Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 3, March 2014, pg.595 – 602

RESEARCH ARTICLE

New Touch Screen Application to Retrieve Speech Information

J.Rajeswari[1], E.Thanga Selvi[2]

PG Scholar, Department of ECE, PSNA College of Engineering and Technology, Dindigul, TamilNadu[1]

Associate Professor, Department of ECE, PSNA College of Engineering and Technology, Dindigul, TamilNadu[2]

rajeswarijeyaraj@gmail.com [1],e.thangaselvi@gmail.com [2]

Abstract- An adaptive speech rate control technology for ultra fast listening that is equivalent to skimming is described. Nowadays, listening to audio books on mobile devices is quite common. People read books at various levels of detail from close reading to skimming. Although a similar feature to skimming is required to efficiently obtain information from audio sources, there is no tool equivalent to skimming for audio playback. Therefore a new speech rate conversion method is developed to efficiently obtain information from audio sources with very fast replay. This algorithm will help not only sighted people to enjoy audio books but also visually impaired people because almost all of their information is obtained from speech. Thus, the implementation of this technology on special audio players for visually impaired people as a new replay function is expected to be useful. Moreover, this technology should be useful for all audio book listeners, not only people with limited sight. A new touch screen application is developed for consumer use.

Keywords- Audio book, Mobile Phone, Speech Rate Conversion, Visually impaired, Text to Speech Conversion

I. INTRODUCTION

We read printed books and magazines at various levels of detail. The slowest case may be close reading, where attention is paid to each word and phrase. On the other hand, the fastest case is skimming, which is used to grasp the overall meaning of text in a short time. There is a similar need when obtaining information from audio sources such as audio books, DAISY (Digital Accessible Information System) talking books, and so forth. However, current audio playback systems have no function equivalent to skimming. Therefore, an algorithm is developed for a method of efficiently obtaining information from speech content that is equivalent to skimming printed books. Moreover, this algorithm is implemented in a popular touch screen application for consumer use. Fig. 1 shows the concept of "speech skimming". The ultimate goal of this research is to develop a fast listening algorithm that everyone can use.

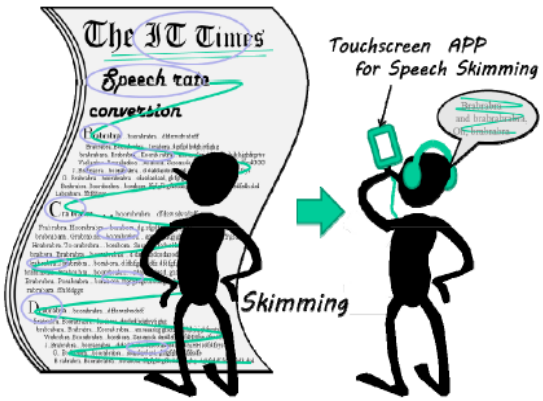

Figure 1: Concept of fast listening method comparable to visual skimming

Therefore, developing an ultrahigh-speed speech playback technology to blind and visually-impaired people, who are more dependent on audible information and obtain a great deal of daily information from speech. They commonly use talking books played at a high speed to obtain information on printed materials such as newspapers and magazines. Speech, however, is a form of time-series data, and many people feel that obtaining an overall picture of information being presented is more difficult with speech than with text-based information. This is often quite a barrier for visually impaired people. For this reason, many people increase the normal playback speed of talking books or screen readers to jump forward, and many people would like to listen at even higher speeds if possible.

II. METHOD

A.Text to speech conversion

A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech

A text-to-speech system is composed of two parts. Front end and back end. The front end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written words. This process is often called text normalization, pre-processing, or tokenization. The front end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called text-to-phoneme or grapheme-to-phoneme conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end. The back end often referred to as the synthesizer then converts the symbolic linguistic representation into sound. In certain systems, this part includes the computation of the target prosody which is then imposed on the output speech.

