Implementation of Advanced Communication Aid for People with Severe Speech Impairment

Ms. M. Shunmugalakshmi (AP/ECE), Sneha Mohan. M, S. K. Sujitha

Department of Electronics and Communication Engineering SNS College of Engineering Coimbatore, India
Department of Electronics and Communication Engineering SNS College of Engineering Coimbatore, India

Abstract: *This paper is mainly aimed at developing an advanced communication aid for the people having severe speech impairment. This method gives rise to a new form of communication aid which recognizes the disordered speech of the user and builds messages which are converted into synthetic speech. While employing this method, the recognition perplexity can be increased. The selected message building technique traded-off various factors including speed of message construction and range of available message outputs. The trial high lightened some issues which limit the performance and usability of device when applied in real usage situations, with mean recognition accuracy of 67% in these circumstances. A field trial by individuals with moderate to severe dysarthia and confined that they can make use of the device to produce intelligible speech output from disordered speech input.*

Index Terms: *Communication aid, speech impairment, synthetic speech, recognition perplexity, dysarthia.*

I. INTRODUCTION

Spoken language communication is a fundamental factor in quality of life, but 1.3% of the population cannot use natural speech reliably to communicate. Generally speech impairment may be due to severe physical disabilities such as motor neurone disease, congenital conditions such as cerebral palsy or acquired neurological conditions such as stroke or traumatic brain injury. Current technological tools generally rely on a switch or keyboard for input. But those communication aids are relatively slow and disrupt eye contact. That user always needs a device which is physically easy to operate. Therefore it is desirable that a new communication aid retains as far as possible, the speed and ideally the naturalness of spoken communication. So that in order to reduce complexity and to improve the efficiency of the spoken language, an advanced communication aid was developed.

II. EXISTING SYSTEM

Commercially available automatic speech recognition systems can work well with mild and even moderate, but these studies show that there is an inverse relationship between the degree of impairment and the accuracy of speech recognition. For people with severe speech impairment, commercial speech recognition systems are not a viable access solution. Moreover, the small-scale laboratory experiments reported which is not representing the range of environmental conditions that are likely to be encountered in realistic usage, which is known to degrade recognition accuracy. Thus, while automatic speech recognition has been used for many years as a method of access to technology by some people with disabilities but unimpaired speech, it received little attention as a potential input channel for other voice output communication aids. But this has been experienced with certain disadvantages as they cannot be a viable access solution. This method is not suitable for environmental conditions and hence recognition accuracy.

III. PROPOSED SYSTEM

A new form of communication device for people with severe speech impairment recognizes the disordered speech of the user and builds messages, which are converted into synthetic speech. System development speech was carried out employing user-centered design and development methods, which identified and refined key requirements for the device .A novel methodology for building small vocabulary, speaker-dependent automatic speech recognizers with reduced amount of training data, was applied. Experiments showed that this method is successful in generating good recognition performance (mean accuracy 96%).On highly disordered speech, even when recognition perplexity is increased. Certain advantages are intelligible speech output can be produced and this method is suitable for impaired speech. On the other hand, new automatic speech recognition device (ASR) with WSN implementation.

IV. SYSTEM DESCRIPTION

The user speaks into a microphone and the speech is processed and recognized by a speech recognizer. The recognized words are passed to a message building module. According to this input, the massage building module will update the screen, potentially supply audio feedback to the user and determine the range of possible

future inputs. When the message is complete it is passed to the speech synthesizer, producing intelligible spoken output.

A .Recognition of speech

ASR and HMM of speech unit models are trained where data are recorded by many speakers. The speech units will be at the level of individual speech sounds, phones. The resultant speaker can be adapted for an individual speaker where a small amount of enrollment speech data from the speaker .Normally speech used in training models for many conventional adaptation techniques to be of assistance. This method introduces a new methodology for building small vocabulary, speaker dependent personal recognizers with reduced amounts of training data. Initial recordings were collected from the user. These were collected using either a headset microphone or a desktop microphone. Signals were sampled at 8Khz.The recordings consist of isolated productions of each of the words that are required for the recognizers input.

