Proceedings of the Institute of Acoustics

A SCRIPT BASED SPEECH AID FOR NON-SPEAKING PEOPLE

Gillian S. Harper, Richard Dye, Norman Alm, John L. Arnott and Iain R. Murray

Department of Applied Computing,
University of Dundee,
Dundee DD1 4HN, Scotland.

1. INTRODUCTION

Physically disabled non-speaking people often use speech aids to communicate. However, the speed with which a disabled person can operate a speech output device is often very slow and this can hinder satisfactory communication performance. This paper describes a software speech aid called ScripTalker which has been developed to help non-speaking people interact more quickly and effectively in everyday situations by using a model of conversation and knowledge of typical interaction patterns. These are embodied in scripts, which hold a structured series of reusable phrases. The phrases are retrieved through a scene-based pictorial interface depicting the current conversational situation. Retrieval is performed by selecting objects within the scene and output via a speech synthesiser. A prompting facility at each stage of the conversation can be used to assist identification of the correct phrase and the system also allows novel phrases to be produced outwith the scripts (and pre-stored if necessary). Preliminary trials have demonstrated that the system has the potential to improve the conversation performance of the user, and a commercial version of the system will soon be available.

2. THE CHALLENGES OF A SPEECH DISABILITY

People with a speech disability often have accompanying physical disabilities. Physical capabilities can vary widely, and many have limited muscle control. Some can only interact with the world by the operation of a switch. In order to communicate, speech-disabled people usually rely on some form of aid to communicate with others. These communication aids may be text- or symbol-based and approximately 60% of non-speaking people have some literacy problem. The communication speed of Alternative and Augmentative Communication (AAC) aid users is typically very slow, usually in the range of 2 to 10 words per minute compared to an average 180 words for a non-impaired speaker [1]. Conversation can therefore be difficult and can result in social isolation. The system described in this paper incorporates a model of conversation patterns within an AAC system to increase the ease and speed of communication by users and to assist interaction in everyday situations.

3. MODELLING CONVERSATION PATTERNS

Many existing communication aids allow a user to communicate by constructing phrases from a set of letters, words or symbols displayed in a grid layout on a static board or computer screen. This approach can provide flexibility in the words and phrases that the user may be able to communicate, but is usually a very slow process. This severe time penalty often effectively isolates the person from situations that require a more rapid response rate.

The ScripTalker prototype is being used to investigate methods of using a model of conversation within an AAC device to store and retrieve sequences of phrases. This method could speed up the

communication process of the user and allow the user to interact in situations where they may previously have been excluded.

Research into conversation and the pragmatics of interaction has shown that much of everyday conversation follows patterns and is thus repeated and "reusable" [2,3,4,5]. Therefore it is possible to predict interaction sequences and use this knowledge within an AAC system [6]. The general structure of a conversation starts with an opening phase, and moves on through topic discussion and feedback to the closing phase. The speech acts involved in the openings, feedback and closing phase are highly predictable, the discussion phase less so, and this is the stage of primary interest.

Many everyday events are structured and the interaction patterns of participants can be described. The success of the interaction often depends on each participant having knowledge of his or her rôle and behaving and speaking accordingly [4,5,7]. Research in the field of artificial intelligence has also investigated such stereotypical or "scripted" interactions [8,9,10], and this knowledge could also be exploited within an AAC system.

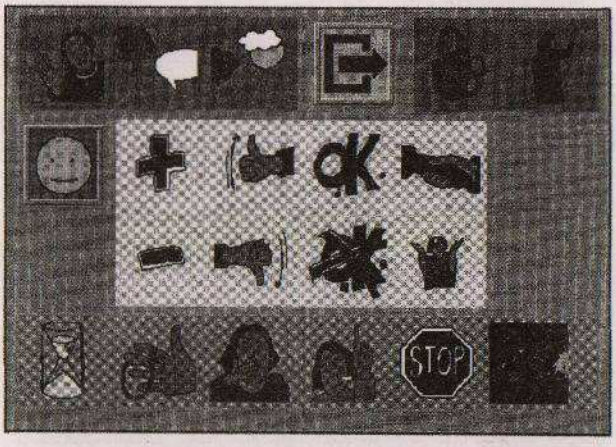
ScripTalker incorporates a model which included reusable phrases, structured for opening and closing the conversation, giving feedback and participating in everyday stereotypical situations through scripts of the user's side of the interaction. These scripts are accessed through a novel pictorial interface. Preliminary trials of the script-type approach for storing and retrieving reusable phrase sequences have shown encouraging results [11,12].

