with wordspotting, is an essential technology for successful mass deployment of ASR into the telephone network.
- Subword units. It is now possible to build a speaker-independent dictionary of models comprised of constituent phonetic (or phoneme-like) statistical models. Therefore, the effort and expense of gathering speech

from many speakers for each new vocabulary word is eliminated, making the development and deployment of new and improved applications simple, quick and efficient.

- Noise immunity and channel equilization. Better speech enhancement algorithms and channel modeling have made speech recognizers more accurate in noisy or changing environments, such as airports or automobiles [2, 7].

The design of an easy-to-use dialogue with a computer system is a significant challenge. We know from experience that is possible to design good human interfaces for computer dialogue systems. Unfortunately, it has also been verified that it is possible to design systems that aggravate people. At this time there are some general guidelines for good human interface designs, but there is no "cookbook" recipe that guarantees a pleasant and easy-to-use system [4]. Thus, the art of speech recognition technology needs to be advanced while waiting for the major research breakthroughs to occur.

3 Applications of Speech Recognition

It is important to bear in mind that the speech technologies described above, notwithstanding advances in reliability, remain error-prone. For this reason, the first successful products and services will be those that have the following characteristics:

- Simplicity. Successful services will be natural to use. For instance, they may use speech recognition to provide menu capabilities using only small vocabularies (less than 10 words), rather than large vocabularies (greater than 1000 words).
- Evolutionary growth. The first applications will be extensions of existing systems. For example, speech recognition as a TouchTone replacement for voice response systems.
- Tolerance of errors. Given that any speech recognizer will make occasional errors, the inconvenience to the user should be minimized. This means that careful design of human factors will be essential in providing suitable systems.

3.1 Cost Reduction vs. New Revenue Opportunities

There are two classes of applications that are beginning to appear. The first, *cost reduction applications*, are those for which a person is currently trying to accomplish a task by talking with a human attendant. In such applications, the accuracy and efficiency of the computer system that replaces the attendant is of paramount concern. This is because the benefits of ASR technology generally reside with the corporation that is reducing its costs, and not necessarily with the end users. Hence, users may not be sympathetic to technology failures. Examples of such applications include: (1) automation of operator services, currently being deployed by many telephone companies including AT&T, Northern Telecom, Ameritech and New Jersey Bell, (2) automation of directory assistance, currently being trialed by Northern Telecom, and (3) control of network fraud currently being developed by Texas Instruments (TI), Sprint and AT&T.

The second class of applications are services which generate new revenues. For these applications, the benefit of speech recognition technology generally resides with the end user. Hence, users may be more tolerant of technology limitations. This results in a *win-win* situation. The users are happy and the service providers are happy. Examples include: (1) automation of banking services using ASR offered since 1981 by NTT in Japan, (2) TouchTone and rotary phone replacement with ASR introduced by AT&T in 1992, and (4) information access services, such as a stock quotation services currently being trialed at BNR.

In general, the most desirable applications are those which provide some real usefulness to customers, not just gimmicks. Since these technologies are in their infancy and have obvious limitations, careful planning and deployment of services must be achieved if mass deployment of the technologies is to occur.

3.2 Automation of Operator Services

Beginning in 1985, AT&T began investigating the possibility of using limited vocabulary, speaker independent speech recognition capabilities to automate a portion of calls currently handled by operators. The exact task studied was the automation of the billing functions. Customers would be asked to identify verbally the type of call they wished to make without directly speaking to a human operator. After extensive field trials in Dallas, Seattle and Jacksonville during 1991 and 1992, AT&T announced that it would deploy VRCP (Voice Recognition Call Processing). This service would automate *collect, calling card, person-to-person* and *bill to third number* calls using wordspotting and barge-in technology. These trials were considered successful not just from a technological point of view, but also because customers were willing to use the service [1]. By the end of last year, it is estimated that over 1 billion telephone calls were automated by the VRCP service.