B. Building of message

The building module constructs the messages, which the user wishes to communicate from the corresponding input words. The ideal message buildings given that word units recognizes each word individually and speak out the same word in a clear voice. The accuracy of speech decreases rapidly as the input of vocabulary size increases for methods with larger perplexity

PARTICIPANTS IN THE USER TRAIL

ID	SEX	AGE	SPEECH DISORDER	AETIOLOGY	CURRENT VOCA
E1	F	30s	Severe dysarthria	CP	None
E2	M	80s	Progressive severe dysarthria	PD	None
E3	F	30s	Severe dysarthria	TBI	Lightwriter
E4	M	50s	Moderate dysarthria	CP	None
E5	F	70s	Progresssive moderate dysarthria	PD	None
E6	F	70s	Progressive severe dysarthria	MSA	Lightwriter
E7	F	40s	Progressive severe dysarthria	FA	Lightwriter

In the first column D denotes a participant who took part in the development stages and E denotes a participant who did not. The abbreviated aetiologies refers to cerebral palsy(CP),Parkinson's disease(PD),Traumatic brain injury(TBI),multisystem atrophy (MsA),and Fredreich's alaxia (FA).Participants used in the user trial may have moderate to severe dysarthria.

TABLE II

STAGES OF THE EVALUATION

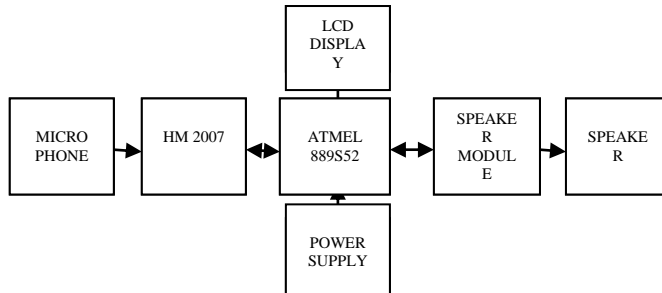| STAGE	DESCRIPTION	APPROXIMATE DURATION
1	Specify input vocabulary and output speech	1-2 weeks
2	Record input vocabulary	2-4 weeks
3	User training	2-4 weeks
4	Device familiarization	2 weeks
5	Trial usage	4 weeks
6	Interview & questionnaire	1 week

Synthesis of speech

A specially compiled version of software called open CV was prepared for the Windows Mobile for Pocket PC operating system. The acceptability of the output was evaluated with potential users. Some users preferred pre-recorded output for their system, either because of the potential for delay introduced by the computationally intensive synthesis process, or the perceived poor quality of the synthetic speech. The two major processes involved in the synthesis of speech are perception and repetition. Speech perception refers to the

process by which humans are able to interpret and understand the sounds used in language. Spoken vocalizations are quickly turned from sensory inputs into motor instructions needed for their immediate or delayed vocal limitations. This occurs independently of speech perception.

V. Hardware Implementaton

Block diagram

```
                    ┌──────────┐
                    │   LCD    │
                    │ DISPLAY  │
                    └────┬─────┘
                         │
┌────────┐   ┌────────┐  ┌────────┐   ┌────────┐   ┌────────┐
│ MICRO  │──▶│ HM 2007│◀▶│ ATMEL  │◀─▶│ SPEAKER│──▶│SPEAKER │
│ PHONE  │   │        │  │ 889S52 │   │ MODULE │   │        │
└────────┘   └────────┘  └───┬────┘   └────────┘   └────────┘
                         ┌────┴─────┐
                         │  POWER   │
                         │  SUPPLY  │
                         └──────────┘
```

POWER SUPPLY

The ac voltage, typically 230V is connected to transformer, which steps the ac voltage down to the level for desired dc output. A diode rectifier provides a full wave rectifier voltage that is initially filtered by a simple capacitive filter to produce a dc voltage. The resulting dc voltage usually has some ripples or ac voltage variation. A regulator circuit can use this dc input to provide a regulated that not only has much ripple voltage.

STEP DOWN TRANSFORMER

The step down transformer is used to step down the supply voltage of 230V ac from mains to lower values. The conversion of these ac values to dc values is done using the full wave rectifier unit.

RECTIFIER UNIT

A diode bridge is an arrangement of four diodes connected in a bridge circuit that connects the polarity of output voltage of any polarity of the input voltage. When used in its most common application, for conversion of ac into dc output.

