Technical journals online.com

Review Article

AN ADVANCE CRITICAL REVIEW & OPPORTUNITIES OF SPEECH RECOGNIZATION & PROCESSING IN MILITARY APPLICATIONS

[1]Amandeep Kaur, [2]Satnam Kaur, [3]Tajinder Singh

Address for Correspondence

[1,3]GIMET, Amritsar, [2]GNDU, Amritsar

ABSTRACT

Speech reorganization and processing is a major area of research. This process depends on different factors on which the speech processing and reorganization depends. This paper presents a study of military applications of advanced speech processing technology which includes review and assessment of current efforts in military applications of speech technology, identification of opportunities for future military applications of advanced speech technology. In this process, Emulating emotion and empathy is on its way right now. Currently, most consultants of artificial intelligent customer response systems for 'call centers' advise that the voice on the other end if coming from a machine, should be easily identifiable by the human calling in as a computer systems with voice recognition features, because humans do not like to be tricked, when they find out, it makes them upset. So the armature users instead of the professionals have ability to do this now. In this paper we are describing the Current effort in military applications of speech technology, which includes narrowband (2400 b/s) and very low-rate (50-1200 b/s) secure voice communication, integration of voice as well as data in computer networks, speech recognition in fighter aircraft, military helicopters, battle management, and air traffic control training systems & noise and interference removal for human listeners.

KEYWORDS: Speech, LPC, HMM, DTW, ATC, Noise.

1 INTRODUCTION

This paper is the result of a study of military application of advanced speech processing technology which has been undertaken with the following goals:

- Review and assess a representative sampling of current efforts in military applications of advanced speech processing technology;
- To identify opportunities for new military applications, or further development of current applications and
- To identify areas where improvements to speech processing technology are needed to address military problems.

The intention is to outline a fairly broad range of applications, opportunities and speech technology areas; however, the coverage is not intended to be fully comprehensive, nor to be very detailed in any particular area

1 Digital Narrowband Communication Systems –

The Linear Predictive Coding (LPC) algorithm was relatively new in 1977. Improvements in technology and the coding algorithm have now led to widespread deployment of digital narrowband secure voice, especially by means of the STU-III (secure terminal unit) family of equipment at 2.4 kb/s. In addition, significant progress has been made in developing practical coders for lower rates using Vector Quantization (i.e., pattern matching) techniques. Linear predictive coding

$$y(n) = \sum_{i=1}^{n} a_i \text{x(n-i)}$$

Where y(n) is the predicted signal value, x(n-i) the previous observed values, and a_i the predictor coefficients. The error generated by this estimate is

$$e(n)=x(n)-y(n).$$

These equations are valid for all types of (one-dimensional) linear prediction. The differences are found in the way the parameters a_i are chosen.

2. **Automatic Speech Recognition** –

Major advances both in CSR and IWR have been made largely through the wide scale development of statistically-based Hidden Markov Model (HMM) techniques, as well as through the development and application of dynamic time warping (DTW) recognition techniques. It MM techniques which were pioneered prior to 1977, have in recent years been further developed at a large number of laboratories, with significant advances both in recognition performance and in efficiency of implementation.

3. Noise and Interference Reduction –

Work in application of digital speech processing to noise and interference reduction was relatively new in 1977, and has progressed significantly since that time. Hardware systems for speech enhancement have been developed and have been shown to improve both speech readability and ASR performance under certain conditions of noise and interference.

2. Current Work in Development of Military Applications of Speech Technology

2.1 Narrowband Secure Voice for Tactical Applications

Most applications of narrowband voice coders at 2.4kb/s (e.g., STU-III) have been in office environments where background acoustic noise and other environmental effects are not major problems. Operational military platforms such as fighter aircraft, helicopters, airborne command posts, and tanks, pose additional challenges since the performance of narrowband algorithms tend to be sensitive to noise and distortion both in talker and listener environments. However, substantial progress has been made in developing the voice algorithm, microphone, and system integration technology for tactical deployment of 2.4 kb/s voice. Examples include the Joint Tactical Information Distribution System (JTIDS) narrowband voice efforts in the U.S. and in the U.K. and the development of the Advanced Narrowband Digital Voice Terminal (ANDVT) family of equipment for a variety of environments.