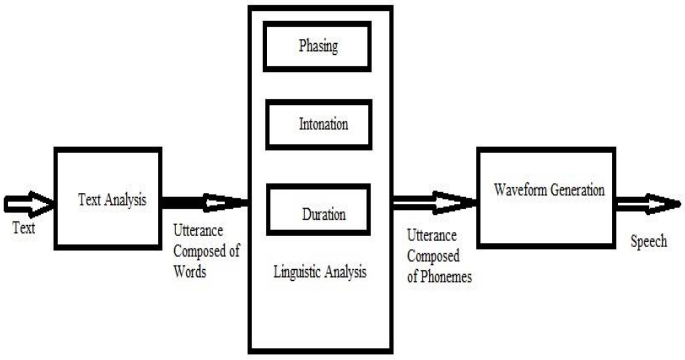

Figure 2: Typical TTS System

III. ALGORITHM

When saying something important, people raise the pitch or power of their voice or both. The basic principle of this algorithm is to retain parts of speech that are likely to have a significant effect on listening comprehension. For example, if both the power and the pitch of a voice are relatively low in a particular section of a continuous utterance, it can be removed because it is expected to be difficult to hear in the original recording, even if played back at normal speeds. So proactively delete it and assign the time to another section of the utterance. Since this operation would not be practical unless it were fully automatic, only specific acoustical features that can be obtained by audio analysis in real time must be used in this process.
From these viewpoints, this section describes the algorithm used in adaptive "SRC" method. This algorithm observes time fluctuations of the power and F0 of the speech signal and deletes sections in which they fall below a certain threshold. Moreover the remaining speech signal also has its speech rate converted not uniformly but adaptively in accordance with the typical and smoothed F0 movement of the utterance. The adaptive "SRC" method was designed on the basis of the following factors:

1) The key to simple listening is to catch the beginning of each utterance correctly.
2) Sections with locally elevated pitch compared with neighboring sections have particular significance.
3) Sections where both the power and the pitch are lower, especially near the ends of sentences, may be less important for comprehension.

The adaptive "SRC" algorithm works by partially repeating or deleting by the pitch period from a speech waveform. In silent or unvoiced portions, a pseudo pitch is also extracted on the basis of the peak of the autocorrelation function, and sections of the corresponding length are repeated in the same way as in voiced portions. Fig. 3 shows the basic principles of the SRC method base on waveform analysis and synthesis. The extension ratio of speech and non-speech are set on demand, and these parameters can be controlled together in a single motion simultaneously.

SRC using the proposed adaptive speech rate function is called "adaptive SRC", and control using a linear rate function that changes the speech rate uniformly is called "linear SRC".

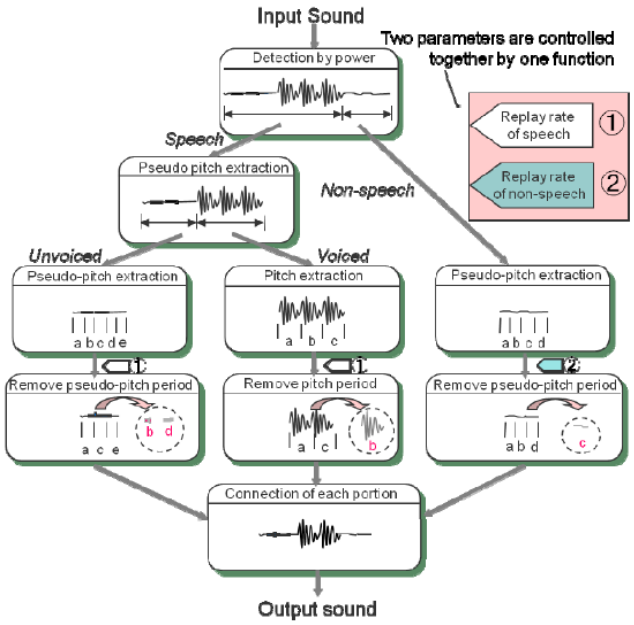

Figure 3: Basic principles of SRC method based on waveform analysis and synthesis

A. Adaptive SRC function

This function gives a variable boosting ratio of *rp*, which is the reciprocal of the required replay rate, and this is carried out by time-series processing. Equation (1) is the proposed adaptive SRC function, which decreases monotonically and smoothly in a given time that is the quasi prediction time of a breath group. *R (t)* changes from *rs* to *re* continuously inside every breath group in a given time.

$$R(t) = r_e + (r_s - r_e)1/2[\cos\{\pi (t-t_0)/T\} + 1.0] \qquad (1)$$

Here, *rs* (>1.0) is the first boosting rate at the beginning of a breath group (*T*=0), and *re* (=1.0) is the rate after *T* [ms] (*T*=2500). However, the duration of the utterance is unpredictable; thus, adopt a fixed time interval *T* on the basis of the typical breath length. When the speech continues after *T*, the rate *re* is applied continuously until the end of the speech. In the case of replaying at *rp* times the normal speed, pitch periods are inserted or deleted so that the length of the waveform *l (n)* [ms] from the beginning of the utterance to the *k*th pitch period *Pl (k)* should be as follows.

