Proceedings of the Institute of Acoustics

MODELLING CONVERSATION PATTERNS TO IMPROVE AUGMENTED COMMUNICATION FOR NON-VOCAL PEOPLE

Norman Alm(1), John Todman(2), Leona Elder(3) & Iain Murray(1)

(1) MicroCentre, Department of Mathematics and Computer Science, University of Dundee
(2) Department of Psychology, University of Dundee
(3) Department of Business Studies, University of Abertay Dundee

1. INTRODUCTION

People who are unable to speak through physical impairment face formidable difficulties in communicating, even with the help of synthetic speech output devices. Typical speaking rates with current systems are 2-10 words per minute, or even less, as measured against the 180 words per minute average of natural speakers [12]. One way to increase communication rate would be to allow the user to store and recall a large number of reusable texts. This method has thus far proved impracticable because of the difficulty users have in remembering what they have stored, and how to access it, within a conversational situation where a quick response is essential for success.

2. MODELLING COMMON CONVERSATIONAL FEATURES

A possible help in increasing the effectiveness of text storage augmentative communication systems would be to have them contain models of conversation patterning, in order to provide the user with predicted sequences and likely predictions for the next thing to say, drawn from the stored texts. We have developed a number of research prototypes based on different aspects of conversational patterning, in order to examine their effectiveness in improving computer-aided communication.

Speech does seem to be infinitely variable. However, it does not follow that it is totally unpredictable. As Fillmore says : 'an enormous amount of natural language is formulaic, automatic, and rehearsed, rather than propositional, creative, or freely generated' [9].

If a complete conversation is considered, at the most simple abstract level it can be said to have three components, in the following order :

(1) Opening the conversation
(2) Conducting topic discussion, and
(3) Closing the conversation

The process of opening a conversation has the following possible elements, in the order given below [13,18] :

(1) Bid for attention
(2) Verbal salute
(3) Identification
(4) Personal inquiry, and
(5) Smalltalk

Just as in opening a conversation we employ a predictable routine, we also tend to follow a set procedure to close the conversation. As Schegloff and Sacks state it, a conversation 'does not simply end, but is brought to a close' [19].

The basic elements, in order, of closing a conversation are [11,19] :
(1) Transition signals
(2) Exchange of phatic remarks, and
(3) Exchange of farewells

Continuous feedback from the listener to the speaker is important in creating the rapport necessary for conversation. This feedback can, as with fillers, merely denote continued attention. It can also convey more information, such as agreement, puzzlement, amusement, shock, or other reactions [23]. It is partly because physically impaired non-vocal people have difficulty in conveying this sort of feedback that they are often mistaken by unfamiliar people as being deaf, uninterested, or unintelligent.

2.1 The CHAT prototype -- opening, closing, and giving feedback

A prototype communication system, called CHAT, has been developed, which allows a user to produce opening/closing sequences and feedback remarks easily. CHAT is an acronym for Conversation Helped by Automatic Talk. The intention of the design is to have the system automate as much of the conversational process as is feasible, and in this way reduce the keystrokes necessary to operate it, substantially increasing the rate of conversational participation that is possible by a non-speaking person.

As well as opening and closing sequences, a range of feedback remarks has been included in CHAT, all of which are available for the equivalent of one keystroke. Having only one keystroke per speech act is particularly important when giving feedback to another speaker, since timing here is important for the remark to have its effect. The feedback remarks were chosen to cover as wide a range of situations as possible.

CHAT operates at the level of a speech act, and not a specific utterance, and is thus able to provide automatically a variation in output, simulating what unimpeded speakers do in avoiding clumsy repetition. The user can opt for CHAT's predicted speech act, or direct it to output another type of speech act. Where CHAT is unable to help, the user always has the option to create unique text (at, of course, a much slower rate).

An extract from the dialogues carried out using CHAT is given below :

CHAT user :	So, what about this weather recently ?	(*Keystrokes needed to produce utterance* : 1)
Conversation partner :	I know, it's awful isn't it. Wish it wasn't quite so bad. It was snowing when I came out of school today.	
CHAT user :	Yeah. Well...	(*Keystrokes needed to produce utterance* : 1)
Conversation partner :	Yeah.	
CHAT user :	All is well with you I trust.	(*Keystrokes needed to produce utterance* : 1)
Conversation partner :	Yeah, except for this cold. I wish I could get rid of it.	

