Proceedings of the Institute of Acoustics

BYPASSING COMMUNICATION DIFFICULTIES TO ALLOW SATISFYING CONVERSATIONAL PARTICIPATION BY A NON-SPEAKING PERSON

N Alm, J L Arnott & I R Murray

The MicroCentre, Dept. of Mathematics and Computer Science,
The University, Dundee, DD1 4HN, Scotland

1. INTRODUCTION

Synthetic speech is now widely used by severely physically impaired non-speaking people, but the problem remains of grossly unequal balance in conversational participation and control between non-speaking users of speech output systems and natural speakers. The difficulty is that lack of speech is normally accompanied by a general lack of motor control, so that even operating customised input switches is a slow and laborious procedure for the person using a speech output system. This puts a ceiling on the achievable input rate to such a system.

Current computer based communication devices for non-speaking people operate on a letter-by-letter or word-by-word basis. While unimpeded conversation proceeds at 120-200 words per minute [5], the best that users of even the latest computer-based communication system can achieve is 2-10 words per minute [8]. The effect is that even with the latest available technical assistance, severely impaired non-speaking people are significantly frustrated in their ability to take a full part in the world around them. A great many of these people are not cognitively impaired in any way (such as the well-known cosmologist Stephen Hawking). This makes their predicament even more difficult.

2. APPLYING PRAGMATICS TO THE PROBLEM OF SLOW COMMUNICATION RATE

2.1 The Pragmatics of Conversation

One way to view this problem is that the non-speaker needs to have an artificial way to produce speech which imitates the ordinary speaking process. Thus far augmentative communication systems have attempted to provide this for non-speakers at the level of producing individual words. Certainly if the process is viewed as one of information transfer, then this model is adequate enough. It does, however, leave the augmented communicator with the very slow communication rate given above, and, given the ceiling on input rate to the system, it does not offer much hope for improvement.

In fact, the word-by-word model of conversation is a simplification of what actually happens when people converse. It is a simplification because communicating at 120-200 words per minute does not allow the speaker consciously to plan the speaking of each word separately. The process is too rapid for this, so a great deal of it is not under conscious control [18]. The psychological reality, for the speaker, consists of forming an intention to communicate something to another person, and putting that plan into action. Through a lifetime's training at producing speech, the details of the implementation are done almost automatically. Thus, the non-speaker with a communication system is at present operating at the relatively low level of word production, which is largely under automatic control for the unimpaired speaker. Of course, the mechanisms of moving from intention to speech production are still imperfectly understood, but their semi-automatic nature, in an unimpaired speaker, is clear.

The word-by-word model is thus an oversimplification of what actually happens when we speak. The simple information transfer model of conversation is also inadequate. Conversation is used to convey a great deal more than the propositional content of what is said. Through conversation we gain a sense of social participation, we express our personalities, we control others, we tease, entertain, and perform many interactional tasks and routines where the amount of new information involved is minimal, or zero. The development of the concept of the speech act began with this central insight [2,16], which had an antecedent in Wittgenstein's concept of language games [20]. As well as carrying information, speech must also be viewed as a behaviour which is intended to bring about certain effects, and this aspect may or may not relate to any propositional content.

It is the pragmatics of communication which takes as its subject this level of verbal interaction. Pragmatics is the study of language as it is used in context [10]. In an automatic language understanding system, for example, the pragmatic stage of analysis would follow phonetic, morphological, syntactic, and semantic stages. A sentence can be interpreted correctly up to the semantic level, and still be incorrect without a pragmatic analysis. An example of an error due to lack of pragmatic processing would be to regard the sentence ' Could you pass the salt ? ' as a request for information about your physical capabilities, rather than a polite request for action on your part. A pragmatic analysis of natural speech thus focuses on the social context of language use. Our choice of vocabulary, and the meaning which can be attached to what we say, depends on who we are addressing (e.g. child or adult), why we are speaking (e.g. to express anger), where we are speaking (e.g. in church), and so on.