In 1989, BNR began deploying AABS (Automated Alternate Billing Services) [5] through local telephone companies in the United States, with Ameritech being the first. For this service, ASR and DTMF technology is limited to automation of only the back-end of collect and bill to third number calls. That is, after the customer places a call, a speech recognition device is used to recognize the called party's response to the question: *You have a collect call. Please say yes to accept the charges or no to refuse the charges.*

NAME	COMPANY	DATE	TECHNOLOGY	TASK
ANSER	NTT	1981	Small vocabulary, isolated word ASR	Banking services
Automated Alternative Billing Services	BNR/Ameritech	1989	Small vocabulary, isolated word ASR	Automation of operator services
Intelligent Network	AT&T	1991	Small vocabulary, wordspotting, barge-in, Spanish	Rotary telephone replacement in Spain
Voice Recognition Call Processing (VRCP)	AT&T	1991	Small vocabulary, wordspotting, barge-in	Automation of operator services
Voice Interactive Phone (VIP)	AT&T	1992	Small vocabulary, wordspotting, barge-in	Automated access to enhancement telephone features
Directory Assistance Call Completion (DACC)	Nynex, AT&T	1992-3	Small vocabulary ASR	Automation of directory services
Flex-Word	BNR	1993	Large vocabulary, isolated word ASR	Stock quotations and automation of directory assistance
Voice Dialing	Nynex, Sprint	1993	Small vocabulary speaker dependent isolated word ASR	Automatic name dialing
Voice Prompter	AT&T	1993	Small vocabulary, wordspotting, barge-in	Rotary telephone replacement

Table 1: Network Based Voice Processing – Based Services

3.3 Voice Access to Information over the Telephone Network

An emerging area of telephone-based speech processing applications is that of Intelligent Networks. For example, a current 800 service might begin with the prompt *Dial 1 for sales information or 2 for customer service.* Until now, the caller input to Intelligent Network services required the use of TouchTone phones. Automatic speech recognition technology has been introduced to modernize the user interface, especially when rotary phones are used. Using the AT&T 800 Speech Recognition service, announced in the beginning of 1993, menu-driven 800 services can now respond to spoken digits in place of TouchTones and will provide automatic call routing.

Intelligent Network services have attracted the interest of international customers. The first AT&T deployment with speech recognition was in Spain in 1991 [3]. The low TouchTone penetration rate in Spain (less than 5%) and the high-tech profile of Telefónica, the Spanish telephone company, were the primary motivating factors. In a sense, Telefónica is trying to leap-frog past TouchTone technology with more advanced technologies.

Subword-based speech recognition has recently been applied to an information access system by BNR working with Northern Telecom [6]. Using the 2000 company names on the New York Stock Exchange, callers to this system can obtain the current price of a stock simply by speaking the name of the stock. Though the accuracy is not perfect, it is adequate considering the possibilities of confusion on such a large vocabulary. The experimental system is freely accessible on an 800 number, and it has been heavily used since its inception in mid-1992. There are thousands of calls to the system each day, and evidence suggests that almost all callers are able to obtain the stock information they desire.

3.4 Voice Dialing

One of the biggest new revenue generating applications of speech recognition technology in telecommunications is voice dialing. Obviously, current ASR technology is not advanced enough to handle requests such as, *Please get me the pizza place on the corner of 1st and Main, I think it's called Mom's or Tom's.* However, most requests to place telephone calls are much simpler than that—for example, *Call Diane Smith, Call home,* or *I'd like to call John Doe.* ASR technology, enhanced with with wordspotting, can easily provide the necessary capabilities to automate much of the current dialing that occurs today.

Voice-controlled repertory dialing telephones have been on the market for several years and have achieved some level of market success in cellular telephones. Recently, Nynex, Sprint and AT&T announced the first network-based voice dialing services.

4 Technical Challenges

While the prospect of having a machine that humans can converse with as fluently as they do with other humans remains the Holy Grail of speech technologists and one that we may not see realized for another generation or two, there are many critical technical problems that we believe we will see overcome in the next two to five years. Solving these challenges will lead to the ubiquitous use of speech recognition and synthesis technologies within the telecommunications industry. The question is only *how* these advances will be achieved. For speech recognition, these research challenges include:

- Better handling of the varied channel and microphone conditions. The telephone network is constantly changing, most recently moving from analog to digital circuitry and from the old style non-linear carbon button type

transducers to the newer linear electret type. Each of these changes affect the spectral characteristics of the speech signal. Current ASR and especially speaker verification algorithms have been shown not to be very robust to such variability. A representation of the speech signal which is invariant to network variations needs to be pursued.

