ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

Use of Speech Synthesis in an Application

Angelien Sanderman, Ellen Bosgoed, Hans de Graaff, Peter van Splunder
KPN Research, P.O. Box 421, 2260 AK Leidschendam
A.A.Sanderman@research.kpn.com

Abstract

In this paper the use of speech synthesis in applications is investigated. It is of high importance to know if speech synthesis needs to be improved to achieve acceptable quality for applications like email reading. Also, it is important to know how user interfaces influence the use of speech synthesis in application. To gain insight into the use of speech synthesis in an application, two available email reading systems are tested. The results of an *expert* and *user* test are presented in this paper. The mean judgment of speech synthesis in these email reading applications is 'insufficient'. The intelligibility is mostly scored as 'not quite sufficient'. Especially, email headers and very short messages are difficult to understand. Also, it appears that the user interface influences the perception of quality of the speech synthesis.

10.21437/Eurospeech.1999-233

INTRODUCTION

Nowadays, there is much interest in the use of speech technology in applications. The speech technology (speech synthesis, speech recognition, speaker verification) required to build automated services is maturing rapidly. In this paper we concentrate primarily on the use of speech synthesis in an email reading application.
In our previous research [1,2], different speech synthesis systems for Dutch were evaluated in a laboratory environment with respect to intelligibility and acceptability. Intelligibility was measured by the ability of listeners to write down correctly semantically unpredictable sentences and acceptability was tested both with five-point semantic scales and with a two-alternative forced choice preference test. The results of these tests showed that the system giving the best intelligibility performed worst in the acceptability tests. It was the only one, which scored below the criterion value of 3 for general quality in the subjective rating test, and also scored lowest in the preference test. This shows that subjective acceptability is not a simple consequence of intelligibility. Also we can conclude that when we compare the results of test [1] with test [2] that speech synthesis for Dutch is continuously improving and that it is becoming more acceptable to use it in applications. However, many aspects are still in need of improvement, in particular prosodic aspects.

The research presented here is focused on the use of speech synthesis in an e-mail reading application, since this type of application has a need for several different kinds of synthesis such as longer texts, names, and short utterances. An email reader gives users the possibility to listen to their email by telephone. The email message is read through a text-to-speech system. To navigate and control the messages some email reading applications uses speech recognition combined with a question-and-answer dialogue style user interface, whereas others uses DTMF (touch tones) with a menu-driven dialogue style.
Although based on our previous research we expect that speech synthesis is suitable for applications, it remains open to question whether users accept speech synthesis in a real application, such as email reading. There are a several factors that can influences their judgement of the quality of speech synthesis, such as context of use, the motivation of the user (e.g. highly dependent of the benefits for the user to use the application), the attention of the user, the amount of use etc. Since we want to test the email reader in a real world environment, we can not influence these factors, however combined effects of these factors on the quality of speech synthesis can be measured. This will provide insight into the use of speech synthesis in email reading.
One of the factors that we can influence is the quality of the user interface. It is our assumption that the user interface for email reading affects the perception of speech synthesis. Therefore, the user interface aspects are studied separately as much as possible. These results are also reported in this paper.

METHOD

1. Email reading systems

To gain insight into the use of email reading and the perception of the quality of speech synthesis two available email reading systems are tested. The systems are operational, but not yet commercially available. Both systems contain an e-mail pre-processor that expands abbreviations, pronounces acronyms, filters headers and footers etc. Also, both contain a different Dutch speech synthesis system. One e-mail reader uses speech synthesis for the prompts as well as for reading the messages, whereas the other one uses pre-recorded voice for the prompts and speech synthesis for reading the

messages. Previous to this research the subjective quality of both synthesis systems had already been tested [2]. One system contains a dialogue in English, while the other uses a Dutch dialogue. One system uses DTMF (touch-tones) to navigate and control the messages, while the other uses English speech recognition in combination with 'barge-in' (the possibility to interrupt the system by speech). With respect to functionality, both systems contain the possibility to read the messages and to reply to a message by voice. One of the two systems also contains the possibility to send a new message. To do this, it was necessary to fill in an address book with names and email addresses on a web page connected to this email reader. Both systems include a web page where users can fill in some preferences, like their experience level, and the elements they want to hear (date, subject, etc.).

