SPEECH PATHOLOGY APPLICATIONS OF AUTOMATIC SPEECH RECOGNITION TECHNOLOGY

Simone Griffin, Linda Wilson, Elizabeth Clark
Speech Pathology
School of Community Health
Charles Sturt University
sgriffin@csu.edu.au, liwilson@csu.edu.au, eclark@csu.edu.au

ABSTRACT: Few studies have investigated the benefits and potential difficulties of automatic speech recognition (ASR) application for people with speech and language impairment. The literature has demonstrated that ASR has the potential to assist individuals with dysarthria and hearing impairment to communicate with computers but the use of ASR with people with other speech and language disorders is less well documented. This paper examines the potential applications of ASR in the domain of speech pathology, including therapeutic and assessment applications, report writing, and as a mode of alternative and augmentative communication (AAC). It also identifies areas in which further research is required before these potential applications can be realised.

AUTOMATC SPEECH RECOGNITION AND SPEECH AND LANGUAGE IMPAIRMENT

The applications of automatic speech recognition (ASR) have increased markedly in the public and professional domains over the last three decades, with ASR now feasible in fields such as aviation, health care, telecommunications, and banking (Fournier, 1996). Numerous studies have identified the benefits of ASR applications for people with physical impairments (Dalton & Peterson, 1997; Ferrier, Jarrell, Carpenter & Shane, 1992; Porter, 1998; Smedman & Chesla, 1999). However, there has been limited research to investigate the impact of a range of speech impairments on the accuracy of ASR (Thomas-Stonell, Kotler, Leeper & Doyle, 1998), despite the fact that one in seven Australians have a speech or language impairment (Speech Pathology Australia, 2000). Further investigation of ASR with this population is important for two reasons: Individuals with speech and language disorders, already frequently marginalised in society (Speech Pathology Association of Australia, 2000), are at risk of being further marginalised with increased use of ASR technology in the public and professional domains; and ASR has the potential to assist speech pathologists and their clients.

SPEECH PATHOLOGY APPLICATIONS

The literature has demonstrated that ASR has the potential to assist individuals with dysarthria and hearing impairment to communicate with computers (Dalton & Peterson, 1997; Ferrier et al., 1992; Porter, 1998; Smedman & Chelsa, 1999). However, there are many other potential applications of ASR in the area of speech pathology and for speech pathology clients. For example, ASR systems may be used for therapeutic applications, assessment applications, report writing and as a mode of alternative and augmentative communication (AAC). Some of these have been suggested in the literature; others have not, and are presented below.

Therapeutic Applications

ASR has the potential to be used in speech pathology as a feedback system (Ferrier et al., 1992; Fried-Oken, 1985; Parsons, 1997). An interesting feature of ASR systems is that they can provide feedback to the user on articulation consistency (Ferrier et al., 1992). This feature of ASR systems has the potential to be used as a therapy tool. For example, when a client produces a particular word, the computer identifies the production as correct or incorrect simply by having the recognised word appear on the computer screen. If the word does not appear, or a different word appears, then the production is incorrect. Ferrier et al. (1992) found that both control and experimental subjects (speakers with dysarthria) had increased articulation precision in response to mis-recognition by an ASR system.

There are, however, current limitations of commonly available ASR systems that are potentially preventing their application as a feedback system in therapy. For example, if a speaker-dependent system was to be used in the scenario above, the speech pathology client would first need to be able

to produce correct productions of the target word to train the computer system; or, if a speaker-independent system was to be used, the highly inaccurate recognition rates achieved currently would need to be considered. Speaker-adaptive programs would also be unsuitable because the purpose of the activity is not to have the computer adapt to incorrect productions, and reinforce non-target responses, but rather accept or reject specific productions.

ASR systems also have the potential to motivate clients in therapy. Motivation is an important aspect of success in therapy, and research has demonstrated that motivation is facilitated when variable or novel stimuli are presented (Parsons, 1997). ASR systems have this potential. The motivational value of computer-assisted intervention has been noted in several reports (Parsons, 1997; Shriberg, Kwiatkowski & Snyder, 1986; Shriberg, Kwiatkowski & Snyder, 1990). In addition, Fried-Oken (1985) found that her subjects displayed increased motivation to learn, and expressed a desire to continue using an ASR system.

