

Voice Technology & Speech Recognition in ELT

Dr.S.Mercy Gnana Gandhi
Sathyabama University
drmerci2010@gmail.com

Prof.R.Augustian Isaac
SRM University
sangam.naadu@gmail.com

Abstract -Understanding the potential of the technology and its impact on society is one of the keys to developing new applications and services. People are able to communicate with the Worldwide Web everywhere and at any time, using mobile devices and other tailored human-machine interfaces. Researchers in the acquisition of L2 phonology have attempted an experimental assessment of the progress of the students after using new methods and softwares for learning . The paper investigates the unlimited access to a speech-recognition-based language learning program that would improve the efficiency of pronunciation of a group of engineering students studying in India. Ten students were given the program Talk to Me as a 100-hour course in Technical English, and were made positive to practice on their own computers. Their progress was compared with a control group of another 10 students who are not provided with software. The author used Talk to Me software which uses speech recognition to afford comfortable language practice and scoring of pronunciation. Students met at regular intervals with a pronunciation tutor who could steer them in the right direction for finding the most important sections to practice their particular pronunciation problems. Students were highly satisfied at the software application and used it for an average of 10 hours. They were tested and the results indicate that practice with the program was beneficial to those students who began the course with a strong foreign accent but that students who began the course with intermediate pronunciation did not show the same improvement.

Keywords: *Technology, applications, softwares, computers,multimedia*

I. INTRODUCTION

The World Wide Web has now reached people in all the corners of the world and computing and Computers add a new dimension to the Web society. The rapid growth of computer technologies and their remarkable applications in numerous educational disciplines has resulted in the use of computer-mediated communication (CMC) for Foreign Language Acquisition and Teaching. Moreover, the variety of technologies used for FL acquisition has become unparalleled, when compared to the technological developments within other domains. Foreign language technologies have also been developed, implemented and researched by FL

specialists. Researchers have been eager to investigate claims that the use of technology can contribute to positive student outcomes.

II. BACKGROUND

Microsoft has been involved in research on speech recognition and text to speech. Speech recognition technology has been used in some of Microsoft's products, including Microsoft Dictation which was also included in Office XP, Office 2003. Voice recognition software has been developed to provide a fastest method of writing on a computer which helps people with a variety of disabilities. Obviously, it can also help those with spelling difficulties. They work by analysing sounds and converting them to text. Speech Recognition software is widely available to everyone at a fairly reasonable price. Hence, teachers need to look at how they can use this type of software to enhance their curriculum.

III. A SOURCE OF LANGUAGE LEARNING

The most inspired uses of Speech Recognition Software is to assist language learning. Presently, plenty of language softwares are available which can check a student's ability to speak languages. For example, a student learning Hindi can be asked to state specific words in Hindi. The computer can then evaluate their ability to speak the words properly. The students are now asked to read the passage silently in their native language so that they can achieve the target language . Thus software can speak to the student in Hindi and then assess the students response to the new statement to determine if it was right. It is suggested that in future, computers may have the ability to evaluate a students' response and then respond back with its own customized response.

IV. BASICS OF SPEECH RECOGNITION

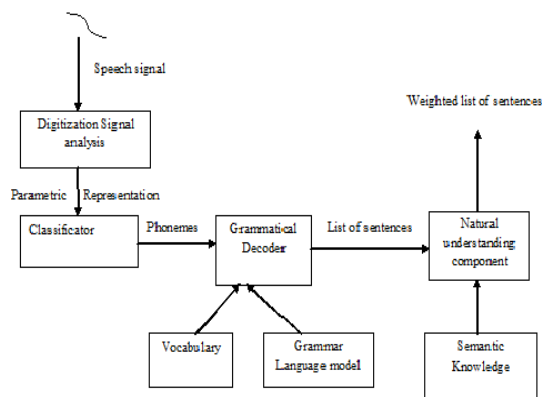
Speech recognition is a wide complex field. Just as written language has letters as units, speech recognition is based on phonemes and words. The speech signal is transformed into frequency spaced representation for further processing. Vowels only generate resonance frequencies in the speech signal which are called as formants. Consonants are produced by turbulent airflow in the human speech

system. From formants, speech recognition system can derive phonemes via signal analysis.

a) Technology of Continuous Speech Recognition System

The system can receive the digitized speech signal as an input and transforms it into the frequency space. This signal analysis represents the spectral distribution of speech signal for a given time frame. This representation of speech signal is given to the classifier, which uses pattern recognition to generate phonemes out of the vector. The grammatical decoder uses the vocabulary and grammar rules to construct a list of possible sentences for the received phoneme sequence. Grammatical words or language words of vocabulary are allowed in a sentence. At last the natural understanding component reads this list and sorts out

Overview Of Continuous Speech Recognition System



the most probable words based on semantic knowledge.