In fact, adult human communicators are so skilled in pragmatics that the meaning of a verbal interaction is often conveyed by what is not said, or by something which is said which need not be made explicit, and therefore gains emphasis by being expressed. This is the powerful role of implicature in human communication [6]. An interesting and relevant example of this is the discomfort which speakers in our culture feel at the presence of silence in a conversation. In many western cultures, the intrusion of silence into an interaction conveys a negative message about the interaction. The participants are likely to feel that too much silence implies disagreement, boredom or unspoken hostility [11].

This has a direct bearing on the problem faced by severely physically impaired non-speakers, since long silences are an unavoidable feature of their slow communication rate. People who interact with them, unless they know them well, experience discomfort from this degree of silence. The result is that the number of potential interactions is reduced, since many people find this effort and stress too difficult. What is in operation here is an ingrained attitude which is difficult to alter, despite the obvious reasons for a new approach being necessary. Conversation partners of augmented speakers can train themselves out of this attitude toward silences, but it does take an effort to do so [4].

2.2 Bypassing Communication Difficulties

A study of the pragmatics of conversation suggests that a different approach to augmented communication may be fruitful. If we see the communication process as beginning with the speaker's intentions, which are carried out by means of the speech process, then current augmentative communication systems are prosthetic devices which replace the word production part of the speech process in order to help the user carry out their intentions. A more direct approach would be to assist the speaker to move straight from an intention to the delivery of an appropriate speech act to the listener, bypassing the time-consuming word production process. If conversation consists of entirely novel material which must be newly created each time people speak with each other, this approach will clearly not be feasible. The user must create all utterances word by word in that case.

However, the pragmatics of conversation, the study of conversation as it is actually used, demonstrates that 'Chomskian linguistics has often over-emphasized the creativity of everyday language. In practice, a significant percentage of conversational language is highly routinized into pre-fabricated utterances' [17]. Fillmore states this more strongly : 'an enormous amount of natural language is formulaic, automatic, and rehearsed, rather than propositional, creative, or freely generated' [cited in 7].

A research group at Dundee University's MicroCentre is investigating the application of these ideas to the design of augmentative communication systems. Prototypes have been developed which employ pragmatic considerations to effectively bypass the parts of the communication process which non-speakers find difficult, and help them to go directly from conversational intention to conversational impact, through synthetic speech output. This paper describes two approaches to this research work which have proved fruitful.

3. CONVERSATION AS PARTICIPATION

At times, it is of prime importance for a speaker to keep involved in a conversational encounter, and it is of less importance precisely what they say to accomplish this. The negative effects of silence, noted above, require that in a group conversation, participants will ensure that they all make contributions from time to time, otherwise their non-participation will be treated as conveying an unintended negative message. This can be seen very clearly in a business meeting where participants represent various working groups. It is often felt to be important that everyone 'has a say', to establish the importance of their group within the general context, and regardless of the specific content of the contribution.

Related to this are ritualistic verbal encounters, where the playing out of the ritual is felt to be important, and the precise verbal content is of less importance. Conversation analysis has demonstrated that conversational opening and closing sequences are important and frequent ritualistic encounters which are quite formulaic in their content [9,14].

The giving of verbal and non-verbal feedback to a speaker by listeners is another important and frequent conversational activity which does not demonstrate a great deal of creativity in the words chosen to express the feedback. Without this feedback, however, conversations tend to become awkward [21]. Conversation is in fact the simultaneous creation of all the participants, not just a matter of speaker and listener alternating roles [3]. The speed necessary for feedback to be relevant, and the fact that the intention (e.g. to convey continued interest) places minimal requirements on the information content of the utterance results in similar utterances being used continually (e.g. 'Uh-huh').

3.1 CHAT - Providing Predictable Conversational Sequences Automatically

A communication system has been developed which embodies the idea that a vital aspect of conversation is to keep involved, and this supersedes in importance the exercising of control over the fine details of what is said. This system was designed to provide the user with predicted opening and closing sequences, the capability for giving rapid feedback to the other speaker, and to control a speech synthesiser as close as possible to the equivalent of one keystroke per speech act [1].

CHAT is an acronym for Conversation Helped by Automatic Talk. The intention of the design is to have the system automate some key elements of the conversational process, 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.

The CHAT system was written in Pascal. It produces satisfactory real-time performance, with no processing delays. The hardware platform is a laptop PC, with a text-to-speech synthesiser. This configuration is widely available at modest cost.

