An Audio System of Electronic Texts for the Visually Impaired People and Perception of Different Speech Rates by the Blind and the Sighted

Meelis MIHKLA a,1, Indrek HEIN a, Indrek KIISSEL a, Margit ORUSAAR b and Artur RÄPP c

a Institute of the Estonian Language
b Estonian Library for the Blind
c North-Estonian Union of the Blind

Abstract. This work presents the audio system of electronic texts and audio books, designed for the visually impaired people, helping them read the news, newspapers, magazines and books and listen to audio books over the Internet. Considered are the possibilities of use and the functional range of the system. For selection of various speech rates perception tests were arranged for the blind and those sighted. The perception tests revealed that the so-called trained blind prefer to listen to the speech of a higher rate than the sighted or the blind, lacking the experience of daily use of computer.

Keywords. audio system, visually impaired people, speech rate, perception test

Introduction

The goal of our joint project is to create an unitary audio system of listening electronic information texts and audio books in Estonian on the basis of the Estonian Library for the Blind. With the help of the audio system of information texts http://ehr.like.ee/ the blind and the visually impaired can have read to them the news, newspapers, magazines and books and listen to audio books. Essential goals of the on-line system are involvement of the broadest possible range of users and the simplicity of use, so that the reading system could be used not only in home computers but also in libraries, Internet public access points, vehicles etc. Handling of the audio system has been made maximally simple for the blind and the visually impaired, so it does not call for special time-consuming training or long experience of use of computer. Selection of press and books, browsing and navigating in the sound library can be performed with the help of only the number keys and some function keys. The software created enables browsing and seeking the material available in the server of the information literature, selecting the reading and listening

1Corresponding Author: Senior Researcher, Institute of the Estonian Language, Roosikrantsi 6, Tallinn, 10119 Estonia; E-mail: meelis.mihkla@eki.ee.
rate and the synthetic voices. Two TTS synthesizers were used: diphone synthesis [1] using an Mbrola engine and unit selection synthetic voices [2] developed by means of the Festival system. Development of possibilities of information technology accommodated to the visually impaired will improve their access to necessary information, create to the visually impaired students new opportunities at teaching effort, enlarge the extracurriculum opportunities of the blind and the visually impaired, help pass leisure time in a more meaningful way by bringing the sound library to home, contribute to job-related development etc.

In the process of the use of the first version of the system it was found that many blind people wish to listen to the news and newspaper articles at a speech rate significantly higher than the normal rate. As the system is server-based, it was not possible to provide to users the arbitrary possibility of change of the speech rate, which makes the system too cumbersome and slow. Hence the need to find some optimal rates to supplement the user menu with two speech rates from the faster-than-normal range, e.g. fast and very fast. Perception of a rapid rate of speech and the limits of its temporal compression has been discussed in several studies (e.g. [3,4]). It has been found that very fast speech is preferred and perfectly understood by trained people, i.e. those with an everyday experience of a screen reader and a speech synthesizer. In the study [5] the brains of blind and of sighted subjects were scanned while they were listening to rapid speech; it was found that in the blind an ultra-fast rate was systematically accompanied by activation in the brain zone that in the sighted is used for processing visual information.

Is such a phenomenon general and how many blind have developed that ability? An endeavour was made, to obtain clarity to this and many other questions related to the speech rate, in the perception test arranged to the blind and the sighted for assessment of different speech rates.

1. Audio system of electronic texts and sound books

1.1. Background

In Estonia there are about 2000 visually impaired people, whose visual handicap prevents their use of traditional devices of procurement of information necessary for daily life. Ca. 500 among them, use actively the computer with the support of available accessories (magnifying and various speech software). The Estonian Library for the Blind has in its holdings ca. 900 sound books. For procurement and exchange of information with the help of computers the visually impaired people mostly use various screen readers (Jaws, Window-Eyes, ZoomText etc.) and speech synthesizers. The screen readers are rather expensive and they must be bought by everyone personally. The goal of our project was to create an unitary audio system of reading the electronic information texts and audio books in Estonian, which would not necessitate buying expensive software. The main goal of the audio system is to create a possibility to benefit from the electronic media also by those blind and visually impaired, who do not use the computer on a daily basis.

Figure 1 presents the website of the system, where it is possible to download the electronic audio system, to listen to the instruction of use, whereas all users can upload into the system various electronic texts (handbooks, news, pieces of information etc.). Such on-line system enlarges the possibilities of access of the blind and the visually
impaired to the necessary information, because it can be used in vehicles, in the wireless Internet area, libraries and public Internet points, etc.

