

**Computer-assisted conversation for nonvocal people using prestored texts.** N.Alm & J.L.Arnott, *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, Vol.28, No.3, August 1998, pp. 318-328. DOI: 10.1109/5326.704557

## Computer-assisted Conversation for Nonvocal People using Prestored Texts

Norman Alm and John Arnott

University of Dundee, Dundee DD1 4HN, Scotland, U.K.

The research described here is published in:

*IEEE Transactions on Systems, Man & Cybernetics, Part C: Applications and Reviews*

Vol. 28, No. 3, August 1998, pp. 318-328. © IEEE. Publisher Item Id: S 1094-6977(98)03904-2.

**ISSN:** 1094-6977

**URL:** The DOI bookmark for the published article is: <http://dx.doi.org/10.1109/5326.704557>

**DOI:** 10.1109/5326.704557

**Abstract:** The challenges faced by people who are severely physically impaired and nonspeaking are demonstrated by the fact that, even with current technology, the word rates that they can achieve using current communication systems tend to range from two to ten words per minute. Ways have been investigated in which prestored reusable texts might assist in improving the performance of such systems. Beginning with a study of the acceptability and effects of including prestored texts in a conversation, this investigation examined a series of methods of increasing complexity that had the system itself perform some of the cognitive and manipulative tasks required to locate suitable texts for a user to include in conversation. These methods have included logging conversational paths for potential reuse in future interactions, using a hypertext structure for storing and navigating through conversational texts, employing fuzzy information retrieval, and building a system based on scripts, plans, and goals. Seeing the possible conversation modes as occurring at varying levels of novelty, from predictable routines through reusable texts to uniquely created utterances, the most likely area for significant improvement then is in the storage and retrieval of large amounts of reusable text. Research needs to discover ways of making the retrieval process as effortless and efficient as possible to free the user to concentrate on the interaction itself.

**Keywords:** Conversation modelling, assistive technology, assistive systems, AAC, augmentative and alternative communication, fuzzy information retrieval, HCI, human-computer interaction, nonspeaking, rehabilitation.

**Contact Address:** N.Alm / J.L.Arnott,  
University of Dundee,  
Dundee DD1 4HN, Scotland, UK.  
{nalm / jarnott} at computing.dundee.ac.uk

## 1. INTRODUCTION

Some people with physical impairments are also unable to speak. Because of the importance of interpersonal communication in human society, a person who is able to participate only marginally in daily communication finds that his or her ability to function as a human being is substantially reduced. Some notable people have made themselves an exception to this rule, but most nonvocal people report that they must often endure being considered unintelligent, uninteresting, immature, and perhaps worst of all, invisible [1,2].

Computer-based speech output systems can help many people with physical impairments to communicate, and there is an active research field developing ways in which portable, low-cost, powerful information technology can ameliorate such communication problems. In engineering terms, the problem can be considered to be one of a highly reduced output bandwidth being available to the physically impaired person. They may be able to close a switch reliably no more than a few times per minute, for example. Speech rates achievable using existing technology usually range from two to ten words per minute [3]. Alternative access methods, such as gaze monitoring, gesture analysis, and other advanced interface systems [4], have roles to play here, although none of these is a panacea for people with a physical impairment. Such impairment may reduce the user's ability to use any technique effectively (e.g., poor control of head or limb movement may reduce the efficacy of eye-gaze or gesture analysis). There is a constant search for new interface methods to give more flexibility to the clinicians and engineers who develop, adapt and prescribe systems for use by persons with disabilities [5].

## 2. CURRENT APPROACHES TO THE PROBLEM

One approach to improving this situation is to devise more efficient coding methods for the input of information into an assisted communication system. Systems are now commercially available that allow the encoding of words by abbreviations [6], multi-meaning icon sequences [7], word dictionaries with optimal search methods [8], and word prediction algorithms [8,9]. Word prediction [9-11] can reduce physical workload and, in some cases, improve word rates for a very slow typist or word-processor user. Cognitive load is an important issue in this context. The cognitive load involved in searching and responding to word prediction lists is such that only very slow users are likely to experience any rate increase in writing text. Researchers are investigating the modeling of human performance to study this aspect of the use of prediction [12]. Word prediction has other advantages for many users, however; it can improve spelling and composition for some persons with impaired language and developmental difficulties [13,14], and it has significant value in this role.

Researchers are also examining reduced-key keyboards [15-17] and the automatic expansion of minimal text into full sentences using sentence "compansion" [18,19]. Sentence compansion is a technique that aims to allow a user to enter a sentence in the form of uninflected content words and have the system output a syntactically correct sentence. The user has less information to enter to encode a full sentence, hence reducing the required number of selection actions. Some users with mild linguistic impairment or developmental problems may also benefit from assistance in forming a syntactically correct sentence. The sentence compansion system makes extensive use of natural language processing techniques, such as a semantic parser, representation translator, and sentence generator.

The techniques referred to above all relate to the process of creating unique text, i.e., writing words or phrases *ab initio*. Certainly, this is something that any communication system will need to be used for, since language can be infinitely varied. It does not follow, however, that language is entirely unpredictable. Human communication is a mixture of unique utterances and language that is pre-packaged, formulaic, and repetitive [20]. A good deal of this sort of language consists of fairly lengthy, re-usable sequences (e.g., commonly told stories, personal biographical material, anecdotes, or family news). It would seem perverse to require that someone who creates their utterances with great effort should have to entirely recreate them every time they want to use them. In fact, we know that people who use assisted communication systems do not usually take part in this form of communication at all because it is impractical. There is increasing evidence, however, that narrative and story-telling play a very important part in the communicative repertoire of all speakers. In particular, the ability to draw on episodes from your life history in current conversation is vital to maintaining a full impression of your personality in dealing with others [21,22].

We have been investigating ways in which prestored re-usable texts might be of assistance in improving the performance of communication systems for nonspeaking people. The study began with an investigation of the acceptability and effects of including prestored texts in a conversation. Next, a series of methods of increasing complexity were developed for having the system itself perform some of the cognitive and manipulative tasks required to locate suitable texts for a user to include in conversation.

### 3. STORING AND RETRIEVING LONGER TEXTS FOR CONVERSATION

The communicational repertoire of a computer-assisted speaker could be improved significantly if the user could store and retrieve reusable narratives, stories, and other long text sequences for use during conversation. The authors are pursuing this approach as part of an overall program that aims to develop a multi-level communication system, offering assistance for unique utterances, re-usable texts, and formulaic sequences [23]. Early examples of computer-based assistive communication systems did include the ability to store and retrieve large numbers of phrases [7,24], and although this approach presents no great problem in terms of storage space on the computer, what has made this method unworkable in practice is the inability of the user to remember more than a small number of stored phrases and how to access them. The phrases are needed in a rapidly moving conversational situation, which imposes the additional problem that they are required urgently by a user who can only access the communication system slowly. The demands made on the design for an effective communication system are thus quite severe. Large amounts of text must be stored and be capable of being called up with minimal effort for inclusion in an activity on which the user must concentrate most of his or her attention while simultaneously looking for the next piece of text.