CHAT provides a predictive facility, which is employed in the opening and closing stages of a conversation, and gives a useful set of feedback remarks which are presented in the topic discussion phase. The user can opt for CHAT's predicted speech act, or direct it to output another type of speech act. The choices as to what to say next are specified in terms of speech acts, not as a menu of specific utterances. Because CHAT operates at the level of a speech act, and not of specific utterances, it is able to provide automatically a variation in output, simulating what unimpeded speakers do in avoiding clumsy repetition. Where CHAT is unable to help, the user always has the option to create unique text (at, of course, a much slower rate).

The user can originate all of the stored phrases, entering them in his or her own time, when input rate is not important. The stored text can be changed or updated simply. The intention is for the phrases to have the personal stamp of the user, and to be modified over time, according to changing needs and preferences. This will help to counteract any tendency towards impersonality when an utterance is automatically selected during a conversation. Through this procedure, all the possible choices will reflect the personality of the user, even though the selection of a particular utterance is made by the computer.

3.2 Evaluation of CHAT

The CHAT system described above was implemented on an MS-DOS lap top computer with portable speech synthesiser and initially pilot-tested with non-impaired subjects. It was then trialled with four physically impaired non-speaking people. Measures taken included communication rate, number and variety of speech acts employed and users' and partners' subjective reactions, as reported in a questionnaire.

These tests were designed to determine the feasibility of the CHAT concept, and focused on the types of conversation in which the speaker can participate employing only openings, smalltalk, feedback remarks, and closings.

The CHAT system was found to provide a much faster method of generating this type of conversational material than existing devices. As has been said, naturally spoken conversation proceeds at 120-200 words per minute, whereas typical rates achieved by communication device users are 2-10 per minute. A simulation of the effect of a range of physical disability in controlling CHAT can be made by using data from a pilot study with non-impaired subjects and adding a nominal delay time for each keystroke. The word rate achievable with CHAT was calculated using this method at 12 to 85 words per minute, depending on the degree of simulated disability. The results are shown graphically in Figure 1. The results from trials of CHAT with four physically impaired non-speakers, who achieved respective average rates of 19, 28, 42 and 54 words per minute, confirmed these calculations.

The fact that all stored phrases are in the user's own style of expression ensures that whatever is said will contribute to projecting their personality. In the questionnaires following trials with non-speakers using CHAT, the conversation partners said they felt the users' personalities and moods

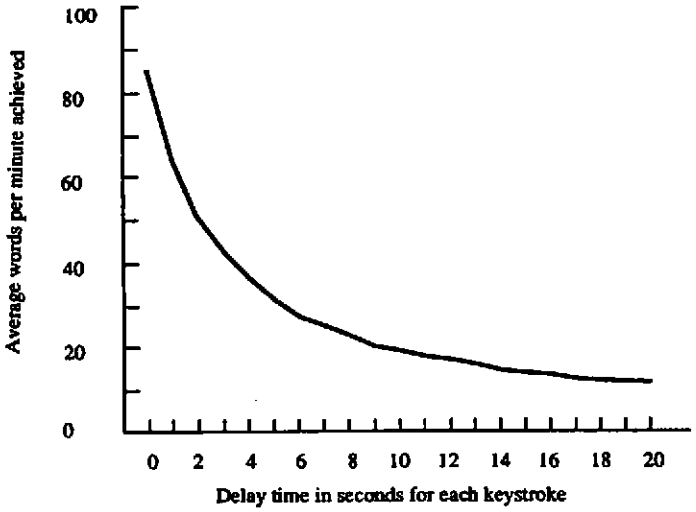

Figure 1: CHAT words per minute rates with a range of delays in keying times

came over more strongly when using CHAT compared to their usual mode of communication. The CHAT system helps a user to make conversation which sounds more natural, and therefore more acceptable to people who may not know the user well, and who may not be aware of the severe limitations on communication which the disability imposes. Even with conversation partners who are quite familiar with the user, the ability to participate more fully in normal verbal interaction was found to be of great value. The strongest negative comment in the questionnaire from non-speaking CHAT users was that it does not help them in creating topic-based discussion. At present, CHAT requires that the user employ whatever conventional methods they already use to convey unique information.

