Review of Speech Recognition for Improvement in Learning

Prachi Poddar

Department of Computer Science, Global Institute of Technology, Jaipur, Rajasthan

ABSTRACT: This paper review literature from 1998 to 2014 inclusively on how Speech Recognition has been implemented to enhance learning. Speech recognition (SR) technology has a burgeoning range of applications in education system. The first motive of this review is to understand how SR technology has been practiced to improve learning over the past sixteen years, and the second is to analyze all research evidence on how SR technology can improve pedagogy. Earlier, SR technology was confined only to specific users i.e. students with physical or cognitive disabilities, Now, it has been used by a broader group of people. This technology is rapidly taking its roots into traditional learning also. There are numerous advantages of implementing SR technology for learning purpose. Moreover, some implications over the SR technology in pedagogical and technological aspect are discussed in the review, such as learning behaviors to use STR-texts, accuracy rate of the SR process and design of technology-based learning activities. Finally, the paper discusses some potential solutions for the future research.

Keywords: improving learning, literature, pedagogy, review, speech recognition, SR technology Weighting with Mean of Load(WD), Weighting with Sum of Mean of Generation Power and Load(WGD)

I.   INTRODUCTION

Present evidences indicate that some limitations and challenges exist in online synchronous and physical pedagogy environment. These problems can be solved by adopting some assistive media-to-text recognition technologies, such as writing-to-text, image-to-text, diagram-to-text, text-to-speech and speech-to-text. SR technology was mostly used to assist specific groups of students (i.e. students with learning or physical disabilities or foreign students) in order to guarantee them the equal access to learning. However, as time passed by, the target users involved into research on SR technology got broader. Now-a-days, SR technology is adopted to assist not only students with special needs but also general population of students for more educational purposes, such as enhancing students' understanding of a presented learning content during and after academic activities as well as offering students guidance to accomplish reflective writing and homework.

Furthermore, due to recent improvements in SR technology, particularly its accuracy rate, the technology is also adopted to support collaborative learning activities with multiple participants speaking simultaneously, such as group discussion or students' oral presentation. Therefore, this study particularly summarized SR development history and its usage by specific group of users. First, this study looks into how SR technology has been used in education over the past sixteen years by reviewing relevant research. Secondly, this study demonstrates how effective SR technology can be to enhance learning for different group of users, such as students with learning or physical disabilities, foreign students, online students and students who study in physical environment. This study further highlights finding on SR technology and proposes several suggestions for future research.

II.   METHODOLOGY

The literature from 1998 to 2014 inclusively were searched using the search terms such as speech-to-text, voice-to-speech, speech recognition, transcription and learning from ACM Digital Library, EBSCO Discovery Service, ERIC, PsychINFO and Social Sciences Citation Index databases. A total of 42 selected articles were reviewed. Primary data source for this review include peer-reviewed journal articles, conference proceedings and frequently cited books. The references provide a complete list of all the articles and journals reviewed for this project. The publications reviewed are organized into four dimensions that addresses (a) potentials of SR technology, (b) its use by specific groups of users in different domains, (c) research findings from studies using quantitative and/or qualitative methodologies, and (d) issues and considerations of applying

SR technology. These categories provide an organizational framework to understand how SR technology has been used in learning, and to explore any research evidence in terms of how Speech Recognition technology can improve learning.

Findings of this review were organized into two particular aspects. The first aspect is SR methodology and approach. That is, this review aimed to understand how the SR technology has been applied to support learning. Findings in this aspect are reported based on SR technological development. In earlier stages, SR technology was not as well-developed as it is now. One major issue was how to generate a satisfactory accuracy rate of transcripts from a speaker's speech. Therefore, earlier attempts were made to apply SR technology only for particular groups of users, such as student with cognitive or physical disabilities. Afterwards, a lot of studies of how computers can assist language learning were carried out with applications of SR technology. Finally, SR technology was developed more mature and reliable; the accuracy rate of recognition voice into text became higher and even STR could transcribe multiple speakers at the same time. Thus, some experts applied SR in the traditional classroom during lectures or collaborative learning activities on other fields of knowledge. The other aspect is potential in findings of SR technology to facilitate learning. That is, this review attempted to analyze all research evidences that how Speech-to-Text Recognition technology can enhance learning. Finding in this aspect centered on applications of the SR technology to support learning of different groups of users in traditional and online learning environments.