1.2. Functionality of the system

An endeavour has been made to make handling of the audio system as simple as possible to the blind. Selection of all news stories, articles or books, browsing of the newspaper and navigating in the sound library can be performed with the help of only the number keys and some function keys. It is only when searching the books and logging in that the letter keys must be used. The pressures of key are doubled with the sound file. When entering the system, the main menu on the screen is read out loud and it is also possible to listen to the instruction of use of the system (see Fig. 2).

Reading of news and certain newspapers is free to everyone. For reading the books and listening to audio books and magazines one must log in, while respective passwords is distributed by the Estonian Library for the Blind to all its registered users. To enter the sub-menus, a respective number key must be pressed; hence the menus may have up to nine sections. For instance, if there are eighteen chapters in the book, chapters one to nine can be selected with the first level menu (number key 1) or chapters ten to eighteen (2), and in the menu of the next level, with the renewed pressure on the number key the concrete chapter can be selected (for instance for selection of chapter 13 one must first press on key "2", and thereafter on key "4"). Movement back to the higher level menu can be effected by pressure of key BACKSPACE or ESC.

The system contains the books in two forms: as an electronic text and as a audio book. The speech synthesizer reads the books as a text, generating the sound files, the sound books have been read by professional speakers and the respective sound files are already in server. The information exchange with computer of user of the server is per-
formed in mp3-format. It is possible to listen to the presentations of new books having arrived in the library in the previous and current years and added to the system, to seek a concrete book by author or part of the title. In the audio system certain function keys have specific meanings:

- F6 – search for a book
- F7 – change of the speech rate
- F8 – selection of voice of the synthesizer
- F9 – logging in
- F10 – programme information
- F11 – reading of the instruction for use
- F12 – information of time of the clock

Besides that, it is possible to navigate in the audio file by special keys: the blank space stops the reading, at renewed pressure on the blank space the reading continues, with the left arrow key < it is possible to move back in the sound file for 20 seconds and with the right arrow key -> 20 seconds forward, BACKSPACE and ESCAPE terminate the presentation of the given sound file and revert to the previous menu. By key F8 it is possible to select various synthetic voices, the system contains one diphone synthesizer [1] and 2-3 different unit selection voices [2]. It is also possible to change the speech rate: 1 – normal, 2 – rapid and 3 – very rapid.

2. Perception of the speech rate

In the process of the use of the first version of the system it was found that many blind people wish to listen to the news and the newspaper articles at a speech rate significantly higher than the normal rate. With the server-based system it is not possible to provide to the user all various values of the speech rate. Hence the need for perception tests of the speech rate arose. Although the on-line audio system is meant in the first place to the visually impaired, the sighted too were involved in the perception test of speech rate for comparison. Such comparative basis helps find an answer to the questions: "What speech rates are preferred by the blind and what are preferred by the sighted?", "Does the myth
of preference accorded by the visually impaired to the ultra-high speech rate hold?”, "Is there an optimum speech rate in existence?".

2.1. Subjects

The test was attended by 58 blinds or people gravely visually impaired (29 women and 29 men, aged 14-79) and 56 sighted (41 women and 15 men, aged 18-58). The native language of all participants was Estonian.

2.2. Material of the test

Stimuli of the perception test of the speech rate were generated on the basis of two sound books (T. Dreiser’s "American tragedy", read by male voice and E. Stein-Fisher’s "Das erste Mal und mehr" read by female voice, both in Estonian translation) and passages of a news story read by synthesizer. The speech synthesizer was the synthesizer text-speech in Estonian basing on diphones [1], using the MBROLA synthesis engine. When generating the synthesized voice two different prosody modules were used: rule-based (SYNT1) and statistical (SYNT2). It was done for verifying the opinion of some visually impaired people that the quality of speech of synthesizer with statistical prosody module deteriorates at higher speech rate.

The natural reading rate of the female voice involved in the test was on average 135 words per minute, of the male voice 122 words per minute. Synthesizers were tuned to meet the speech rate of the female voice. For every voice, there were generated eight 35-55-second speech samples of different speech rate (see Table 1).

The limits of time compression of the speech were previously coordinated with "trained persons", who have every day the listening practice of the synthesized speech. Upon their opinion the speech rate twice as high as the normal is the maximum rate, at which they could be listened to literature. In the process of preparation of the test some elderly visually impaired people expressed an opinion that they would listen to some sections also at a lower rate than the normal one. Therefore there were additionally generated two lower-rate speech stimuli (0.8 and 0.6 times slower than the normal rate). The rate of the samples from audio books was regulated by means of signal processing program Adobe Audition 3 for high precision time compression with time stretch (preserves pitch).