2. Tests

To evaluate these systems we performed two tests: an expert test and a user test.

2.1. Expert test

Eight experts on speech synthesis in applications were asked to perform several tasks with both email readers, 4 experts started with system 1 and 4 started with system 2. The experts were asked to give their feedback on a questionnaire. This questionnaire contained mainly questions about the user interface and the speech synthesis. Also, the experts had to compare both systems at the end of the test, for instance on speech synthesis, type of navigation, functionality, ease of use. To do this test we made an email test box, which contained a random set of email messages. The experts did not have the possibility to connect their own email box to the email reader, but they were allowed to send their own email to the test box.

2.2. User test

To join the user test, users had to meet the following criteria: 1) they travel around often, and 2) they make use of email. The users were asked to use one of the two email readers for at least one month in their daily life with their own email box. In total 10 users used one of the two systems. The results were gathered by means of questionnaires and analysis of the email reader log files. The questionnaires contained comparable questions to those of the experts.

RESULTS

Since the results of the expert test and user test are very comparable, we will not discuss the results separately. In general, the overall score given by the users and experts of both email reading systems is 4.2 on a 10-point scale. The results described below, show that this relatively low score is due to a combination of both the speech synthesis and the user interface used. The mean judgment of speech synthesis in these email reading applications is 'insufficient'. The intelligibility is mostly scored as 'not quite sufficient'. Especially, email headers and very short messages are difficult to understand, which is certainly the case for email headers. The user interface of both systems is quite different. The mean judgement varies between 'quite sufficient' and 'insufficient'.

1.Synthesis

Users and experts were asked to score several items of speech synthesis on 5-point scales for both systems. The quality of system 2 is scored a little lower than that of system 1. The results are summarized in Table 1. Also, in the comparison test done by the expert's, system 2 is preferred over system 1 by most of them (6 of the 8).
The questionnaires contain several open questions. With respect to speech synthesis we can summarize the following disadvantages mentioned by users:

- difficult to understand, intonation insufficient, strange accents
- pronunciation sounds handicapped
- pausing is bad, especially in long sentences.
- difficult to understand, especially in the car, since a high level of concentration is required
- English words are not always pronounced very well within Dutch text

Table 1: mean qualitative answers on speech synthesis questions

Question	Answer system 1	Answer system 2
The general quality of the speech synthesis is	Insufficient	Bad / Insufficient
To understand the message	It takes some effort to understand the message	It takes some effort to understand the message
Are there words that are difficult to understand	Often	Sometimes
The intelligibility of the system is	Not quite sufficient	Not quite sufficient
Are there disrupting sounds	Yes, a little disrupting	Yes, disrupting
The speech rate is	A little slower than expected / Neutral	A little slower than expected / Neutral
The voice is unpleasant – pleasant	2.8	2.5
The voice is unnatural – natural	2.3	2.3
The voice is impolite – polite	3.3.	3.8

- date is not pronounced in a suitable way
- bad speech quality

Also there were positive reactions, like:
- Most of the text is understandable for the way I want to use it
- Pre-recorded voice for the prompts is pleasant and clearly distinguishable from the speech synthesis (this was the case for just one of the two systems)
- In general, speech synthesis is good enough. Only time, date and telephone numbers are difficult to understand and read incorrectly

Users mentioned that the email headers were difficult to follow. It was often difficult to understand the name. The date was pronounced in a strange way and also hard to understand. Also, they noted that for system 2 Dutch headers were pronounced in English, which later turned out to be a bug in the system.
Furthermore, users said that sometimes abbreviations, punctuation marks and codes (like those present in messages from schedulers) were occasionally not well detected and processed by the email pre-processor, which resulted in messages being difficult to understand (consequently leading to confusion and a lack of information).