4. THE DEVELOPMENT OF A NEW SPEECH AID

The ScripTalker user interface has three main components: rapidly produced speech acts, the scripts and a unique text facility. The rapid speech act component is used in the opening, feedback, and closing portions of a conversation, and consists of groups of speech-act buttons. The scripts component is used in the discussion or transaction phase of a conversation, and consists of a set of scripts with which the user can interact. The unique utterance component is used when there are no appropriate pre-stored utterances available, and consists of a virtual on-screen keyboard and a word prediction mechanism. Due to the limited literacy of many AAC users, the interface is, as far as possible, pictorially based.

The rapid speech acts interface provides access to a range of stored speech acts that are used in the opening and closing phases of a conversation. This facility is based on previous research on a rapid speech act system, which provided users with just these speech act possibilities. The system provided a reasonable conversation rate and users and partners reported positively on the impact of the system on the quality of the interaction [13].

The interface used for the rapid speech act facility in the ScripTalker prototype consists of a button for each speech act being modelled. The speech-act buttons are arranged into 5 function groups (see Figure 1). Each button contains a set of 5 phrases that are worded to achieve the same effect. When the user selects a button, one of the five phrases is randomly selected from the phrases for that category and output via a speech synthesiser. This provides variability and helps prevent the user sounding repetitive.



Group	Speech-act buttons
I'm listening:	Uh-huh
Openers:	Greetings, Responses and Small talk
Closers:	Wrap-up, Farewells
Feedback:	Yes, No, Good, Bad, Agree, Disagree, Thank you
Control:	OOPS, Please wait, Tell me more, Can I interrupt?, You misunderstand

FIGURE 1 - Rapid speech act user interface

A mood button allows the speech act selected to reflect the users' mood. The mood is currently conveyed in the wording of the phrase. Further work may enable mood to also be reflected in the speech synthesis output. The mood button allows 1 of 4 moods to be selected. The moods available for selection are polite (default), informal, humorous and angry.

The system currently has seven scripts for the following activities: Daily Living, the Restaurant, the Doctor, Meeting People, Shopping, on the Telephone, and Communicating Emotions. These are the main situations that were identified during consultation with a group of potential users.

The content of a script is derived from a structured analysis of the situation. Each phrase that may be required in the given situation is recorded and placed into a sequence. This script sequence is further divided into a sequence of scenes and each phrase in a scene is mapped onto a relevant interface object.

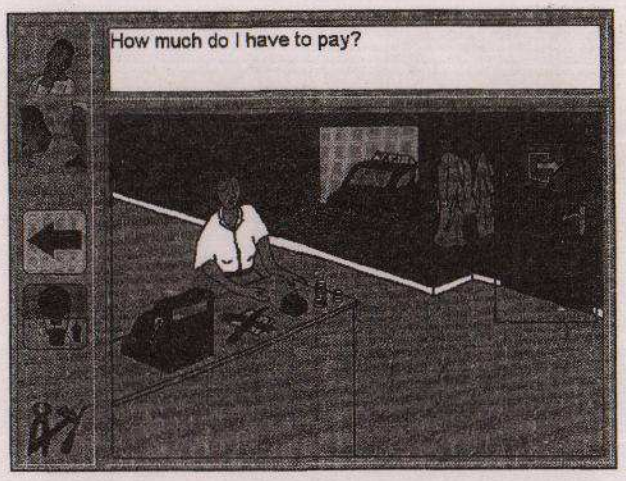


FIGURE 2 - The interface showing a scene within a script

The script interface uses a pictorial scene representation. Each script is structured in a series of scenes. Scenes contain phrases for that stage of the interaction. Each coloured object in the scene will cause an action such as retrieving a phrase or navigating to the next scene in the script. The objects within the scene are relevant to the phrase; for example, selection of the cash register object in the scene shown in Figure 2 would retrieve the phrase "How much do I have to pay?" The objects are arranged in a natural scene rather than a grid layout because the natural arrangements of objects within a context helps the user to recognise the situation for which the scene is intended, and to recognise and recall objects within the scene [14,15].