Let l(0)= 0 (2)

$$l(n)= r_p^{-1}.\sum_{k=1}^{n} pl(k).R(l(n-1))$$

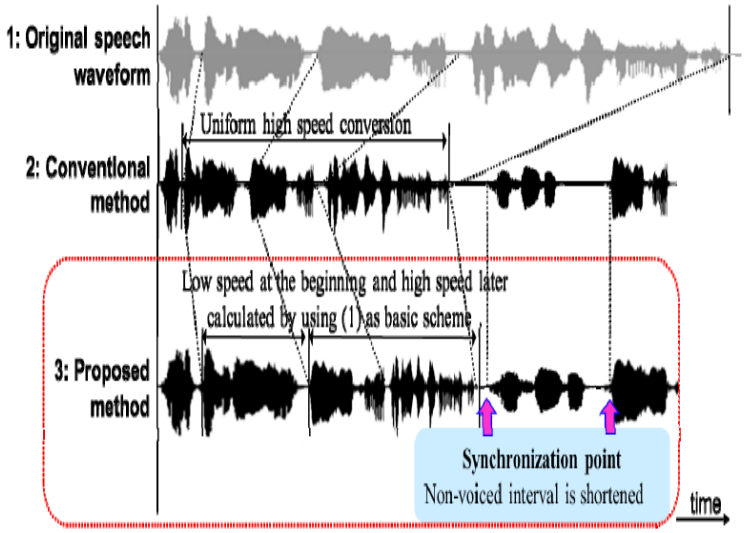

Figure 4a: Example of basic operation of adaptive technology

The beginning of each utterance is detected on the basis of the duration of the preceding pause in the speech. Here, the threshold of the original speech was set at 200[ms], the start of speech following a pause of at least this duration is regarded as the beginning of the utterance. An example of the basic operation of the adaptive SRC technology is shown in Fig. 4a. Both contracted waveforms (2 and 3) were converted at twice normal speed. The upper waveform was converted by the linear SRC function, and the lower one was converted by the adaptive SRC function. These waveforms clearly have the same total duration.

B. Adaptation

Equation (1) is used to give the impression of slower speech at the beginning of each breath group. Additionally, emphasized portions of the speech are slightly slowed down (Fig. 4b).Therefore; increase the degree of expansion in portions where F0 increases abruptly, as detected by comparing the current value of F0 with the moving average from the beginning of the utterance of each breath group. This rate of increase *rinc* is defined by equation (3) as a percentage of *x* [%].

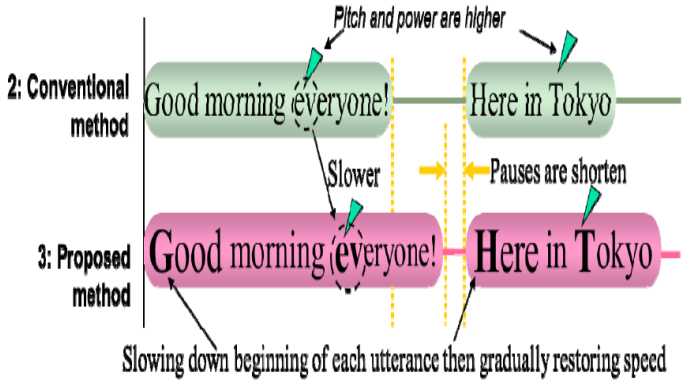

Figure 4b: Example of additional operation of adaptive SRC technology

$r_{inc} = 1 + x/100$ (3)
$(10 \leq x \leq 20)$

The adaptive SRC applies an expansion rate *R'(t)of* up to 20% to the default rate *R(t)* in emphasized portions (at time *t0*) for several pitch periods of α [ms] after t0. *R' (t)* is defined by equation (4).

