Working Paper Series
ISSN 1170-487X

Voice input and information exchange in asynchronous group communication

by Robert J. McQueen

Working Paper 92/3

March, 1992

Department of Computer Science
The University of Waikato
Private Bag 3105
Hamilton, New Zealand

Voice Input and Information Exchange in Asynchronous Group Communication

Robert John McQueen
University of Waikato
Hamilton, New Zealand
bmcqueen@waikato.ac.nz

Abstract

Existing computer supported co-operative work (CSCW) systems for group communication typically require some amount of keyboard input, and this may limit their usefulness. A voice input prototype system for asynchronous (time separated transactions) group communication (AGC) with simulated conversion to text was developed and an experiment constructed to investigate if advantages over conventional keyboard input computer conferencing were possible for the information exchange task. Increases in words used and facts disclosed were higher for voice input compared to text input, which implies that voice input capability could be advantageous for future asynchronous group communication systems supporting information exchange.

Asynchronous group communication

Group communication can take place through a number of media. The most common and widely used is the face-to-face meeting, which can be characterized by a high bandwidth (amount of information transferred per unit of time), high interactivity (turn sharing), multi-channel (both forward and reverse channels between speaker and listeners), and synchronous (all participants temporally linked).

The last characteristic, time, can be used to divide group communication media into two major categories:

a) Synchronous (all participants interacting at the same time)
- -simultaneous face-to-face meeting
- -simultaneous video & voice teleconference
- -simultaneous telephone conference call
- -computer supported group decision systems
- -remote "whiteboard" systems

b) Asynchronous (participants connect at different times; contributions stored and retrieved/replayed)
- -the voice input Moot prototype discussed below
- -computer conferencing
- -electronic mail
- -exchange of paper documents (copy lists)

Synchronous small group communication, using the medium of face-to-face meetings, is a well researched area[1]. Technologies which support synchronous group communication from remote locations, such as telephone conference calls, videoconferencing, and more recently remote

[1]Bales, 1972, is widely cited in small group work.

whiteboard systems, have appeared over the last few decades, but have not yet attracted widespread popularity compared to the dominant face-to-face meeting..

Asynchronous group communication requires two important shifts from the face-to-face meeting medium. First, some kind of technology must come between speaker and listener, to enable time delay between contribution and review of comments (capture, storage, replay). Second, the words used to convey information and ideas are likely to be in a textual form rather than oral form, for reasons of efficiency of transmission, storage, and review, and therefore require typing. While previous studies [Hiltz 1984] have shown that typing speed was not a barrier to the level of commitment to use of a text based conferencing system, a competent typist would certainly have an advantage over non-typists in the ease of preparing keyboarded contributions.

These shifts would be seen by most participants as diminishing the richness, ease of use, and interactivity of face-to-face group communication. If asynchronous group communication is to flourish, then ways must be found to minimize these barriers to use, while maximising the benefits that may be obtained.

Asynchronous group communication, supported by computer, is an existing and well studied area[2]. A number of electronic media communication systems have provided a basis for examining the factors involved, and a foundation on which successive iterations of system designs can be built. To date, most of these systems are based on keyboard input of text. The extensive literature on small group interaction and a growing body of work on the effects of computers on interpersonal communication patterns is supportive of further research in this area.

Comparison of face-to-face and present AGC

The present "typed input" AGC environment is quite different than the environment of a face-to-face meeting. AGC is asynchronous - the interaction between group members is not in "real time". This means participants have more time to reflect before contributing. They are able to seek other sources of information and opinion off line, with no apparent disruptive delay, before responding. Conversely, there is a greatly reduced visual content for AGC participants compared to that experienced by face-to-face participants, which results in a lack of instantaneous support/dissent indicators from other participants. Typed input also results in greatly reduced emphasis and intonation signals readily available in face-to-face discussions.

Information exchange

Information exchange is one of the activities that take place in a normal face-to-face meeting, along with decision making, socializing, concensus gathering and similar functions. In a study of users of the COM computer conferencing system, Adrianson and Hjelmquist [1988] found that information exchange played the key role in motivating users. When asked why they *started* to use COM, 53% (the highest ranked reason) gave "need for information" as the reason. When asked how COM was useful for work related activities, 83% (again the highest) gave getting information as the reason. Decision making ranked near the bottom as a suitable activity for COM.

