Computer Aided Conversation for Severely Physically Impaired Non-speaking People

Norman Alm
MicroCentre
University of Dundee
Dundee, Scotland, U.K.
Tel: +44 382 23181
nalm@uk.ac.dund.mcs

John Todman
Psychology Department
University of Dundee,
Dundee, Scotland U.K.
Tel: +44 382 23181

Leona Elder
Business Studies Department
Dundee Institute of Technology
Dundee, Scotland, UK
Tel: +44 382 308000

A. F. Newell
MicroCentre
University of Dundee
Dundee, Scotland, U.K.
Tel: +44 382 23181
afn@uk.ac.dund.mcs

ABSTRACT This paper reports the development of a computer-aided conversation prosthesis which is designed for severely physically impaired non-speaking people. The research methodology was to model aspects of conversational structure derived from the field of conversation analysis within a prototype conversational prosthesis. The prototype was evaluated in empirical investigations which also suggested successful strategies for carrying out satisfying conversation using such a system. Two versions have been built and tested, one using an able-bodied operator to test the feasibility of creating conversation from prestored material, the second being used by a physically impaired non-speaking operator. The prototype demonstrated the advantages of this interface design in helping the user to carry out natural sounding and satisfying conversations.

KEYWORDS: human-computer interaction, user study, interface design, user observation, dialogue design, discourse analysis, user interfaces, retrieval models, search process, selection process, disability, speech synthesis

INTRODUCTION

People can be so severely physically impaired, either through paralysis or lack of muscle control, that they are unable to speak. Many such individuals can communicate via a keyboard (or special switches) connected to a speech synthesiser. However, the general lack of muscle control which usually accompanies loss of speech means that they are very slow in controlling any sort of input device. The result is that, even with currently available technical help, users typically only achieve 2-10 words per minute [7]. Unimpeded conversation, by contrast, usually proceeds at 120-200 words per minute [5], and slowness of communication and long silences tend to be interpreted negatively by listeners [12,16]. Non-speaking users of current communication systems also have great difficulty in taking the initiative in conversations, and retaining conversational control [9].

The fact that most people whose physical impairments keep them from speaking are cognitively intact, and can understand what is communicated to them, makes their situation even more frustrating. The development of high quality speech synthesis and the widespread availability of portable computers would seem to offer a potential solution to this problem, but the progress thus far has been modest. Current systems either allow the user to build utterances word by word, giving the typical communication rate already cited of 2-10 words per minute, or allow the storage and retrieval of phrases and sentences. Although this could offer much higher communication rates, in practice users found it impractical to remember all but a small part of what they had stored, and how to access it [1].

TOWARDS A MORE INTELLIGENT COMMUNICATION AID

A more promising approach to this problem, however, is to have the communication system itself keep track of the user's conversation, and offer help in selecting the next thing to say. Rather than just passively storing material, such a system might use information about conversational structure and the pragmatics of dialogue to retrieve conversational items depending on what is currently happening in the conversation, and what would be likely candidates for the next thing to say.

Although, of course, conversation can be unpredictable, it is a highly rule-governed activity [3,17], and, 'an enormous amount of natural language is formulaic, automatic, and rehearsed, rather than propositional, creative, or freely generated' [Fillmore, cited in 6]. Thus the content of many conversations is to a certain extent predictable.

With Fillmore's observation as a starting point, a reasonable focus is to consider ways of aiding a user to store and retrieve reusable conversational material. Following the development of a predictive system which helps a user to open and close conversations and give feedback to the other speaker [2], work has been proceeding on developing ways to incorporate predictive and conversation modelling techniques into a system which would help a user to take part in topic discussion, drawing on a store of reusable conversational material.

The way that an interface is designed between the user and this store of reusable material should reflect the pragmatics of the conversation. The pragmatics of a language concerns how the language is used to accomplish interactional purposes. It is on the top of the hierarchy of language analysis which has the usual stages: lexical, syntactic, semantic, and pragmatic. To give an example, the utterance 'Could you pass me the salt?', stripped of its pragmatic component, is simply a request for information. More than semantics is needed to interpret this phrase as a polite request for action.

Pragmatic analysis is important both in interpersonal communication, and in human-computer interaction. [11,18]. In the application described here pragmatic considerations are important, both at the user-computer and user-listener interface. The interface with the system needs to be sensitive to what the user is wanting to accomplish in the dialogue with the other person, rather than operating at the low level of individual words which are being used. Such sensitivity would increase the speed of the communication and give a transparency to the computer operations which would allow the user to concentrate on the interaction. Interpreting the interaction with the other person in terms of pragmatics suggests a communication system which helps the user to accomplish certain interactional goals, rather than simply to send messages.