One aspect of the design must be to keep the cognitive demands that are placed on the user within reasonable bounds. The emphasis must lie on making accurate selection of a very small number of phrases that can be displayed for the user to assimilate and choose from. If the cognitive demands on the user are excessive, the user's response will be slow and the conversation will not progress well.

The process of developing a communication system includes identifying which aspects of the communication task should be performed by the user and which can be done by the system. The system should have facilities to perform those parts of the communication process that can be automated within technological limitations, while the user remains in overall control. This requires the application of language processing and dialogue modelling techniques for the purposes of predicting the user's communication requirements as well as good interface design to give the user efficient access to, and control over, the communication facilities that the system offers. We have, therefore, been investigating ways in which the communication system itself could take on many of the cognitive and manipulative tasks required to locate and produce suitable texts for the user. This can be partially achieved if the system is able to model aspects of conversational interchanges. Early work resulted in a prototype communication system called CHAT, which can assist the user by modelling the opening and closing phases of a conversation, and, for the central portion of the conversation, can provide the user with readily available feedback remarks with which to react to the other speaker [25].

Related research has explored a number of different approaches to providing predictive help in the topic discussion phase of a conversation. A system called TalksBack [26,27] was developed, which predicts sentences from a large prestored collection, with the prediction being based on a model of the user's interests, the interests and social profile of frequent conversation partners, and the topic currently under discussion. Another system called PROSE was developed to facilitate the telling of stories [28], an activity which is an integral and essential part of all human communication. A great deal of time in conversation is spent in relating narratives to each other, and, as indicated above, there is evidence that the telling of stories is the primary way in which we build and maintain a projection of our personality with others. Features from CHAT, TalksBack and PROSE were integrated into a single communication system [29].

Such techniques may be useful for people with cognitive as well as physical problems, such as occur with dysphasia following a stroke. The system would prompt the user with appropriate selections from the database of personal conversational material and, thus, act as a kind of "cognitive prosthesis". The techniques have thus far

been tried out with some people who have dysphasia, with promising results, and a fuller investigation is under way [26,30].

#### 4. EFFECT OF INCORPORATING PRESTORED TEXT INTO AIDED CONVERSATIONS

Our work on prestored texts began with an investigation of whether, if prestored texts were readily available, their inclusion in computer-aided conversation would improve the quantity and quality of conversational output achieved by a system user with impaired communication. In order to make the task of the user as simple as possible in the first instance, the system was developed to deal with one topic of conversation, with eight subtopics, each subtopic having twelve associated texts. A user interface was devised by a process of iterative development with the active participation of a nonvocal person with motor impairments who advised on, and tested, the interface throughout its development.

##### A. Including Prestored Texts to Increase Conversational Participation and Control

The prototype system was given the name “Floorgrabber,” one of its intended functions being to increase the user’s conversational control [31]. After experimenting with various types of interface, a dashboard metaphor was chosen. This allowed for a tightly packed presentation that was still clear in its layout and meant that all of the controls and displays could be put on one screen, with no need to go to other screens or menus. There is always a trade-off between complexity and ease-of-use in interface design, but given the movement difficulties that most potential users experience, it was thought preferable to have one screen that contained all of the functions they would need. Three types of on-screen buttons were used, which had the effect of 1) speaking the text in the box pointed to, 2) speaking a quick comment, and 3) going to another subtopic, which was brought up with one click on the appropriate topic button. The interface is shown in Figure 1.

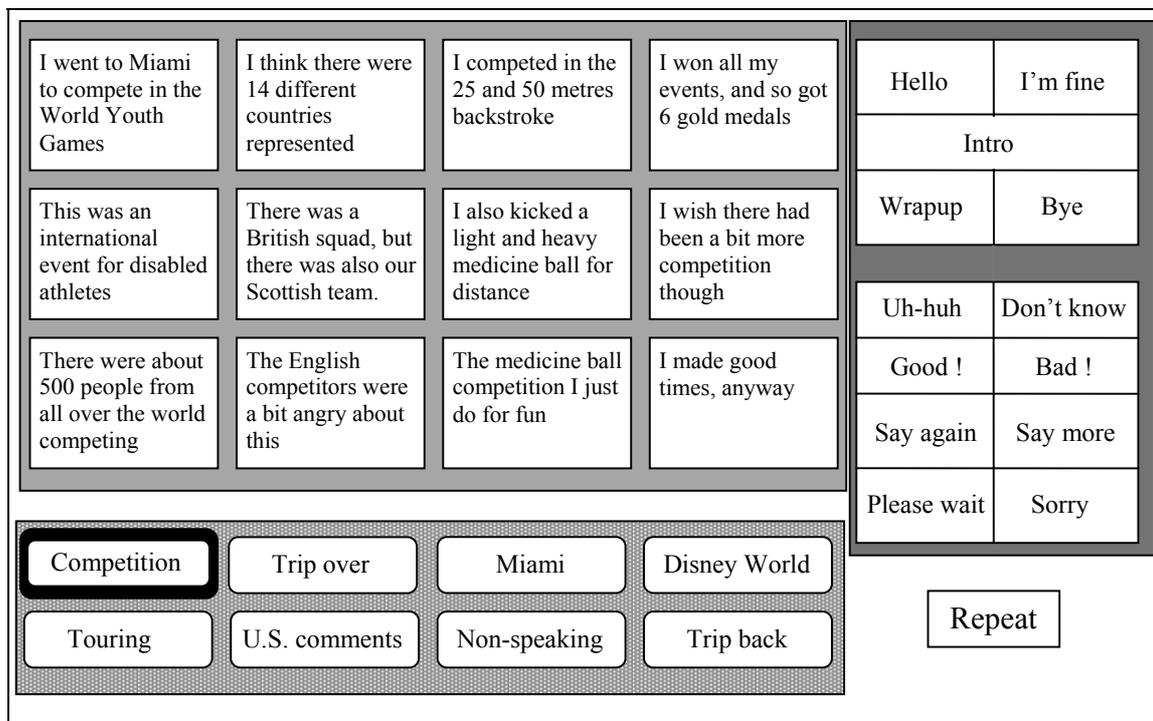


Figure 1. Interface for the Floorgrabber system.

The nonvocal user produced the textual material that made up the content of Floorgrabber. The text related to one topic, a journey abroad he made for an international disabled swimming competition. This topic was chosen because he is often asked about this experience, and he has difficulty describing it adequately with his current communication techniques. His usual method of communication is a 400-word chart, supplemented with a synthetic speech output word storage device, plus gestures and some vocalizations.

He conveyed anecdotes and comments about the swimming competition to volunteers by using his conventional communication methods. These methods are quite time consuming, requiring a great deal of ingenuity by the non-speaker and guesswork by the listener. This material was then typed into the system by the volunteers. A typical communication from this user in creating the text for speaking might proceed as follows. The user points to the word "August" on his word chart, then to the word "country," and the letter "S" (the helper then guesses "Sweden?" and this is confirmed with a vocalization and a smile). The user makes a swimming gesture and points to the word "game." The helper checks out by saying "Swimming competition in Sweden in August?" and gets a confirmation. The user points to the words "I," "win," "one," and "brown" on the chart. The helper feeds back "You won one bronze medal there?" The user confirms this with a vocalization and a head nod. The finished sentence is entered into the system: "I won a bronze medal at the swimming competition in Sweden in August."