2.2 Digital Narrowband Secure Voice -the STU-III

The STU-III represents a marriage of a sophisticated speech algorithm, the Linear Predictive Coding (LPC) technique at 2.4 kb/s, with very large-scale integration (VLSI) digital signal processor (DSP) technology to allow development of a secure terminal which is small enough and low enough in cost to be widely used for secure voice communication over telephone circuits in the United States. The primary

factor which has allowed its widespread application has been progress in VLSI technology. Although the 2.4 kb/s LPC algorithm in STU-III produce intelligible speech, it is not toll quality and current efforts are focused on providing improved quality for secure voice, while maintaining the ability to transmit over standard telephone circuits. Modern technology has evolved to the point where 4.8 kb/s is now generally supportable over the dial network. Hence, recent efforts have focused, with some success (see [66]) on the development of 4.8 kb/s voice coders with higher quality than LPC. Based on this work, the Code-Excited LPC (CELP) technique has been proposed as a standard for4.8 kb/s secure voice communication.

2.3 Voice/Data Integration in Computer Networks

The widespread development of computer networks using packet switching technology has opened opportunities for a variety of applications of speech technology, including: packet voice communications with efficient sharing of network resources for voice and data; advanced intelligent terminals [39,104] with multi-media communications; multi-media conferencing and voice control of resources and services (such as voicemail) in computer networks Since data communications using packet systems is becoming widely used in military systems, integration of voice and data on these networks provides significant advantages. Applicable technologies are speech coding, speech recognition, speech synthesis, and multiplexing techniques including Time-Assigned Speech Interpolation (TASI), which take advantage of the burst nature of speech communications.

3. APPLICATIONS

3.1 Speech Recognition Systems in High-Performance Fighter Aircraft

The pilot in a high-performance military aircraft operates in a heavy workload environment, where handstand eyes are busy and speech recognition could be of significant advantage. For example, the pilot could use speech recognizer to set a radio frequency or to choose weapon, without moving his hands or bringing his gaze inside the cockpit. This would allow the pilot to concentrate more effectively on flying the airplane in combat situations. The potential improvement in pilot effectiveness could be extremely significant in critical situations.

3.2 Speech Recognition Systems in Helicopter Environments

The opportunities for speech recognition systems to improve pilot performance in military helicopters are similar to those in fighter aircraft. In a hands-busy, eyes busy, heavy workload situation, speech recognition (as well as speech synthesis) could be of significant benefit to the pilot. Of course, the problems of achieving high recognition accuracy under stress and noise pertain strongly to the helicopter environment as well as to the fighter environment. The acoustic noise problem is actually more severe in the helicopter environment, not only because of the high noise levels but also because the helicopter pilot generally does not wear a facemask, which would reduce acoustic noise in the microphone.

3.3 Speech Recognition Systems in Battle Management

Battle management command centers generally require rapid access to and control of large, rapidly changing information databases. Commanders and system operators need to query these databases as conveniently as possible, in an eyes-busy environment where much of the information is presented in display format. Human machine interaction by voice has the potential to be very useful in these environments. A number of efforts have been undertaken to interface commercially available so late-word recognizers into battle management environments.

3.4 Training of Air Traffic Controllers

Training for military (or civilian) air traffic controllers(ATC) represents an excellent application for speech recognition systems. Many ATC training systems currently require a person to act as a "pseudo-pilot", engaging in a voice dialog with the trainee controller, which simulates the dialog which the controller would have to conduct with pilots in a real ATC situation. Speech recognition and synthesis techniques offer the potential to eliminate the need for a person to act as pseudo-pilot, thus reducing training and support personnel Air controller tasks are also characterized by highly structured speech as the primary output of the controller, hence reducing the difficulty of the speech recognition task. The U.S. Naval Training Equipment Center has sponsored a number of developments of prototype ATC trainers using speech recognition. Generally, the recognition accuracy falls short of providing graceful interaction between the trainee and the system. However, the prototype training systems demonstrated significant potential for voice interaction in these systems, and in other training applications. the recognizer is constrained in vocabulary, one of the goals of the training programs is to teach the controllers to speak in a constrained language, using specific vocabulary specifically designed for the ATC task. Recent research in France on application of speech recognition in ATC training systems, directed at issues both in speech recognition and in application of task-domain grammar constraints.