An aspect of designing for people with disabilities is that the requirement that a system be as easy to use as possible is not an optional, but an essential feature of a system where most users will need to exert a great deal of effort to make any sort of directed movement. In fact, we are all disabled in a number of ways from the point of view of a computer, particularly in the bandwidth by which we are able to communicate with it - it is just that some of us are more disabled than others [10].

Given the extreme slowness with which a severely impaired person would be able to input text for speaking, the assumption made in this research was that the items to be spoken would be entered in advance, in their own time, by the user.

The interface design began with a simulation using text written on index cards. In this simulation, a researcher took the role of a non-speaker, and prepared on file cards plausible contributions to a conversation on 'job interviews'. Conversations were then conducted with volunteers by means of the researcher pointing at a card whenever she wished to communicate what was on that card. The content of the card was then read out by another volunteer. Repeated runs of this system lead to the addition, deletion, and alteration of cards as problems became apparent through breakdowns in the conversation. The participants reported that they found the conversations reasonably natural, and control passed satisfactorily between the card user and the conversation partner.

LESSONS FROM CONVERSATION ANALYSIS

In addition to lessons learned from the above simulation, the interface design was also based on a number of aspects of conversational structure which have been established by conversation analysis research. The aspects which were incorporated were the predictability of opening and closing sequences, the alteration of speaking turns and perspectives, the orderly movement through topics in a conversation. and the importance of rapid and continuous feedback to the other speaker.

Opening and closing sequences

Opening and closing sequences are an important and frequently performed conversational feature. People do not just start talking about something with each other. They work at getting started, and they work at finishing the conversation gracefully [8,15]. If steps are left out of these sequences, the impression is created of rudeness, or lack of social training. A common opening sequence is GREETING & SMALLTALK. A common closing sequence is WRAPUP REMARKS & FAREWELL. For this prototype, the opening sequence was GREETING & INTRODUCTION, and the closing sequence was WRAPUP REMARKS & FAREWELL. Because the system was to be tested in a laboratory experimental situation, it was thought that introductory remarks were more appropriate than smalltalk.

Alternating perspectives

Conversation tends to consist of alternating turns, with alternating perspectives on the topic being expressed [4]. For the conversational content of the system it was decided to enable the user to have a dialogue on the subject of 'holidays'. The prototype allows the user to change perspectives and select an appropriate new set of texts by choosing three aspects of the topic of holidays: person, time, and orientation. The person aspect is either ME or YOU. The time aspects are PAST, PRESENT, FUTURE. The orientation aspects are WHERE, WHAT, HOW, WHEN, WHO, WHY. For instance, selecting ME-WHO-PAST would produce a set of candidate texts all on the subject of people who, in the past, accompanied the user on holiday, or who they met on past holidays (see Figure 1).

Topic shifts

The way speakers tend to move through topics in a conversation is governed by rules of relevance. It is usual to change topic in a 'step-wise' fashion, by small incremental moves [14]. The system presents the user with a set of candidate texts for speaking, which have been selected on the basis of the perspective chosen. The user may speak the texts in sequence, which produces a coherent narrative, or select and speak any text or any sequence of texts within the chosen set. If the user wishes to change topic, the selection of a new choice for one of the aspects will bring up a new set of texts, which share the other two aspects with the previous set. The user is free to reset two, or all three aspects if they wish.

Feedback remarks

Rapid and continuous feedback from listener to speaker is important in a conversation. The model of an active speaker and a passive listener is a simplification. Conversation is a simultaneous creation of all the participants. The speaker relies on a continuous stream of feedback from the listeners, and will tend to modify what they are saying and how they say it accordingly [20]. This is a particularly difficult problem for physically impaired nonspeakers, who find it very difficult, or impossible, to give appropriate feedback to another speaker. (As a consequence they can be wrongly thought of as deaf, unintelligent, or not paying attention.) The importance of timing of feedback remarks suggests that it is better to say something which is more or less appropriate than to take too long to say a precise remark. The system thus incorporates a feature developed in an earlier experimental prototype, which gives immediate output of randomly selected exemplars of a selected speech act [2]. For instance, selecting AGREE would result in phrases such as "I"m with you on that.', and 'Yes, I agree with you there.' Trials of the earlier prototype suggested that, with speech acts such as feedback remarks, users might accept a trade-off between speed and specificity [2]. This is not surprising, since, for many non-speaking people, the communication rate problem simply precludes them from using these speech acts effectively.