To ensure that the final version was accurate, the material was checked over with him several times and modified until he was completely satisfied that it represented the way he would like to express himself. No guidelines were given to him as to the content of the material he produced. He was asked to produce text that he might like to use in a conversation telling someone about the swimming competition. This procedure is, of course, far from the way in which natural conversation is produced. It was the intention of this experiment, however, to determine whether and how such pre-planned material might be used effectively as part of a free-flowing conversation.

Trials of the prototype were performed by using a single-case experimental design, in which the user conducted 12 conversations with 12 different people on the chosen topic. The 12 sessions followed an ABAB sequence, in which the user employed only his current communication methods during the baseline sessions (A), but used the prototype Floorgrabber system as an additional mode of communication during the intervention sessions (B). Each conversation lasted for 15 minutes. Half of the conversation partners knew the user and had previously communicated with him using his conventional communication methods. The other half did not know him and had no experience of communicating with a nonvocal person. The user's instructions were to use whatever communication mode was most comfortable and effective at any point throughout the dialogues. The conversation partners were asked to have a 15-minute conversation with the user about the swimming competition. All of these dialogues were videotaped and subsequently transcribed and coded for analysis.

The effectiveness of the prototype system was assessed in terms of its ability to assist the user to participate more fully in a dialogue and to have more control over its direction. These measures were chosen because they are often cited as key factors in the unsatisfactory nature of much communication by nonspeaking people [3], [32]. In order to measure the amount of conversational participation and the degree of control by the user over the conversation 1) all of the words produced by each partner in each dialogue were counted and 2) two types of conversational move that were of relevance in conversational control were defined as follows:

RESPONDER: An answer to a question or a feedback (back-channel) comment to the other speaker.

INITIATOR: A question or a statement that is not a responder.

All occurrences of these conversational moves by both partners were counted from the transcripts.

The average number of words used by the user in the dialogues increased by 273% ( $p < 0.01$ ), from 143 words/dialogue to 534, when Floorgrabber was available to the user. The number of words used by the natural speakers in the conversation also increased slightly when Floorgrabber was present, although not to a significant extent. For the Floorgrabber user, the mean number of initiators per dialogue increased by 170% ( $p < 0.05$ ), from 10 to 27, and the mean number of responders was reduced, but not to a significant extent. While the natural speakers

used virtually the same number of initiators (48-50) when Floorgrabber was available to its user, they used many more responders (36 instead of 19), because they had much more opportunity to respond to the increased number of initiators issued by the Floorgrabber user. These results showed that Floorgrabber enabled its user to increase his control over the dialogues by using significantly more words and initiators than before, while the output of the other speakers was not limited or impaired. The overall balance of conversational participation and control was improved significantly by the use of Floorgrabber. There was certainly a cognitive load for the user involved in selecting text for speaking from Floorgrabber. The increased amount of material produced, however, indicates that the trade-off here may be an acceptable one. It must be remembered that a high cognitive load will likely be the price paid for any augmented communication, since the effort in producing it is far from the nearly unconscious fluency that natural speakers enjoy. The question may be how much communication is actually produced for a given amount of effort by the user. In this case having relatively large chunks of prestored material may provide significant assistance.

### *B. Listeners' Perceptions of Conversation Content*

An interesting finding has emerged from related work being done by Alm within another research collaboration. In this work, a prototype similar to Floorgrabber was used to evaluate listener's impressions of the content of computer-aided communication based on prestored texts, as compared to naturally occurring dialogues. The computer-aided conversations involved volunteers and one of the research team, who simulated a nonspeaking person using the prototype augmentative communication system. The user was able to use only prestored texts to conduct the conversations. Most of the text was material about one subject (holidays). A number of rapidly accessible comments and quick feedback remarks were also available. The unaided conversations were between pairs of volunteers who were asked to converse together on the topic of holidays. Transcripts of randomly sampled sections of the conversations and audio recordings of re-enactments of the samples with pauses removed were rated for social competence on a six-item scale (co-efficient  $\alpha = 0.83$ ) by 24 judges. This scale was devised, taking into account Light's distinction between socio-linguistic and socio-relational skills [33]. The judges were asked to make judgments on both of these skill areas using the categories of the conversation's naturalness, cohesiveness, fluency, balance, liveliness, and friendliness. The content of the computer-aided conversations was rated significantly higher than that of the unaided samples ( $p < 0.001$ ). The judges also rated the individual contributions of the computer-aided communicator and her unaided partners on how "socially worthwhile and involving" these appeared. There was no significant difference between the ratings of their respective contributions ( $p > 0.05$ ) [34].

This finding came as something of a surprise to the researchers, since the purpose was to establish whether conversations using prestored material would simply be able to equal naturally occurring conversations in terms of quality of content. Of course, the pauses in actual computer-aided conversation do have an effect on listeners' impressions of the quality of the communication, but this finding is still of interest, since it suggests that in some ways augmented communication could have an edge over naturally occurring talk. A plausible explanation for this finding is that naturally occurring talk is full of high-speed dysfluencies, mistakes, substitutions, and other "messy" features that listeners, using their ability to infer what the speaker is intending to say, tend to discount. Prestored material is selected because, by its nature, it may be of particular interest, and it is expressed more carefully than quick-flowing talk. It may therefore appear more orderly and dense with meaning than natural talk.

## 5. HYPERTEXT AS A HOST FOR A STORE OF CONVERSATIONAL MATERIAL

Having established that prestored text had a role to play in an assistive communication system, the next step in the research was to explore methods for storing and retrieving larger amounts of conversational material. It was important to retain an easy-to-use user interface that did not require a large effort from the user to produce appropriate texts. What was being sought was a form of conversational prediction that would be able to automatically produce appropriate choices for the next thing to say. It was decided that hypertext was worth exploring as a structure that could facilitate this type of predictive system, since hypertext is a method for storing and navigating through information that purports to be based on the way the human mind stores and accesses infor-

mation using highly flexible associative links [35]. Any cross referencing in documents can be considered as a simple form of hypertext, but the provision of a rich network of such associations on a computer with interactive capabilities gives hypertext its real character. Hypertext, thus, might be able to model the flexible fashion in which the mind stores and recalls conversation items and the manner in which conversational items are introduced into a dialogue in a way that maintains the coherence of the conversation. The well-known problem of getting lost in a hypertext structure suggests that it might be a more suitable environment for browsing than for directed information finding. There is an interesting parallel here with the way informal conversation wanders through a number of topics, much as a user might browse through a richly connected information system. Some work on hypertext design based on cognitive flexibility theory has suggested that, as a mechanism for conveying how information is inter-related, hypertext environments need to be quite complex. If they are too simple, they do not convey the context of the information sufficiently [36]. Again, this suggests that there may be a trade-off between effectiveness and ease of use where hypertext is being used to convey information. Where hypertext is being used as a store for conversational material through which a user is allowed and, in fact, encouraged, to “wander,” this will not be as much of a problem.

