Voice-Based Group Support Systems

Milam Aiken
University of Mississippi, USA

INTRODUCTION

One of the primary reasons large meetings utilizing group support systems (GSS) are more efficient and effective than traditional meetings is because the former are based upon typed comments and opinions while the latter are based upon voice input (Nunamaker, Briggs, Mittleman, Vogel, & Balthazard, 1997). Using a keyboard, participants can submit comments to the group anonymously and simultaneously, but in an oral meeting, participants must take turns speaking to avoid confusion. Further, group members in an electronic meeting can skim recorded typed comments easily, while most traditional meetings do not have complete transcripts available as the discussion progresses. Even analyzing new comments is more efficient with a GSS. Most people can read faster than they can listen, and in an oral meeting, the rate of input is limited by the current speaker's voice.

Although many people are now familiar with typing, most are more comfortable with speaking and can generate more words per minute by voice. Are there more efficient means of generating ideas in a GSS session (Briggs, Nunamaker, & Sprague, 1998)? To improve the productivity of electronic meetings, the systems should be made as "typewriter-less" as possible (Gray & Olfman, 1989). Integrating automatic speech recognition (ASR), or simply, speech recognition (SR), with a GSS might improve the rate of comment generation.

Accuracy has been the major barrier to greater SR acceptance, but high accuracy might not be needed in a GSS meeting. For example, if one comment is not understood, there are likely to be other similar, if not redundant, adjoining comments that might be clear or could aid the understanding of the earlier comment. In addition, a participant can submit a new comment asking for clarification from the group.

This paper summarizes research conducted using SR technology during electronic meetings. Results of these voice-based GSS (VGSS) studies show that SR transcription accuracy generally is low due to background noise in these face-to-face meetings. Distributed VGSS meetings are likely to be more efficient and effective.

BACKGROUND

An electronic meeting is one form of e-collaboration that typically involves exchanging comments via a computer-based network (Fjermestad & Hiltz, 2000). The principal reason that these meetings are superior to traditional, oral meetings when sharing ideas among many people is that comments are typed, allowing all comments to be recorded as they are written, anonymous submission of ideas, and simultaneous generation of text.

At any time, GSS group members can read or skim through old comments very quickly. While there is no standard test for measuring reading speed (reading material varies in length, complexity, style, etc.) and tests for reading comprehension are subjective, average readers generally read around 200 words per minute (wpm) with a typical comprehension of 60% (Speed reading test online, n.d.) and average college students can read fiction and non-technical materials between 250 and 350 wpm (Suggestions for improving reading speed, n.d.). Thus, we will assume the average GSS participant can read the public comments at about 300 wpm.

But generating text is much slower than reading. While the maximum typing speed in English has been recorded as high as 212 wpm (Glenday, 2005), most people, of course, type far slower. A student often can type 13 to 41 wpm, a good typist can generate from 61 to 90 wpm, and an excellent typist may produce between 85 to 112 wpm, assuming five characters per word (Cooper, 1983), and typical undergraduate Business students can type 36 "easy" words per minute (commonly occurring words with few syllables) and type 24 "difficult" words per minute (Rebman, Aiken, & Cegielski, 2003). Part of the difference in student typing rates between the two studies can be explained by the far greater use of computers now; many more people are familiar with typing, and many type every day.

However, the rates above were not adjusted for errors. Over an hour-long period, one typist was able to produce 149 wpm, with a 10-word penalty per error (McWhirter & McWhirter, 1973). When simply transcribing text, knowledge workers at IBM were able to generate 32.5

corrected words per minute (cwpm)—time was taken to backspace and type over errors (Karat, Halverson, Horn, & Karat, 1999). In addition, extra time is needed when composing fresh ideas, and the same workers were able to generate only 19 cwpm when thinking of new ideas. Although GSS participants must compose fresh ideas, they do not often take extra time correcting mistakes, as most transcripts have many grammatical (e.g., lack of capitalization and poor punctuation) and spelling errors. For example, in one study (Aiken, Vanjani, Martin, Young, & Govindarajulu, 1994), 30% of all comments typed by undergraduate business students in a GSS meeting had at least one grammatical or spelling error.

Assuming a typical reading speed of 300 wpm and a typing speed of 20 wpm, a participant in an electronic meeting should be able to keep up with 15 other group members typing simultaneously, assuming these other participants are not also spending time reading comments. But GSS participants typically spend 60% of the total meeting time reading others' typed comments (Aiken & Vanjani, 1996). Thus, each minute, 180 words per minute are read and 8 words per minute are typed, on average, by each participant, and each should be able to keep up with the generation of new text until the group reaches a size of 23 people. However, participants do not necessarily need to read everything (especially since a recorded transcript is available for later review), and they could instead decide to skim and selectively read for greater detail, increasing the maximum group size.

One way to address this imbalance is to generate comments faster, perhaps through automatic speech recognition. People speak much faster than they type. The maximum rate of speech may be as high as 637 wpm (Glenday, 2005), but when composing fresh ideas, people typically speak at about 100 to 150 wpm (Lenneberg, 1967). Further, there is no need to worry about spelling or some grammatical errors such as lack of capitalization and punctuation. Assuming a typical speech rate of 120 wpm and a typical typing rate of 20 wpm when composing fresh ideas, it might be possible to generate about six times more text during a GSS meeting with SR, or alternatively, have a meeting last only 1/6 as long.

Figure 1 illustrates the rates of text generation possible in oral, GSS, and SR/GSS meetings. Because group members in a traditional, oral meeting must take turns speaking, the rate of text generation per minute remains constant, such as 120 wpm. However, in a GSS meeting, each group member can type at the same time. Thus, the rate of text generation per minute increases with the group size. Assuming a typical typing rate of 20 wpm, a six-person GSS meeting can generate as much text as a six-person oral meeting. However, integrating SR into a GSS meeting could allow far greater rates of text generation, perhaps as high as 510 wpm for a six-person meeting, assuming each is able to generate 85 wpm using the speech recognition software.

SPEECH RECOGNITION

There are definite advantages to sharing ideas via written text in an electronic meeting, but people often are more comfortable with speaking. Automatic speech recognition has the potential to bridge this gap. The speech recognition process includes several steps (Markowitz, 1996):

1. **Audio input:** The human voice is transmitted through a microphone connected to a microcomputer with a standard sound card.
2. **Acoustic processor:** The acoustic processor converts the captured audio into a series of phonemes.
3. **Word matching:** The software attempts to match the phonemes to the most likely words. First, it uses acoustical analysis to build a list of possible words that contain similar sounds. Then, the software uses contextual information to predict what words should come next, helping the system to distinguish among homonyms, for example.
4. **Decoder:** The decoder selects the most likely word based on the rankings assigned during word matching and assembles the words in the most likely sentence combination. It then transfers the sentence to the word processing application.

To train the software to recognize the speaker's voice, a process known as enrollment is used. During this initial phase, the user reads one or more pre-selected passages of text on the computer screen while the software matches the words with the speaker's distinctive vocal patterns. Although more training usually results in greater accuracy, fairly good results often are achieved within as little as five or 10 minutes.

As one would expect, speech with SR can generate words very quickly, and one study (Rebman & Aiken, 2000) showed that undergraduate Business students can

Forecasting Smart Tourism Visitor Flows Leveraging Big Data Technology Assistance
Guoqiang Tong (2024). *International Journal of e-Collaboration (pp. 1-24).*
www.irma-international.org/article/forecasting-smart-tourism-visitor-flows/346809