Communication Aid for Severe Speech Impairment

Aswathi N G[1], Juliet Joju M[2], Judith Sen E[3]

[1,2,3]**Jawaharlal College of Engineering and Technology, Calicut University**
Palakad, Kerala, India

Abstract

Communication aid is a device that helps the user to construct messages from the disordered speech and is converted into synthetic speech. The main aim of this device is to interpret the speech of a severe dysarthria patient without the help of his/her relatives. The system consists of a user centered design and is developed to satisfy the primary requirements. In order to reduce the amount of training data, a novel methodology for building small vocabulary, speaker-dependent automatic speech recognizers, was applied. Even though the recognition perplexity is increased; this method is successful in generating good recognition performance on highly disordered speech. The aid helps the persons with moderate to severe dysarthria by producing intelligible speech output from disordered speech input.

Keywords: Augmentative and Alternative Communication, Baum-Welch Algorithm, Hybrid Synthesis.

1. Introduction

Speech is a way of communication to express ideas, emotions and for better understanding. It is one the simple way of communication between humans that it does not require any skill, tools or education. But for a few, due to some disabilities caused either due to accident or naturally acquired by birth, they find it difficult to communicate in a natural way. This disability may be due to the neurological conditions occurred to the person such as a cerebral palsy or acquired neurological conditions occur due to traumatic brain injury or stroke.

It is difficult for the person having disabilities to communicate with unfamiliar people. So such case, augmentative and alternative communication aids (AAC) helps them to bridge their gap in communication. With the advancement in technology such aids have been more helpful to people for interaction. Voice-output communication aids (VOCAs) which basically works on a switch or keyboard for input. The user find it difficult and tiring and hence it does not provide a natural way of communication. So they are found to be relatively slow and can disrupt eye contact.

A new communication aid need to be developed for dysarthria patient considering in mind to maintain the speed and naturalness of their speech. This arise the need for a new communication aid which is easy to operate by the patient at different conditions and environment. This new communication system is capable of eliminating a major communication gap between the vocally disabled with common community.

2. Literature Survey

One of the most commonly acquired disorders of communication is dysarthria. Alternative communication method or improved vocalization is found to improve the ability of communication with severe dysarthria patients [1]. Different communication devices are available for people with severe speech disabilities depending on age and situations. But these aids will usually not preferable as the patients opt to speak rather than use the aid, even their speech is more unintelligible, as their natural instinct is to talk.

Automatic speech recognition system is a method implemented by the technology to help the people with disabilities in speaking. But this method or system has not been a solution to a far extend. This method is not adaptable with environmental conditions and so it can affect the accuracy of detection. The proposed system can recognized and interpret an individual's disordered speech and deliver the required message in clear synthesized form [2].

Augmentative and alternative communication (AAC)

systems [3] are usually used by adults with acquired neurogenic communication disorders. The neurological conditions is such that the person can gradually or suddenly lose their abilities to speak and is hence forced to depend on AAC system to meet their needs. With the help of computer, a person with severe speech and motor impairment can get himself more involved into the world. So computer provides a different aspect to people with disabilities, 'friend' is one such communication aid for people with motor impairment [4]. Motor impairment is defined to be a partial or complete disorder of motor neurons leading to disabled muscle neurons. Dr. Stephen Hawkins is one such person whose having motor impairment and also developed one such software to communicate to the outside world with the help of computers.

Many of the motor impairment patients are not in a condition to afford and use the appropriate software according to their need. Because a person having severe speech impairment along with motor impairment want to be able to communicate either through vocal or by sign language. So there comes a need to develop a flexible system which can be available to users within a wide range of physical and mental disabilities. The main advantage of the system is that the interfaces in the aid are user centered. Here a single system can provides different interfaces suitable for different user than a single interface.

3. System Description

The system requirement considers both communication and user and the voice output communication aids. The system consist block of speech recognition, message building and speech synthesis.

The input which is a user speaking into microphone and that speech is recognized by a speech recognizer [5]. The message building module, receives these recognized words from the recognizer. The message building module will update screen, supply audio feedback to user and determine the range of all possible future input depending on this input. This message building is an iterative method. One the message is complete it is passed to speech synthesizer giving a clear and intellectable output from the speaker.

3.1 Speech Recognition

Speech understanding and speech recognition are two different aspects but related speech process. Understanding a speech involves the proper recognition and responses given by a person to an utterance of a speech. Whereas speech recognition is just to recognize the speech without understand what is normally delivered. A variety of techniques were used among which HMM was the one found to have comparative better performance.

A speech recognition mainly have 3 stages: a preprocess stage that converts the speech signal to a sequence of observation symbols, a language model to predict the occurrence of words in string and acoustic model to predict the pronunciation of this word string.

The first stage of recognizer involves the decision regarding an input acoustic data. This is implemented with the help of pattern recognition algorithms on the speech signals. The speech is recorded or collected from the person using microphone and sampled on Bluetooth audio channel at 8 KHz. Using this a vocabulary needed for recognizer is done by recording the produced output words.