Voice input and speech recognition

The desire to use speech recognition for computer applications has preceded the ability of this technology to support the common modes of speaking. Human beings (and other sentient animals!) can recognize speech which is connected word (no pauses between words) and speaker independent (understand a variety of people), but this has proved to be a difficult task for computers. However,

---

[2]See Rice, 1987

some research has been previously done which investigates human communication tasks which might be mediated or supported through computer technology.

The problem solving task was the subject of an early investigation by Weeks & Chapanis (1976). They looked at the effect of communication channels (teletypwriter, telephone, CCTV and face to face) on a synchronous problem solving task for pairs of subjects, and found that channels using voice were much faster, and far more verbose, than those using typed word channels. They also found that the addition of a visual channel, through either a camera/display or glass window did not appreciably decrease solution times.

For the task of document preparation, Gould et al. [1983] investigated the effect of a simulated listening typewriter without having an automatic speech recognition system in operation. They located a subject in one room and a human typist in a second room, and experimented with both isolated and connected word speech input to their simulated listening typewriter. They were able to draw conclusions about the application of speech recognition to the problem of interest, without actually having the technology operating.

Newell et al. [1990] followed Gould's study with an examination of a voice input word processing system. The "man behind the curtain" used a machine shorthand transcription system (potential speed 180 wpm; actual operator speed 120 wpm) rather than Gould's 80 wpm typist, allowing the possibility of faster input speeds if required. The subjects used voice input for editing commands as well as raw text. In a partial replication of Gould's experiment, word composition rates were lower (7.9 wpm vs Gould's 11.5 wpm) despite the faster speeds. Efficiency rates (words on final document vs words spoken) were low (39%), reflecting the large number of words required for "hands off" editing, capitalization and punctuation functions.

Both Gould and Newell were concerned with the required result being a "perfect" typed document, in the same way that a business letter is only sent in perfect form. Their intent was to provide in their simulations the ability to give punctuation and capitalization instructions through speech input, as well as the raw words for the text itself.

Using speech input for commands to a computer software package is another potential use of the technology. Martin [1989] describes a study using speech input for a VLSI design package, as an alternate to keyboard command entry. They found that speech input is faster than typed input, as well as increasing user productivity by providing an additional response channel.

Office applications of speech recognition are reviewed by Noyes & Frankish [1989]. They discuss potential applications in the areas of voice messaging, word processing, data entry, information retrieval, and environmental control. They conclude that currently available speech recognition technology has shortcomings which may be critical for projected office applications. They comment that "there is a suggestion here that the superficial attractiveness of speech recognition technology has cast it as a solution in search of a problem". However, they argue that research should follow application investigation as well as research into voice input interface design.

Several organizations are active in developing speech recognition devices that have potential application in the office environment. Kurzweil, Dragon Dictate and IBM have large vocabulary speaker dependent products presently available on the market.

The area of automatic speech recognition is both developing rapidly and failing to meet inflated expectations of worthwhile applications. From the work that has been undertaken on listening typewriters, it may be concluded that voice input as the exclusive input and control mechanism for both commands as well as text input is simply too clumsy and inefficient compared to alternative methods. However, a combination of voice input for the words and a direct manipulation pointer for command execution and editing may be a happy compromise.

An experiment to compare voice input and text input AGC [3]

We wished to compare the influence that the two media types (voice input and text input) would have on the ability of group members to exchange information amongst themselves. Other important group processes such as discussion, deliberation, consensus gathering, and decision making were not of primary interest in this study. Hypotheses were constructed based on the following research questions:

a) How is the number of words contributed by participants related to the medium of voice input compared to typed input?

b) For an information exchange task, will the number of facts disclosed by a participant to the group differ when voice input is compared to typed input?

c) Where a participant holds a large proportion of the facts held by all group members, how is the number of words and the number of facts contributed by that participant related to the medium of voice input compared to typed input?

d) Where a participant holds a small proportion of the facts held by all group members, how is the number of words and the number of facts contributed by that participant related to the medium of voice input compared to typed input?

The Moot voice input AGC system

For the purposes of this research, we wished to test the effect of voice in, text out capability on information exchange in an AGC system, rather than the actual mechanics of speech recognition. Therefore, we avoid the difficult problem of speech to text recognition, by using a manual (secretary) method to transcribe the voice files to text form. If voice input were an option for AGC, it might be a potential leveller of variation in contribution quantity by removing the typing problem, and also present a communication media similar to the normal mode of voice contribution in face-to-face meetings.

