USING A SPEECH SYNTHESIS SYSTEM FOR CASUAL CONVERSATION -- A SERIOUS APPLICATION

Norman Alm*, Alan F. Newell*, John L. Arnott*

ABSTRACT

We have developed a prototype communication system which helps non-speakers to perform communication acts competently, rather than concentrating on improving the production efficiency for specific letters, words or phrases. The system, called CHAT, is based on patterns which are found in normal unconstrained dialogue. It contains information about the present conversational move, the next likely move, the person being addressed, and the mood of the interaction. The system can move automatically through a dialogue, or can offer the user a selection of conversational moves on the basis of single keystrokes. The increased communication speed which is thus offered can significantly improve both the flow of the conversation and the disabled person's control of the dialogue.

INTRODUCTION

One application for speech technology which has a great potential impact for the people concerned is in helping severely disabled non-speaking people to communicate with an artificial voice. The availability of relatively inexpensive speech synthesisers has stimulated the production of a number of such devices. There are a number of difficulties still to be overcome, however. One is the very low communication rate possible. This is because almost always with lack of speech goes lack of all motor controls. It is not uncommon for a disabled non-speaker to have to employ a mouth-held stick to use the keyboard, or to be unable to use the keyboard at all. Keyboard substitutes include software that scans possible letters or words until the user signals with a one-way switch. Clearly, the speed constraints for individuals with this level of handicap are considerable. The rate of communication achievable is typically 2-10 words per minute, compared with unimpaired conversation, which proceeds typically at a rate of 180-200 wpm (ref 1).

CHAT -- A PROTOTYPE CONVERSATION AID

The CHAT system is an attempt to provide a severely disabled non-speaker with the chance to take part in genuine conversational interaction. The CHAT software is based on a model of casual conversation (refs 2,3,4). The contour of a typical, complete conversation is given in Figure 1.

*Microcomputer Centre, University of Dundee, Dundee DD1 4HN
CHAT is designed to increase the communication rate by acting as automatically as possible, issuing appropriate contributions to a conversation with a single keystroke. Its name is an acronym of Conversation Helped by Automatic Talk. In its most effortless mode, CHAT will output speech on the repeated pressing of the same key, which might be labelled 'Make an appropriate remark here'. (More precise control of the output is always possible, but at the cost of increased key selections.)

The automatic production of conversational contributions is a possibility because of the predictability and repetitiveness of a large amount of natural conversation. An automatic system which was too obvious in its operation, and too dictatorial, however, would probably be unacceptable in the very human process of conversational interaction. In order to counter-balance this tendency, CHAT has an ability to simulate the natural variations and unpredictability of expression that occur in unimpeded utterances. It is also designed to take on the personal expressive style of the user.

In order to avoid repetitions of exactly the same phrase in any category, CHAT makes a random choice each time from a list of suitable alternatives for the current conversational move. (This is, in fact, a constrained random choice, that does not produce exactly the same phrase until the other possibilities have all been used.) Thus, one person might be greeted by 'Hello there. How's it going?' whereas the next person might be offered 'Hi. How are you?'

Spoken language is full of words, expressions, and vocalisations that carry no intrinsic information, but are nevertheless important elements of speech. They have a number of uses, including holding the floor while composing the next thought, signalling continued attention to a speaker, or simply making speech more rhythmic. CHAT allows the user to output a FILLER at any point, and ensures that it is an appropriate one for the conversational context.
The expression of personality and mood is an important purpose of conversation. CHAT allows the user to select a mood, and then all subsequent utterances will be given in that mood. So far we have experimented with POLITE, INFORMAL, HUMOROUS, and ANGRY moods. (POLITE is the default setting!)

CHAT also contains a list of known people, whose names can be inserted automatically into the phrases as appropriate. A mood setting can be linked to any name, so that, for instance, close friends would always be spoken to in the informal mood. For easy changing and updating, the names and all of the phrases stored in the system are accessible through a simple editor. Our intention is for the user to put in phrases in these different moods to reflect their own personality and style. Thus, if an INFORMAL FAREWELL is requested, the system makes a random choice from a list of appropriate phrases, each of which reflects the user’s style. This simulates what happens naturally, when we might have any number of ways of expressing the same thought, but a significant percentage of them will be qualitatively different from the way another person’s expressions. The first set of phrases for CHAT were put in by a researcher with an American background. In operation, this reflects his personal style to such an extent that it provokes a humorous reaction from visitors when it is demonstrated by a British member of the team.

We have piloted the system using volunteers to interact with a member of the research team, who only uses CHAT to speak. Two extracts from the recordings of these sessions are given below:

A: Hi, Mary, how are you?
B: I'm fine, thank you.
   How are you?
A: Not bad, thanks, Mary. We're moving along as usual here.
B: That's good.
A: How have you been getting on?
B: Very well, thank you.

Time taken: 25 seconds     Mood: Polite

A: Hi, Bill. How's it been going with you?
B: Pretty well, and yourself?
A: Just dandy thanks, Bill. What have you been up to recently?
B: The naughty things I couldn't tell you about.
A: Aye.

Time taken: 26 seconds     Mood: Informal

We were very encouraged in general by the naturalness of the dialogues. It is interesting that, where CHAT did produce something inappropriate, the dialogue continued more or less intact. This is to expected from studies of the way natural
conversations cope well with errors and omissions. If we produce transcripts with no clue as to who is the speaker using CHAT, as in the above examples, it normally takes a careful study by an outside observer to decide which partner in the conversation is using CHAT.

What has given us most encouragement however is the timing of these pilot runs. The examples above are typical. Because the aided speaker is producing an entire speech act for each keystroke, the rate is of the order of 70-80 words per minute (both speakers).

**DIALOGUE DESIGN FOR SPEECH SYSTEMS**

The development reported here has shown that conversational rules can be built into a prosthesis for the physically handicapped in such a way as to enable such people to conduct realistic conversations, with a minimum of control signals.

Although the system which has been developed requires (minimal) human intervention, fully automatic systems could be designed based on similar principles. Such systems would have potential in developments for speech input-output systems for the non-handicapped. An example is a telephone data base query system for use by the general public. Other situations where there is a requirement for a computer generated natural language system are also potential application areas for these techniques.

**CONCLUSIONS**

A conversation aid based on modelling conversational patterns in the opening and closing stages has shown promise. We have begun work on including a facility for a range of feedback remarks to the other speaker. The next major part of our research will be to apply a similar approach to modelling the central part of a conversation: topic discussion.

**REFERENCES**


This work was funded in part by the Science and Engineering Research Council.