In trials the CHAT system was found to provide a much faster method of generating this type of conversational material than existing devices. The results from trials of CHAT with physically impaired non-speakers, showed words per minute rates ranging from 19 to 54, compared with the current achievable rates of 2-10 words per minute. Users and conversation partners reported generally positive impressions of the conversations produced using CHAT [3]. This system has been incorporated into a communication device for non-vocal people, called Talk:About^{TM} which is now commercially available [25].

2.2 Approaches to topic discussion

Although topic discussion contains some predictable sequences and routines, it has a much less obvious structure than the opening and closing sections of a conversation. Research in discourse analysis has only just begun to map out the wide and varied area of topic discussion [10, 14, 22]. It may be possible however to take Fillmore's view about the amount of non-original material included in everyday conversation as a starting point, however, and to conclude that what can be predicted about topic discussion is that it is often repetitive.

One method therefore which may be of help to increase conversational effectiveness of non-speaking users of communication systems is to develop a technique for storing and retrieving large amounts of reusable conversational texts. The method developed would need to have the 'intelligence' to offer the user suggested ways of navigating through this text for producing conversation, while at the same time providing a very simple interface, which does not impede the user from the main task at hand, which is social interaction. To examine the feasibility of this approach we have been carrying out a number of different studies. Each involves the development of a prototype system which accesses a large database of prestored conversational material. This requires the participation of volunteers who have created these personal conversational databases for testing purposes. We have been working both with able-bodied and non-vocal physically impaired volunteers.

One characteristic of conversation is moving amongst various perspectives of the topic currently being discussed, and then moving in a coherent way to another topic. Topic movement is normally handled in a step-wise fashion by conversationalists [7]. In this way, the coherence of the conversation is cooperatively maintained, while each speaker is able to take part in controlling the conversational content. For the non-speaking person using communication system, this is particularly problematic. It has been shown that a great deal of such communication consists of single word or short phrase responses, where the topic has been set by the unaided speaker [15]. Where the communication system user might like to introduce a new direction in the conversation, the time it takes to create and output a suitable contribution usually means the conversation has moved on.

Here we describe three methods which are currently under investigation as possible ways to structure a topic movement facility in an augmentative communication system. The methods are modelling perspective switching in a conversation, and using two techniques for modelling topic-change: hypertext and fuzzy information retrieval.

2.2.1 Perspective switching. A prototype system, called TALK is being developed for the purpose of exploring a number of ways to improve augmented communication, including automatic perspective switching within a topic. We have attempted to provide for transitions between perspectives in two ways. First, the user can select a "Me" or "You". Second, choices are displayed for the selection of a "Time" perspective (Past, Present, Future) and an "Orientation" perspective (Where, What, How, When, Who, Why). On the screen, current menu selections are highlighted, making it possible to select any combination of perspectives with a maximum of three mouse clicks.

Although a deliberate choice of conversational perspective - as with preplanning possible conversational items - is not an activity that natural speakers need to do very often, it is part of the "overhead" needed to operate this system. It seems unavoidable that some such effort would be necessary for users of a communication system. This prototype, in fact, attempted to close the gap between current systems, which require conscious planning at the word and letter level, and natural speaking, in which many of the details of speech organization and production are not at a conscious level.

Trials of the TALK prototype were conducted with an able-bodied person using it to communicate to a number of other people on a set topic. Conversations on this topic were also held by pairs of natural speakers for comparison. All conversations were transcribed and coded. An extract from each is given below :

Two natural speakers

Speaker 1 : They don't seem to have heard of speed limits.

Speaker 2 : Oh that's the least of their problems I think. I don't think they've heard of junctions.

Speaker 1 : Or rules about how to overtake.

Speaker 2 : Yes, this is true.

Speaker 1 : They say the food is a lot better than the driving.

Speaker 2 : Yes, yes. The food was good when we went. It was excellent. Sitting in the cafes, it was nice.

Speaker 1 : That sounds great.