Assessment Applications

There are several potential applications of ASR to the assessment of speech pathology clients. One potential application is transcription. Transcription forms a large part of any assessment of speech and language. To accomplish this, the clinician must transcribe, in phonetic symbols or orthographic text, exactly what the client has said for subsequent analysis and identification of specific speech or language impairment. However, this can be time consuming. Computer Aided Speech and Language Analysis (CASALA) (Serry, Blamey, Spain & James, 1997) is a program that can be used to perform a variety of analyses of speech and language; but the samples must be transcribed first, and it has been recognised that manual transcription is the most time-consuming part of using this program (Serry et al., 1997). ASR has the potential to reduce the time spent manually transcribing speech and language samples. For example, instead of audio-recording a speech or language sample and transcribing it manually after the session, a microphone placed near a client and attached to a computer with ASR, has the potential to transcribe the sample automatically.

However, there are current limitations of the ASR systems that are potentially preventing this application. First, most ASR systems do not currently transcribe in phonetic symbols. Although a system may, at some point in the recognition process, identify the sound-to-phonetic symbol relationship, statistical prediction and pattern matching usually influence recognition. However, this aspect is important to evaluate because if a client presents with a speech disorder (e.g., omitting final sounds of words) and the computer statistically predicts what the client has said from the incomplete word recognition, then the speech impairment may be mis-represented by the transcription output. Similarly with language transcription (which is usually transcribed using orthographic gloss): the current statistic prediction of ASR systems may obscure the true language impairment of a client. Therefore, the use of ASR for speech and language transcription might require more reliance on acoustic-phonetics and not utilise statistical prediction or language modelling. However, until research is conducted into the accuracy of speech and language transcription by ASR systems with people with speech and language impairments, it is unclear to what extent ASR will be a useful assessment tool.

Report Writing

Report writing is an application that could be used with current ASR systems. Specific vocabulary for speech pathology terminology and symbols may need to be trained into the systems; however, report writing could be used by speech pathologists in a similar way to how it is used by law and medical professionals for dictation. This has the potential to reduce the time spent writing reports and file notes. Also, there are various computer and macro based speech pathology report writing templates available (Flynn, Parsons & Shipp, 2000; Gupta, Nathan, Fisher & Bruce, 2000; McLeod, 1993) that have the potential to be combined with ASR and further reduce the time spent report writing.

Augmentative and Alternative Communication (AAC)

ASR presents interesting possibilities for use as an AAC device (Goette & Marchewaka, 1994). An AAC device is a form of communication offered to assist those who cannot speak or whose communication abilities are reduced (Morris, 1997). These devices can be as simple as a notebook and pencil or more technically advanced like a computers with synthetic voice output (Mann & Lane, 1991).

Stevens and Bernstein (1985) compared intelligibility to human listeners, and accuracy of computer ASR of hearing impaired speakers. They found that three of the five hearing impaired speakers were better understood by the recognition device than by human listeners. They suggested that these three speakers might be good candidates for an AAC device (Stevens & Bernstein, 1985). Similar conclusions were drawn by Coleman and Meyers (1991) for individuals with dysarthric speech. If an ASR system can recognise dysarthric speech with accuracy, then it might be used as an AAC device which could translate the difficult-to-understand speech of the speaker to a more recognisable form, either written or in the form of synthetic voice output (Coleman & Meyers, 1991). In addition, ASR offers the distinct advantage as an AAC device of producing speech and/or printed output almost immediately. This avoids the serious rate problems inherent in most present AAC systems (Mann & Lane, 1991).

POTENTIAL AREAS OF RESEARCH

If these potential speech pathology applications are to be realised, further research needs to be conducted in the following areas.

Spectrographic Characteristics of Impaired Speech

Farmer (1997) indicated that characteristics of almost every speech and language disorder have been identified and studied via the spectrograph, including stuttering, dysarthria, apraxia and aphasia, phonological/phonetic disorders, speech of the hearing impaired, and voice disorders. Ferrier et al. (1992) noted that there is potential to apply information of this kind to research into the impact of specific spectrographic characteristics of various speech and language impairments on ASR accuracy and application. It appears that this research is yet to be conducted. In other words, further research is needed in the area of the specific spectrographic characteristics and parameters associated with various speech and language impairments and to determine whether these parameters are also associated with low recognition scores (Doyle et al., 1997). A related issue requiring further investigation relates to which characteristics of speech affect recognition accuracy (Thomas-Stonell et al., 1998). Research has indicated that certain speech characteristics, such as voicing constraints, nasalisation, and vowel height, may be critical in degrading speech intelligibility (Doyle et al., 1997). The identification of these specific characteristics would be of assistance to speech pathologists for guiding the changes in speech production that should be targeted by individual speakers in order to improve the accuracy of ASR (Goette & Marchewaka, 1994).

Investigation of a Range of Speech and Language Impairments & ASR

There has been limited research to investigate the impact of a range of speech impairments on the accuracy of ASR (Thomas-Stonell et al., 1998). The two groups studied to date are individuals with dysarthric speech and individuals with hearing impairment. However, further research is needed to identify the accuracy of ASR with people with other speech and language disorders.

