Survey of Continuous Speech Recognition Software Usability

Hsin Eu and Alan Hedge
Cornell University
Department of Design & Environmental Analysis
MVR Hall, Ithaca, NY 14853.

Abstract

This research examines the usability of English language continuous speech recognition software programs. Continuous speech recognition programs (CSRP) were brought to market at the end of 1997, with claims that they are capable of recognizing users' continuously spoken speech and translating this directly into text processing software. A comprehensive usability survey questionnaire was administered via the web to collect information from both CSRP users and non-users. Survey respondents were recruited via newsgroup announcements, email referrals, and email-lists. Over a two month period, 162 respondents answered the survey, and 80 were CSRP users. Results from the survey profile patterns of general computer usage and opinions on CSRP. Results from the CSRP users indicates user behavior pattern, the pros and cons of current CSRP, new design opportunities, and levels of user satisfaction. Results from the survey are relevant to ergonomic guidelines for improving interface design for future CSRP.

Introduction

The pace of development in microprocessor development continues to follow Moore's law, and recent generations of computer chips have the power to meet the processing demands of speech recognition. Improvement in computer technology coupled with new approaches to speech recognition algorithms has recently produced several commercially available software programs that claim to support continuous speech recognition. The use of continuous speech may provide for a more natural and user-friendly interface. Continuous speech recognition programs (CSRP) also may obviate the need for extensive use of the computer keyboard and mouse, thereby reducing risks of musculoskeletal discomfort and repetitive strain injuries.

In 1997, Dragon Systems released "Naturally Speaking", the first general-purpose speech recognition system claimed to allow users to dictate to a computer in

a conversational manner (ZDnet, 1998). In 1998, two leading continuous-speech recognition programs were tested by PC Labs (ZDnet, 1998). Both continuous speech recognition programs (Dragon Systems NaturallySpeaking and IBM's ViaVoice) offer apparently impressive features: their abilities to transcribe continuous speech, to decipher seemingly difficult words, to access large lexicons, to allow speech control of most computer operations. However, neither product achieved the 95 percent accuracy rate, which users apparently would like to see in a mainstream consumer product. PC Labs also found that making corrections or formatting documents using a speech interface could be painstaking.

Speech technology is more than just a way of transmitting words or ideas; it conveys the essence of human emotion, moods, and personality (Westall, Johnston, & Lewis, 1996). In 1997, Dillon and Norcio conducted on a research on novice and expert nurses performed a hands-busy and eyes-busy task using a continuous speech-input interface. The results suggested that factors that improve the user's performance include expertise in the domain, experience with the interface, and the use of a small vocabulary. Also, experience with the interface corresponded with a higher degree of acceptance.

In sum, although some researchers have investigated human performance aspects of speech recognition programs, no published studies have evaluated user reactions to all aspects of CSRP or the actual usability of the software interface. The present study reports results from a survey of users and non-users of CSRP that focuses on detailed usability issues. Non-users of CSRP were surveyed to elucidate reasons for not using CSRP.

Research Method

Subjects

One hundred and sixty two respondents participated in this research. Eighty respondents were CSRP users.

Survey instrument

An electronic survey questionnaire was developed for administration via the world-wide web. The questionnaire was comprised of 2 parts: a section of 13 questions

asking for responses to 45 items for completion by all respondents, and a section of 31 questions asking for responses to 201 items for completion only by CSRP users. For most questions responses were made by selecting appropriate radio buttons or check boxes. For some questions, respondents selected answers from a pull-down menu or typed a short answer into a text box.

Procedure

A sample of users and non-users of CSRP was obtained by targeting potential respondents with an electronic research request announcement, which described the project and provided a hyperlink to the Web survey. This request was posted to selected newsgroups and email-lists (Table 1). Respondents were also encouraged to forward the request to colleagues.

The first section of the Web survey for all respondents took 3-5 minutes to complete. The second section for CSRP users took around 20 minutes to complete. Respondents received confirmation of successful submissions. Survey data were collected automatically for subsequent statistical analysis.

RESULTS

General Computer Usage

The demographic profile of respondents shows that 67.3% were male; 68.6% were aged 26-50 years; 93% used a computer >3 years; 79% used a computer 6 to 7 days a week; 51.8% used computers 7-9 hours a day; 25.9% use 4-6 hours a day; and 66.7% use a windows (Win95/98/NT) platform.

Eighty four percent indicated that they knew about computer continuous speech recognition programs, such as Dragon NatuallySpeaking, IBM Via Voice, and L & H Voice Xpress (Kurzweil) and 11.7% had not heard of these. 51.2% of respondents think CSRP will replace keyboards for 'composing documents' and 24.0% think CSRP will replace keyboards for 'computer image manipulation'.

Usability of Computer Continuous Speech Recognition Programs (CSRP)

The demographic profile of CSRP users showed that 73.8% were male; 70.1% were aged 26-50; 85% use their CSRP on a 'Desktop' computer; and 97.5% use CSRP on a windows platform ('Win95/98/NT'). 48.8% users have a 200-300 MHz CPU; and 56.1% users had 128 Mbytes RAM in their computers. Sixty-nine CSRP users (86.3 %) were native English speakers; seven were not; four people did not specify.

More than half CSRP users in this survey use Dragon NaturallySpeaking (DNS): 22 users (27.5%) use Dragon NaturallySpeaking Preferred 3.0, and 19 people (23.8%) use Dragon NaturallySpeaking Professional. Other subjects use other versions of DNS or IBM Via Voice or L & H VoiceXpress (Kurzweil). When asked what programs they previously used: 26 respondents used Dragon NaturallySpeaking Preferred 2.0; 19 used 'IBM Via Voice Gold'; 17 used Dragon NaturallySpeaking Professional, and 13 people used Dragon NaturallySpeaking Preferred 3.0. Thirty-four users have/had used their CSRP for '1-6 months'; twelve have used for '7-11 months'; twenty-seven have used for '1-2 years'. About 59% users reported that they use/used their continuous speech recognition programs 5-7 days a week. The total time of using CSRP was calculated by summing usage time for different locations. Users use CSRP 3.04 hours a day on average.