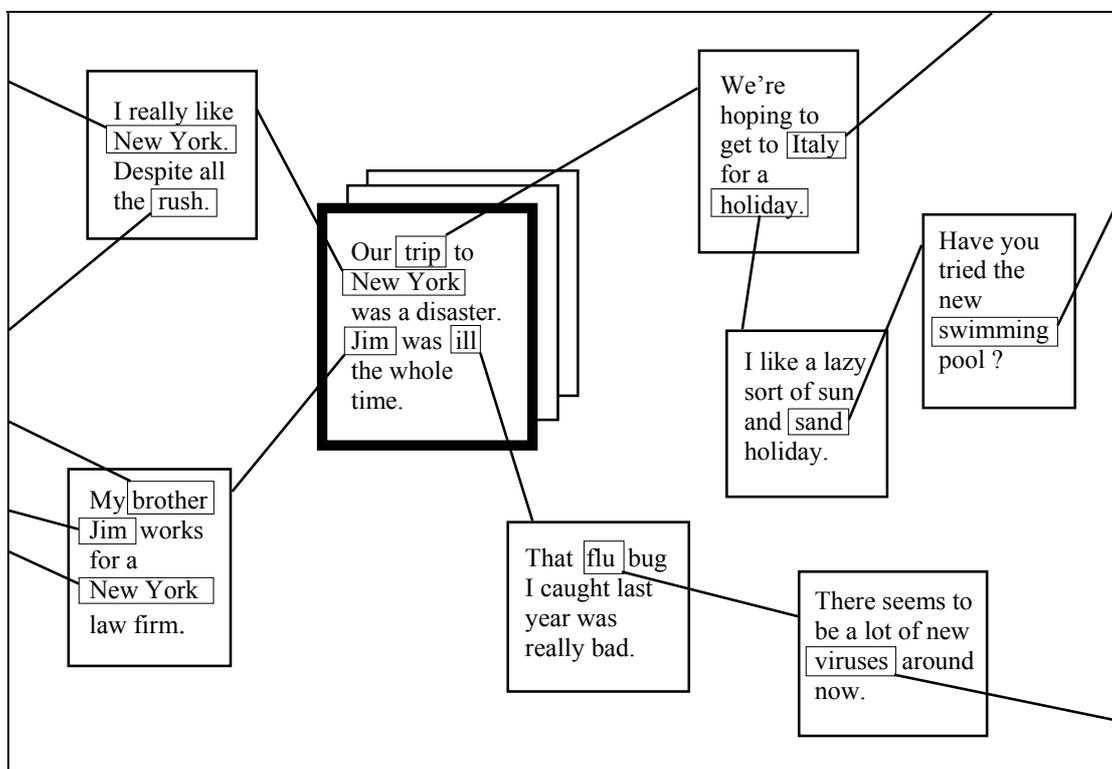


Figure 2. Hypertext links in a store of reusable conversation items.

The diagram in Figure 2 shows the way stored text items might be linked in a hypertext system. The text in bold outline (“Our trip to New York ...”) is the one that the user is currently speaking. The user can carry on with the next part of this particular narrative (indicated by the stack of pages under the current one) or he or she may choose to branch off to related texts (shown in the other text boxes). It should be noted that the linking is not being done by means of keywords, as would be done in a conventional database. The words are shown in boxes only to make visible the hypertext link.

The diagram suggests a high degree of interconnection between stored text items. There are obvious difficulties involved in achieving such a degree of interconnection, but it gives a goal toward which the research can progress, with the hope that partial implementations will prove useful as the work continues.

## 6. MAKING LINKS BETWEEN TEXT ITEMS IN A HYPERTEXT SYSTEM

The trials of the prototype Floorgrabber system demonstrated that it was possible to give a nonvocal person increased conversational control by providing them with prestored text. Expansion of the system would inevitably involve the inclusion of a much larger number of topics, however, and it would therefore be necessary to find a way to avoid long searches through the stored material. A method for offering the user predictions of the next thing to say would be required. Such predictions would represent “jumps” to other parts of the conversation store, but the problem was how to provide such a facility in a way that would produce the kind of natural conversational sequences that the user would want to use.

### *A. Providing Conversational “Paths”*

One possible method was derived from work in the philosophy of language on treating language as rule-guided action, rather than as an outward expression of mental phenomena. Computational design based on this approach should focus on developing empirically accurate rules for using a phrase, rather than on building a model of language storage and production that is somehow taken to mimic human “internal” mental events.

Such an approach was adopted in the development of a prototype system called the MODular Social Communicator (MOSCO) [37], in which the prediction of an appropriate utterance was performed by viewing the preceding utterance as an example of “target” behaviour. Previous conversations in which the same target statement had occurred were reviewed by the system, and statements that were adjacent to the target in these previous conversations were assumed to be contextually relevant to the current conversation and appropriate for use at that point in the conversation. The system then offered these statements to the user as possible next utterances.

The “targeting” process in MOSCO was a relatively simple one based on the assumption that, if Statement B had followed Statement A in a previous conversation, Statement B might be an appropriate utterance to follow Statement A in another conversation. No semantic analysis was performed during this process, but MOSCO did have a basic parsing ability that enabled it to code statements in a syntactically reduced form. This made it easier for the system to compare statements that differed only in syntactic detail (e.g., tense) and, hence, find close matches for the current target statement.

This approach, which might be called “path-tracking,” could also have a simpler implementation that logged all of the user’s moves through the material. The system would offer, at any point in the conversation, one or more “paths” that the user might like to take based on their previous paths through the material [26]. This path-tracking algorithm would clearly be appropriate when an anecdote was involved. To tell the entire anecdote, the user need only pick the first text item and then continue to ask for the next text item in the chain until the story was told. It seemed likely that shorter chains of text items might also prove useful, particularly if more than one path was offered at any point.

Both the parsing with path-tracking and the simple path-tracking methods were tested using MOSCO and other simpler implementations in a series of conversations with speaking conversation partners. The transcripts of these conversations were compared with transcripts produced in ordinary unconstrained dialogues and were found to be equivalent in terms of coherence.

### *B. Modelling Step-wise Topic Shifts*

Modelling the type of “step-wise” topic shifting that is typical of unconstrained conversation [38] was identified as the next research priority. The requirement was for a method of providing the user with texts that had a link to the current text, but were not on exactly the same topic. A fully functioning hypertext system with all the words in the text as potential linking buttons would be the ideal interface to implement such linkage, but this is clearly a long-term task, with a number of major problems still to be solved. It was decided to start by devising ways of linking whole text items with as much flexibility as possible, the intention being to provide the user with a set of

linked texts that would permit moves from the present topic to similar ones while still lending a degree of coherence to the way the conversation was moving. This would be a start in supplying a hypertext-style interface, where “nearby” items in the information space were displayed together on the screen, but a number of “jumps” to other parts of the information space were also offered. In this case, the jumps would be to text items that were not directly related to the current text, but which would be appropriate ones to say next in a conversation.

In a simple database, each item can be tagged with a number of descriptors drawn from a larger set and a search for related items involves comparing these descriptors. Although we have experimented with using such a tagging method [39], this is not an entirely successful approach to use in a conversational database. It is difficult to decide how many descriptors would be needed for a large database, and even with a large number, there will be frequent problems of deciding whether to assign a particular item to a given descriptor. If too few descriptors are used, there will be too many candidates after any search and further searching will be needed to identify suitable ones. If too many descriptors are used, the possibilities increase of getting a zero result from a search. An additional problem is that the process of deciding on a list of descriptors requires advance knowledge of the material to be stored in the database, and the purpose for which it is required. This is not appropriate for a conversational database, which will need to grow with each individual and be used by them flexibly in a large number of different situations.