Training

Alongside the identification of the specific characteristics of the various speech and language impairments, further research is needed to determine how many training sessions are required to achieve maximal adaptation of an ASR system for individuals with speech and language impairment (Thomas-Stonell et al., 1998). Preliminary research has identified that ASR with individuals with a speech impairment was characterised by initially steep increases in correct recognition with more gradual increases noted during later sessions (Doyle et al., 1997). However, data to confirm a specific

time frame or learning curve to achieve maximal adaptation are not yet available (Doyle et al., 1997). This information has the potential to be applied in the identification of goals and the planning of therapy for speech pathology clients for whom ASR may be suitable.

CONCLUSION

The literature has demonstrated that ASR has the potential to assist individuals with dysarthria and hearing impairment to communicate with computers. However, there are many other potential applications of ASR in the area of speech pathology and for speech pathology clients. For example, ASR systems have the potential to be used for therapeutic applications, assessment applications, report writing and for AAC.

However, research is needed before these potential speech pathology applications can be realised. Consequently, there are several potential areas of research relating to speech pathology and ASR technology. Among these are the investigation of ASR as a tool for phonetic and orthographic transcription; the use of ASR as a feedback and motivating system in speech therapy; the identification of a vocabulary specific to speech pathology reports for ASR, and the ability to combine this with computer and macro based speech pathology report writing templates; and the potential uses of ASR for AAC. Also, further research could be directed towards the identification of specific spectrographic and acoustic characteristics associated with speech and language impairments and low ASR levels, the investigation of ASR in a range of speech and language impairments, and the level of training required to achieve maximal adaptation of ASR system for individuals with speech and language impairment.

REFERENCES

Anderson, J. F. (1998) *Transcribing with voice recognition software: A new tool for qualitative researchers.* Qualitative Health Research, 8, 718-723.

Coleman, C. L., & Meyers, L. (1991) *Computer recognition of the speech of adults with cerebral palsy and dysarthria.* AAC: Augmentative and Alternative Communication, 7, 34-42.

Curtin, M. (1994) *Technology for people with tetraplegia, Part 1: Accessing computers.* British Journal of Occupational Therapy, 57, 376-380.

Dalton, J. R., & Peterson, C. Q. (1997) *The use of voice recognition as a control interface for word processing.* Occupational Therapy in Health Care, 11, 75-81.

Doyle, P. C., Leeper, H. A., Kotler, A., Thomas-Stonell, N., O'Neil, C., Dylke, M., & Rolls, K. (1997) *Dysarthric speech: A comparison of computerized speech recognition and listener intelligibility.* Journal of Rehabilitation Research & Development, 34, 309-316.

Farmer, A. (1997) *Spectrography.* In M. J. Ball & C. Code (Eds.), Instrumental Clinical Phonetics. Whurr: London.

Ferrier, L. J., Jarrell, N., & Carpenter, T., Shane, H. C. (1992) *A case study of a dysarthric speakers using the DragonDictate voice recognition system.* Journal of Computer Users and Speech and Hearing, 8, 33-52.

Flynn, M. C., Parsons, C. L., & Shipp, L. (2000) *An interactive computerized report writer for speech pathology.* ACQ, 2(2), 66-68.

Fournier, R. S. (1996) *Developments in voice-input technology.* Journal of Education for Business, 71, 241-245.

Fried-Oken, M. (1985) *Voice recognition device as a computer interface for motor and speech impaired people.* Archives of Physical Medical Rehabilitation, 66, 678-681.

Gupta, J., Nathan, T., Fisher, M., & Bruce, T. (2000) *Computerization of modified barium swallow reporting.* Paper presented at Speech Pathology Australia National Conference, South Australia.

Goette, T., & Marchewaka, J. T. (1994) *Voice recognition technology for persons who have motoric disabilities.* Journal of Rehabilitation, 60, 38-41.

Mann, W., & Lane, J. P. (1991) *Assistive technology for persons with disabilities: The role of occupational therapy.* The American Occupational Therapy Association: Marylands.

McLeod, S. (1993) *Speech pathology report templates* (2nd ed.). The University of Sydney: Sydney.

Morris, D. (1997) *Dictionary of communication disorders* (3rd ed.). Whurr: London.

Parsons, C. L. (1997) *Communication with computers: The use of communication technology in speech-language pathology.* Australian Communication Quarterly, Spring, 9-15.

Porter, S. G. (1998) *Watch you language -- The PC is listening.* Women in Business, 50, 42.

Serry, T., Blamey, P., Spain, P., & James, C. (1997) *CASALA: Computer aided speech and language analysis.* Australian Communication Quarterly, Spring, 27-28.