There are certain skills and characteristics needed by a student to use speech recognition software particularly his language, reading, spelling and speech abilities and above all strong determination to do any type of language task. The victory of this program relies on the quality of training and support which is available for the student. Fortunately, some students are successful in using the program with very limited training when some others require constant training and support to make effective use of the software. Moreover, there is a need for good language processing skills including the capability to speak in complete sentences or at least long phrases. It may be more valuable for few students with learning disabilities to use in such distinct speech software since it is intended to take word-by-word input rather than speaking obviously in longer sentences.

b) Challenges of Implementation

With any novel technology tool, students have become more comfortable with using speech-to-text, along with the training to recognize their voices. Very soon, he gains experience with an innovative way of script, knowing the difference between speaking and writing and thereby correcting errors within the text. The availability of the flourishing softwares has improved to certain extent, but on the other hand speech-to-text programs are not capable of distinguishing the voices all the time. Speaking for writing is an action which requires diverse skills than speaking in discussion. Students must be conscious of the distinction between the two. It is truly challenging for early writers, who had not made that distinction. Using speech recognition, technology can be associated with instructions on various writing strategies, brainstorming, drafting and arranging the ideas. There is a need for error correction and observing of misrecognized words. So students should be aware of errors which are unrecognized by the program. Another important execution challenge is that the software necessitates a good deal of memory and must be saved on a single server folder. These voice files could improve the accuracy of language pronunciation and so it is imperative that students work in their own saved file. This assistive technology is not always transferable. Schools and colleges have overcome this challenge by assigning students the brand new laptops with advanced software installed or storing files on a networked server that can be accessed from anywhere on campus.

V. HOW IT STARTED

Tentative research in speech recognition technology originated in industrial research labs in the United States in the 1950s. As computers came with limited computing power till the late 1970s, they were not equipped to analyse a continuous speech pattern. Major research advances in the mid-1980s finally made it possible for a desktop personal computer to recognise a vocabulary of 20,000 words — a meagre capacity compared to the extensive vocabulary of 250,000-300,000 words in contemporary speech recognition software. Speech recognition has come a long way from the 1990s, when initial offerings required a lot of additional hardware to operate. Now speech recognition software can run on standard platforms such as Windows XP and Windows Vista. Programs permit dictation of text directly into a voice-aware application that collaborates with the word processor. An online error correction feature allows words to be transcribed based on the context in which they are used. Efficiency may vary depending on user accent, but typically there is a learning feature that enables the system to improve its recognition of users' speech with regular use. It's also very fast — high quality speech recognition

software can type out 160 words per minute depending on your computer's processing speed.

Speech recognition technology is finding new applications in a variety of fields. Today speech recognition is also being used in mobile phones and personal digital assistants. For instance, in mobile phones it allows you to train the phone to call someone by just uttering his or her name. "Earlier, speech recognition technologies that came in from the West were attuned to Western pronunciation and had trouble deciphering the Indian accent," says Dr Aniruddha Sen, senior researcher at Tata Institute of Fundamental Research (TIFR). "However, the scenario is changing as more IT companies and research firms are working to develop speech recognition technologies compatible with Indian English and regional Indian languages." IBM recently launched speech recognition in Hindi that is even sensitive to variations in dialect. The software that has a vocabulary of 75,000 words can enable semi-literate and physically challenged people to access information through voice-enabled ATMs, kiosks and other such devices.

VI. CHALLENGES OF SPEECH RECOGNITION

Since 1950's, a lot of research has been carried out to achieve speech recognition systems with accurate recognition rates which are acceptable for every day use. Moreover, such systems are not provided for the generic case of speaker independent and task independent speech recognition. Speech recognition has some serious challenges and the recognition accuracy is constrained by the following problems.

(1) Environmental Noise :

Obviously, environmental noise has a remarkable impact on the recognition rate. Critical environments such as Public places, offices, factories, Trains etc., affect the speech recognition system very much. Still these systems can be adopted through proper training and filtering to special environmental noise.

(2) Line and Digitization Noise :

Systematic distortion such as echoes during overseas calls are damping of high frequencies are introduced by regular analog lines. Short interruptions on digital lines such as GSM create systematic distortions.

(3) Vocabulary ambiguity

Humans use context information to solve this kind of problem. Words like 'show' and 'snow' sound similar and it is difficult for the speech recognition system to recognize the correct word. It is easy for human to decide that the sentence must be

'show me your house' and not 'snow me your house', but for a computer this is a tedious task.

(4) Vocabulary Size

The accuracy of speech recognition depends on the vocabulary size which varies inversely with the accuracy and efficiency. Moreover, most of the words introduce more ambiguities and require more time to process.

(5) Channel Quality

Channel Quality depends primarily on the microphone, while different microphones create different speech signals. If the channel uses a digital line such as GSM, the digital processing of these channels will have an effect on the recognition rate.

(6) Task and Language constraints

Speech recognition is much simplified when the system uses limited dictionaries and grammar. For example, an Airline information system can restrict its grammar to say, 1000 possible combinations of words. This increases the recognition speed and accuracy dramatically and systems with word recognition rate higher than 70% are using this technique.