## 7. FUZZY INFORMATION RETRIEVAL

The possibility of using fuzzy information retrieval methods to provide suitable link items has been examined. Applied to an information retrieval system, a fuzzy-set-based approach would allow for more flexible storage and retrieval methods [40,41]. The relatedness of items in the database could be captured without the need for similar items to share a number of descriptors from a given set. With a database of potential conversational material, a fuzzy set retrieval system would have the advantage that, given one item, it would always produce a set of the most closely related items in the database. It would never return from a search to report that zero items had been found and also could be used to limit the result of a search that could otherwise return many possible items.

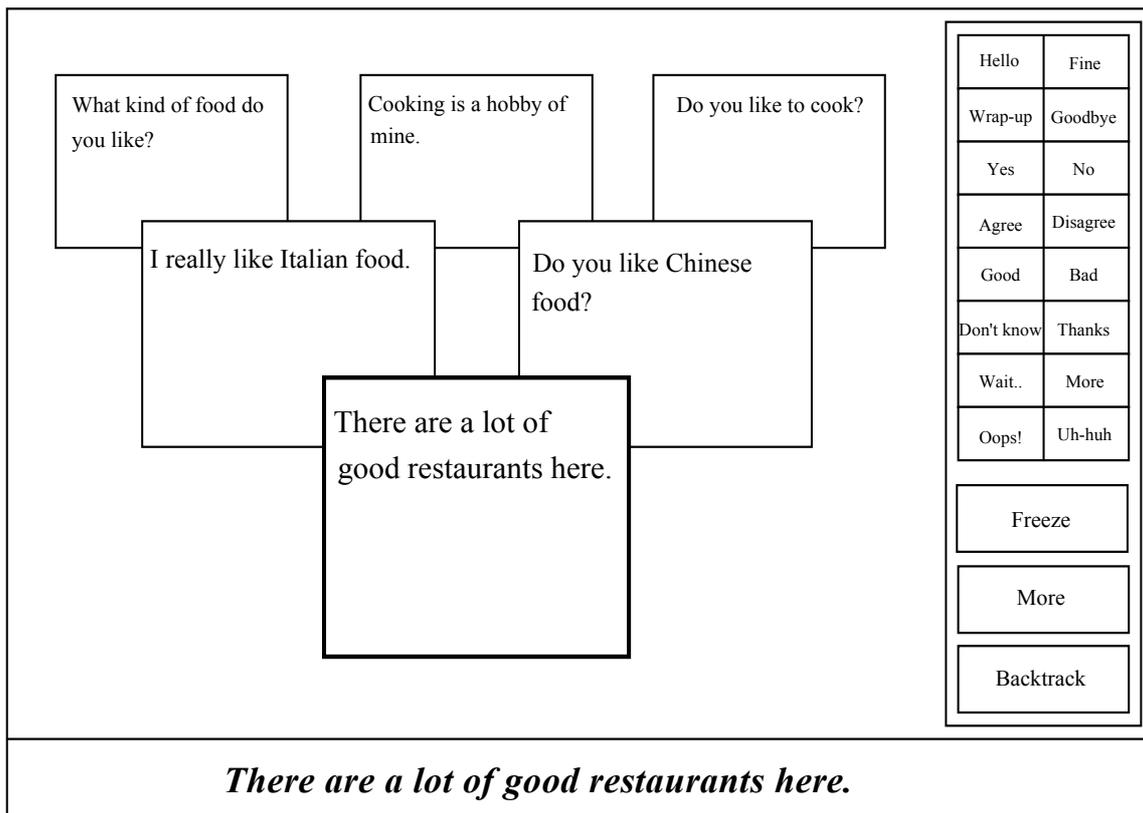
### *A. Prototype Fuzzy Retrieval System*

A prototype system was developed to test out the feasibility of using fuzzy retrieval methods in a conversation aid [42]. Conversational texts for the prototype were taken from the Floorgrabber system, and a set of descriptors were selected that could apply, to varying degrees, to all items in the database. Two types of descriptor were used, representing either the semantics of an item (its subject) or the pragmatics (its purpose in a dialogue, i.e., speech act). The system used 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 had a vector associated with it that described its degree of belonging to these 13 categories. The values for this vector were set individually for each item by the user. This was clearly a time-consuming process, and originally it was thought that devising a way of automating the process, with, for instance, some sort of content analysis of the text, would be desirable. However, in using the system, it became apparent that the values on these vectors were quite individual to the particular user. For instance, if the user’s job was connected to sport, the link between work and sport items might be quite close, while for others they would be more distant. Perhaps some automatic help in setting up default values would be welcome. In use, of course, the items will be added one at a time, over an extended period, and it may not be too much of an overhead to assign values to each item as it is created, particularly with an easy-to-use interface, such as a set of on-screen slider controls.

A version of the system was created for the purpose of evaluating the concept against an equivalent system based on Boolean search and retrieval methods. This version contained the same items, but depended on conventional database searching to locate them. As expected, the conventional system often produced no texts that matched a given text, whereas the fuzzy set system always produced a full set of candidate texts. A comparison was then done with a more advanced version of the Floorgrabber prototype, which had been expanded to include 850

stored text items, retrievable through a hierarchical search method (topic → subtopic → text item). Each system was equipped with the same text items, and conversation sequences that had been performed with the fuzzy set system were then reproduced using the Floorgrabber system. Fewer input activations (20%-35% fewer) were needed to produce the conversation using the fuzzy set system. The Floorgrabber system occasionally required a large number of input activations to produce an utterance, whereas the fuzzy set system tended to consistently need only one or two. There was also a difference in the cognitive task involved in that the Floorgrabber system required the user to formulate and execute a search plan, while the fuzzy set system simply presented appropriate material automatically.



**Figure 3.** Interface for a conversation aid based on fuzzy information retrieval.

### B. GUI for Fuzzy Retrieval of Text Items

A further development of this prototype was to devise an interface that graphically conveyed the relationship between potential conversational text items. The interface is shown in Figure 3. With this interface, every time the user selected an item to speak, the most closely connected other items were presented as a cluster around it. The on-screen buttons on the right produced quick-fire remarks in the appropriate category. The Freeze, More, and Backtrack buttons allowed the user, respectively, to keep the set of candidate texts fixed, or get a completely new set, or to back-up through the selections they had seen thus far. The quickest way to navigate through texts was just to click on the wanted text, which was then spoken. It became the text at the bottom of the display, and the five closest matches to it were shown above it.

The ease of use of this interface was tested in a series of conversations held by an able-bodied person using only the system to communicate with volunteer conversation partners. The sessions were recorded, and transcripts

produced. The conversations varied in length from 5 to 7.5 minutes. The average number of mouse clicks that the user needed to speak one item was 2.1. The average length of the text items was 8.2 words, which gives a rate of about four words per mouse click. An average of 90.1 words per minute was achieved by the able-bodied user. This was a promising result, since this was using quite short text items, and even extrapolating the rate to simulate a user who could only manage (for example) six activations per minute, the speaking rate would still be about 24 words per minute, which is an improvement on current performances [43].