The stimuli of the speech rate were presented to the people tested in random sequence by voice series. The suitability of the speech rate was to be assessed in the five grade system (5 – the most suitable speech rate, 4 – the good speech rate, 3 – the speech of bearable rate, 2 – the unpleasant speech rate, 1 – the unsuitable speech rate, i.e. incomprehensible speech or the rate too low/too high).

<table>
<thead>
<tr>
<th>81</th>
<th>108</th>
<th>135</th>
<th>162</th>
<th>189</th>
<th>216</th>
<th>243</th>
<th>270</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>80</td>
<td>100</td>
<td>120</td>
<td>140</td>
<td>160</td>
<td>180</td>
<td>200</td>
</tr>
</tbody>
</table>
2.3. Results

Fig. 3 presents the average scores of the blind and the sighted for speech rates of different voices. The left-hand graph provides the assessments to the male voice and synthesized speech SYNT2, the right-hand graph assessment to the female voice and synthesized speech SYNT1. On the basis of the graphs of darker line it may be said that the blind prefer mostly the rates in the interval 1.1–1.4, which are 10–40% higher than the normal rate. The maximum of the preferences of the sighted keeps within the confines 1.0–1.2, i.e. either the normal speech rate (135 words per minute) or up to 20% higher than the normal rate.

In case of the male voice both the sighted and the blind assessed as the best the speech rate 1.2. Such assessment evidently derives from the normal rate of the male voice, which is about 10% slower than other voices (122 s/min versus 135 s/min).

Although the participants in the test were requested to assess only the suitability of the speech rate, not the pleasantness of the voice, the synthesised speech nevertheless scored almost one grade less than the human speech. However there were no noticeable
differences in grades caused by synthesizer’s prosody module at higher speech rates in SYNT1 and SYNT2. Thus the results fail to support the idea that the quality of the SYNT2 voice might deteriorate at quicker rates. It was revealed in the contrary that in case of SYNT2 the higher speech rates scored higher (compare the graphs SYNT1 and SYNT2 in Fig. 3).

More generally, the results of the test showed that the average assessments of the blinds and the sighted to speech rates differ much less than one could envisage when designing the test (compare the two left-side diagrams of Fig. 4). Among the blind however there were many of those, who use the computer rarely or never do it. Hence they lack the experience for listening to the synthesized speech of different speech rate. Three right-hand diagrams of Fig. 4 present the average assessments to speech rates in dependence on previous experience in years of use of computer and screen reader of the visually impaired. The results clearly evidence the tendency that together with the growth of the so-called trainedness or longer listening practice of the synthesized speech the visually impaired develop the ability to understand the high rate speech. However that ability is not manifested in everybody. Among the test group of the blind there were 11 people (among 19% of the visually impaired tested), who had a clearly developed ability to understand the high rate speech. The speech rates preferred by them were 2.0 (twice as high as the normal rate) or 1.8 times speech rate. The experience of use of computer and the screen reader by all those "trained blind", exceeded five years. A separate research is needed to find out whether a certain brain area activates in them, in order to understand the high-rate speech.

The so-called “optimum speech rate” could not be found by analysing the test results either for the blind or the sighted, because the preferences to the speech rate were very individual and depended on very many circumstances. It was decided, however to add to the selection of the speech rate of the audio system the higher speech rate 1.3 (176 words per minute) and a very high speech rate 1.6 (216 words per minute). The very high speech rate applied may not always satisfy those “trained” persons who can understand the ultra-high speech rate. Since those persons numbered below one fifth, the offered speech rate is the so-called compromise value between the regular and skilful computer users. Because the speech rates lower than the regular rate (0.6 and 0.8) scored rather low by both the blind and the sighted, it was decided not to apply in the audio system those speech rates.

3. Conclusion

Audio system enlarges access of the visually impaired to necessary information. Simplicity of use enables, through Internet, to benefit of various electronic texts and sound library also by those blind not having daily experience of use of the computer. At the perception test of the speech rate it was not possible to obtain unequivocal answers to all questions posed. The preferences of the blind and the sighted to the speech rate displayed certain differences however the size of difference does not depend so much on the visual impairment as rather on experience of use of computer and screen reader. The ability to understand the high rate speech may generate by daily use of computer and listening to synthesized speech after the fifth year. Among our test group of visually impaired people such capacity was manifested in every fifth blind person. Although we did not manage
to establish the optimum speech rate suitable to all in the perception test, we applied in the audio system besides the regular rate two higher speech rates: high rate, 30% higher than the normal rate and very high rate 60% higher than the normal rate.

Acknowledgements

This work is supported by National Programme for Estonian Language Technology, grant ETF7998 and project SF0050023s09.

References