Institutional Review Board General Application

Audiovisual Distinctive-Feature-Based Recognition of Dysarthric Speech

Hasegawa-Johnson, ECE Department, Univ. Illinois at Urbana-Champaign
Adrienne Perlman, SHS Department, Univ. Illinois at Urbana-Champaign
Jon Gunderson, Disability Resources and Educational Services, UIUC
Heejin Kim, Beckman Institute, Univ. Illinois at Urbana-Champaign

Mary Sesto, Trace Center, University of Wisconsin at Madison (serving as liaison for researchers at University of Illinois at Urbana-Champaign)

1. Abstract

In the past ten years speech recognition systems have improved tremendously and continuous speech recognizers with high levels of accuracy are available to the general public. Many people with gross motor impairment, including some people with cerebral palsy and closed head injuries, have not enjoyed the benefit of these advances, because their general motor impairment includes a component of dysarthria: reduced speech intelligibility caused by neuromotor impairment. These motor impairments often preclude normal use of a keyboard. For this reason, case studies have shown that some dysarthric users may find it easier, instead of a keyboard, to use a small-vocabulary automatic speech recognition system, with code words representing letters and formatting commands, and with acoustic speech recognition models carefully adapted to the speech of the individual user. Development of each individualized speech recognition system remains extremely labor-intensive, because so little is understood about the general characteristics of dysarthric speech. We propose to study the general audio and visual characteristics of articulation errors in dysarthric speech, and to apply the results of our scientific study to the development of speaker-independent large-vocabulary and small-vocabulary audio and audiovisual dysarthric speech recognition systems. We propose to record talkers with dysarthria, and age-matched normal control subjects, using an existing array of four video cameras and eight microphones developed in the principal investigator's laboratory. Interactive phonetic analysis will seek to describe the talker-dependent and, if possible, the general talker-independent characteristics of articulation error in dysarthria. Based on interactive analysis, speech recognition models will be developed, and will be evaluated to determine the effectiveness of each model as a user interface option to computing technologies and the ability of the models to generalize to the general population of people with dysarthric speech.

2. Study Design and Methods

Subjects with dysarthria will be recruited through personal contacts established with clients of the Waisman Center and the Communication Disorders Department at the University of Wisconsin. Julie Gamradt (Waisman Center) and Jamie Murray Branch (Communication Disorders) will send an announcement letter to each candidate subject, by e-mail or by U.S. postal mail, inviting interested subjects to fill out a short on-line web form giving their contact information to all four investigators named on this application. Julie Gamradt and Jamie Murray Branch will not give subject names to any investigators from outside the University of Wisconsin; only the subject himself or herself will be allowed to provide such contact information to non-UW investigators.

Subject Initial Contact with Investigators

Subjects will be invited to respond by filling out a short on-line web page, providing us with their contact information, age, gender, an evaluation of their own degree of speech pathology, and an evaluation of their own past experiences with automatic speech recognition or contacting us by phone. Subjects will be informed that this web page provides subject contact information to University of Illinois researchers Prof. Mark Hasegawa-Johnson, Prof. Adrienne Perlman, Dr. Jon Gunderson, and Dr. Heejin Kim. Dysarthric subjects matching the criteria for inclusion will be contacted by e-mail or telephone immediately after we receive their contact information, and we will forward to them a copy of the consent form and information about the study, and will suggest possible dates and times when we might schedule a recording session.

Inclusion Criteria and Number of Participants

Our goal is to recruit and record 50 subjects with spastic dysarthria. To the extent possible, subjects will be selected in order to ensure equal representation of both genders and all ethnicities, however it is quite likely that there will be more male than female talkers, because spastic dysarthria has a higher rate of incidence among male talkers. Talkers with dysarthria will be recruited based on self-report of a history of spastic dysarthria. Specifically, talkers will be asked to briefly describe the history of their speech/ language pathology; talkers whose self-reported history and speech characteristics match the broad parameters of spastic dysarthria, in the opinion of the investigator, will be considered qualified for inclusion. Besides self-reported history of speech pathology, the only other criterion for inclusion is age: in order to avoid ambiguity in a subject's privacy preferences, we will not record subjects below the age of 18 years.

Role of Participants

Recording sessions will be scheduled for two hours (120 minutes), but most subjects are anticipated to complete the sessions within 60-90 minutes. The first part of each session will be allocated for four tasks, all conducted with the recording equipment off. First, the purposes of the study will be explained to subjects. Second, our privacy protections will be explained. Third, the recording hardware will be explained. Fourth, the consent form will be explained.   Subjects who choose to sign the consent form will then be recorded. Subjects will read prompt text from a computer display while being videotaped using a multimodal audiovisual sensor array including a digital video camera and an array of microphones (seven one-millimeter microphones arranged in an array, total array dimension is roughly 12cm X 4cm). Recordings will take place while subject is seated comfortably in front of a desktop personal computer. At least two experimenters will be present with the subject at all time. Subjects will read one word at a time from a list of phonetically balanced words and single word commands (i.e. Alpha, Bravo, Charlie,…). Each word will be recorded several times. No deception will occur.