III.   USAGE OF SPEECH RECOGNITION IN LEARNING

This section analyzes findings from other studies regarding how the SR technology was applied in learning in the past sixteen years. Finding in this section are classified into the following categories:

3.1   Students with cognitive or physical disabilities

Students with hearing impairments rely on either reading lips or watching an interpreter to access what the instructor spoke. It is extremely difficult for these students to focus their visual attention on note-taking and the instructor simultaneously. Therefore, it was suggested to apply assistive technologies, such as a speech-to-text support service, to enhance computer-assisted learning for students with different types of disabilities.

3.2   Online students

Network traffic congestion can cause poor quality of audio communication in a synchronous cyber classroom. Under such condition, students are not able to hear a speaker clearly. This issue was viewed as one technological challenge. It negatively affects online teaching and learning activities as it hinders students' understanding of a delivered speech, and it also hampers students from engaging in classroom participation and interaction. Thus, the SR technology could "learn" a speaker's voice and terminology for specific field during the training and then achieve a certain level of accuracy rate when being applied in the activities. A speaker's speech was transcribed by the SR technology into text which was displayed simultaneously to students on their computer screens. Thus, the students could listen to a speaker and read the transcripts at the same time. More importantly, SR-generated text was saved for future revision to fix some recognition errors, and the students could obtain a nearly verbatim transcript to study it after the activities and to complete summary writing tasks.

3.3   Students in traditional learning environment

The adoption of the SR technology in traditional learning environment has several benefits. One of them is to improve teaching methods and to enhance learning opportunities. For example, by using the Speech Recognition, teachers can take a proactive, rather than a reactive approach to teach students with different learning styles. It provides educators with a practical means of making their teaching accessible and improve the quality of instruction in the process.

3.4   Non-native speakers

SR technology has a potential to be an instructional support mechanism, and there were a number of perceived benefits associated with the SR use. Most non-native speaker students, due to their language barrier and mishearing some important parts on the instructor's speech, admitted that Speech Recognition is useful during lectures to follow the instructor and to clarify and understand lecture content. Transcripts generated from speeches of second language speakers by using SR were with significantly higher accuracy rate. These results were consistent in line with native speakers'' scores; that is, the highest accuracy score were achieved at the lower level of analysis, the word level, and the lowest scores at the t-unit, or sentence level of analysis.

IV. LITERATURE REVIEW

In [1] Decadt, Jacques, Daelemans, Walter and Wambacqdescribes a method to enhance the readability of the textual output in a large vocabulary continuous speech recognition system when out-of-vocabulary words occur. The basic idea is to replace uncertain words in the transcriptions with a phoneme recognition result that is post-processed using a phoneme-to-grapheme converter. This technique uses machine learning concepts.

In [2] Verena Stein, Robert NeBelrath and Jan Alexandersson states that blind people often use text-to-speech synthesis when they work on the computer. Using a Braille line to convert digital texts into embossed writing is slow and requires extra hardware. Relying more on TTS technology, however, lessens blind people's exposure to the written form of words which is especially important when studying another language.

In [3] RustamShadiev, Wu-Yung Hwang, Nian-Shing Chen and Yueh-Min Huang enumerates various ways in which Speech Recognition technology can enhance learning for different group of users.