In addition to the scenes, some special interface modules were developed and integrated. These include clock and calendar interfaces with which the user can communicate time-related phrases. A notepad facility allows users to construct additional phrases which can be stored and retrieved at a later time.

The ScripTalker system has been designed to be usable for a number of languages, initially English, German and Dutch. The system can interface to various speech synthesisers such as DECTalk TextAssist, Eurovocs, Apollo and InfoVox. This flexibility in the synthesisers that can be used was incorporated because the quality of available speech synthesis output in different languages is variable. The content of the system can also be easily localised for different languages.

5. PROTOTYPE EVALUATION

A pilot evaluation of the ScripTalker prototype was performed using a single case study design. The general aim of the pilot study was to perform an initial evaluation of the system interface and the concept of conversational scripts within an AAC system. Two developed scripts were used and tested in detail to assess their suitability for the interactions in which a non-speaking person would use them. The results gained from the study have been used to influence the development of the system and script design, and further evaluation with a larger number of subjects is planned.

The subject was a moderately literate 25 year-old male with cerebral palsy without the ability to communicate solely through natural speech. The subject normally uses a word board to communicate. (The word board displays approximately 500 words in a grid layout; the communication partner must read each word pointed at by the user.)

The study contained eight sessions over a six-week period. Each session had a duration of approximately one and a half hours. Four training sessions were used to train the subject to use the prototype and two scripts, one script for visiting a restaurant, and the other for consulting a doctor. Four trial sessions were used to carry out structured role-plays of the interaction situations. These sessions were videotaped to enable further examination and transcription.

The subject was given two training sessions using the two scripts. The trial sessions used structured role-plays of the relevant situation. The role-play interaction partners were recruited from staff and students. The interaction was role-played by the subject using his current communication aid and, after a short break, the situation was role-played using the ScripTalker prototype. Each participant, subject and role-play partners, completed a questionnaire to give subjective feedback on each interaction. The first trial session was used to evaluate the restaurant script and the second session to evaluate the doctor script. The subject found the restaurant script fast and easy to use, but had more difficulty using the doctor script and locating the correct phrases quickly. Therefore, the study concentrated on the doctor script for the remaining sessions. The subject was given two further training sessions using the doctor script and participated in two further trial sessions.

The results of the trials of the prototype system were encouraging. The subjective feedback gained from the subject and the role-play interaction partners was generally very positive. The system enabled the user to communicate the majority of the required information. Where the system did not provide the required phrases, the user was able to use gestures and vocalisations to communicate. These situations were identified and the scripts were augmented to include the required phrases. Communication performance measurements were taken using transcriptions of the interactions taken from the video recordings and are summarised for three of the role-plays in Figure 3. The measurements included the total number of words produced by the subject during each role-play, the word rate in terms of number of words produced per minute of subject conversation turns, and the total number of words produced per conversation turn by the subject. In each of these indicators the performance of the subject using the ScripTalker prototype was higher than when the subject used his existing communication word board.

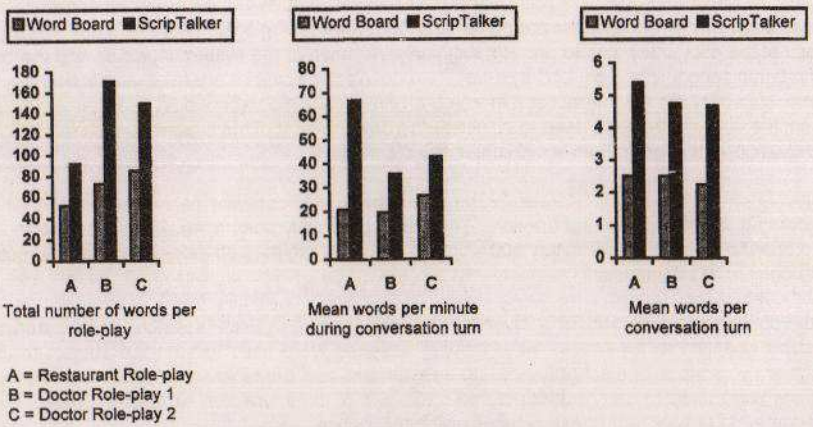


FIGURE 3 - Pilot evaluation results

The performance of the subject during the role-plays suggests that ScripTalker has the potential to improve the communication performance of the user. The subject found the system to be easy to understand, efficient and it handled the presentation and update of scripts and scenes very quickly. Therefore, indications from the pilot evaluation are that scripts within an AAC system are useful for communication by a disabled non-speaking person in a transactional interaction. Further evaluation of the ScripTalker system with a larger number of subjects who use different AAC devices, in more than one country, will be undertaken as part of the continuing project.