A possible objection to the use of fuzzy retrieval in this application is that the “prompting” nature of the fuzzy set system is inimical to real conversation, in that we could never know what the user might have said if working with a totally free choice. It can be argued, however, that the most important aspect is that the user has the opportunity to use some stored material rather than having nothing at all to say and the user can always decide to create some unique text if the reusable material is not suitable for the current situation. It can also be argued that the order in which the material is introduced into a conversation matters less than the fact that continuity of the conversation is maintained. The criterion is thus not a direct comparison with unaided speech, but with successful versus unsuccessful communication, however it is accomplished. Certainly, it will be important to optimize the user interface so as to reduce the cognitive load. Paying attention, for instance, to Miller’s “7 plus-or-minus 2” rule [44] for elements that the user must pay attention to will be important.

Further work will need to be done to establish whether the cognitive load of using a system such as this might outweigh its advantages. As with the other prototype system discussed here, the question may be one of granularity. If the elements of a system are chunks of text, rather than individual words or letters, any cognitive load involved in their selection may be outweighed by the amount of conversation actually accomplished.

## 8. GOALS, PLANS AND SCRIPTS

Many dialogues take place to accomplish a particular purpose, which means that there is an element of predictability about them. This is true of commercial transactions, such as dialogues in shops, and of discussions with professional people, such as doctors. Any situation in which the speaker is trying to accomplish a particular concrete goal is likely to be predictable because such situations are rarely unique. They repeat themselves regularly and make up part of the person’s routine. Just as conversational elements are reusable, so the sets of utterances we use to accomplish a particular purpose are likely to be reusable and could be said to form a kind of “script.”

The theory of scripts that Schank and Abelson developed was based on the observation that people do not enter into every situation completely unprepared for it [45]. There is a certain amount of stereotypical conversation in many situations that we encounter in our daily lives; indeed some situations consist of nothing but this form of communication. Scripts have been suggested as a useful technique for augmentative communication [46], and it is clearly possible that a script-based communication system could provide a valuable form of prediction in situations where the user is trying to achieve a particular functional goal.

Schank and Abelson developed the hierarchy of

GOALS → PLANS → SCRIPTS

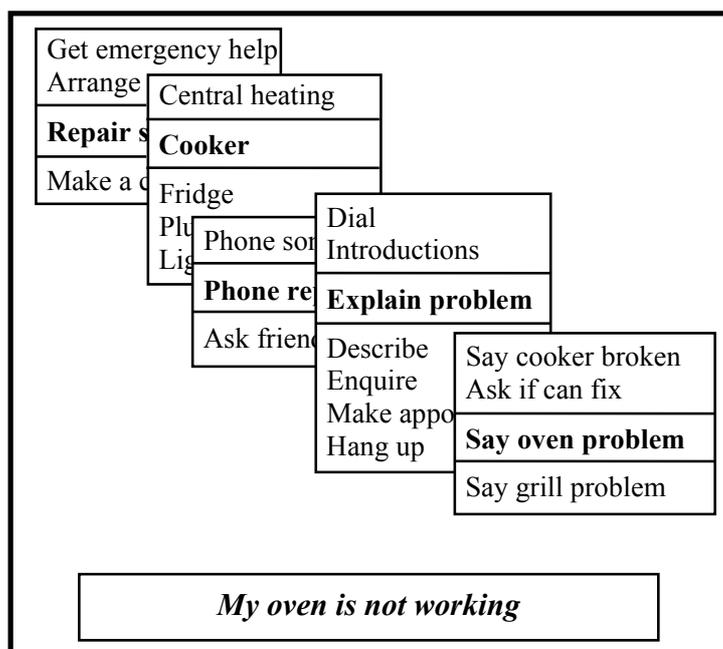
which provides a method of coping with breakdown, or unforeseen events that result in a premature exit from a script. If a script user should encounter a breakdown, one technique for recovery would be to move up the hierarchy to select an alternative script, or to restate at that point the plan or goal that the user is trying to accomplish. This might in itself elicit the appropriate help or response from the other speaker.

### *A. Experiments with a Script-Based System*

A prototype script-based system was developed to facilitate experimentation with a goals, plans and scripts structure in computer-assisted communication [47]. The user interface, part of which is shown in Figure 4, allowed the user to move easily up and down a goals-plans-scripts hierarchy, selecting text items for output. The type of conversation studied in the tests was that of the user telephoning to get something accomplished. This was chosen because there is a clear goal, with a number of possible plans and scripts to achieve it. Also, for non-speaking people who are living in their own homes, use of the telephone, particularly with people who do not

know them, presents great difficulties, and has clear implications for their independence. The intention was to determine whether a goals-plans-scripts-based system could be used in a realistic dialogue situation, with a minimal number of conversation breakdowns and with adequate information transmission between participants.

An experiment was devised wherein the experimenter used the prototype script system to attempt to complete a set communicational task over the telephone. The task was to request assistance in the repair of a cooker (kitchen stove). The experimenter assumed the role of the person requesting assistance, and forty volunteer subjects were asked to play the role of an appliance repair person. The experimenter spoke to the subjects by telephone, using his natural voice with the first group of subjects and synthetic speech driven by the prototype script system with the second group. Forty conversations were conducted in total, one with each subject. The prototype system was deliberately configured so that the experimenter could not type in unique text; all contingencies had to be dealt with using prestored scripts. This imposed an exacting test on the script system's ability to support the user in the task. The computer-mediated conversations took longer than the natural conversations, as expected, but the average duration was only 50% greater. [The mean duration was 102 seconds (s.d. 14 seconds) for natural voice, and 150 seconds (s.d. 26 seconds) for synthetic (script-based) speech.] On average, it took about five keystrokes to produce an utterance. Information transfer and effectiveness in accomplishing the given task were not significantly different between the two groups. Almost all of the subjects were positive about the system, in a questionnaire survey, although some of the subjects had difficulty understanding the synthetic speech.



**Figure 4.** Example of user interface for the script-based system. The user moves down through a goal-plan-script hierarchy to finally select the wanted utterance.

There were short delays while the experimenter selected scripted items from the system, and these caused some difficulties for a small number of the subjects. The delay time before the “disabled” person (the experimenter) spoke tended to vary between 2 and 5 seconds, depending on whether the desired text for speaking was immediately available on the current screen. This delay time is in the region of the 3-second “awkwardness limen,” identified by McLaughlin and Cody [48] as being the boundary between acceptable and unacceptable delays in conversation. The fact that the script-based prototype came close to a threshold established for natural speakers was encouraging. This evaluation established that a script-based approach to goal-oriented interactions shows much promise as a means of helping a nonspeaking person to communicate effectively in such situations.

## 9. DISCUSSION

The principal aim of this research is to enable people with impaired communication ability to participate effectively in conversations. This is a substantial challenge; people without communication impairments can talk at speech rates up to 200 words per minute [49], whereas many communication impaired people are limited to rates in the region of two-to-ten words per minute using current augmentative communication systems [3]. There is a requirement, therefore, for augmentative communication techniques that can radically improve effective communication rates while facilitating the sort of conversational moves and strategies that are typical of unimpeded conversation.