INPUT FILTER

Capacitors are used as filters. The ripples from the dc voltages are removed and dc voltage is obtained. The primary action performed by the capacitor is charging and discharging. It charges in the positive half cycle of the ac voltage it will discharge in negative half cycle. So it allows only ac voltage and does not allow the dc voltage. This filter is fixed before the regulator.

REGULATOR

7805 is a voltage integrated circuit. It is a member of 78xx series of fixed linear voltage regulator IC's. The voltage regulated IC maintains the output voltage at a constant value.

OUTPUT FILTER

This filter is fixed after the regulation circuit in filter any of the possibly found ripples in the output received finally.

ATMEL 89S52 MICROCONTROLLER

The AT89S52 is a low power, high-performance CMOS 8bit microcomputer with 8K bytes of flash programmable erasable read only memory. The device is manufactured using Atmel's high-density non volatile memory technology and is compatible with the industry-standard MCS-52 instruction set and pin out.

LIQUID CRYSTAL DISPLAY (LCD)

LCD is a type of display used in digital watches and many portable computers.LCD displays utilize two sheets of polarizing material with a liquid crystal solution between them. An electric current passed through the liquid causes the crystals to align so that light cannot pass through them.LCD technology has advanced very rapidly since its initial inception over a decade ago for use in laptop computers. The liquid crystals can be manipulated through an applied electric voltage so that light is allowed to pass or is blocked. By carefully controlling where and what wavelength of light is allowed to pass, the LCD monitor is able to display images. A

backlight provides LCD monitor's brightness. The three major kinds of LCD used are Smectic, Mnematic and Cholestric.

HM2007

HM2007 is 48-pin single chip CMOS voice recognition LSI circuit with on-chip analog front end.HM2007 is a speech recognition kit. It consists of an IC chip which is a kind of voice recognition circuit in varied specifications. This module comes with an analog front end, recognition process, sound control functions and voice analysis. When combined with microprocessor, these IC chips can be turned into an intelligent recognition system. The speech recognition system is a completely assembled and easy to use programmable speech recognition circuit. Programmable, in the sense we can train the words or vocal utterances. It has 8 bit data out which can be interfaced with any microcontroller for further development. A 40 word isolated-word voice recognition system can be composed of an external microphone, keyboard, SRAM and a few other components. When combined with a microprocessor, an intelligent recognition system can be built. The speech recognition system is a completely assembled and easy to use programmable speech recognition circuit.HM2007 is a 48-pin single chip CMOS voice recognition LSI circuit with on-chip analog front end, voice analysis, recognition process and system control functions.

Flat panel speakers use voice coils or other apparatus mounted to flat panels to act as the source of sound. Standard flat panel electrodynamics loudspeaker has been difficult to make because it is difficult to vibrate the entire flat surface evenly while creating good frequency response. Thus other speaker types have evolved to try to make a flat speaker.

A microphone is an acoustic-to-electric transducer or sensor that converts sound into an electrical signal. The microphones generally use electromagnetic induction, capacitance change, piezoelectric generation or light modulation to generate an electrical voltage signal from mechanical vibration. Hence they convert acoustical energy (sound waves) into electrical energy (the audio signal).When the diaphragm vibrates; it causes other components in the microphone to vibrate. These vibrations are converted into an electrical current which becomes the audio signal. At the other end of the audio chain, the loudspeaker converts the electrical energy back into acoustical energy. Hence the conversion of both electrical energy into acoustical energy and acoustical energy into electrical energy are done in a microphone.

SPEAKER MODULE

The VEX speaker module is used to generate output sounds from the VEX cortex microcontroller.

VI.   Software Implementation

SPEECH

Speech is the vocalized form of human communication. It is based upon the syntactic combination of lexical and names that are drawn from very large(usually about 10,000 different words).Each spoken word is created out of the phonetic combination of a limited set of vowel and consonant speech sound units. These vocabularies, the syntax which structures them and their set of speech sound units differ, creating the existence of many thousands of different types of mutually unintelligible human languages. Most human speakers are able to communicate in two or more of them, hence being polyglots. The vocal abilities that enable humans to produce speech also provide humans with the ability to sing.

Speech is a complex activity; as a result, errors are often made in speech. Speech errors have been analyzed by scientists to understand the nature of the process involved in the production of speech.