6. THE FUTURE

Further development work on the ScripTalker system is being undertaken in order to realise a practical AAC device. Further evaluation of the system with a larger number of subjects will be performed. A script authoring facility will be developed to assist in the development and integration of new scripts and the integration of third party environmental control facility will be carried out. A basic framework for the addition of emotion to speech synthesis is in place and further development of this facility may be undertaken.

7. CONCLUSION

Disabled non-speaking people often rely on speech aids to communicate, but inability to operate systems quickly can reduce the speed of communication. This can restrict the user's opportunities for interaction. The prototype described has been used to investigate the potential of using a conversation model within a speech aid to increase the volume and speed of communication that can be achieved. The ScripTalker model incorporates a rapid speech act facility for opening, closing and

feedback stages in a conversation and conversation scripts for use during common situations. The content of the system incorporates sequenced reusable phrases which can be retrieved and output via a speech synthesiser using a novel pictorial scene interface. Preliminary trials have shown that the ScripTalker system has the potential to improve the performance of the user in terms of volume and speed of communication. Further work is underway to carry out more extensive evaluation and system development.

8. ACKNOWLEDGMENTS

The development of ScripTalker was supported by the TIDE Programme of the Commission of the European Union under the ALADIN project (TP 1035). The project partners were The University of Dundee (UK), Instituut vor Revalidatievraagstukken (The Netherlands), Kompagne BV (The Netherlands), and IGEL GmbH (Germany). Further work in this field is being conducted under the REACT project (Telematics project DE 4207), involving these four partners and Instituut voor Doven (The Netherlands).

9. REFERENCES

[1] D BEUKELMAN & P MIRENDA, "*Augmentative and alternative communication: management of severe communication disorder in children and adults*", New York: Paul Brookes Publishing Co. (1992).
[2] E A SCHEGLOFF & H SACKS, "Opening up closings", *Semiotica*, **7(4)**, pp. 289-327 (1973).
[3] J R SEARLE, *Speech acts*, Cambridge University Press (1969).
[4] E GOFFMAN, *Relations in public*, Penguin Press, London (1971).
[5] M ARGYLE, *The social psychology of everyday life (2nd Ed)*, Penguin Books, Harmondsworth (1992).
[6] N ALM, J TODMAN, L ELDER & I MURRAY, "Modelling conversation patterns to improve augmented communication for non-vocal people", *Proc. Institute of Acoustics, Autumn Conference Speech & Hearing*, 16(5), pp. 381-389 (1994).
[7] J GAHAGAN, *Social interaction and its management*, Methuen, London (1984).
[8] R SCHANK & R ABELSON, *Scripts, plans, goals, and understanding*, Lawrence Erlbaum, New Jersey (1977).
[9] R C SCHANK, *Dynamic memory*, Cambridge University Press (1982).
[10] J M MANDLER, "Categorical and schematic organization in memory", in C R PUFF (Ed.), *Memory organisation and structure*, Academic Press, New York, pp. 259-299 (1979).
[11] N ALM, A, MORRISON & J L ARNOTT, "A communication system based on scripts, plans and goals for enabling non-speaking people to conduct telephone conversations", *Proc. IEEE Conference on Systems, Man & Cybernetics*, Vancouver, Canada, pp.2408-2412 (1995).
[12] P B VANDERHEYDEN, P W DEMASCO, K F McCOY & C A PENNINGTON, "A preliminary study into schema-based access and organization of re-usable text in AAC", *Proc. RESNA '96 Conference*, Salt Lake City, Utah, USA, pp. 59-61 (1996).
[13] N ALM, J L ARNOTT & A F NEWELL, "Prediction and conversational momentum in an augmentative communication system", *Communications of the ACM*, **35(5)**, pp. 46-57 (1992).
[14] J M MANDLER & R E PARKER, "Memory for descriptive and spatial information in complex pictures", *J. Experimental Psychology: Human Learning and Memory*, **2**, pp. 38-48 (1976).
[15] J M MANDLER, *Stories, scripts and scenes: aspects of schema theory*, Lawrence Erlbaum Associates, Hillsdale (1984).