(i) Hidden Markov Model:

A Hidden Markov Model (HMM) is a mathematical tool to model time series. It provide algorithm to estimate state and parameter and automatic performance and time warping of signals that are reduced or stretched [6]. HMMs all based on Markov Chain, here it is consist of HMM with 11 states. The acoustic vectors were 12 Melfrequency cepstral coefficients derived from a filter bank with an analysis window. HMM implement stochastic models from the already known utterance and compares both known and unknown utterance given by each model.

(ii) Baum-Welch Algorithm

The Baum-Welch Algorithm determinates the parameters that best matches the observation data and hence gives HMM the best observed data.

3.2 Message Building

The message building module constructs the message that user needs for communication. The message that is built according to the recognized words. And produced

the same word in synthesize form. The design of the system is based on the user and condition of this disability. This helps to interact with the persons that he/she may be not familiar to or either in any emergencies.

Here we have to consider the speed of speech also the correction of errors if occurring. The error occurred while speech recognition is comparatively high than other methods of giving input method. Different message building method showed that large input vocabularies give greater communication rate. Similarly communication rate falls off rapidly with decreasing recognition rate which may be due to the time cost of correcting errors.

Combining the model of communication rate with information on recognition for a variety of ranges of input vocabularies, the communication rate for message building methods for different levels of dysarthria. Severe dysarthria patient needs an efficient method with a smaller input vocabulary for message building technique. So the components of message building method, the need of the user decides the type of message building method and the input and output vocabularies.

3.3 Speech Synthesis

Speech synthesis demands the need when the produced speech is not suitable for the application. So a separate amplifier and speaker is also needed along with the output and hence the system becomes voice input voice output communication and in turn controlled by automatic speech recognition system.

For designing such system the user needs to have prerecorded and synthetic output which appear similar to natural sound. In order to complement both of this idea in design, few techniques were introduced. Among the various introduced techniques concatenative and hybrid methods were considerably efficient and widely used method.

A high quality synthesis method using a festival system to synthesis the voice outputs and save this as speech wave form and played when required.

(i) Concatenative synthesis

The implementation of prerecordered samples is used in synthesis to generate natural speech. Using these technique parts of prerecordered natural speech is used to generate speech. This method is usually a data base driven synthesis method. The present set of synthesizers use Concatenative method, though the formant synthesis was dominant for a long time. A number of different methodologies are used by concatenative synthesis such as TDPSOLA, PSOLA, MBROLA and Epoch Synchronous Non Overlapping Add (ESNOLA).

Based on the application of speech data base used Concatenative synthesis can be classified as unit selection synthesis which uses a large set of speech data base, diphone synthesis, which uses minimum speech data base consisting of all diphones of a given language. Domain specific synthesis is also a method that Concatenates prerecordered words and phrases to create output. But the disadvantage is that the synthesizer output is limited to only a particular domain.

(ii) Hybrid Synthesis

One of the disadvantages of producing a natural sounding speech output is that some segment combination of synthesized speech is discontinues at segment boundaries. The other method of producing speech allows a control over fundamental frequency but with voice sounds more synthetic. So to avoid the problem we introduce a hybrid system which combines time and frequency domain methods.

(iii) Festival

Festival is aimed speech synthesis for at least 3 level of users. First, for people who need to have high quality speech form unknown text with less effort. Second, for those who want develop language system along with synthesis output. Third, involves developing and testing different and new synthesis method.

Festival includes accommodating build speech synthesis with various models. It is also multi-lingual hence is helpful to develop in different language.

4. Conclusion

The system discusses the different methods implemented as a solution for the different levels of speech impaired persons, mainly dysarthria patients. It also takes account of the condition and environment of the user, giving solution

for training. Hence more words can be involved in vocabulary and trained for generating output at the speaker. The development of a 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 small set of input "words". Such communication aids are useful in situation for people with speech disabilities where speech and intelligibility are necessary.

Acknowledgments

References

[1] Pam Enderby, Joyce Emerson, "Speech and language therapy: Does it work?", BMJ vol.312, June

[2] Ibrahim Patel, Dr. Y. Srinivasa Rao, "Automated Speech Recognition Approach to Continuous Cue-Symbols Generation", International Journal of Power Control Signal and Computation (IJPCSC) Vol.2 No.1

[3] David R. Beukelman, Susan Fager, Laura Ball and Aimee Dietz, "AAC for Adults with Acquired Neurological Conditions: A Review", Augmentative and Alternative Communication, September 2007 Vol. 23 (3), pp. 230 –242

[4] Pradipta Biswas, Debasis Samanta, "Friend: A Communication Aid for Persons with Disabilities", IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 16, No. 2, April 2008

[5] Mark S. Hawley, Stuart P. Cunningham, Phil D. Green, "A Voice-Input Voice-Output Communication Aid for People With Severe Speech Impairment", IEEE Transactions on Neural Systems and Rehabilitation Engineering , Vol. 21, No. 1, January 2013

[6] Ir. P. Wiggers, Dr. drs. L.J.M. Rothkrantz, "Automatic Speech Recognition Using Hidden Markov Models", Real-time AI & Automatische Spraakherkenning September 2003

[7] Simi Oommen, Greeshma Liz Jose,"Communication Aiding System for People with Speech Impairment", International Journal of Engineering Research & Technology, Vol.3, October 2014