Data Recording and Storage

Data will be recorded using a multimodal array consisting of one videocameras and seven microphones. Audio signals from the seven microphones will be recorded using seven of the eight channels on a digital audio recording interface. The eighth channel, and one audio channel on the mini-DV video tape, will both record a synchronization signal generated by a touch-tone telephone. Each prompt screen will be coded with a number between 0 and 9; when the experimenter advances the prompt screen to a particular word, the corresponding synchronization tone will be recorded to both the video and audio recordings. Video and audio data will be transferred to hard disk, and uploaded to the password-protected video database, using a desktop system consisting of a mini-DV cassette system and a MacIntosh with Adobe Premiere software. Video and audio data will be automatically edited to exclude all utterances not enveloped by touch-tone synchronization keypresses, including all spontaneous subject comments and conversation. After video data have been uploaded to the password-protected database, all videotapes and signed consent forms will be permanently stored in a locked file cabinet in the investigator’s office at the University of Illinois. Copies of the participant consent forms completing research at the Trace Center will be stored in a locked cabinet at the Trace Center, University of Wisconsin.

Anonymity of Subjects

No subject identifying information, other than recorded images of the subject’s face and voice, will be recorded in any location other than the subject’s signed consent form. Video and audio files and videotapes will be identified by: (1) a unique subject number, (2) the gender of the subject, and (3) the subject’s privacy preferences, as specified below. Unique subject number of each subject will be noted on the subject’s consent form, so that the principal investigator may contact each subject at a later date if necessary.

Consent Forms

In the consent form (attached), subjects will be asked to specifically grant or withhold permission for each of the following possible uses of the data:

1) Presentation of audio recordings at professional conferences by one of the experimenters,
2) Presentation of video recordings at professional conferences by one of the experimenters,

3) Distribution of audio recordings to researchers from other laboratories and at other universities who request a copy of these recordings,
4) Distribution of video recordings to researchers from other laboratories and at other universities who request a copy of these recordings.

If a subject does not grant permission for the presentation of his or her recordings at any professional conference, then no video or audio recordings of his or her face or voice will be published or presented to persons outside the research group of the principal investigators. In this case, the data provided by the subject will be used in two ways: (1) to develop automatic speech recognition models; the statistical parameters of such models (including either talker-dependent or talker-independent models) may be published outside of the University of Illinois at Urbana-Champaign and the University of Wisconsin at Madison, provided that all such models are in a form from which it is impossible to reconstruct the original voice or face of the talker, (2) for the purpose of aggregate multi-talker statistical or signal processing analyses, the results of which may be published in any form to any audience, provided that it is impossible to reconstruct the original voice or face of the talker from any of the published materials.

If researchers at other laboratories or at other universities request a copy of the recorded data, they will be mailed physical media (CDROM, DVD, or similar media) containing only those recordings whose distribution was specifically authorized by the subjects being recorded. Recordings distributed to other laboratories will not include any subject identifying information: instead, each file will be identified only by (1) a unique subject number, (2) the gender of the subject, and (3) the subject’s privacy preferences.

The password-protected master database will reside on a server controlled by the principal investigator. Access to the password-protected master database will be granted only to University of Illinois and University of Wisconsin faculty, staff, and students working on the subject of audiovisual speech recognition under the direct supervision of one of the investigators named in this proposal.

Compensation

Subjects will be compensated $50 for completing a recording session regardless of how much time it takes them to complete the session. Subjects who fail to complete the recording session will be compensated for the time they spend, at a rate of $20/hour. Subjects who must travel more than 10 miles to the recording site will also be compensated for travel, at a rate of $0.44/mile.

Sites

The proposed research is already taking place at the University of Illinois, Urbana-Champaign, under the auspices of NSF grant IIS 05-34106 and the enclosed University of Illinois IRB approval. We propose to also record Wisconsin subjects; this IRB application seeks permission to do so.

Measurement Procedures

Subjects will be prompted from a phonetically balanced word list that will be presented one word at a time on a standard computer screen. A researcher will manually advance to the next word after the subject says the current word. The word lists are presented in blocks to allow subjects to rest between lists. Subjects can rest as long as they want between reading word lists.

3. Risk/Benefit Assessment

The proposed research poses two potential risks:

1) Subject privacy
2) Subject discomfort due to prolonged talking during the recording session.

The experiment has been designed so that neither of these risks is greater than similar risks ordinarily encountered in daily life. The risk to subject privacy will be minimized using password protection of the database and subject privacy preferences as specified in section 2, "Study Design and Methods." The risk of subject discomfort during the recording session will be minimized in the following ways. First, prompts will be presented in the form of short texts, of at most 100 words per prompt screen. Before beginning each block of text, the experimenter will press a key that will record a synchronization tone to an auxiliary channel on the tape recorder; the end of each block of text will be similarly marked. Because of this annotation procedure, subjects will be free to take a break of any length, between any two sequential prompt screens; the annotation scheme will also be designed so that a subject may repeat any prompt screen as often as desired, in case a subject wants to correct his or her reading. At least once per 30 minutes, the investigator will suggest that the subject take a 5-minute break in order to have a drink and walk around. If any talker with dysarthria is unable to complete all recording material by the end of two hours, the experiment will be concluded at that time, and the subject will be paid $50 for participation.

There is no immediate benefit to the subject of the proposed research, but the subject may eventually benefit from improved automatic speech recognition software developed using these recordings. The benefit to society of the proposed research is the development of improved speech recognition for human-computer interaction systems designed for subjects with spastic motor disorders including dysarthria.

Because there is no immediate benefit to the subject, data acquisition procedures have been specifically designed to minimize or eliminate risk to the subject. Specifically, (1) the subject can control his or her own recording environment, in order to minimize discomfort, and (2) the risk of subject privacy violations is under the specific control of the subject, and may be adjusted by the subject to satisfy his or her personal privacy preferences.

4. Consent Forms

Protection of subject privacy will be ensured as described in Section 2., “Study Design and Methods.” Consent form is attached.