A prototype voice input AGC prototype system, nicknamed Moot, was developed as a basis to investigate these areas. The Moot system allowed participants to contribute voice messages, and read the text equivalent of those messages on the screen. From the user's perspective, it was indistinguishable whether the voice messages are automatically "speech recognized" and converted to text by the computer, or whether a manual transcription process is taking place offline.

The Moot system operated on a 386 PC under DOS and Windows. For the experiment, the Moot system was located in a private office. On entering, the user would see a microcomputer (screen and processor box) located on the desk in the office. Plugged into the microcomputer is a headset, consisting of earphones and an attached microphone, which connect to an internal speech digitizing board. A mouse, also attached to the computer, is on the desk. A keyboard is present, but has been placed out of the way, instead of its normal position in front of the screen. The user sits in the chair, and put on the headset. At signon, the user is automatically connected to their group discussion (in the experiment, each subject was only a member of one discussion). The first unread message put in by other group members since that user's last sign-on is displayed on the screen. If the voice message has been transcribed, then the text of the message will appear in the box on the screen. If the transcription has not yet been completed, then the words "There is no text translation for this message yet" appears in the box.

---

[3] A complete description of the experiment, the Moot prototype and the statistical results is given in McQueen, 1991.

The Experimental Methods

The objective of the experiment was to measure information exchange in a group discussion, using the medium of voice input asynchronous group communication, and compare this with the typed input medium.

To provide a basis on which information exchange could be structured and observed, "information profiles" of three hypothetical prospective university students were developed, and then the profile information subdivided into five parts. Twelve groups of 5 subjects each were formed. Each group used the three media (text, voice, and face-to-face) in turn to discuss one of the three prospective students.

The task given to each experimental subject, and their group, was to formulate advice to this prospective student on which degree should be undertaken, and what institution should be attended. Each individual group member held seemingly unconnected pieces of information, which formed parts of the whole body of information about the prospective student. Individuals did not know which of the pieces of information they held were shared by other group members.

The underlying need, as perceived by group members, was to exchange their information (whatever they saw as relevant to the discussion) with the others in the group, discuss the suggestions and opinions of what should be recommended to this prospective student, and finally agree on a consensus piece of advice for that student.

The discussions of group members would be recorded, and subsequently analyzed, with respect to which of the information facts they held was actually disclosed through information exchange to the other group members. The actual advice offered was of no interest.

Experimental Results

Table 1 gives the overall means of contributed words and facts by role. Roles 1 and 2 had the greatest number of facts available (33 and 28) while roles 3 and 4 had the fewest (12 and 10). Role 5 had a small number of facts, but had a responsibility to record the discussion results, and thus became involved with some discussion facilitating activities. CoSy was the text input conferencing system, and Moot was the voice input system.

Table 1 Overall Means - Words and Facts

	-----facts-----		-----words-----		--role means--	
role	cosy	moot	cosy	moot	facts	words
1	12.5	15.5	690	1055	14.0	872
2	6.8	9.6	441	934	8.2	688
3	5.3	4.4	651	592	4.9	621
4	5.2	7.3	468	821	6.3	644
5	3.9	3.3	556	970	3.6	763
Media means	6.8	8.1	559	883	7.4	721
high fact(1&2)	9.5	12.4	560	992		
low fact(3&4)	5.3	5.9	559	706		

Table 2 presents a different view of the data by calculating means for facts and words by group.

Table 2 Means by Group for All Participants

grp	n	cosy facts	moot facts	ratio M:C	diff M-C	cosy words	moot words	ratio M:C	diff M-C
1	4	3.75	4.50	1.20	0.75	304	983	3.23	679
2	5	5.00	5.60	1.12	0.60	538	900	1.67	362
3	5	7.20	9.40	1.31	2.20	735	757	1.03	22
4	5	2.20	7.20	3.27	5.00	214	1153	5.39	939
5	5	7.40	4.80	0.65	-2.60	934	592	0.63	-342
6	3	8.33	14.00	1.68	5.67	143	958	6.70	815
7	4	5.00	5.75	1.15	0.75	393	458	1.17	65
8	5	10.60	10.20	0.96	-0.40	598	882	1.47	284
9	5	6.80	6.80	1.00	0.00	791	1499	1.90	708
10	5	5.60	7.40	1.32	1.80	475	692	1.46	217
11	5	8.60	12.60	1.47	4.00	317	770	2.43	453
12	4	11.00	10.00	0.91	-1.00	1129	926	0.82	-203
mean		6.790	8.188	1.336	1.398	548	881	2.325	333
max		11.0	14.0	3.27	5.67	1129	1499	6.70	939
min		2.2	4.5	0.65	-2.60	143	458	0.63	-342

Compared to text input, voice input resulted in an increase of the number of words contributed by all test subjects by 151% ($p<.02$) and an increase in the information facts disclosed by 38% ($p<.1$). For subjects holding a high proportion of available facts, voice input resulted in increases of 204% ($p<.02$) in words and 39% ($p<.1$) for facts. No conclusion was reached on subjects holding a low number of facts. Effective input speed using voice was 150 words per minute.

