Associated Colleges of the South

Teaching with Technology Fellowship Program Proposal

Summer 2006

The human voice - computer interface

Dr. Dan Boye

Department of Physics

Davidson College

1. Background

While the primary computer user input device is the keyboard and the output device is a visual screen, these interfaces may not always be available or the most practical and efficient. A great amount of communication between humans occurs through speech and hearing and so one might assume that such modes would find easy application in computers. After all, sound cards can readily record and play CD quality music. However, the voice-computer interface has been applied with only moderate success. Current specialized applications include machine control, voice telephone dialing, human-machine interaction over the telephone and dictation. Each of these applications has a unique set of user and computer requirements and tolerances for error. Only in the most recent versions of Microsoft Office have speech tools been available to the general PC user.

Two-way communication occurs through Automatic Speech Recognition (ASR) and Text-To-Speech (TTS) software routines. In ASR routines, spoken words enter the computer through the microphone and are recorded by the sound card. The ASR routine tries to match characteristics of the sound to a template of basic sounds. At present, the ASR routine must be "trained" by each speaker. The software routine starts with identifying basic sounds like vowels and consonants and compares them to a dictionary of possible words. More successful ASR routines extend beyond single word identification to the use of contextual rules of grammar. The most probable word is entered into a text editor, e.g. Microsoft Word. The current ASR word error rate is about 10 times that of human-human speech recognition. In the end, a human is still needed to check the transcription for errors. TTS routines work in the opposite order, taking a previously written text file, identifying syllables and basic sounds associated with the letter combinations, stringing together the basic sounds of a prescribed voice and sending the output to the sound card and the computer speakers. TTS routines are not able to reproduce all of the content of the author's intentions because inflections and emphases are not currently coded.

An added benefit of exploring the voice–computer interface at this time is that purportedly there will be improved TTS and ASR capabilities in the upcoming release of the new Windows operating system called Vista. Performing the exercises described in this proposal will provide a baseline for judging the degree of improvement of the new software which should be the campus standard by the fall of 2007.

Liberal arts curricula provide many opportunities for developing disciplined, observant, creative and socially conscious minds. Core requirements within the curriculum are designed to expose students to areas outside of their major track of study. The curricular material that will be developed in this project is to be used in a non-science major course currently being offered by the physics department and which satisfies a core requirement in the sciences. This course challenges students to become directly involved in scientific study and it seeks to demonstrate the importance of scientifically literate leaders and members of society. A formal laboratory period is not associated with this course and so conceptual discovery and quantitative experimentation must take place outside of the classroom. Because of the physiological and psychological nature of communication, the proposed exercises may find application in several psychology courses. The maturing technology of the voice-computer interface provides a rich area of investigation to help realize the purpose of these courses.

2. Description

The voice–computer interface affords many opportunities for explorations because it is *not* a mature technology. There are still many issues that need to be solved in order for the interface to work seamlessly. This Fellowship will provide support for the development of several exercises that will investigate the successes and shortcomings of ASR and TTS routines. The exercises will be designed to help non-science majors explore concepts in much the same way that a scientist does.

A beginning exercise that exposes the students to the basics of speech acoustics will tie in well with current course material which examines the voice as a musical instrument. Students will record the basic sounds of speech, phonemes, and determine the frequency spectrum for each phoneme using a freeware sound analysis program called Audacity, with which the students will already be familiar. They will then identify the characteristics that are common to the phoneme production of the other members of the class. The degree to which common characteristics are easily identified will help them understand the process and less-than-perfect success rate of the ASR routine. It is the variability of phoneme characteristics between speakers that is the biggest stumbling block to ASR. Even an individual speaker introduces variability depending on the inflection of the word, the running together of words (coarticulation), the amount of disfluences (uhs and ums) and the day-to-day changes in voice quality.

Subsequent exercises would expose the students to the normal workings of the ASR and TTS software and the manner in which each routine accomplishes its task.

Finally, exercises would be written that would measure the error rate under different conditions. For example, one student could train the ASR. Then, the error rate when that student speaks could be compared to the error rate of another student using the same template. Another exercise would test the intelligibility of the computer's TTS voice under different ambient noise conditions. Errors created by the presence of ambient noise are a very important consideration for using TTS and ASR in the lab, in the field or in an industrial environment.

In a typical exercise, the class of 30 students will be divided into ~10 small groups of students. Each group may choose to use their own or a physics lab Windows-based computer or be supplied with a Tablet PC should they wish to take advantage of a mobile computer.

The work for this coming summer will fall into three areas:

(i) Writing clear descriptions of the fundamentals of speech acoustics and the ASR and TTS routines

(ii) Determining a list of experimental conditions that students may wish to investigate

(iii) Writing and testing the procedures for doing the exercises and performing the error analysis (most of the effort).

3. Timeline

The proposed work will be accomplished in the period from mid-May through August 2006. First full use of the new curricular material will be in the Musical Technology (Physics 115) course to be taught in the spring of 2007.

4 & 5. Technology and Other Support

Davidson College is supporting this work by providing the investigator with the necessary equipment (lab computers, Tablet PC's, microphone) and Microsoft Office software. Students may choose to use their own computers. The proposed work will be carried out in collaboration with Davidson's Instructional Technology Group. They have a significant supply of Tablet PCs and employees that have experience using the ASR and TTS software and Microsoft's speech SDK. The Tablet PCs are equipped with most of the quality features of a desktop PC: an onboard soundcard, microphone, speaker, and jacks for external microphones and speakers.

6. Learning Outcomes

One purpose of the proposed curricular material is for the student to become engaged in the same processes that a scientist uses. According to recent physics education research, active engagement with course material in and out of the classroom or laboratory provides a powerful learning process. Students will use several new pieces of ASR and TTS software that have now become generally available. Upon completion of the material described here, the student will be familiar with important tools that expand the use of computer technology in their lives.

7. Curriculum

The proposed curricular material will be used at Davidson College the next time Musical Technology (Physics 115) will be taught, which is in the spring of 2007. In addition, the material developed through this fellowship would be applicable in Cognitive Psychology (PSY 276) and Perception and Attention (PSY 301) taught by a member of the Davidson Psychology Department. The material could be used in courses on speech and hearing, and on ergonomics as well. While Davidson does not offer these courses, the web page that will be created as part of this project will be a resource to other ACS schools that do offer these courses.

8. Assessment

The individual exercises will be assigned and graded in a manner similar to end-of-chapter homework sets. The grade for this portion will depend upon the students' written responses. In order for the project as a whole to be evaluated by the students, a section of the end-of-course student evaluation form will be dedicated to the new material. The principal investigator has

worked with students in Davidson's Industrial and Organizational Psychology (PSY 254) class to develop assessment tools for past ACS Technology Fellowship projects. This will be the case for the proposed project.

9. Dissemination

At the local level, the results of this work will be shared with the Davidson College community via the Instructional Technology Seminar Series sponsored by the Instructional Technology Group. In the fall of 2007, the results will be presented at the North Carolina Section meeting of the American Association of Physics Teachers. This venue is an ideal place to pass along the results of this work to many ACS colleagues and other educators in physics (high school, 2-year and 4-year college and graduate school faculty) throughout the state in the form of a talk and/or a workshop. In order to have the broadest impact with the curricular materials, a webpage will be created for the free distribution of this material to faculty at Davidson, to ACS colleagues, and to higher educators in general. In addition, the new material may be presented at an upcoming meeting of the Acoustical Society of America.