There are a number of facets to the problem, ranging from improving the efficiency of the system user in encoding information at the word and character level through managing the storage, retrieval, and use of prestored texts to the provision of predictable chatting sentences that convey little or no new information but increase the bonding and understanding between conversation partners. It is therefore possible to conceive of the augmentative communication system functioning at a number of different levels [50], with the management of prestored and reusable topic-oriented text lying at mid-level between the low information chatting functions and the high-information word or character encoding. Chatting functions can be provided relatively easily in an augmentative communication system based on current computer technology [25] and provide valuable assistance to a user in keeping a conversation going. They cannot, however, fulfil all conversational requirements. The efficiency of word- and character-encoding can only be improved to a certain extent, given that there will always be a (typically quite low) ceiling on the number of activations per minute that a user can perform.

A significant gain (unfortunately very difficult to achieve) is only possible at the middle topic-oriented level, where improvements in efficiency and strategy could radically improve the ability of the system user to participate effectively in conversation. The research described in this paper is addressing this problem in a number of ways, developing and evaluating different methods of using prestored texts that can collectively contribute to the overall goal of enabling a communication-impaired person to converse sufficiently rapidly and responsively to maintain the interest and attention of a conversation partner. A hypertext structure may well be useful for this purpose, in that it offers the possibility of modelling the way in which conversational contributions are linked in a coherent sequence. The important research questions lie in devising ways to implement a hypertext style linkage between stored items, which is simple enough to be practical yet complex enough to model the variety of natural conversation. The research has also established that a goals-plans-scripts hierarchy is useful in enabling the user of a speech output system with prestored utterances to successfully accomplish a useful task over the telephone. Although tested initially by a non-disabled person, the technique showed sufficient significant improvement over existing communication methods to suggest that it is worthy of further development.

A number of encouraging results have therefore been achieved so far. The Floorgrabber system improved the balance of control in a conversation by increasing the total number of words used in a given period and the number of initiators used by the communication impaired person relative to their conversation partner. The Floorgrabber user employed approximately 50% as many initiators as the conversation partner, as compared to 20% without Floorgrabber. Experiments with the fuzzy text retrieval system showed that up to 35% fewer switch activations were needed to retrieve appropriate topic text from storage for use during conversation; it was estimated that a user who could achieve six switch activations per minute would be able to output text at around 24 words per minute. A prototype script-based system developed to assist telephone enquiries resulted in experimental computer-aided dialogues that required an average of about five keystrokes to produce an appropriate utterance and, when employed by an able-bodied user, were only 50% longer in duration than unaided dialogues. Related research has discovered that the content of computer-aided conversations can be perceived as being as good as, or better than, the content of unaided conversations, because the use of prestored text in conversation brings the additional benefit that the material is well-organized and composed. These various results indicate that there is much to be gained from the inclusion of prestored texts in computer-assisted conversation and that progress can be made in reducing the conversation handicap faced by people with communication impairments.

## 10. CONCLUSIONS

It is evident that computer-aided conversations using prestored texts can be effective and successful, if the user can access the system rapidly enough to guide it in the management and selection of texts. Significant research is required to increase the communication rates that can be achieved by techniques described here, to investigate new techniques that can contribute to communication performance, and to integrate the various approaches into combined systems. It is probable that all of these techniques will have roles to play in the augmentative communication system of the future, with different components providing different aspects of the necessary processing. The required research will reach into the fields of human-machine interaction, natural language processing, dialogue design, and adaptive and predictive systems. It will require expertise from areas such as psychology, computer science, cognitive science, speech and language therapy, linguistics, discourse and conversation analysis, and rehabilitation engineering, in what must be a multidisciplinary approach to the wide range of problems that exist. The field is a challenging one, with many areas of knowledge incomplete, but computer-assisted conversation has much unrealized potential that should be developed and harnessed for the benefit of people with communication impairments.

## ACKNOWLEDGEMENTS

This work was supported by Digital Equipment Corporation, The Remedi Foundation and the Scottish Home & Health Department.

## REFERENCES

- [1] J. Eulenberg, Ed., *Conversations with Non-speaking People*. Toronto, Ontario, Canada: Canadian Rehabilitation Council for the Disabled, 1984.
- [2] R. Sienkewicz-Mercer and S. Kaplan, *I Raise My Eyes to Say Yes*. New York: Houghton-Mifflin, 1989.
- [3] D. Beukelman and P. Mirenda, *Augmentative and Alternative Communication: Management of Severe Communication Disorders in Children and Adults*. New York: Paul Brookes, 1992.
- [4] A. Y. Cairns and A. F. Newell, "Toward gesture recognition for the physically impaired," in Proc. RESNA'94 Annual Conference, Nashville, TN, pp. 414-416.
- [5] A. Y. Cairns, W. Smart, and I. Ricketts, "Alternative access to assistive devices," in Proc. RESNA'94 Annual Conference, Nashville, TN, pp. 397-399.
- [6] G. Vanderheiden and D. Kelso, "Comparative analysis of fixed vocabulary communication acceleration techniques," *Augmentative and Alternative Communication*, vol. 3, no. 4, pp. 196-206, 1987.
- [7] B. Baker, "Minspeak," *Byte*, vol. 7, no. 9, pp. 186-202, 1982.
- [8] W. Woltosz, "Stephen Hawking's communication system," *Communication Outlook*, vol. 10, no. 1, pp. 8-11, 1988.
- [9] A. Swiffin, J. L. Arnott, J. A. Pickering, and A. F. Newell, "Adaptive and predictive techniques in a communication prosthesis," *Augmentative and Alternative Communication*, vol. 3, no. 4, pp. 181-191, 1987.
- [10] W. Beattie, A. McKinlay, J. L. Arnott, and P. Gregor, "An investigation of the use of recency in word predictive algorithms," in Proc. ISAAC'94 Conference, Maastricht, The Netherlands, pp. 496-498.
- [11] S. Tyvand and P. Demasco, "Syntax statistics in word prediction," in Proc. ECART 2, 2nd European Conference on the Advancement of Rehabilitation Technology, Stockholm, Sweden, Paper 11.2, 1993.
- [12] H. Horstmann Koester and S. P. Levine, "Modeling the speed of text entry with a word prediction interface," *IEEE Trans. Rehabilitation Engineering*, vol. 2, pp. 177-187, Mar. 1994.
- [13] C. Morris, L. Booth, I. W. Ricketts, N. Alm, and A. Newell, "Evaluation of a syntax-driven word predictor for children with language impairments," in Proc. RESNA'93 Annual Conference, Las Vegas, NV, pp. 423-425.
- [14] A. Newell, L. Booth, and W. Beattie, "Predictive text entry with PAL and children with learning difficulties," *British Journal of Educational Technology*, vol. 22, no. 1, pp. 23-40, 1991.
- [15] J. L. Arnott and M. Y. Javed, "Probabilistic character disambiguation for reduced keyboards using small text samples," *Augmentative and Alternative Communication*, vol. 8, no. 3, pp. 215-223, 1992.
- [16] M. T. King, C. A. Kushler, and D. A. Grover, "JustType - Efficient communication with eight keys," in Proc. RESNA'95 Annual Conference, Vancouver, BC, Canada, pp. 94-96.