Removal of Noise from Noise Degraded Speech Signals

There are a variety of military and non-military applications where removal of noise and interference from speech signals is important, and a significant amount of work continues to be devoted to this area, both InTechnology development and in applications. Application areas identified in the study include:

- two-way communication by voice
- transcription of a single, important recording
- transcription of quantities of recorded material.

The focus of the study is on speech processing to aid the human listener. The panel concluded that, although some noise reduction methods appear to improve speech quality in noise, intelligibility improvements had not been demonstrated using closed response tests such as the Diagnostic Rhyme Test (DRT).

3.5 Speaker Recognition and Speaker Verification

Automatic speech processing techniques for identification of people from their voice characteristics have number of military and non-military applications, these applications include:

- Security, where the task is to verify the identity of an individual (e.g., for control of access to a restricted facility), and where the subject can often be instructed to speak a required phrase (this is referred to as "text-dependent" speaker verification)
- Surveillance of communication channels [where the tasks to identify a speaker from samples of unconstrained text ("text-independent" speaker recognition); and (3) forensic applications.

The generic systems include:

- Integrated multi-rate voice/data communications terminal
- Interactive speech enhancement system
- Voice controlled pilot's associate system
- Advanced air traffic control training system
- Battle management command and control support system with spoken natural language interface
- Spoken language translation system.

4 OPPURTUNITIES FOR THE APPLICATION

4.1 Integrated Multi-Rate Voice/Data Communications Terminal

Advanced speech processing will play a very important role in meeting the multiple and time-varying communications needs of military users. For example, a commander in a fixed or mobile command center will require communication over a variety of networks at a variety of conditions of stress on the networks. An integrated, multi-rate voice/data terminal could be developed to support the commander's needs under normal and stressed conditions as follows:

- Under normal conditions, the terminal would provide secure digital voice, low-rate digital video, and graphics;
- Under heavily stressed conditions with network jamming and damage ,the terminal would be limited to stylized data messages;
- Under more favorable but degraded network conditions, more interactive communications would be provided, including very-low-rate secure voice using speech recognition and synthesis techniques.

4.2 Interactive Speech Enhancement Workstation

Advances in speech enhancement technology, coupled with the growing availability of high-performance graphics workstations and signal processing hardware, offer the opportunity for the development of an advanced, interactive speech enhancement workstation with multiple military applications. Such a system would include:

- Real-time speech I/O, including the capability for simultaneous handling of inputs from multiple microphones or sensors
- High capacity digital speech storage and playback facilities
- A user-selectable library of noise suppression, interference suppression, speech transformation, and filtering