- Better noise immunity. While advances have been made over the past few years, we are a long way away from recognition systems that work equally well from a quiet home or office to the noisy environments encountered at an airport or in a moving car.

- Better decision criteria. For a long time, researchers have mainly considered the speech recognition task as a two-class problem, either the recognizer is right or it is wrong, when in reality it is a three-class problem. The third possibility is that of a *non-decision*. Over the past few years, researchers have begun to study the fundamental principles that underline most of today's algorithms with an eye towards developing the necessary metrics that will feed the creation of robust rejection criteria.

- Better out-of-vocabulary rejection. While current wordspotting techniques do an excellent job of rejecting much of the out-of-vocabulary signals that are seen in today's applications, they are by no means perfect. As our basic research into more advanced, large vocabulary systems progresses, better out-of-vocabulary rejection will continue to be a focusing point. With all this activity being directed to this issue, I am sure we will see a steady stream of new ideas aimed at solving this problem.

- Better understanding and incorporation of task syntax and semantics and human interface design into speech recognition systems. This will be essential if we are to overcome the short-term deficiencies in the basic technologies. As ASR-based applications continue to be deployed, the industry is beginning to understand the power task-specific constraints have on providing useful technology to its customers.

- Recognition of multiple languages. Globalization of ASR technology is here. An increased ability to identify which language is being spoken, and then to recognize what is spoken is essential.

5 Conclusion

Looking to the future, we believe that speech recognition over telephone lines will continue to be the most important market segment of the voice processing industry, both in terms of the number of users of this technology and in terms of its economic impact. It is safe to predict that "simple" applications of speech recognition will become commonplace. By the year 2000, it is likely that more people will get more information via voice dialogues than will by typing commands on TouchTone keypads to access remote databases. However, despite some advances in language modeling, the speech understanding capability of computers will remain far short of human capabilities until well into the next century. Applications that depend on language understanding for *unrestricted* vocabularies and tasks will remain a formidable challenge, and will not be successfully deployed for mass consumption in a telecommunications environment for several decades.

One thing is very clear: sooner than we might expect, applications based on speech recognition technologies will touch the lives of every one of us.

References

[1] V. Franco. Automation of Operator Services at AT&T. In *Proceedings Voice '93 Conference*, San Diego, Cal., March 1993.

[2] H. Hermansky, N. Morgan, A. Buyya, and P. Kohn. Compensation for the effects of communication channel in auditory like analysis of speech. In *Proc. of Eurospeech '91*, pages 1367–1370, September 1991.

[3] T. E. Jacobs, R. A. Sukkar, and E. R. Burke. Performance trials of the Spain and United Kingdom Intelligent Network automatic speech recognition systems. In *Proc. IEEE Workshop on Inter. Voice Tech. for Telecomm. Appl.*, Piscataway, NJ, October 1992.

[4] C. Kamm. User interfaces for voice applications. In *Comm. Between Humans and Machines*. NAS Press, in press.

[5] M. Lennig. Putting speech recognition to work in the telephone network. *Computer*, 23(8):35–41, August 1990.

[6] M. Lennig. Automated bilingual directory assistance trial in Bell Canada. In *Proc. IEEE Workshop on Inter. Voice Tech. for Telecomm. Appl.*, Piscataway, NJ, October 1992.

[7] H. Murveit, J. Butzberger, and M. Weintraub. Reduced channel dependence for speech recognition. In *Proc. DARPA Speech & Natural Language Workshop*, pages 280–284, Harriman, NY, February 23–26, 1992.

[8] J. R. Pierce. Whither speech recognition. *JASA*, 46(4):1029–1051, 1969.

[9] *Proceedings DARPA Speech & Natural Language Workshop*. Harriman, NY, January 1993.

[10] Rabiner and Juang, "Fundamentals of Speech Recognition,"

[11] D. Roe and J. Wilpon. Whither speech recognition – the next 25 years. *IEEE Trans. on Communications*, to appear.

[12] R. Rose and E. Hofstetter. Techniques for task independent word spotting in continuous speech messages. In *Proc. ICASSP '92*, March 1992.

[13] J. G. Wilpon, L. Rabiner, C-H. Lee, and E. Goldman, Automatic recognition of keywords in unconstrained speech using hidden Markov models. *IEEE Trans. on ASSP*, 38(11):1870–1878, November 1990.