$R'(t) = r_{inc}\, R(t)$ (4)
$\alpha = 400\, r_{inc}$
$(t_0 \leq t \leq t_0 + \alpha)$

As illustrated in Fig. 4b, the speech rate is reduced at the beginning, and gradually increases in the latter half of the phrase. In addition, segments in the speech where F0 and the power are higher are extracted and slowed down while segments where F0 and the power are lower are speeded up.

In this process, the cumulative time delay was observed at all times. The short time power is used to judge when it is permissible to curtail a portion of speech in the event that a time delay has accumulated. Sections of speech with both low F0 and low power are considered to contribute less to the intelligibility of the speech especially near the ends of sentences. When the speech is too understand (around three times the normal speed), curtail such sections, even though they were not wholly silent, so as not to increase the time delay. An example of curtailing operation of adaptive SRC technology is shown in Fig. 4c.

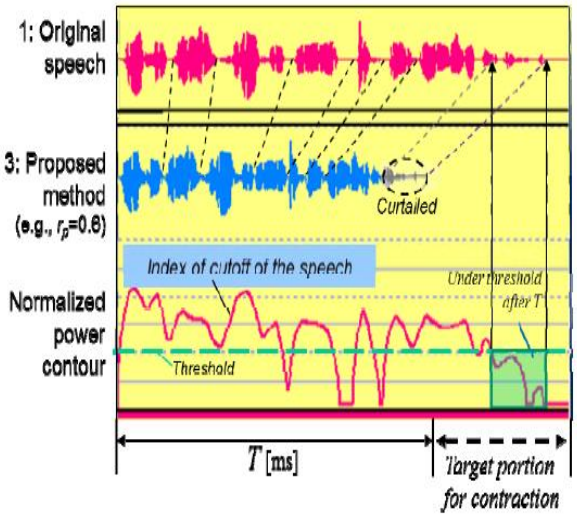

Figure 4c: Example of curtailing operation of adaptive SRC technology

The threshold (in Fig. 4c) or which the speech was set to 50% of the moving average of the power. In the case where the moving average of the under the threshold after a predetermined time T (e.g., 2000 [ms]), the corresponding part is curtailed adaptively. *T* and the threshold are not necessarily single values. These values should be flexible to meet requirements of various listeners. Nevertheless, if the cumulative delay exceeds the set value, *rs* is temporarily decreased. Conversely, if the cumulative remains below the set value, *rs* return to its original value.

IV. APPLICATION

The number of operational parameters was limited to two (Fig. 5): "play rate" and "start rate." If the start rate is set at 1.0, the speech rate is controlled uniformly by the value of the play rate, the same as in the conventional method.

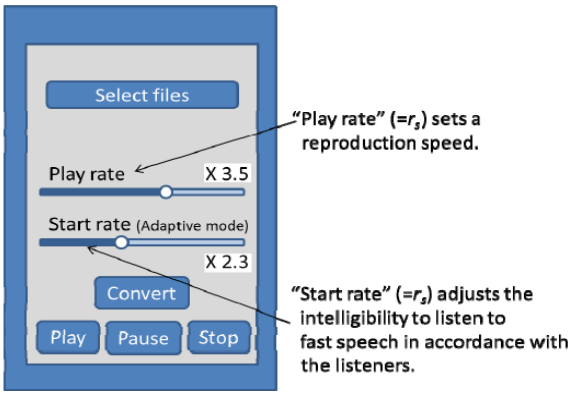

Figure 5: Two simple operational parameters

If the reproduction rate is set to 2.0, the speech is reproduced in half the time of the normal replay. That is, the play rate determines the total replay time. On the other hand, the start rate broadly adjusts the intelligibility for listening to ultra fast speech. The high value of the "start rate", is large than the difference between the faster portions and slower portions. That is, when this mode is used together with the play rate, generated fast and slow portions are automatically in the utterance. These two parameters can be combined easily on demand depending on the content that the user wants to listen to. This technology is developed as a touch screen application.

Of course, this is not assuming itself will become popular with visually impaired immediately. It is intend to provide this algorithm for their special audio players as a useful fast replay functions. However, the touchscreen devices a very convenient tools as a part of multimedia information terminals. Therefore, a convenient touch screen interface for this application will soon be developed for visually impaired.