One speaker using TALK

TALK user : I went to France last year, to Marseille. (*Mouse clicks needed to produce utterance* : 2)

Conversation partner : I've never visited Marseille. I've sort of driven round the outskirts, but never actually gone to Marseille.

TALK user : Surprisingly, it's really beautiful. (*Mouse clicks needed to produce utterance* : 1)

Conversation partner : Really? I just imagine it as sort of a port, and just like any other large city, with nothing particularly interesting.

TALK user : You expect a major port to be fairly grotty, don't you. (*Mouse clicks needed to produce utterance* : 1)

Conversation partner : [LAUGH] That's right. [LAUGH].

In the case of the person using the TALK system, conversation partners reported that the conversations seemed natural, and that a favourable image was projected by the TALK user. The pause times needed by the TALK were significantly reduced from those usual with users of communication systems, and a speaking rate of about 67 words/minute was achieved. Of course, it must be taken into account that the eventual users of such a system will have physical impairment which will slow down their activation rate of the system, but the even with this taken into account, the improvement offered was of such an order that it would still be significant.

Comparisons of the computer-aided conversations with the natural conversations showed that the average length of time spent by each speaker in conversation was about the same. Given the general lack of participation which characterises most augmented communication, this was encouraging. An interesting feature of the TALK user's conversational style was the use of topic changes as a method of repairing conversational problems. The TALK user in fact made more topic changes than the conversation partners. This was also encouraging because it indicated that a system such as this might be able to offer increased conversational control by a non-speaking person, in an acceptable and conversationally interesting way [5, 20, 21]

2.2.2 Hypertext structures. Hypertext is a method for storing and navigating through information which purports to be based on the manner in which the human mind stores and accesses information, using highly flexible associative links [8]. A hypertext system consists of nodes, each of which can be linked to any other node to create an association between them. Any cross-referencing in documents can be considered as a prototype of hypertext, but the provision of a rich network of such associations on a computer with interactive capabilities gives hypertext its real character. Hypertext could thus be a good candidate for hosting a text-based conversation aid, by modelling the flexible way in which the mind stores and recalls conversation items [2].

Hypertext's capacity for creating associative links between items could also be useful in modelling parts of the conversational process itself, thereby assisting the augmented speaker by anticipating their conversational needs. In topic based conversation, a participant moves to another topic in the conversation, by making either a 'boundaried' or a 'step-wise' topic shift [7]. A boundaried shift is made with key phrases which indicate that an abrupt change to an unrelated topic is about to be made. A step-wise move is more common, whereby a move is made to a topic which is clearly related to the present one. In an augmentative communication system, boundaried shifts of topic may be assisted with a store of ready remarks appropriate to this sort of move. We are exploring the potential of a hypertext system for the more difficult task of helping an augmented speaker to effect step-wise topic moves efficiently within a store of conversational material.

As a first step, a prototype system has been developed with the working name 'Floorgrabber', one of its intentions being to increase the user's conversational control [4]. The interface design was a cooperative effort between one of the authors and a non-speaking person. This person also was the user of the system in the trials. The interface consists of text boxes and on-screen buttons which are activated by pointing and clicking with a mouse. Three types of buttons are used, which have the effect of (1) speaking the text in the box pointed to (2) speaking a quick comment, (3) going to another topic. The text choices on the screen are all on a particular topic. They could be spoken in sequence (thus 'holding the floor') or could be used individually, or in any subsequences as appropriate. Moving to another topic is accomplished by clicking on the appropriate button. Topics could be chosen in any sequence, or the user could move on to the next closest topic to the current one.

The system was trialled by the non-vocal volunteer, in a series of conversations with a variety of conversation partners, which were videotaped and transcribed. An extract is given below. Note that the Floorgrabber user continues to employ other methods of communication, using Floorgrabber to produce an extended utterances :

Floorgrabber user :	I went to Miami, to compete in the World Youth Games.
Conversation partner :	Uh-huh.
Floorgrabber user :	This was an international event, for disabled athletes.
Conversation partner :	Uh-huh.
Floorgrabber user :	There were about 500 people from all over the world competing. I think there were 14 different countries represented.
Conversation partner :	So, there was, like, loads of people speaking different languages?
Floorgrabber user :	Eh. (*Vocalisation*)
Conversation partner :	God, I couldn't handle that, I'd ...
Floorgrabber user :	Ehhh ... (*Vocalisation*) (*Gesture: hand-waggle*)
Conversation partner :	Was it all right ?
Floorgrabber user :	Eh. (*Vocalisation*) There was a British squad, but there was also our Scottish team.