Outcomes of the study

One clear outcome of the study was that **voice input to asynchronous group communication is feasible**, and usable for its intended purpose. However, the present shortcomings in available speech recognition, which are outside the scope of this research, may seriously limit its applicability to group communication systems in the short term.

The second clear outcome is that **voice input capability results in a large increase in the number of words contributed to a discussion**, above that of keyboard input systems, for the systems tested, and is strongly supported by the results obtained. Further work may be required to determine whether this apparently desirable effect for computer support of group communication is in fact of value.

The third outcome is that **the number of facts disclosed by participants is moderately increased when a voice input capability is incorporated**, based on the test environment, and is moderately supported by the results obtained. This effect is seen for all participants in the test, and for those with a high proportion of the facts available for information exchange. This outcome was not proved for the subcategory of participants with low facts available.

The fourth outcome is that the prototype, using voice input and a direct manipulation mouse driven user interface, was successful for this type of application, and that **"keyboardless" user interfaces for this type of application seem to have merit, and should be further investigated and developed**.

Implications

a) Information exchange.

Information exchange has been identified as an important component of business meetings, and this research has shown that this function can be supported as or more effectively by voice input asynchronous group communication than by existing typed input methods.

Information exchange is an important group meeting component. These results suggest that because it appears that more words, and more facts are likely to be contributed if an AGC system has voice input capability, further work is justified to construct systems and techniques which may exploit the apparent advantages of voice input as a medium for information exchange through AGC.

b) Asynchronous group communication

The benefits and shortcomings of computer mediated asynchronous group communication compared to face-to-face communication have been raised elsewhere. However, the question raised by the results of this research is whether this new technique of inputing information into a group discussions will be the significant breakthrough to overcome present preference to "doing it in the flesh".

Will there be the expectation that AGC can undertake all tasks usually associated with a face-to-face meeting, such as brainstorming, decision making, and particularly socializing, as successfully as it appears that the information exchange task might be performed? And if these expectations cannot be satisfied, will a good performance in the information exchange area alone be sufficient to sustain interest?

Perhaps we can find a compromise solution in using particular media for the tasks best suited: for example, using voice input AGC for a first phase of fact gathering (over a few weeks, from diverse geographical locations), followed by a second phase face-to-face gathering where decisions are made and social relationships are established or confirmed, followed in turn by a third phase typed input follow-up and cleanup. No one medium may be sufficient, or suitable, for all of these phases.

c) Keyboardless computer systems

The mouse-icon style of user interface continues to grow in popularity, from its birth at Xerox PARC, through its successful introduction with the Macintosh, and the rapid acceptance of the Windows 3 product from Microsoft. Most applications, however, still require some movement between mouse and keyboard. The development and testing of the Moot prototype has shown that potential exists for completely keyboardless applications supporting interpersonal communication. Eyes do not have to be shifted from keyboard to screen continually. Hands don't have to be shifted from keyboard to mouse and back. Voice input allows gaze to be continually on the screen, hand continually on the mouse, and voice input used to input the "data".

The demonstration in this research that voice input can be effectively used in user interfaces may provide incentives for further investigation of keyboardless applications.

d) Speech recognition

This research was intended to look beyond the present limitations of speech recognition, to investigate whether an *application* using speech recognition could be successfully developed and applied, before the necessary speech recognition technology was available to support it.

It has been shown that large vocabulary, connected word speech recognition (perhaps with speaker independence thrown in!) could provide advantages over present typed input. Therefore, this may

provide incentive for speech recognition technology developers to investigate potential products for this use.

Present speech recognition technologies are based mainly on real time segmentation of speech into small intervals (typically 100 milliseconds), pattern matching those segments into phonemes or other basic building blocks, and then assembling those phonemes into words and sentences. Output from the speech recognition process is typically provided instantly.