NEURAL NETWORK

Neural networks cannot do anything that cannot be done using traditional techniques. But they can do some things which would be otherwise very difficult. In particular, they can form a model from their training data (or possibly input data) alone. A great deal of research is going on in neural networks worldwide. This ranges from basic research into new and more efficient learning algorithms to networks which can respond to temporarily varying patterns, to techniques for implementing neural networks directly in silicon. Already one chip commercially available exists, but it does not include adaptation. Edinburgh University have implemented a neural network chip, and are working on the learning problem. Production of a learning chip allows the application of this techno.ogy to a whole range of problems where the price of a PC and software cannot be justified. There is particular interest in sensory and sensing applications: nets which learn to interpret real-world sensors and learn about their environment.

In computer science and related fields, artificial neural networks are computational models that are capable of machine learning and pattern recognition. They are usually presented as systems of interconnected neurons that can compute values from inputs by feeding information through the network.

For example, in a neural network for handwriting, recognition, a set of input neurons may be activated by the pixels of an input image representing a letter or digit. The activations of these neurons are then passed on, weighted and transformed by some function determined by the network's designer, to other neurons, etc., until finally an output neuron is activated that determines which character was read.

OPEN CV
Open source computer vision library is a library of programming functions mainly aimed at real-time computer vision, developed by Intel, and now supported by Willow Garage and Itseez. It is free for use under the open source BSD license

PROGRAMMING LANGUAGE
Open CV is written in C++ and its primary interface is in C++, but it still retains a less comprehensive though extensive older C interface. There are now full interfaces in Python, Java and MATLAB/OCTAVE (as of version 2.5). The API for these interfaces can be found in the online documentation. Wrappers in the older language such as C#, Ch, Ruby have been developed to encourage adaptation by a wider audience. All of the new developments and algorithms in open CV are now developed in the C++ interface. A CUDA based GPU interface has been in progress since September 2010.An open CV based GPU interface has been in progress since October 2012, documentation for version 2.4.5 can be found here.

OS SUPPORT
Open CV runs on windows, android, maemo, free BSD, open BSD, Ios, Black Berry10, Linux, and OSX. The user can get official releases from Source Forge, or take the current snap shot under SVN from there. Open CV uses CMake.

FEATURES
- Image data manipulation
- Image and video I/O
- Matrix and vector manipulation and linear algebra routines
- Various dynamic data structures
- Basic image processing
- Structural analysis
- Camera calibration
- Motion analysis

APPLICATIONS
- Egomotion estimation
- Facial recognition system
- Gesture recognition
- Human-computer interaction(HCI)
- Mobile robotics
- Motion understanding

VISUAL STUDIO
Microsoft visual studio is an integrated development environment (IDE) from Microsoft. It is used to develop console and graphical user interface applications along with windows forms or WPF applications, web services and also windows store apps in both native code together with managed code for all platforms supported by Microsoft Windows, Windows Mobile, Windows CE, NET Framework, NET Compact Framework and Microsoft silver light.

FEATURES
- Code editor
- Debugger
- Designer
- Extensibility

ARTIFICIAL NEURAL NETWORK

An artificial neural network is an interconnected group of nodes. They are usually presented as systems of interconnected "neurons' that can compute values from inputs by feeding information through the network. Like other machine learning methods, neural networks have been used to solve a wide variety of tasks that are hard to solve using ordinary rule-based programming, including computer vision and speech recognition. A statistical models may be called "neural" if they

    1. Consist of sets of adaptive weights i.e., numerical parameters that are tuned by a learning algorithm, and
    2. Are capable of approximating non-linear functions of their inputs.

VII.   Conclusion

This system gives an exact solution for the trained words irrespective of its environmental conditions. Hence more number of words can be trained and pre programmed for generating the voice output at the speaker. This system has also described the development of portable, voice output communication aid controlled by automatic speech recognition. The device can be configured to enable the user to create either simple or complex messages using a combination of a relatively small set of input "words". User consultation suggests that such a device would be acceptable and would be useful in communication situations where speech and intelligibility are crucial.

VIII.   Discussion

A voice output communication aid controllable by automatic speech recognition has been successfully produced and tested. The development of the device followed a user-centered iterative process; where by a group of users evaluated each stage of the development and this lead to modifications and improvements to the device. Intension behind this system is to provide a sophisticated environment for the people having speech impairment with ease. Speech recognition technique will build on previously successful development of speech-based various methods for speech translations are being evaluated. In real time, highly disorder speech can also be build into continuous messages by training the databases in neural networks. This system does not impose any limitation of words. On the other hand, word perplexity can also be increased.