Users were asked how well the CSRP worked for various tasks (1 indicates 'very poorly' and 4 is 'very well'). Use for 'Dictating Letters' was ranked first (average score of 3.16) and 'Navigation between Programs' was ranked last (average score of 2.06). When asked what made them choose to use a CSRP: 22.4% reported that this was 'an alternative to typing' and 21.8% said 'to improve personal productivity'. When asked how they chose their program: 36.7% responses said this was based upon a 'Product review on Internet or computer magazine' and 32.7% on 'Other' (e.g. tried all programs, the software came with hardware purchase, from work). Users were asked about the importance of various CSRP software features that affected their choice of CSRP ('1' is very unimportant and '4' is very important): 'Product accuracy' had the highest average score (3.79) over the other aspects and 'Price' has lowest score (2.49). When asked which characteristics users thought were most important for a CSRP <u>after</u> they had used their chosen CSRP: 21.8% responses were 'Product

accuracy', and 15.8% responses were 'Dictation speed'. Users were asked how long they had trained her/his CSRP before starting to use this: 31 subjects (40.8%) replied '30-59 min'. 44.6% of users did not feel annoyed with CSRP training, mostly because this was seen as 'very necessary'; 32.6% also said the training was not annoying and that it helped to familiarize them with the program.' When asked how long they had read their manual before they started to use the CSRP software 38.8% subjects answered '<30 min'; 25.0 % users said '30-59 min'; and 23.49 % people said '0 min, because I didn't have to'.

Users were asked to rate the various aspects of user's continuous speech recognition program on a 4 point scale (1 was negative attribute and 4 was positive attribute): 'Ability to expand vocabulary' had highest average score of 3.39 and 'User (technical) support' had the lowest score of 2.45.

When asked which input advice users' would prefer to finish the following tasks: 45 CSRP users (56.3%) preferred to use 'Voice' for *composing documents*; 22 subjects chose 'Voice & Keyboard' & 17 chose 'Keyboard & Mouse' for *correcting mistakes in documents*; 21 subjects chose 'Voice & Keyboard' & 15 chose 'Keyboard & Mouse' for *editing documents*; 34 subjects chose 'Keyboard', 19 chose 'Voice' and 17 preferred 'Voice and Keyboard' for *database input*; 23 subjects preferred 'Mouse' and 18 preferred 'Keyboard & Mouse' for *computer image manipulation.* 18 people selected 'Keyboard' and 18 selected 'Mouse' for *searching & browsing information*; 21 people preferred 'Mouse' for *navigating within a program*; and 20 users chose 'Mouse' and 17 chose 'Voice' for *navigating between programs.*

Users were asked which symptoms, if any, they had experienced that they attributed to the sustained use of CSRP: 45.3% responses reported 'Dry throat' and 30.2% said 'Vocal strain'. 32 subjects (40 %) had experienced that a bad cold affected the performance of their CSRP. 27.5% of users said that CSRP had increased their productivity, and 19.6% indicated other benefits (e.g. reduce RSI risk, saved money).

Users indicated several current limitations of CSRP: 29% said 'The reliability is not consistent while the task complexity increases' and 24.3% said 'Performance might be sensitive to voice changes, such as 'catching a cold'. Problems that users experienced when they first started to use CSRP were as follows: 29.2% said that

there original computer capabilities 'didn't satisfy the minimum equipment requirement of the program' and 15.4% said that they didn't know 'what kind of soundcard they should get'. To solve these problems, 24 out of 43 respondents upgraded or bought a new computer. Users reported several usability problems with their CSRP: 29.1% said 'I tried to improve the accuracy but there was not much improvement'; 23.3% found that their environment was 'so noisy that it decreased the accuracy of speech recognition', and 19.8% reported cognitive difficulties concerning the fact that they could not think while talking to the computer. Over one-third of users reported that either they still haven't solved problems or that they have 'uninstalled the program'.

Fifty-one subjects (63.8%) reported that to effectively use the CSRP they had to change their way of speaking. Forty-five users (56.3%) said they were satisfied with the documentation that came with their CSRP. Sixty-two users (77.5%) are planning more extensive use of their CSRP in the future.

Comparison of Dragon NaturallySpeaking (DNS) Preferred 3.0 and DNS Professional

The two major programs used by this sample of CSRP users were Dragon NaturallySpeaking (DNS) Preferred 3.0 (22 users) and DNS Professional (19 users). Comparison of reactions to these programs showed significant differences between these in 'Varieties of functions' (Chi-square = 4.18, df = 1, $p < 0.05$) and 'Dictation speed' (Chi-square = 8.58, df = 1, $p < 0.05$). Users of DNS Professional rated these two dimensions significantly higher than users of DNS Preferred 3.0.

References

Dillon, T.W., & Norcio, A. F. (1997, October). User performance and acceptance of a speech-input interface in a health assessment task. International Journal of Human-Computer Studies, 47(4), 591-602.

IBM Voice Type (1998).
http://www.software.ibm.com/workgroup/voicetyp/us_menu.html

Speech Technology Timeline. (1998, March).
http://www5.zdnet.com/products/content/pcmg/1705/283053.html

Watch What You Say. (1998).
http://www.zdnet.com/pcmag/features/speech/intro.html

Westall F. A., Johnston R. D., & Lewis, A.V. (1996, January). Speech technology for telecommunications. BT Technology Journal, 14 (1), 9-27.