Consider the implications of voice input AGC on speech recognition requirements. Output is not needed instantly; in fact, it may be acceptable to have delays of hours or even days between when a voice message was input, and when the text translation is required. This then implies that the speech recognition itself need no longer be necessarily be done in real time, allowing other techniques to be applied to the task. To start, many more iterations can be applied to whatever phoneme pattern matching algorithm is used, to converge on potentially better solutions. A new approach to non-real time speech recognition might now be able to use phonemes, both before and after the one presently being decoded, to narrow the choices available. Multiple passes over a sentence or phrase of connected speech might start with only a few phonemes or words being recognized in the first few iterations, but converge on the correct translation after perhaps millions of iterations.

Perhaps connected word, large vocabulary speech recognition is within reach for voice input AGC applications, where more time, and computing cycles over time, are available.

Conclusions

a) AGC and information exchange

The fit of asynchronous group communication to the information exchange component of group communication seems to be appropriate, and the experimental results show that this function can be performed adequately on either of the asynchronous group communication media tested. Other components of business meetings (decision making, socialising) may not be as well served by AGC. Further testing is required to quantify how well AGC performs the information exchange task, and these other business meeting tasks, when compared to a face-to-face meeting.

b) Voice input and AGC

The use of voice input for AGC systems in general seems to be worthy of further investigation, and development of more advanced operational systems based on the prototype Moot design would seem to be feasible. The fit of voice input to the particular information exchange task on AGC also seems to have merit. While there are present speech recognition shortcomings, this research has demonstrated that there is potential in this area.

c) Limitations of this research

The conclusions drawn about word and fact differences between a voice input system, and a text input system for asynchronous group communication are based on two specific software systems, namely the Moot prototype and the CoSy computer conferencing system. Caution should be used in generalizing these results to encompass generic voice input or generic text input systems, without further study to confirm the results, using perhaps different software systems and different testing environments.

d) Real systems

This study has developed and tested a voice input asynchronous group communication system prototype that simulates the availability of connected word, large vocabulary speech recognition to convert the voice messages into text. How feasible, and how soon will such technology be available?

A major attraction of present voice messaging systems (without speech recognition) is that they can be accessed from any telephone. Future AGC systems taking advantage of voice input might consider:

-interfaces to permit remote telephone access
-voice synthesis of system output
-use of the touch tone keypad for command or password entry

This research has demonstrated the feasibility of a "keyboardless" human-computer interface incorporating voice input. The combination of a mouse and voice seems natural, intuitive, and easy. User eye contact is at one place (on the screen, not the screen + keyboard). However, further work is necessary to work around the drawback of users having to put on a headset to use this type of system.

REFERENCES

Adrianson, L. & Hjelmquist, E. Users' experiences of COM - a computer mediated communication system. *Behaviour and Information Technology* (7:1), 1988, pp. 79-99.

Bales, Robert F. How people interact in conferences, in *Communication in Face-to-face Interaction*, Laver, J & Hutcheson, S. (eds.) Penguin Books:England, 1972, pp. 364-373.

Gould, John D, Conti, John & Hovanyecz, Todd. Composing letters with a simulated voice typewriter. *Communications of the ACM* (26:4), 1983, pp. 295-308.

Hiltz, Starr Roxanne. *Online Communities - A Case Study of the Office of the Future*. Ablex: Norwood NJ, 1984.

Martin, Gale L. The utility of speech input in user-computer interfaces. *International Journal of Man-Machine Studies* (30), 1989, pp. 355-375.

McQueen, Robert J. *The Effect of Voice Input on Information Exchange in Computer Supported Asynchronous Group Communication*. DPhil Thesis, Department of Computer Science, University of Waikato, Hamilton, New Zealand, 1991.

Newell, A.F., Arnott, J.L., Carter, K. & Cruickshank, G. Listening typewriter simulation studies. *International Journal of Man-Machine Studies* (33:1), 1990, pp. 1-19.

Noyes, J.M & Frankish, C.R. A review of speech recognition applications in the office. *Behaviour & Information Technology* (8:6), 1989, pp. 475-486.

Rice, Ronald E. Computer mediated communication and organizational innovation. *Journal of Communication* (37:4), Autumn 1987, pp. 65-94.

Weeks, G.D. & Chapanis, A. Cooperative versus conflictive problem solving in three telecommunications modes. *Perceptual and Motor Skills* ((42), 1976, pp.879-917.