Analysis of the transcripts showed that, when the prototype was added to the user's communication modes, he was able to increase the total number of words he used in each conversation to a significant degree. The output of the other speakers was unaffected, if anything being slightly higher, which indicates that the system user having the ability to introduce text did not create more passive behaviour on the part of the other speaker.

Conversational control by the system user was also increased, as measured by his increased use in initiators and decrease in responders. Again, the natural speakers retained their level of initiators even when the system user increased his, indicating a dialog which was in general more lively.

2.2.3 Fuzzy information retrieval. A useful technique for a predictive conversation system would be to simulate natural topic movement through a conversation by means of a retrieval system which allowed for an imprecise command such as "Find a conversational text which is something like the present one". In this way the system would provide sensible predictions as to the next text to be spoken, and, by taking a great deal of the cognitive burden off the user, and allow them to participate more naturally in an interaction. We have made a start at examining the practicability of designing such a retrieval system for a conversation helper based on fuzzy set theory.

Fuzzy set theory is an extension to conventional set theory which is particularly appropriate for modelling vaguely defined systems where it is not possible to classify the components of the system into discrete sets [24]. The theory is mathematically rigorous, but takes as its starting point the central concept that membership of any set, instead of being a binary property (yes or no, 1 or 0), is describable as a real number between 0 (definitely not in the set) and 1 (definitely in the set). Thus, instead of 'x belongs to set Y', we might have 'The membership value of x for the set Y is 0.146', which gives a relative value to how strongly x belongs to set Y.

This theory has been applied successfully in control systems of various sorts [16]. If applied to an information retrieval system, the theory allows for more flexible storage and retrieval methods. The similarities between items in the database can be captured without the need for similar items to share a number of descriptors from a given set. From the point of view of a conversational database, a fuzzy set retrieval system has the advantage that, given one item, it will always produce a set of the most similar items in the database. It will never return from a search with no items found [17].

In order to test out the feasibility of using fuzzy set retrieval methods in a conversation aid, a prototype system was developed [6]. For labelling the stored items, two types of descriptor were used. Based on previous work with a text database, the descriptors either represents the semantics of an item (its subject) or the pragmatics (its purpose in a dialogue, i.e. speech act) [1]. The system uses eight subject descriptors (travel, music, sport, driving, communication, work, family, friends) and five purpose descriptors (opening, elaboration, question, joke, conclusion). Each item in the fuzzy set database has a vector associated with it which described its degree of belonging to these thirteen categories.

A number of experiments were carried out to determine an efficient and accurate way to assign these values. The optimum design was an analogue display on the computer screen (a sliding control) with which values could be set for each item. It is expected that the values set will always have a certain degree of subjectivity. This is acceptable, however, and even a desirable characteristic, in a system which should model the conversational style of the user.

To evaluate the performance of the system against an equivalent system based on boolean search database retrieval methods, a version of the system was created which used the same stored text items but which depended on conventional database searching to compare stored items. As expected, the conventional system often produced no texts which matched a given text, whereas the fuzzy set system always produced a full set of candidate texts.

Fewer mouse clicks were needed to produce the conversation using the fuzzy set system (20% to 35% fewer). Occasionally, the conventional database system required a large number of input activations (e.g. mouse clicks) to produce an utterance, whereas the fuzzy set system tended consistently to need an average of about two. Also there was a difference in the cognitive task involved. With the conventional system, the user had to form a search plan and execute it. The fuzzy set system simply presented appropriate material automatically.

Two examples of the system in action are given below. In the first it is used to produce a monologue. This shows that coherent stepwise topic moves can be made with minimal activations from the user. The second extract is of the system being used in a dialogue with a natural speaker.