In [4] B. Duerstock, R. Ranchal, Y. Guo, T. Doughty, J. Robinson, and K. Bain illustrates on innovative Speech Recognition technology and how it is utilized to assist students acquiring more complete and accurate lecture notes. SR-assisted note taking can be accomplished in different ways to allow students to devote more attention to understand course material than manually record the instructor's lecture content. Lecture note taking was physically challenging for many students with disabilities who cannot take notes for the bulk of lecture notes themselves and must rely upon hired note takers for class notes, which may not be readily available or feasible.

In [5] B. Arons wrote an article describing techniques for structuring, filtering and presenting recorded speech, allowing a user to navigate and interactively find information in the audio domain. It describes the Speech Skimmer system for interactively skimming speech recordings. Speech Skimmer uses speech-processing techniques to allow a user to hear recorded sounds quickly, and at several levels of details.

In [6] M. Wald states that deaf and hard of hearing people can find it difficult to follow speech through hearing alone or to take notes when lip-reading or watching a sign-language interpreter. Lectures can be digitally recorded and replayed to provide multimedia revision material for students who attended the class and a substitute learning experience for students unable to attend.

V. CONCLUSION

It is fairly clear that SR technology was implemented to aid learning in different way based on the progress of SR technological development. That is, the past studies applied this technology only to assist learning of particular group of users, such as students with physical and cognitive disabilities, due to low accuracy rate and a delay in SR process. However, over the years SR technology became more reliable. Then, SR was applied to aid pedagogy of students in a traditional learning environment during and after collaborative and individual learning activities. Secondly, the literature considered SR technology beneficial to broaden imparting of knowledge during and after learning activities. There is widespread consensus in the literature about numerous distinct advantages of SR-text, such as helping students to better understand content of academic activities, to confirm missed parts of a speech, to take notes, to complete homework and to prepare for exams. However, some arguments over the SR technology considerations that limit educational value of the technology still exists. The literature review showed how those considerations can be addressed by implementing various approaches to increase the effectiveness of SR application on learning.

Given the present finding, the following are important issues to address in future SR related studies. First, researchers should start theorizing the cognitive processes that occur through learning with SR technology. In addition to it, SR technology needs to be more effective by employing relevant pedagogical principles. Second, there is a need to use reliable and well-established outcome measures in future SR studies. More research needs to be conducted in communicative and dynamic settings, such as collaborative learning and teaching with multiple participants speaking to the SR system simultaneously. Also, it must be investigated whether there will be different learning effect when the SR is applied to learning environment with students of different cultural backgrounds or languages. In the future, SR technology can be extended by combining it with other technology like automatic translation, to simultaneously generate text from a speech and translate it into many languages. This would allow students and teachers to have instant audio-lingual interpretations using their own native languages.

REFERENCES

[1]. Decadt, Bart, et al. "Phoneme-to-grapheme conversion for out-of-vocabulary words in large vocabulary speech recognition." Automatic Speech Recognition and Understanding, 2001.ASRU'01.IEEE Workshop on.IEEE, 2001.

[2]. Senthamarai, Mrs R., and V. Khanaa. "Manuscript Info Abstract." International Journal 1.10 (2013): 518-523.

[3]. Shadiev, Rustam, et al. "Review of Speech-to-Text Recognition Technology for Enhancing Learning." Journal of Educational Technology & Society 17.4 (2014): 65-84.

[4]. Duerstock, Bradley S., et al. "Assistive Notetaking Using Speech Recognition Technology." Proc. Festival of International Conferences on Caregiving, Disability, Aging and Technology (FICCDAT): RESNA/ICTA3, Toronto, Canada. 2011.

[5]. Arons, Barry. "SpeechSkimmer: a system for interactively skimming recorded speech." ACM Transactions on Computer-Human Interaction (TOCHI) 4.1 (1997): 3-38.

[6]. Wald, Mike. "Captioning for deaf and hard of hearing people by editing automatic speech recognition in real time." Computers Helping People with Special Needs.Springer Berlin Heidelberg, 2006.683-690.