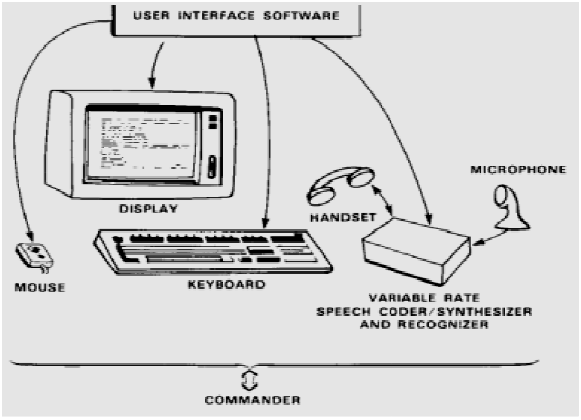

Figure 4.1: Integrated Multi-Rate Voice/Data Communications Terminal

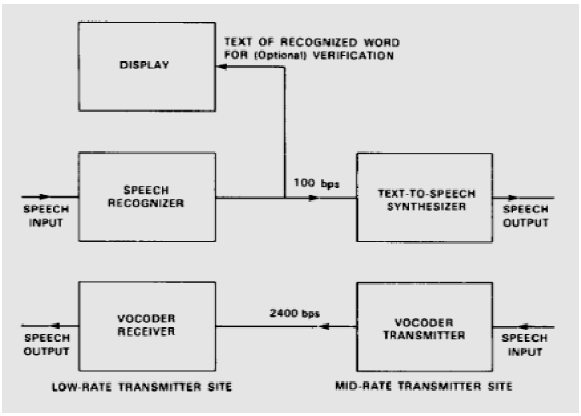

Figure 4.2: Concept of system for speech Recognition

4.3 Voice-Controlled Pilot's Associate System

Pilots in combat face an overwhelming quantity of incoming data or communications on which they must base life or death decisions. In addition, they are faced with the need to control dozens of switches, buttons, and knobs to handle the multiple avionics functions in a modern military airplane cockpit. Especially for the case of a single- seated military aircraft, substantial benefit could be achieved through the development of a voice controlled "pilot's associate", which reduces the pilot's work load, assisting the pilot in controlling avionics system and in keeping track of his changing environment. The concept of the pilot's associate was developed as part of the planning for the DARPA Strategic Computing Program as a paradigm for the development of intelligent "personal associate" systems which could have significant benefits in a variety of human-controlled ,complex, military systems. The pilot's associate would ultimately consist of an ensemble of real-time natural interface system and expert knowledge-based systems.

4.4 Advanced Air Traffic Control Training System

Automated training systems can use computer speech recognition and generation to expedite training and to reduce the load on training personnel in a variety of applications. Speech recognition and synthesis would be very helpful in hands-busy, eyes-busy training situations, for example in training personnel to maintain complex mechanical equipment. Here the individual could request information from an "automated instruction manual "while continuing to carry on a manual task, and while maintaining his view of the equipment (e.g., a complex jet engine). voice-interactive systems are perhaps most attractive

for training in tasks which require voice communication as an integral part of the operational task, such as air traffic control (ATC).Previous efforts in the application of speech technology n ATC training systems have achieved only limited success, but advances in speech technology, simulation technology, expert systems for automated instruction, and performance measurement offer significant potential for major advances in ATC training systems.

4.5 Spoken Language Translation System

Automatic translation of spoken natural language certainly represents one of the "grand challenges" of speech and natural language technology, as well as along-term opportunity for advanced speech technology. Applications of military relevance include: automatic interpreters for multi-language meetings, NATO field communications, a translating telephone, and translation for cooperative space exploration activities. The impact of automated spoken language translation would clearly be enormous; however, the problem is considerably more difficult than either voice-operated natural language dictation machines or machine translation of text; both of which are unsolved problems requiring much future research. It should be noted, however, that progress continues to be made in dictation systems and new initiatives in machine translation of text are being proposed and developed including application of the powerful statistical techniques which have been successful in speech recognition.

5 FURTURE OF VOICE AND SPEECH RECOGNITION

Experiencing the new modern paradigm shifts in technology will require humans to become one with the technologies they create and the ability to interface in real-time. Perhaps, the greatest step towards this goal is Voice Recognition Technology, where humans can talk and communicate in a way that is natural to them through their evolutionary process - vocal cords and speech. Thus, they are able to interface with the tools they have created.

Today, when you buy a new computer with the Microsoft Vista Operating System pre-loaded, it comes with Windows Speech Recognition, but even if you have an older computer, there are many good Voice Recognition products available such as Nuance's Dragon Naturally Speaking. Currently, this article is being written using the version 8.1, soon, It will upgrade to the next version 9.1, said to be even more accurate, within 99%. This is a huge improvement from my first try at speech recognition software, IBM Voice Type Dictation 3.0, I purchased back in 1995.

Over the past couple of years," I have indeed worn off the letters on three different laptop keyboards, perhaps I do not trim my fingernails as often as I should, or perhaps it has something to do with the fact that I write 4,000 to 14,800 words a day, pounding out articles on those plastic keys. Either way, for me Voice Recognition Software ranks up amongst the top greatest inventions of mankind. Technologies that increase productivity and efficiency are the most significant and at this point.

5.1 Future Advances for Voice Recognition:

1. Body Language + Facial Expression + Voice Recognition:

Currently, there are robotic android projects in the works in Japan and in the US; facial expression or mirroring, is very popular. The goal is for the human that interfaces with the system to create an emotional bond with the machine. Voice Recognition systems that also read body language and facial expression can also be used for threat assessment at let's say airports, border crossings and replace human workers at those locations or choke points.