IMPLEMENTATION

The system was developed in Hypercard on a Macintosh computer with an external speech synthesiser. The intention was to have as simple a control procedure as possible, thus leaving the user free to concentrate on the interaction and not spend a great deal of time looking at the computer screen. Two types of user actions were envisaged : to speak, and to request more choices. The user employs a mouse to point and click over the relevant areas of the screen. An example of the interface in use is shown in Figure 1. When the user clicks on any of the text windows in the centre of screen the displayed text is spoken through the speech synthesiser. The displayed texts are replaced when the user selects a new perspective from the left part of the screen. On the right are a set of feedback

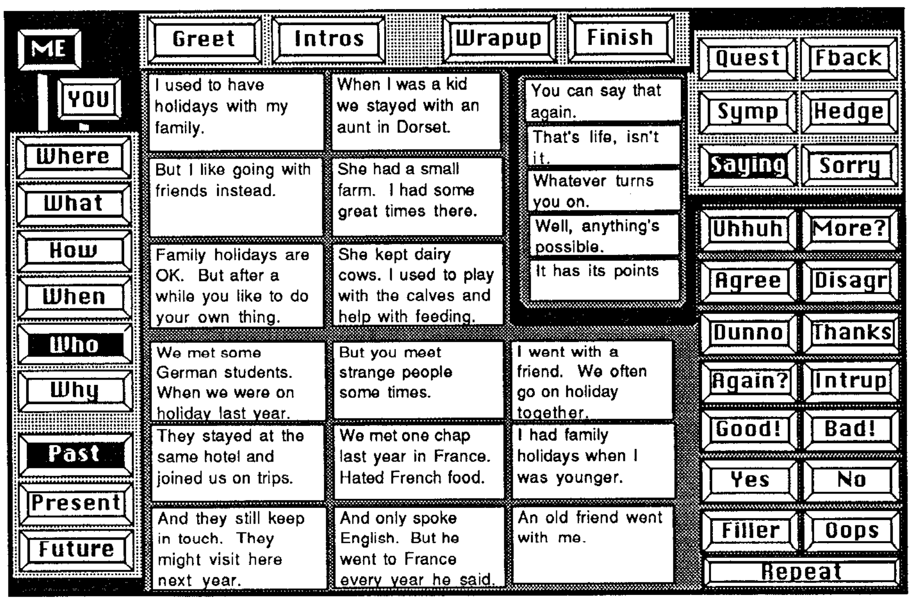

Figure 1 : Example of control screen in use

remarks. When one of these is selected, a suitable speech act of that category is spoken immediately through the speech synthesiser. Opening and closing sequences are selected from the top area of the screen. Also available are comments, which are intermediate between very general feedback remarks and fully specific content remarks. Thus, comments are phrases which are more effective if selected depending on the immediate context. A quick feedback remark might be to THANKS (e.g. 'Thanks very much.'). A comment might be a SAYING (e.g. 'Well, anything's possible, I suppose').

EVALUATION WITH AN ABLE-BODIED USER

The purpose of the initial evaluation was to determine the feasibility and observe the process of using such a system in conversation. Evaluations were done in the first instance with one able-bodied user and one physically impaired non-speaking user.

The system was first tested by a researcher loading it with conversational material about holidays. In all 1600 conversational items were put into the system. The items were stored in structure which mapped the categories of utterance described above, i.e. they were stored as openers/closers, comments, quick feedback remarks, and substantive contributions grouped by perspective.

Conversations were held with eight student volunteers on the subject of holidays. No time limit was put on the conversations. They varied from 5 to 15 minutes in length. For comparison purposes three conversations were held between two student volunteers talking about the same subject. All conversations were audio recorded and transcribed.

Below is an extract from one of the computer-aided conversations :

Computer-aided speaker :
I went to France last year, to Marseilles.
Natural speaker :
I've never visited Marseilles. I've sort of driven round the outskirts, but never actually gone into Marseilles.
Computer-aided speaker :
Surprisingly, it's really beautiful.
Natural speaker :
Really? I just imagine it as sort of a port, and just like any other large city, with nothing particularly interesting.
Computer-aided speaker :
You expect a major port to be fairly grotty, don't you ?
Natural speaker :
[LAUGH] That's right [LAUGH].
Computer-aided speaker :
We also visited other places on the coast, but we decided to give St. Tropez a miss.
Natural speaker :
That's one place I'd like to go.
Computer-aided speaker :
I've heard it's pretty grotty now, and crowded.
Natural speaker :
Really ? Oh, well, maybe give it a miss then.
Of course, there's always the chance to see Brigitte Bardot there.
Computer-aided speaker :
Whatever turns you on.