Extract from monologue produced with the fuzzy system
(Each utterance preceded by the number of mouse clicks needed to produce it)

1: The system which I am using to speak is an experimental one that I am developing.
4: Much of the work at the MicroCentre concerns helping people with special requirements use computers.
2: I came to Dundee University to do an M.Sc. in Computer Science, and enjoyed it so much I decided to stay on.
2: Computers keep developing so fast. It's really odd working with equipment that five years ago would have been pure Science Fiction.
2: Jean, one of my flatmates, works on really high tech equipment at the MicroCentre. Things like Eyetrackers
2: I share a flat in St. Andrews with two other people.
1: Liz, one of the people with whom I share a flat, is also my fiancée.
1: My fiancée studied Medieval and Art History at St. Andrews and is now doing a PhD.
2: I am finally getting married in June to my fiancée of the last four years.
1: Although I do not drive, both of my flatmates own cars which is often very handy.
1: One day I would like to learn to drive.
3: Before I came to Dundee University I studied philosophy and history at St. Andrews.
3: Having studied medieval history it would be nice to see the magnificent cathedrals and castles on the Continent.
1: The medieval monastery at Cluny is supposed to be absolutely amazing.
1: Have you ever been to another country?
5: What sort of music do you like?
2: Do you have a lot of tapes records or compact discs?
3: My main hobbies are reading swimming and music.
2: Do you like swimming?
2: I really like swimming. It's a sport I can do at my own pace.
2: Often the three of us in the flat go swimming together.

Extract from dialogue between fuzzy system user and natural speaker

System user:	I got a wok for Christmas so I'm experimenting with making Chinese food.
Conversation partner:	Oh, right. I haven't got a wok.
System user:	I love eating Chinese food.
System user:	What kind of food do you like?
Conversation partner:	I think my favourite food at the moment is Indian but I tend to go through phases of liking different kinds of food.

It could be objected that the 'prompting' nature of the fuzzy set system was inimical to real conversation, in that one could never know what the user might have said if working with a totally free choice. Against this, it can be argued that with stored reusable material, the order in which it is said matters less than its successful and appropriate introduction into a conversation. The criterion is thus not a direct comparison with unaided speech, but with successful versus unsuccessful communication, however accomplished.

3. SUMMARY AND CONCLUSIONS

The prototype systems described in this paper are all attempts to improve the performance of an augmentative communication system by in some ways simulating the patterns of unimpaired conversation. They are part of a larger research effort at the MicroCentre, which has as a long term goal the development of a communication system which can assist a user to converse on a number of different levels, depending on the discourse goal being pursued. Such a system should be able to deal at the one end with highly predictable, formulaic conversational moves, and at the other with unique, newly created utterances. Between these two communication modes is the type of conversation which involves reusable material. The large scale storage and retrieval of such material in a way which is suitable when the primary task is interacting with another person, and not searching through a database, is the problem addressed by these prototypes

4. ACKNOWLEDGEMENTS

This work was funded with the assistance of the European Community Social Fund, The Remedi Foundation, The Scottish Home and Health Department, the University of Dundee Research Initiatives Fund, and a charitable donation from the Digital Equipment Corporation.