4. CONVERSATION AS PERSONALITY PROJECTION

As well as conveying information, conversation is a continual means of projecting our personality to others, and this is an essential aspect of social interaction [15]. In fact, we are so accustomed to continuously reading meaning into other people's behaviour that it is impossible not to communicate. Even a deliberate attempt to withdraw from communicating will carry a message [19]. Operating in an environment with other people involves all of us in providing others with stable impressions of our identity [15]. This is accomplished in a number of verbal and non-verbal ways, but our speech is a very important indicator of personality. Scherer has pointed out : 'If there is one clearly established finding in this field, it is the fact that people tend to agree strongly on their inferences of personality from speech' [15, p.179]. One method by which we express and confirm our personality and our point of view is through telling narratives to each other. Here, narratives is used in a wide sense, to cover stories, anecdotes, reported speech from parties not present, accounts of the activities of parties not present (i.e. gossip, both positive and negative). Not surprisingly, a great deal of conversation consists of such narratives, since they provide a way of forming and making sense of our present experience and of relating significant past experiences to each other [13].

A promising direction for augmentative communication system development would therefore seem to be helping users to introduce narratives into conversation. As well as facilitating topic discussion, and the improved expression of personality, as suggested above, an added bonus could be increased conversational participation and control merely from the ability to produce an extended conversational turn when appropriate. A communication system called Floorgrabber, which has a narrative capability, was developed to explore these possibilities. The system also included conversational features which had proved successful in the CHAT prototype.

4.1 Floorgrabber - Including Extended Texts Into Aided Conversation

A prototype narration system was developed using the Hypercard software on a Macintosh computer, with output through a text-to-speech synthesiser. The system has been given the working name 'Floorgrabber', one of its intentions being to increase the user's conversational control. The interface consists of text boxes and 'buttons' which are activated by pointing and clicking with a mouse. Three types of buttons were used, which had the effect of: (1) speaking the text in the box pointed to, (2) speaking a quick comment, (3) going to another topic.

Because the user of the system in these trials had been involved from the start in the system design, there was no need for a period of training in its use. The user was a young man of 20, who had been non-speaking from birth through cerebral palsy. He had a fair degree of controlled movement, and could just manage to operate a mouse, although double-clicking it was not possible, and it was occasionally difficult to produce only one press of the button. The software was written so that the on-screen buttons were large enough targets to manage easily. Only single clicks of the mouse were required, and the software was written to detect and ignore extra activations from tremor.

For this first experiment, the user produced textual material about one topic, a trip abroad he did for an international swimming competition. This topic was chosen because he was often asked about this interesting experience. His usual method of communication was a word chart with 400 words on it, which he pointed to, supplemented with a portable speech output device for single words and short phases, plus gesture and some vocalisations. Using these methods, he conveyed to a volunteer anecdotes and comments about the swimming trip. As he communicated, the volunteer stored this material into the system for him. The user's problems with literacy meant that this sort of mediation was necessary. To ensure that the words were truly his, the material was all checked with him several times, and modified until he was completely satisfied with it as representing the way he would like to express himself.

4.2 Evaluation of Floorgrabber

To evaluate the prototype in use, trials were conducted using a single-case experimental design. The user had 12 conversations with 12 different people on the chosen topic. The 12 sessions followed an ABAB pattern, with the baseline sessions (A) consisting of the user taking part in the dialog with his current communication methods, and the intervention sessions (B) differing by the inclusion of the prototype system as an additional mode of communication. Each conversation took 15 minutes. Half of the conversation partners were familiar with the user and had communicated with him using his current communication methods. The other half did not know him, and had no experience of communicating with an AAC user. The user's instructions were to use whatever communication mode was most comfortable and effective throughout the dialogs. The conversation partners were asked to have a 15-minute conversation with the user about his swimming trip. All the dialogs were videotaped and transcribed.