In the person-person conversations, the mean rate of speech was 144.4 words per minute (standard deviation = 4.9). The mean rate in the computer-aided dialogues was 88.2 words per minute (standard deviation = 20.6), with the individual rates for the computer user and her partners being 67.4 words per minute (standard deviation = 14.7) and 132.9 words per minute (standard deviation = 22.2) respectively. For the computer-aided dialogues, the mean length of utterance by the computer user (mean = 6.1 words) and that of her speaking partners (mean = 6.7 words) did not differ significantly in a one-tailed related t test ($t\ (df=7) = 0.74$; $p > 0.05$). In order to ensure that the speaking partner was not accommodating to the computer-aided speaker, a comparison was made between the normal speakers' mean utterance length in the dialogues with a computer-aided speaker (6.7 words) and in the person-person dialogues (6.0 words). An unrelated t test showed no significant difference between these means ($t\ (df=12) = 0.85$; $p > 0.05$).

The interface was designed so that the number of mouse clicks needed to output an utterance varied from 1 to 4. The minimum number occurred when the desired utterance was on the screen and merely had to be clicked on to be spoken. The maximum number occurred when the user made changes to all three aspects of the perspective and then clicked on a displayed text. The usage statistics showed that selections involving more than two mouse clicks were rare (2 out of 422 total selections).

In debriefing interviews, the volunteer conversationalists evaluated the communication system and the competence of the user positively and several expressed surprise at how normal and satisfactory the conversation had seemed, apart from the use of an artificial voice and the longer pause times preceding computer-user turns at speech (median value = 3.4 seconds). Preliminary analysis of ratings of the content of the conversations, obtained in a follow-up experimental study, suggests that the content of the computer-aided conversations was viewed at least as favourably as that of the person-person conversations. Further support for the 'normality' of the computer-aided conversations comes from sequential dependency analyses showing considerable similarity between the 'speech act structures' of the aided and unaided conversations (Chi-square (df=1) = 122; $p < 0.0001$) [19].

EVALUATION WITH A PHYSICALLY IMPAIRED NON-SPEAKING USER

Having established that holding a satisfying conversation with prestored material was possible, a version of the system was tested by a non-speaking physically impaired volunteer. The volunteer was a 20-year old young man with cerebral palsy, who had been non-speaking from birth. This person had a number of communication methods which he employed to try to convey his meaning. His primary means of communication was a word board with 400 words which he pointed to. He augmented this with some gestures and a few vocalisations for simple responses like 'Yes' and 'No'. He also had a portable speech output word and phrase storage device, which he used primarily for names of people and places and specialised words which were not on his board.

He was able to operate a mouse and to make single clicks with it. (For physically-impaired people who have more difficulty operating a mouse, a number of adaptations and emulators are available). In common with many people who have had no expressive language abilities, however, he had problems with reading. This meant that the interface needed to be simplified somewhat for ease and speed of use. A smaller number of text windows was used, and the perspective switching feature was changed to a simpler range of buttons to change to another topic.

The system was tested with this person not as a replacement, but in addition to his other devices and methods. The intention was to measure its effectiveness as a means to include longer narratives and conversational contributions than were possible with his existing methods, while leaving him free to use these as much as necessary. A series of conversations were held on the subject of a trip abroad for a disabled swimming competition which the person had done in the previous year. He held six conversations using his normal communication methods, and six with the addition of the system. Of his conversation partners, half were known to him, and the other half were strangers, who had never communicated with a non-speaker before. They were distributed equally between the system in use and system not in use conditions. A single-case study design was employed with an ABAB sequence, to allow for any learning effects. The conversations each lasted 15 minutes. They were all video-recorded and transcribed. Video recording was necessary in order to capture the non-speaker's gestural communication. All utterances in the transcriptions were coded using a classification system derived from Prutting and Kirchner [13]. Of particular interest were 'Initiators' - utterances which indicated a control over the interaction, which is normally difficult for non-speakers, and 'Responders' - responses to the other speaker's utterances.

The results were encouraging. It seemed subjectively clear that the non-speaker was making good use of the system during the times it was available, and that this had the effect of helping him to take a fuller part in the conversations. The objective measures, from the coded transcripts bore this out. The results are shown in Table 1. The mean number of words spoken by the speaking partners in the conversations is about the same, whether the system is in use or not. Using the system increases the non-speaker's mean number of words nearly 3½ times. Conversational control, as measured by the balance between 'Initiators' and 'Responders', also improved when the system was being used. The improvements in number of total words used and number of Initiators used were statistically significant

One unexpected outcome of the research was that the non-speaking volunteer has used the prototype system as a lecturing giving aid (Giving a talk, followed by answering questions from the audience). Using the system, he has given lectures in Scotland, the U.S. and Canada on the subject of his experiences as a non-speaking person. He has also been interviewed on local radio using the system.