5. REFERENCES

[1] N. ALM, J.L. ARNOTT & A.F. NEWELL, 'Database design for storing and accessing personal conversational material'. *Proceedings of the 12th Annual Conference of the Rehabilitation Engineers Society of North America.* J. Presperin, ed. New Orleans, Louisiana. The RESNA Press. pp 147-148 (1989).
[2] N. ALM, J.L. ARNOTT & A.F. NEWELL, 'Hypertext as a host for an augmentative communication system'. *Proceedings of the European Conference on the Advancement of Rehabilitation Technology. ECART.* Maastricht, The Netherlands. pp 14.4a-14.4b (1990).
[3] N. ALM, J.L. ARNOTT & A.F. NEWELL, 'Prediction and conversational momentum in an augmentative communication system'. *Communications of the ACM*, Vol 35 No 5, pp 46-57 (1992).
[4] N. ALM, J.L. ARNOTT & A.F. NEWELL, 'Evaluation of a text-based communication system for increasing conversational participation and control'. *Proceedings of the 15th Rehabilitation Engineers Society of North America International Conference.* Toronto, 6-11 June 1992. Washington, D.C.: The RESNA Press, pp 366-368 (1992).
[5] N. ALM, J. TODMAN, L. ELDER & A.F. NEWELL, 'Computer aided conversation for severely physically impaired non-speaking people'. *Proceedings of INTERCHI '93.* Editors S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, T. White. Amsterdam : Association of Computing Machinery. pp 236-241 (1993).
[6] N. ALM, M. NICOL & J.L. ARNOTT, 'The application of fuzzy set theory to the storage and retrieval of conversational texts in an augmentative communication system'. Binion, M. (Ed.) *Proceedings of RESNA '93.* Washington, DC : The RESNA Press. pp 127-129 (1993).
[7] G. BUTTON & N. CASEY, 'Generating topic: the use of topic initial elicitors'. In Atkinson J, and Heritage J., *Structures of Social Action - Studies in Conversation Analysis.* London. Cambridge University Press. pp 167-190 (1984).
[8] J. CONKLIN, 'Hypertext: an introduction and survey'. *IEEE Computer.* Vol 2, No 9. pp 17-41 (1987).
[9] J. GUMPERZ, *Discourse Strategies.* Cambridge University Press, London, p133 (1982).
[10] R. HOPPER, 'Interpretation as coherence production', in R. CRAIG & K. TRACY, (eds.) *Conversational Coherence -- Form, Structure, and Strategy.* Sage Publications, Beverly Hills and London (1983).
[11] M. KNAPP, R. HART, G. FRIEDRICH & G. SHULMAN, 'The rhetoric of goodbye : verbal and non-verbal correlates of human leave-taking'. *Speech Monographs.* Vol 40, pp 182-198 (1973).

[12] A. KRAAT, *Communication Interaction Between Aided and Natural Speakers: A State of the Art Report.* Toronto: Canadian Rehabilitation Council for the Disabled. (1985).
[13] J. LAVER, ' Communicative functions of phatic communion'. In J. LAVER, *Semiotic Aspects of Spoken Communication.* Edward Arnold, London (1974).
[14] S. LEVINSON, *Pragmatics.* Cambridge University Press, London (1983).
[15] J. LIGHT, 'Interaction involving individuals using augmentative and alternative communication systems: state of the art and future directions'. *Augmentative and Alternative Communication.* Vol 4, No 2, pp66-82 (1988).
[16] E. H. MAMDANI 'Application of fuzzy set theory to control systems : a survey', In M. GUPTA. G.N. SARIDIS, B.R. GAINS (Eds) *Fuzzy Automata And Decision Processes*, North-Holland, New York (1977).
[17] C.V. NEGOITA, 'On the application of fuzzy set theorem for automatic classification in information retrieval systems', *Information Science*, Vol 5, pp279-286 (1973).
[18] M. ROSNER, 'Three strategic goals employed in conversational openings'. Working Paper No. 46. Fondazione Dalle Molle, Institut pour Les Etudes Semantiques et Cognitives, Universite de Geneve, Geneva (1981).
[19] E. SCHEGLOFF & H. SACKS, ' Opening up closings'. *Semiotica.* Vol 8 , pp289-327 (1973).
[20] J. TODMAN, N. ALM & L. ELDER, 'Computer-aided conversation: a prototype system for non-speaking people with physical disabilities'. *Applied Psycholinguistics.* Vol 15, pp45-73 (1993).
[21] J. TODMAN, N. ALM, L. ELDER & P. FILE Sequential dependencies in computer-aided conversation. *Journal of Pragmatics.* Vol 21, pp141-169 (1994).
[22] W. VON HAHN, 'Pragmatic considerations in man-machine discourse'. In *Proceedings of the Eleventh International Conference on Computational Linguistics (COLING 86).* Association for Computational Linguistics, Bonn, pp520-526 (1986).
[23] V. YNGVE, 'On getting a word in edgewise'. Papers from the Sixth Regional Meeting of the Chicago Linguistic Society. Chicago. Chicago Linguistic Society (1970).
[24] L.A. ZADEH. (1965) Fuzzy sets. *Information and Control*, Vol 8, pp338-353.
[25] DON JOHNSTON INCORPORATED, 1000 Rand Road, Building 115, Wauconda, Illinois, U.S.A.