If you smile at a robotic android and it smiles back at you, while you are having a conversation, this ups the emotional value of the conversation to the human. Perhaps the system might start complimenting you. If you are persnickety to the system, maybe it will mirror those responses or reciprocate an angry response or work to diffuse the situation, of course it all depends on its programming, but you can see the advances, potential applications and the trends going forward.

If you will recall Hal the famous science fiction computer, it said: "I sense hostility in your voice Dave." Perhaps since this was once in a science fiction work, human scientists today are trying to make it so. Right now, we are there, with have this technology, CRM Voice Recognition Software can sense emotion, hesitation, aggression, hostility, anger, etc. So, within five years we will see these features in more and more applications.

2. Emulation of Emotion and Empathy:

Emulating emotion and empathy is on its way right now. Currently, most consultants of artificial intelligent customer response systems for 'call centers' advise that the voice on the other end if coming from a machine, should be easily identifiable by the human calling in as a computer systems with voice recognition features, because humans do not like to be tricked, when they find out, it makes them upset. Of course, with the advent of emotion emulation or empathy it is possible and we have the ability to do this now.

Indeed, artificial intelligent computers have been used to go online and participate on forums and can participate for 15 threads or more, without detection. In voice recognition, if the voice sounds legitimate, a full conversation can go on for a while, without the human realizing it is talking to a machine.

With a call center system handling a complaint, a computer system might side with the customer and listen to them and even say;

"I know how you feel, I am so sorry this has happened, let me see what I can do" or;

"yes, I understand, this is very urgent, let me have you talk to my supervisor"

then pass the customer off to a real human or perhaps another voice system, with a more authoritative voice? The customer on the other line may never know that they are talking to a computer or computers. Indeed, this does not sit too well with many in the industry but it a place where software professionals of voice recognition are thinking and discussing now, surely you can see applications for this.

3. Understanding a Joke and Responding with another One:

Artificial Intelligence is getting better all the time, soon, AI software engineers will create joke recognition systems, where the computer will understand irony and know when the human is telling a joke, then reciprocate with a joke of their own, perhaps creating a joke from scratch. The system would be pre-loaded with all the jokes common to human interaction in all cultures. It will be able to pick one that has most likely not been heard by the human they are working with at the time; also put in memory that it has been told to that individual so it does not repeat it.

Wow, this is getting complicated fast isn't it, and this is exactly why it has not been fully achieved. Humor has been a huge stumbling-block for human voice recognition and artificial intelligence systems, yet it is something that humans have a knack for. Still, they are working on this challenge and we will see it within 5 to 10 years

4. Vocal Cord Vibration Recognition + Current Voice Recognition

Currently, there is advanced research in the US Military that allows vocal cords to be read, without actual speech or voice, these systems are working now. This is done with a device near the larynx that picks up sensitive vibrations, which is coupled to a transmitter for sending. The receiver or other special force member has a tiny ear piece so they can hear that speech, all silent to those nearby, within six inches of those using the system. This is getting pretty close to mimicking thought transfer, but in essence it is a form of voice recognition, hooked to a communication device.

These systems will get much better and soon the secret service members, special forces, SWAT teams, will no longer have little cords coming out of their ears, but they will communicate without notice. The larynx vibration speech recognition might be mounted inside a "clip tie" and no one will be the wiser. There are many applications for this if you think on it.

6 CONCLUSIONS

There has been much progress in recent years in speech technology and in application of this technology in military systems. This progress has brought some applications of speech technology into operational use, has brought other applications into the development and test stage, and has brought closer the opportunities for advanced applications. To realize the necessary technology for advanced applications, and to bring advanced applications into practice, much work is needed in basic speech algorithm technology, speech system implementation, and iterative test and improvement of fielded systems.

This paper has presented in overview form:

- A review and assessment of current military applications of speech technology;
- An identification of a sampling of opportunities for future military applications of advanced speech technology; and
- Identification of problem areas for research to meet applications requirements, and of promising research thrusts.