2. User Interface

The results show that users made use of the e-mail reader only when they had no access to their regular e-mail box on a PC. Users mentioned that it is of importance to hear the name of the sender and the subject of the email, next to the actual message. Also, they mentioned that they mainly wanted to listen to new mail. Reply is scored as important, but is not used very often. In general, the amount of functionality of the tested systems was sufficient. From the comparison test done by the experts we know that the user interface of system 2 was preferred over that of system 1. The preferences for some user interface aspects (ease of use, navigation, the way of reply and the way of interrupting the system) are represented in Figure 1. The way of access to the email reader is judged equal for both.
In general, the use of speech recognition is preferred over the use of DTMF. Especially, users appreciated the barge-in function. However, in noisy environments the users wanted to have fallback to DTMF, since speech recognition did not always work sufficiently. Also, it has to be mentioned that the choice of the touch-tones was inconsistent in comparison with the expectation of users on the basis of other telecommunication services.
Most users only wanted to listen to new messages, and did not like to go through a deep menu structure (too many levels) or question and answer dialogue before listening to new messages. The path to the new messages of both systems was criticized as too long and too many questions were asked.
Also, the users and experts mentioned that the prompts needed to be as short as possible. In general they were too long in both systems.
Another problem is the difficulty to find a specific message, e.g. from a particular date or person, although both systems have navigation features to select a specific message. However, they were inefficient to use and not intuitive. Fast navigation through headers is scored as important, but cannot easily be accomplished.
The type and amount of feedback was sufficient for both systems. The feedback rate in system 1 was perceived as too slow. Users like the possibility to use the expert or the naïve user level. The amount of explanation in the naïve user level is sufficient.

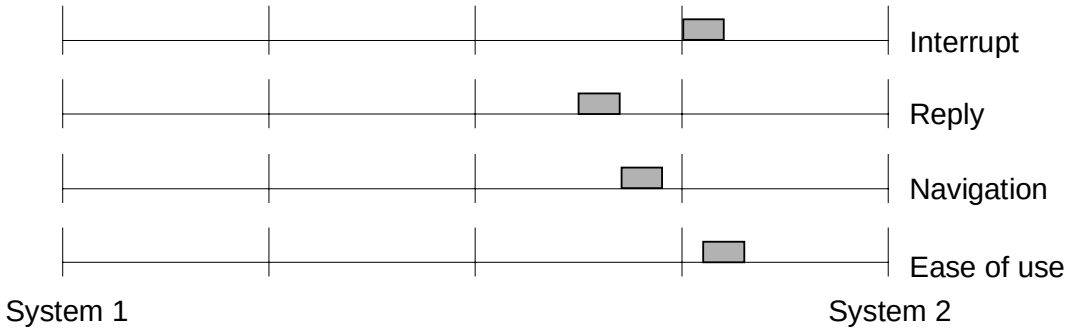

Figure 1: Mean preference score for system 1 and 2 on a 5-point scale

CONCLUSIONS AND DISCUSSION

In general, the results show that users made use of the e-mail reader only when they had no access to their regular e-mail box on a PC. The speech synthesis quality in this application is scored lower than in previous research by the users. As expected, the perception of the quality of speech synthesis in a real application is not completely comparable to the laboratory situation. Users would prefer to hear different voices being used for the prompts compared to the actual message, since this would make it is easier to process the messages, and would avoid the user from getting lost. Also, it appears that they preferred the pre-recorded voice for the prompts. With respect to navigation and ease of use, most users preferred speech recognition to DTMF. Further, we can conclude that ease-of-use is a very important issue. Users did not like the menu-driven dialogue and the time needed to receive their messages. An acceptable level of ease-of-use can, to some extent, be achieved given the present technology, by using a limited amount of functionality.
When we compare the two systems, we can conclude that in general system 2 was preferred over system 1. The user interface of system 2 was preferred, whereas the speech synthesis of system 1 was given preference. Apparently, speech synthesis is not the only dominant factor. User interface clearly influences the perception of quality of the email reader. However, the exact importance of speech synthesis and user interface, and the interaction between these two needs to be investigated separately.

REFERENCES

[1] Rietveld, T., Kerkhoff, J., Emons, M., Meijer, E., Sanderman, A. Sluijter, A. (1997). Evaluation of speech synthesis systems for Dutch in telecommunication applications in GSM and PSTN networks. Eurospeech'97 (Rhodos).

[2] Sluijter, A., Bosgoed, E., Kerkhoff, J., Meijer, E., Rietveld, T., Sanderman, A., Swerts, M., Terken, J. (1998). Evaluation of speech synthesis systems for Dutch in telecommunication applications. In Proceedings of the third ESCA/COCOSDA International Workshop on Speech Synthesis, Jenolan Caves, Australië, 1998

.