The first measure applied to the material was to assess the effectiveness of the prototype system, in

terms of its ability to help the user take a fuller part in a dialog, and to have more control over the direction of the conversation. To measure this, a count of all the words produced by each partner in the dialogs was made. Secondly, two conversational moves which were of relevance in conversational control [12] were defined as follows:

RESPONDER : An answer to a question, or a feedback comment to the other speaker

INITIATOR : A question, or a statement which is not a responder.

From the transcripts, all occurrences of these conversational moves by both partners were counted. The results are summarised in Table 1. 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, which indicates that the AAC user having the ability to introduce text did not create more passive behaviour on the part of the other speaker. Conversational control by the AAC 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 AAC user increased his, indicating a dialog which was in general more lively.

	Not Using System	Using System
Mean number of words		
Natural speaker	917	999
System user	143	534
Mean number of initiators		
Natural speaker	48	50
System user	10	27
Mean number of responders		
Natural speaker	19	36
System user	87	71

Table 1 : Results of trials of narrative telling prototype

5. SUMMARY AND CONCLUSIONS

Pragmatic considerations were applied to the problem of designing communication system for severely physically impaired non-speaking people using synthetic speech. Two prototypes examined different ways in which help might be offered. It seems that designing systems in terms of the pragmatics of conversation, which is a higher level approach than is normally the case, significant improvement can result in the rate of communication using augmentative communication systems, and in the degree of conversational participation and control possible for the non-speaker. There will, naturally, always be the need for users of such systems to create completely new utterances word by

word, with the time penalty which this implies. However, for portions of conversation which need not be freshly minted each time, a significant improvement can be achieved by helping the user proceed straight from conversational intention to conversational impact, in effect bypassing the parts of the communication process which they find difficult.

6. REFERENCES

1. 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.
2. Austin, J.L. How to Do Things with Words. Clarendon Press, Oxford, 1966.
3. Blau, A. Communication in the back-channel: social structural analyses of nonspeech/speech conversations. Ph.D. Dissertation, City University of New York, N.Y., 1986.
4. Eulenberg, J. (ed.) Conversations with Non-speaking People. Canadian Rehabilitation Council for the Disabled, Toronto, Canada, 1984.
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, 83-87.
6. Grice, H. Logic and conversation. (From William James Lectures, Harvard University, 1967). In Cole, P. and Morgan, J. (eds.) Syntax and Semantics -- Vol 3 - Speech Acts. Academic Press, New York, 1975.
7. Gumperz, J. Discourse Strategies. Cambridge University Press, London, 1982.
8. Kraal, A. Communication Interaction Between Aided and Natural Speakers: A State of the Art Report. Canadian Rehabilitation Council for the Disabled, Toronto, 1985, p 82.
9. Laver, J. Communicative functions of phatic communion. In Laver, J. Semiotic Aspects of Spoken Communication. Edward Arnold, London, 1974.
10. Levinson, S. Pragmatics. Cambridge University Press, London, 1983.
11. Newman, H. The sounds of silence in communicative encounters. Communication Quarterly. 30, 2 (Spring 1982), 142-149.
12. 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.
13. Quastoff, U., Nicholas, K., What makes a good story ? Towards the production of conversational narratives. In Flammer, A., Kintsch, W. (eds.) Discourse Processing. North-Holland Publishing Co., Oxford, 1982.
14. Schegloff, E. and Sacks, H. Opening up closings. Semiotica. 8 (1973), 289-327.
15. Scherer, K. Personality markers in speech, in Scherer, K., ed. Social Markers in Speech. Cambridge University Press, London, 1979.
16. Searle, J. Speech Acts: An Essay on the Philosophy of Language. Cambridge University Press, London, 1969.
17. Stubbs, M. Discourse Analysis -- The Sociolinguistic Analysis of Natural Language. Basil Blackwell, Oxford, 1983.
18. Vygotsky, L. Thought and Language. MIT Press, Cambridge, Massachusetts, 1962.
19. Watzlawick, P., Beavin, J., Jackson, D. Pragmatics of Human Communication. Faber and Faber, London, 1967.
20. Wittgenstein, L. Philosophical Investigations Basil Blackwell, Oxford, 1953.
21. Yngve, V. On getting a word in edgeways. Papers from the Sixth Regional Meeting of the Chicago Linguistic Society, Chicago Linguistic Society, Chicago, 1970.