Applications in narrowband speech communication and speech enhancement are seen to be at hand, and opportunities as described for advanced voice/data workstations, based on extension and integration of current technology. In the speech recognition and spoken language understanding areas, a number of current applications are described which are generally in the development and test stage. A number of opportunities for advanced applications in these areas are described; these generally will require significant advances in speech recognition technology and in the integration of speech recognition into systems which will require advanced natural language processing and careful attention to system integration and human factors. Current efforts which are underway in speech recognition, natural language processing, and system development technology should lead, over the next several years, to significant technology advances in speech technology, and to significant progress toward realizing these and other opportunities for military applications of advanced speech processing.

REFERENCES

1. M.G. Abbott, "The Use of Speech Technology to Enhance the Handling of Electronic Flight Progress Strips in an Air Traffic Control Environment," Proc. Voice Systems Worldwide, London, Media Dimensions, pp. 126-134, May 1990
2. L.R. Bahl, et al., "Large Vocabulary Natural Language Continuous Speech Recognition," in IEEE Proc. Int. Conf. Acoustics, Speech and Signal Processing 1989, pp. 465-468, May 1989.
3. J.K. Baker, "The Dragon System - - An Overview," IEEE Trans. Acoustics, Speech and Signal Processing, Vol. ASSP-23, No. 1, pp. 24-29, February 1975.
4. B. Beek, "Overview of State-of-the-Art, R&D NATO Activities, and Possible Applications -Voice Processing Technology," Proc. AGARD Conf. No. 329 on Advanced Avionics and the Military Aircraft Man/Machine Interface, April 1982.
5. B. Beek and R.S. Vonusa, "General Review of Military Applications of Voice Processing," ProcAGARD Lecture Series No. 129 on Speech Processing, June 1983.
6. R. Bell, M.E. Bennett, and W.E. Brown, "Direct Voice Input for the Cockpit Environment," Proc. AGARD Conf. No. 329 on Advanced Avionics and the Military Aircraft Man/Machine Interface, April 1982.
7. P.E. Blankenship and M.L. Malpass, "Frame-Fill Techniques for Reducing Vocoder Data Rates," Technical Report 556, Lincoln Laboratory, M.I.T., February 1981.
8. S.F. Boll and D.C. Pulsipher, "Suppression of Acoustic Noise in Speech Using Two-Microphone Adaptive Noise Cancellation," IEEE Trans. Acoust., Speech and Signal Processing, ASSP-28, pp. 752-753, 1980. 446.
9. G. Booth, Chairman, Session on Training and Air Traffic Control in Proc. Speech Tech 90 (Media Dimensions), April 1990.
10. A.G. Bose and J. Carter, "Headphoning," U.S. Patent No. 4,455,675, June 19, 1984.
11. J.S. Bridle, R.M. Chamberlain, and M.D. Brown, *"An* Algorithm for Connected Word Recognition," Proc. ICASSP'82 (Paris).
12. P. Brown, et at., "A Statistical Approach to French/English Translation," Proc. RIAO 88 Conf. on User-Oriented Content-Based Text and Image Handling, Massachusetts Institute of Technology, March 1988, pp. 810-823.
13. J.P. Campbell, V.C. Welch, and T.E. Tremain "An Expandable Error-Protected 4800 bps CELP Coder (U.S. Federal Standard 4800 bps Voice Coder)," Proc. ICASSP'89, Glasgow, pp. 735-738.

14. A. Chapanis, "Interactive Human Communication," *Scientific American,* Vol. 232, pp. 36-49, March 1975.
15. E.J. Cupples and J.L. Foelker, "Air Force Speech Enhancement Program," Proc. Military Speech Tech 1987 (Media Dimensions) November 1987.
16. E.J. Cupples and B. Beck, "Application of Au-Computing Technology: A Strategic Plan for Its Development and Application to Critical Problems in Defense," 28 October 1983.
17. Proceedings of the DARPA Speech Recognition Workshop, Science Applications International Corp., Report No. SAIC-86/1546, February 1986.
18. Proceedings of the DARPA Speech Recognition Workshop, Science Applications InternationalCorp., Report No. SAIC-87/1644, March 1987.
19. Proceedings of the February 1989 DARPA Speech and Natural Language Workshop (Morgan Kanf- man), February 1989.