V. SIMULATION

Hardware Simulation is done in Proteus Software and Voice conversion is done in Visual Basic6. In hardware simulation instead of ARM7 Processor Atmega16 controller is used and instead of ps/2 keyboard text is typed to virtual display port in serial communication and displayed in LCD using microcontroller and speech is done through visual basic. Virtual serial port emulator software is used to emulate the serial port for simulation. Text is received in VB through serial port.

From microcontroller data's are transferred to serial port in Isis software and from that serial port using virtual serial ports emulator data's are received in visual basic and the text is converted into speech. COM2 port is configured both Isis and VSPE and VB.

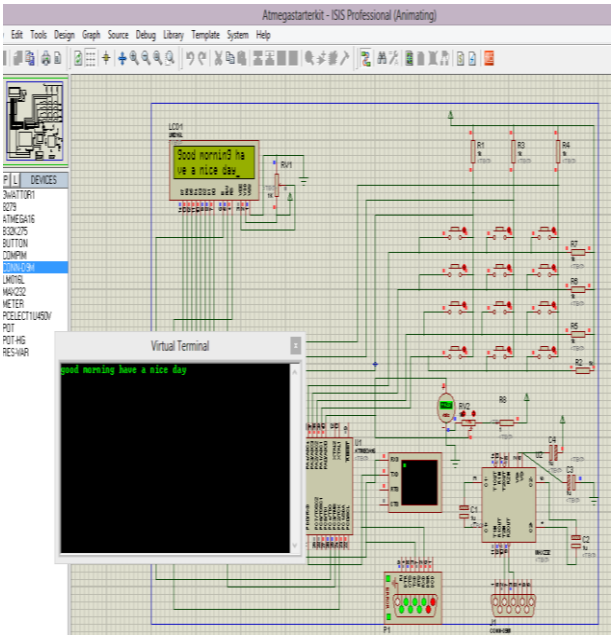

Figure 6a: Simulation Result 1

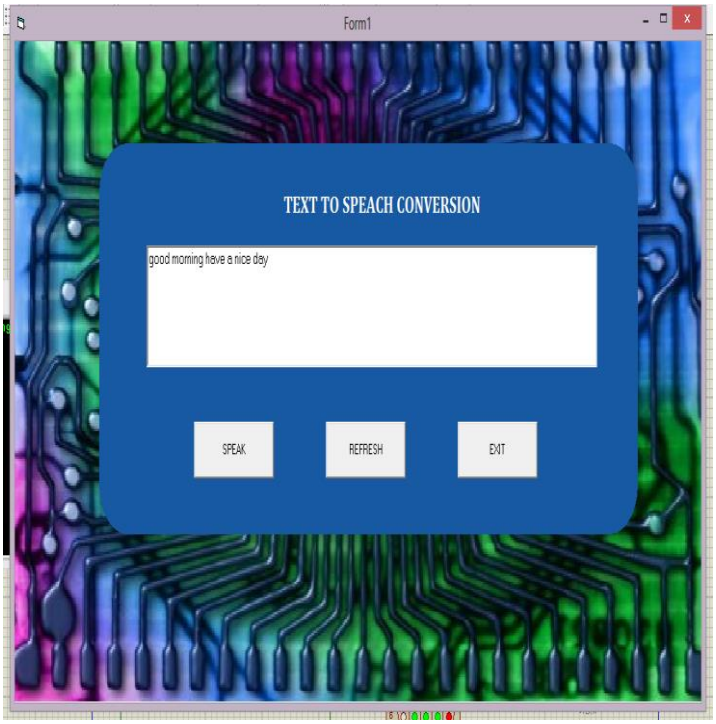

Figure 6b: Simulation Result 2

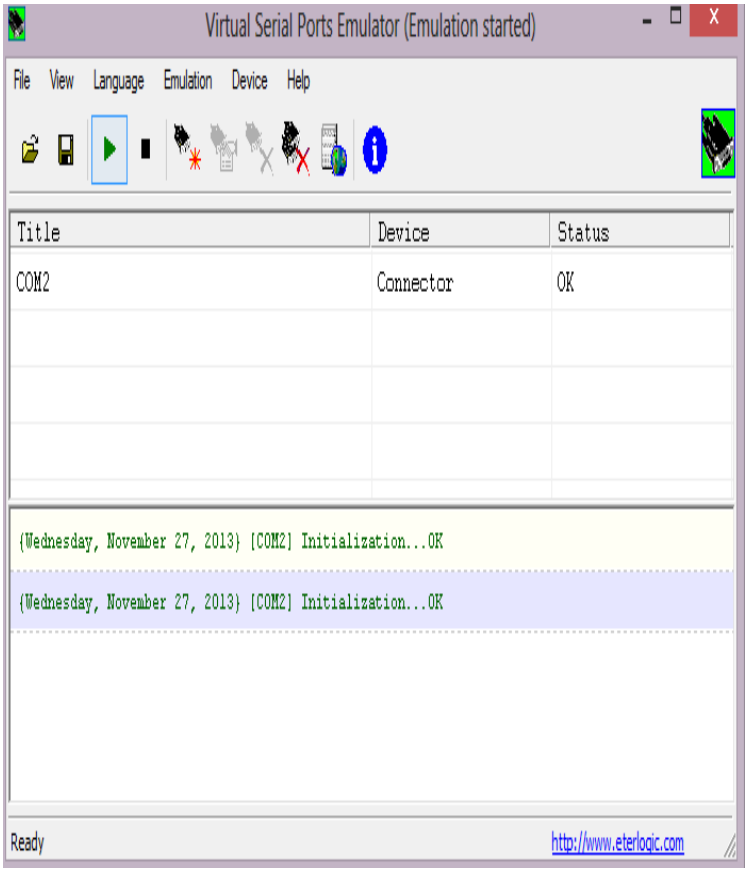

Figure 6c: Simulation Result 3

VI. CONCLUSION

An algorithm for intelligible high speed listening is investigated and developed a new touch screen application to retrieve speech information efficiently. The result shows the text to speech conversion method. The major part is text to speech conversion. In future, implementing the concept in hardware and developing a touch screen application for consumer use. The application is useful for visually impaired people.

An adaptive speech rate conversion method gives users an impression of slower speech even at playback speeds similar to those used in conventional methods and that the proposed method is effective for speech-rate factors greater than two interms of "simple listening".

REFERENCES