	Non-speaker without system	Non-speaker with system
Mean number of words		
Natural speaker	917	999
Non-speaker	143	534
Mean number of initiators		
Natural speaker	48	50
Non-speaker	10	27
Mean number of responders		
Natural speaker	19	36
Non-speaker	87	71

Table 1 : Performance with and without the computer-aided conversation system

CONCLUSIONS

This work has indicated that it is possible to hold reasonably satisfying conversations using a computer-based communication system with prestored material, if appropriate material is available to the user in a suitable form. By applying principles derived from conversation analysis, a prototype system has been developed which helps a user to take part in a conversation in a way which is a marked improvement over existing systems, both in terms of amount of participation, and degree of conversational control exerted. The next step in this research will be to expand the conversational range of the system to cover multiple topics. In order to help the user cope with what will become a very large store of material, predictive techniques will be explored which will offer the user the continual opportunity to move out of the current topic area into other parts of the conversational database.

ACKNOWLEDGEMENTS

This research was supported by a grant from the research Initiatives Committee of Dundee University, and by a charitable donation from the Digital Equipment Corporation.

REFERENCES

1. Alm, N. Towards a conversation aid for severely physically disabled non-speaking people. Ph.D. Thesis, Dept. of Mathematics and Computer Science, University of Dundee, Dundee, Scotland, U.K., 1988.

2. Alm, N., Arnott, J. L., Newell, A. F. Prediction and conversational momentum in an augmentative communication system. *Communications of the ACM*. 35,5 (May 1992), pp 46-57.

3. Beattie, G. *Talk -- An Analysis of Speech and Non-verbal Behaviour.* Open University Press, Milton Keynes, 1983.

4. Duncan, S. Some signals and rules for taking speech turns in conversations. *Journal of Personality and Social Psychology*. 23, 2 (1972), pp 283-292.

5. Foulds, R. Communication rates for non-speech expression as a function of manual tasks and linguistic constraints. In *Proceedings of the International Conference on Rehabilitation Engineering* , (Toronto, Canada, 16-20 June). Rehabilitation Engineers Society of North America, Toronto, 1980, pp 83-87.

6. Gumperz, J. *Discourse Strategies.* Cambridge University Press, London, 1982, p 137.

7. Kraat, A. *Communication Interaction Between Aided and Natural Speakers: A State of the Art Report.* Canadian Rehabilitation Council for the Disabled, Toronto, 1985, p 82.

8. Laver, J. Communicative functions of phatic communion. In Laver, J. *Semiotic Aspects of Spoken Communication.* Edward Arnold, London, 1974.

9. Light, J. Interaction involving individuals using augmentative and alternative communication systems: state of the art and future directions. *Augmentative and Alternative Communication.* 4, 2 (June 1988), pp 66-82.

10. Macmillan, W. Computing for users with special needs and models of computer-human interaction. Proceedings CHI, 1992, (Monterey, California, May 3 - May 7, 1992) ACM, New York, 1992, pp 143-148.

11. Marcus, A. and van Dam, A. User-Interface Developments for the Nineties. *Computer*. 24,9 (1991), pp 49-57.

12. Newman, H. The sounds of silence in communicative encounters. *Communication Quarterly.* 30, 2 (Spring 1982), pp 142-149.

13. Prutting, C. and Kirchner, D. Applied pragmatics. In Gallacher, T. and Prutting, C., (eds.) *Pragmatic Assessment and Intervention Issues in Language.* College Hill Press, San Diego, 1983.

14. Sacks, H. Lecture 5, Spring 1972. University of California at Riverside, pp 15-16. Cited in Atkinson, J. and Heritage, J., (eds.) *Structures of Social Action - Studies in Conversation Analysis.* Cambridge University Press, London, 1985, p 198.

15. Schegloff, E. and Sacks, H. Opening up closings. *Semiotica*. 8 (1973), pp 289-327.

16. Scherer, K. Personality markers in speech, in Scherer, K., ed. *Social Markers in Speech.* Cambridge University Press, London, 1979.

17. Stubbs, M. *Discourse Analysis -- The Sociolinguistic Analysis of Natural Language.* Basil Blackwell, Oxford, 1983.

18. Suchman, L. *Plans and Situated Actions*. Cambridge University Press, London, 1986.

19. Todman, J., Elder, L., Alm, N., File, P. Sequential dependencies in computer-aided conversation. *Journal of Pragmatics*. Under review.

20. Yngve, V. On getting a word in edgeways. *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, Chicago Linguistic Society, Chicago, 1970.