REFERENCES

[1]   D.Beukelman and P.Mirenda, Augumentative and Alternative communication, 3rd ed.Baltimore, MD: Paul H.Brookes,2005.
[2]   P.Enderby and L.Emerson, Does Speech and Language Therapy Work, London, U.K.: Singular, 1995.
[3]   C.L.Kleinke,"Gaze and eye contact: A research review, Psychol.Bull" vol.100, no.1, pp.78-100, 1986.
[4]   B.O'Keefe, N, N.Kozak, and R.Schuller,"Research priorities in augmentative and alternative communication as identified by people who use AAC and their facilitators,"Augumentative Alternative Commun., vol.23, no.1, pp.89-96, 2007.
[5]   J.Murphy,"I prefer contact this close: Perceptions of AAC by people with motor neurone disease and their communication partners,"Augumentative Alternative Commun., vol.20, pp.259-271, 2004.
[6]   B.WisenburnandD.J.Higginbotham,"An AAC application using speaking partner speech recognition to automatically contextually relevant utterances: Objective results;"Augmentative Alternative Commun., vol.24, no.2, pp.100-109, 2008..
[7]   L.J.Ferrier, H.C.Shane, H.F.B.allard, T.Carpenter, and A.Benoit,"Dysarthric speaker's intelligibility and speech characteristics in relation to computer speech recognition, "Augmentative alternative commun, vol.11, no.3, pp.165-175,1995.
[8]   N.Thomas-stonell, A.L.Kotler, H.A.Leeper, and P.C.Doyle,"Computerized speech recognition: Influence of intelligibility and perceptual consistency on recognition accuracy "Augmentative alternative commun. vol.14, no.1, pp.51-56, 1998.
[9]   R.N.Bloor, K.Barret, and C.Geldard,"The clinical application of microcomputers in the treatment of patients with severe speech dysfunction,"IEE Colloquium High-Tech Help Handicapped,pp.9/1-9/2,1990.
[10]   U.Sandler and Y.Sonnenblick,"A system for recognition and translation of the speech of handicapped individuals, "in 9th Mediterranean Electrotech.Conf.(MELECON'98),1998,PP.16-19.
[11]   B.Wisenburn and D.J.Higginbotham,"An ACC application using speaking partner speech recognition to automatically produce contextually relevant utterances: Objective results, "Augmentative Alternative Tully relevant utterances: Objective results, "Augmentative Alternative commun. vol.24, no.2, pp.100-109, 2008.
[12]   M.S.Hawley et al.,"A speech-controlled environmental control system for people with severe dysarthria, Med.Eng.Phys" vol.29, no.5, pp.586-593, 2007.
[13]   H.V.Sharma and M.Hasegawa-johnson,"State-transition interpolation and MAP adaptation for HMM-based dysarthric speech recognition, "in NAACL HLT Workshop Speech Language Process. Assistive Technol., 2010, pp.72-79.
[14]   R.Palmer, P.Enderby, and M.S.Hawley,"A voice input voice output communication aid: What do users and therapists require?"J.Aaaistive Technol., vol.4, no.2, pp.4-14, 2010.
[15]   S.Young et al., Cambridge University Engineering Department the HTK book, 2006.
[16]   P.D.Green, J.Carmichael, A Hatzis, P.Enderby, M.S.Hawley, and M.Parker,"Automatic speech recognition with sparse training data for dysarthric speakers, "in Eur.Conf.Speech Commun.Technol.(Euro speech),2003,pp.1189-1192.
[17]   L.R.Rabiner and B.H.Juang,"An introduction to hidden Markov models,"IEEE ASSP Mag., Jun.1986.
[18]   R.Palmer, P.Enderby, and S.P.Cunningham,"The effect of three practice conditions on the consistency of chronic dysarthric speech,"J.Med.Speech-Language Pathol., vol.12, no.4, pp.183-188, 2004.
[19]   J.N.Holmes and W.J.Holmes, Speech Recognition and Synthesis, 2nd ed.London, U.K.: Taylor Francis, 2001.
[20]   S.L.Glennon and D.C.DeCoste, Handbook of Alternative and augmentative communication.san Diego, CA: Singular, 1997.