- [17] B. J. Oommen, R. S. Valiveti, and J. R. Zgierski, "Correction to 'An adaptive learning solution to the keyboard optimization problem,'" *IEEE Trans. Systems, Man & Cybernetics*, vol. 22, pp. 1233-1243, Sept./Oct. 1992.
- [18] P. Demasco and K. McCoy, "Generating text from compressed input: An intelligent interface for people with severe motor impairments," *Communications of ACM*, vol. 35, no. 5, pp. 68-79, 1992.
- [19] K. McCoy, W. McKnitt, D. Peischl, C. Pennington, P. Vanderheyden, and P. Demasco, "AAC-user therapist interactions: Preliminary linguistic observations and implications for compansion," in *Proc. RESNA'94 Annual Conference*, Nashville, TN, pp. 129-131.
- [20] J. Gumperz, *Discourse Strategies*. London, U.K.: Cambridge Univ. Press, 1982.
- [21] D. Polkinghorne, "Narrative and self-concept," *Journal of Narrative and Life History*, vol. 1, no. 213, pp. 135-153, 1991.
- [22] K. Nelson, "Remembering and telling: A developmental story," *Journal of Narrative and Life History*, vol. 1, no. 213, pp. 109-127, 1991.
- [23] A. F. Newell, N. Alm, and J. L. Arnott, "An integrated strategy for improving communication systems for severely physically impaired non-speaking people," in *Rehabilitation Technology - Strategies for the European Union, Studies in Health Technology and Informatics - Vol. 9*. Amsterdam, The Netherlands: IOS, 1993, pp. 249-253.
- [24] J. Montgomery, Ed., *The Assisted Communicator: A Monograph on the Use of the Phonic Mirror HandiVoice*. Mill Valley, CA: Phonic Ear Inc., 1982.
- [25] N. Alm, J. L. Arnott, and A. F. Newell, "Prediction and conversational momentum in an augmentative communication system," *Communications of ACM*, vol. 35, no. 5, pp. 46-57, 1992.
- [26] E. Broumley, "Talksback: The use of social knowledge in an augmentative communication system," Ph.D. dissertation, University of Dundee, Dundee, Scotland, U.K., 1994.
- [27] E. Broumley, J. L. Arnott, A. Y. Cairns, and A. F. Newell, "TalksBack: An application of artificial intelligence techniques to a communication prosthesis for nonspeaking people," in *Proc. 9th European Conference on Artificial Intelligence*, Stockholm, Sweden, 1990, pp. 117-119.
- [28] A. Waller, "Providing narratives in an augmentative communication system," Ph.D. dissertation, University of Dundee, Dundee, Scotland, U.K., 1992.
- [29] Don Johnston Incorporated, *Talk:About™*, Don Johnston Inc., Wauconda, IL, 1994.
- [30] F. Dennis, A. Waller, A. Y. Cairns, J. Brodie, A. F. Newell, and K. Morrison, "Investigating the use of a predictive conversation aid with dysphasic adults," in *Abstracts of ISAAC Communication Matters National Symposium*, Mansfield, Nottingham, U.K., 1993, p. 24.
- [31] N. Alm, J. L. Arnott, and A. F. Newell, "Evaluation of a text-based communication system for increasing conversational participation and control," in *Proc. RESNA International '92 Conference*, Toronto, Ont., Canada, pp. 366-368.
- [32] J. Light, "Interaction involving individuals using AAC communication systems: State of the art and future directions," *Augmentative and Alternative Communication*, vol. 4, no. 2, pp. 66-82, 1988.
- [33] J. Light, "Toward a definition of communicative competence for individuals using augmentative and alternative communication systems," *Augmentative and Alternative Communication*, vol. 5, no. 2, pp. 137-144, 1989.
- [34] J. Todman, L. Elder, and N. Alm, "Evaluation of the content of computer-aided conversations," *Augmentative and Alternative Communication*, vol. 11, no. 4, pp. 229-234, 1995.
- [35] J. Conklin, "Hypertext: An introduction and survey," *IEEE Trans. on Computers*, vol. C-20, pp. 17-41, Jan. 1987.
- [36] R. Spiro and M. Jacobson, "Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation," *J. Educational Computing Research*, vol. 12, pp. 301-333, 1995.
- [37] A. McKinlay, "Using a social approach in the development of a communication aid to achieve perceived communicative competence," in *Proc. RESNA'91 Annual Conference*, Kansas City, MO, pp. 204-206.
- [38] G. Button and N. Casey, "Generating topic: The use of topic-initial elicitors," in *Structures of Social Action - Studies in Conversation Analysis*, J. Atkins and J. Heritage, Eds. London, U.K.: Cambridge Univ. Press, 1984.
- [39] N. Alm, J. L. Arnott, and A. F. Newell, "Database design for storing and accessing personal conversational material," in *Proc. RESNA'89 Annual Conference*, New Orleans, LA, pp. 147-148.
- [40] L. A. Zadeh, "Fuzzy sets," *Information & Control*, vol. 8, no. 3, pp. 338-353, 1965.

- [41] C. V. Negoita and P. Flonder, "On fuzziness in information retrieval," *International Journal of Man-Machine Studies*, vol. 8, pp. 711-716, 1976.
- [42] N. Alm, M. Nicol, and J. L. Arnott, "The application of fuzzy set theory to the storage and retrieval of conversational texts in an augmentative communication system," in *Proc. RESNA'93 Annual Conference*, Las Vegas, NV, pp. 127-129.
- [43] N. Alm, "Fuzzy set theory in augmentative communication," *Applied Computer Studies Division*, University of Dundee, Dundee, Scotland, U.K., Rep. Scottish Home Health Dept. Project K/RED/6/31/5/2, 1994.
- [44] G. A. Miller "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychological Review*, vol. 63, no. 2, pp. 81-97, 1956.
- [45] R. Schank and R. Abelson, *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Lawrence Erlbaum, 1977.
- [46] A. Kraat, Speech-Language Pathologist, Morton Roberts Center for Augmentative Communication, Queens College Speech Hearing Center, Flushing, NY, private communication, 1987.
- [47] N. Alm, A. I. Morrison, and J. L. Arnott, "A communication system based on scripts, plans and goals for enabling nonspeaking people to conduct telephone conversations," in *Proc. IEEE Conference on Systems, Man & Cybernetics*, Vancouver, B.C., Canada, 1995, pp. 2408-2412.
- [48] M. McLaughlin and M. Cody, "Awkward silences: Behavioral antecedents and consequences of the conversational lapse," *Human Communication Research*, vol. 8, pp. 299-316, 1982.
- [49] A. C. Downton, "Simultaneous transcription of machine shorthand for the deaf," Ph.D. dissertation, University of Southampton, Southampton, U.K., 1982, p.7.
- [50] N. Alm, A. Newell, and J. L. Arnott, "An integrated, multi-level communication system for physically impaired non-speaking children and adults," in *Proc. First International Conference of the Saudi Association for Handicapped Children*, Riyadh, Saudi Arabia, 1992, pp. 926-933.