COMMUNICATION AIDS AND VOICE SYNTHESIS

B. Brophy*, J. L. Arnott*, A. F. Newell*,

ABSTRACT

Good quality speech synthesis has an important application in the design of communication aids for the speech impaired. The quality of voice in currently available communication aids is generally poor, which imposes abnormal strains on the listener and speaker and distorts their natural patterns of conversation. The application of speech synthesis in this field is in its early stages, but higher quality voice synthesis should improve the situation in two major areas.

a) the listener would be able to use his/her natural patterns of communication
b) the non-vocal person would have the opportunity to learn the normal patterns of communication

This paper will discuss the applications of high quality speech synthesis in communication aids with particular emphasis on:

i) the listeners perception of the voice synthesis
ii) how speech synthesis might change the interaction between able bodied persons and the user
iii) some outcomes of clinical applications of communication aids and voice synthesis

SUMMARY OF RECENT IMPROVEMENTS IN SPEECH TECHNOLOGY FOR SPEECH AIDS

Sheahan et al (ref 1) describe the applications of voice synthesis for the disabled and include a review of communication aids with voice synthesis. They note the need for speech aids
a) to be highly intelligible, to allow the use to communicate with strangers, have natural intonation, and convey emotion
b) be appropriate to the users age and sex.

Currently aids using pre-coded speech forms rather than text to speech, phoneme based synthesis are more intelligible to a stranger, although these aids are relatively less successful in producing phrases. Recent developments in voice synthesis use the encoding of information about pitch and the transitions between phonemes (co-articulation) to improve intelligibility. However these forms of synthesis are not readily available to communication aid users due to expense and portability.

*Micro-computer Centre, University of Dundee, DD1 4HN
LISTENERS PERCEPTION OF VOICE SYNTHESIS

Greene et al (ref 2) compared eight text to speech synthesisers using sentence and single word tests. They suggest that interpreting synthetic speech requires more effort from the listener than interpreting natural speech.

The authors tested DECTALK, a voice synthesiser which has improved the quality of synthetic voice by using co-articulation and found it to be the most easily understood synthetic voice. It was nearly as intelligible as natural voice both for single word and semantically correct and anomalous sentences. The authors analyse the influence of synthetic speech and pose questions for further research:

i) How is listener comprehension affected by acoustic input?
ii) Does the listener enjoy listening to the voice or does he become bored, lose attention, concentration?

In a further study the authors examined the effect of intelligibility on clients with a language deficit (ie non-native English language speakers or children with cognitive disability). They concluded that synthetic speech interacts with task difficulty to affect the client's performance. The extra cognitive load of interpreting poor quality synthetic speech makes word recognition more difficult in both children and adults, and adults with incomplete grammatical knowledge (second language learners) have more difficulty than native speakers.

These findings have implications for speech aid users. The quality of their interactions with their listeners may be affected by the quality of the voice synthesis for these reasons:

a) The listener's difficulty in understanding the voice will demand more frequent repetition than in natural communication.

b) The listener may react negatively to the voice quality, losing attention and concentration, and so restricting or avoiding communication with the user.

c) If the listener is also communicatively disabled the synthetic speech will add to their difficulty in understanding.

HOW SPEECH SYNTHESIS MIGHT CHANGE THE INTERACTION BETWEEN ABLE-BODIED AND USER.

Advances and improvements in voice synthesis have improved the quality of the units of speech production. Already there are advances in the developments of voice synthesisers which will allow users to add intonation and other "suprasegmental" linguistic features (prosody,
mood, emotion) to speech which will influence the perception of synthetic speech in several ways:

a) Intelligibility will be increased, as listeners use intonation to segment passages of continuous speech into units to be processed and understood.

b) Listeners' attention and concentration will improve if the speech is more enjoyable to listen to.

c) Improved voice synthesis might have benefits educationally. Children with oro-motor difficulties have difficulty in learning to spell because they are unable to articulate sound. Accurate voice synthesis might contribute to learning letter-sound associations in the absence of vocal speech.

Voice synthesis might also contribute to conversational communication. Communicating is more than exchanging words. Models of communication include chronemics (timing), proxemics (position) and kinesics (movement) in the control of conversations. These controls are as important to communication efficiency as rate and intelligibility. Voicing plays a part in the control of conversations by feeding back to the listener information about the speaker's identity (age, sex) and state (mood). Current voice synthesis devices give very little feedback to the listener about the speaker. Better devices might contribute to enabling non-vocal people to learn these interaction techniques by imitation of natural speakers. Aid users could then craft their communication in the same medium as their models.

iii) SOME OUTCOMES OF CLINICAL APPLICATIONS OF COMMUNICATION AIDS AND VOICE SYNTHESIS

Farrier Yorkston and Beukelman (ref 3) examined the conversational control exhibited by communication aids users. They defined conversational control as the manner and extent that a communicator directs and restrains communicative interaction, which was marked by measures of turn-regulation, etc. The study used normally speaking people who agreed to attempt to communicate via an aid. Despite the communicative abilities these subjects enjoyed, the authors found that aid users had minimal conversational control and frequently relinquished control of the communication to the speaking partners, in order to maintain conversational efficiency. Speaking partners lost interest or attempted to accelerate conversation by interrupting to complete sentences for the non-speaker. The listener needed to receive feedback and to be involved in the message creation process. This is possible with devices which do not use voice synthesis i.e. word-boards, or continuous display devices. Devices using voice synthesis demand that the listener wait for the user...
to create his message but do not occupy the listener visually or auditorily while s/he is waiting.

Buzolich (in ref 3) supports the need for listener involvement in communication aid use. She studied the rates of communication of alphabet board and Handivioce users (a communication aid with voice synthesis) and compared the results with the user's preference for each mode of communication. She found that measures of rate were the same for each mode of communication. The users preferred the alphabet boards because the listener had to be directly involved in decoding the message.

There are many undocumented reports of reactions to voice synthesis, (ie how users identify with the synthesised voices) which need to be balanced against the potential benefits that using voice could bring. Even excellent synthesised voice may not be accepted by an aid user as a solution to his communication impairment. This report has briefly underlined the error in assuming that using moderately good voice substitution will mimic the benefits of natural voice. There are many user and listener considerations to take into account which extend beyond the provision of 'natural voice'. It may be that having no voice is easier to manage than having a moderate one.

REFERENCES