(7) Isolated, connected and Continuous speech

Isolated words are easily distinguished from each other whereas continuous speech recognition is very difficult if the words are not defined clearly. As we all know, poor articulation of words creates ambiguities and stronger co articulation causes problems. In spontaneous continuous speech, the sentences are often grammatically incorrect and contain lot of hesitations or press noises such as 'hmmmm'. It is difficult to recognize the word boundaries in connected word recognition.

(8) Speaker dependent / independent

As single speaker speech recognition systems perform much better than the systems dealing with different speakers, because once speaker performance is always better since, there is a possibility of training him quickly and efficiently.

With incessant progress of hardware support and speech recognition algorithms, speech recognition will naturally increase its usability and will be widely applied in new areas in the future.

VII. METHOD

a). Framework

This study involved two groups of students, the control group taking a course in the summer of 2014, and the experimental group taking the same course the following term, Autumn 2014. The experimental group received Talk to Me English (1) as

supplemental courseware. The course was a 100-hour, ten-week course in Technical English for Engineering students offered at Sathyabama University in India. In the summer of 2014, students followed the normal course plan and were tested for the purposes of future comparison. A three hour pronunciation tutoring practice was assisted by a software program in which they did not receive their own pronunciation program. But in the Autumn of 2014, students were offered a chance to deal an hour's tutoring for Talk to Me on their own computers. The course design and content for the experimental and control groups, was generally the same but with different teachers.

b).Subjects

The students in the course were engineers from different language backgrounds. The students were admitted to the course based on their performance in a placement test, who possessed varying skills in English ranging from advanced beginner to upper intermediate.

VIII. RESULTS

The course at the end was enjoyed and the students stated that the program was full of fun to use and thought it benefited their English course. This paper concentrates in highlighting the sub-score for pronunciation, which replicates the hearing models in the speech recognition system. It is thus a reflection of phonetic accuracy.

IX.CONCLUSION

Talk to Me provides language learning potential by highlighting the 'worst' word in an utterance and giving scores on pronunciation. The users also require power in finding the necessary information to correct their mistakes. The software could make the student become aware of that he had troubles in understanding a specific phoneme. The student being actively engaged in the diagnosis of his pronunciation difficulties is pedagogically desirable. The ideal automatic system would also point the student in the direction of the appropriate remedial activities. The fact that the four students showing the greatest improvement had used the program the least on their own could indicate that it was important that a large proportion of the learning time was spent in conjunction with the human tutor. The results of this study show just how important it is that the

pedagogical tools be appropriate to a given student's level of development in the target language.

X. ACKNOWLEDGEMENTS

The author expresses her hearty thanks to the management of Sathyabama University and the students and faculty for their valuable help and participation in this study.

REFERENCES

1. Cucchiari, C; Strik, H; & Boves, L. "Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms" in *Speech Communication* 30, 109-119, 2000.References:
2. Chun, D. (1998). Signal analysis software for teaching intonation. *Language Learning and Technology*, 2(1), 74-93.
3. Chun, D. (1994). Using computer networking to facilitate the acquisition of interactive competence. *System* 22 (1): 17-31.
4. Higgins, E.L., Raskind, M.H. (2004). Speech recognition-based and automaticity programs to help students with severe reading and spelling problems. *Annals of Dyslexia*, 54(2), 365-392;
5. Honeycutt, L. (2003). Researching the use of voice recognition writing software. *Computers and Composition*, 20, 77-95.
6. Chun, D. (2002). *Discourse Intonation in L2: From theory and research to practice.* Philadelphia, PA: John Benjamins.
7. Hincks, R. "Using speech recognition to evaluate skills in spoken English". in *Papers from Fonetik 2001*, Lund University Department of Linguistics, 2001, pp. 58-61.
8. Warschauer, M., Turbee, L., & Roberts, B. (1996). Computer learning networks and student empowerment. *SYSTEM*, 24, 1-14.
9. Warschauer, M. (1997). Computer-mediated collaborative learning: Theory and practice. *Modern Language Journal*, 81(3), 470-481.
10. Zechner, K., Higgins, D., Xi, X., & Williamson, D. (in press, 2009). Automatic Scoring of Non-Native Spontaneous Speech in Tests of Spoken English. *SpeechCommunication* 51 (10). 883-895.
11. Precoda, K; Halverson, C. & Franco, H.. "Effects of Speech Recognition-based Pronunciation Feedback of SecondLanguage Pronunciation Ability" In *Proceedings of InStil 2000 Dundee, Scotland, 2000.*, pp.102-105.

Author Profile



Dr.S.Mercy Gnana Gandhi, working as a Professor of English Language Teaching at Sathyabama University, India has done her Ph.D in English Language Teaching with a Project on Multimedia Applications in English Language Teaching and on Computer Assisted Language Teaching. She has participated and presented research papers and articles in more than 25 National and International conferences and has around 20 Journal publications. Her Specialized areas are English Language Teaching, Multimedia and CALL. She teaches for IELTS, TOEFL, Cambridge BEC Exams apart from academic curriculum.