[1] M. Furini, "Fast Play: A Novel Feature for Digital Consumer Video Devices," *IEEE Transaction on Consumer Electronics*, Vol.54, No.2, pp. 513-520, May. 2008

[2] A. Nakamura, N. Seiyama, A. Imai, T. Takagi and E. Miyasaka, "A New Approach to Compensate Degeneration of Speech Intelligibility for Elderly Listeners," *IEEE Transaction on Broadcasting*, Vol.42, No3, pp. 285-293, Sept.1996

[3] N. Tazawa, S. Torihara, Y. Iwahana, A. Imai, N. Seiyama, and T.Takagi, "Rapid Listening of DAISY Digital Talking Books by Speech Rate Conversion Technology for People with Visual Impairments", in *Proc. ICCHP(1), 2010, pp.62-68*

[4] S. Torihara, "Oblique Listening System – Speed-reading System for the Visually Impaired using Syntactic Information, Technical Report of IEICE, 5th Meeting of the Technical Committee on Well-being Information Technology", Nov. 2000. (In Japanese)

[5] C. Asakawa, H. Takagi, S. Ino, and T. Ifukube, "The Optimal and Maximum Listening Rates in Presenting Speech Information to the Blind, Journal of Human Interface Society", Vol.7, No.1, pp. 105-111, 2005. (In Japanese)

[6] T. Takagi, N. Seiyama and E. Miyasaka, "A method for pitch extraction of speech signals using autocorrelation function through multiple window-lengths", IEICE vol. J80-A No.9 pp.1341-1350 Sept. 1997. (In Japanese)

[7] A. Imai, R. Ikezawa, N. Seiyama, A. Nakamura, T. Takagi and E. Miyasaka, "An Adaptive Speech-Rate Conversion Method for News Programs without Accumulating Time Delay," IEICE Transactions A, Vol. J83-A, No. 8, pp. 935-945, Aug. 2000. (In Japanese)

[8] M. Sugihara, "A speech metrical theory of standard Tokyo dialect." The NHK Monthly Report on Broadcasting Research, pp.76-90 Apr.2011