Shriberg, L. D., Kwiatowski, J., & Synder, T. (1986) *Articulation testing by microcomputer.* Journal of Speech and Hearing Disorders, 52, 309-324.

Shriberg, L. D., Kwiatowski, J., & Synder, T. (1990) *Tabletop versus microcomputer-assister speech management: Response evocation phase.* Journal of Speech and Hearing Disorders, 55, *635*-655.

Smedman, L., & Chelsa, A. E. (1999) *Brace new yackety-yak.* Inside Ms, 17, 50-56.

Speech Pathology Australia (2000*) One in seven.* National Stop Press, March, 4.

Speech Pathology Association of Australia. *What's it like to have a communication disability?* <http://www.speechpathologyaustralia.org.au/factsheet1.4.html> (Accessed 17 July 2000)

Stevens, G., & Bernstein, J. (1985) *Intelligibility and machine recognition of deaf speech.* Proceedings of the Eighth Annual Conference on Rehabilitation Technology, 308-310.

Thomas-Stonell, T., Kotler, A., Leeper, H. A., & Doyle, P. C. (1998) *Computerized speech recognition: Influence of intelligibility and perceptual consistency on recognition accuracy.* AAC: Augmentative and Alternative Communication, 14, 51-56.

ON PREDICTING PATIENT'S VOICE AFTER SURGICAL OPERATION

Cheolwoo Jo*, Daehyun Kim*, Moojin Baek**, Sugeon Wang***
*Changwon National University, **Inje University,
***Busan National University
KOREA

ABSTRACT: This paper describes a procedure to predict a patient's voice after surgical operation. To do this, the voice before and after surgical operation is collected from the same patient. Collected voice is analyzed to obtain differences affected by surgical operation. To measure the change of acoustical characteristics of voice, jitter, shimmer and other spectral domain and time domain parameters are computed and compared. According to the result, it is shown that the factors that change are caused not only by vocal fold components but also by vocal tract. One method to implement the predictive synthesis of voice after surgery, residual excited PSOLA method is applied. The resulting voice is compared to the voice after surgery in terms of spectral and perceptual similarity.

1. INTRODUCTION

Recently interest on the human health is increasing. And speech is a basic mean to human communication. In many cases, diseases at the vocal folds cause the change of voice quality. This is caused by tumor, folip, swelling, hardening of vocal folds and its vicinity etc. In some cases, patients are asked to take surgery. Patients, who have distorted voice, generally want to know how much their voice can be improved after surgery. Also doctors want to let them know the possible improvements and make them have peace in mind. This process very important.

The relationship at the quality change of the voice before and after surgery is not well known. Some doctors can predict the status of patient's neck by hearing speech only. Based on this fact, there are researches to diagnose patient's voice only by acoustic signal. (Jo & Kim,1998) (Koike, Takahashi , Calcatera,1977) (Issiki, Okamura, Tanabe , Morimoto,1969) (Murry, Singh , Sargent,1977) (Hammarberg, Fritzell , Gaffin,1967) Some of them are based on the perceptual distinctions and subjective tests, but there are some which are based on the objective or numerical methods. Predicting voice after surgery is not reported much yet on the literature.

In this paper changes of various parameters from voices before and after surgery. Based on the observed phenomena, Predictive synthesis of voice after surgery is conducted. And the procedures are introduced.

2. PATHOLOGICAL VOICE DATA

Voice data is collected from the patients at ENT department of hospital. Range of patients' age is 50-60. Names of diseseas are mostly polyp with two cases of pholyposis. All the patients are diagnosed to be benign by several tests and taken surgery afterwards. Collection is performed in silent room. 5 vowel sounds (/a/, /e/, /i/, /o/, /u/) and one sentence sound are collected. Patients are asked to pronounce the vowels consecutively for 3 seconds with proper silence between utterances. Microphone is positioned at 15 cm front of the mouth. Spoken voice is recorded using Sony's DAT(DTC-59ESJ) recorder. Collected materials are stored in wav file format. Pathological data is typically noisy and husky. It is easily recognizable voice after surgery is less noisy and clearer.

3. ANALYSIS OF VOICE

Speech materials are analyzed by Kay's MDVP software package. Total 14 different parameters are computed. Those are as follows. In terms of short-term and long-term frequency disturbance, Jita, Jitt, RAP, PPQ, sPPQ, vF0. In terms of short-term and long-term amplitude perturbation, ShdB, Shim, APQ, sAPQ, vAm. In terms of noisy components, NHR, VTI, SPI.

Table.1 shows relative changes of each parameters. It is shown that voice after surgery has less jitter and shimmer in general. It is a quite predictable result based on perceptual difference.