3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

A PORTABLE DIGITAL SPEECH-RATE CONVERTER AND ITS EVALUATION BY HEARING-IMPAIRED LISTENERS.

Yoshito Nejime, Toshiyuki Aritsuka*,
Toshiki Imamura**, Tohru Ifukube**, and Jun'ichi Matsushima****

* Central Research Laboratory, Hitachi Ltd., Kokubunji, Tokyo 185, JAPAN.
** Research Institute for Electronic Science, *** School of Medicine,
Hokkaido Univ., Sapporo 060, JAPAN.

ABSTRACT

A new portable device that reduces speech speed in real time has been developed to aid hearing-impaired listeners. The device employs real-time digital signal processing to expand the time scale of the speech signal without changing the pitch. Seven out of ten elderly hearing-impaired listeners have shown improvements in word discrimination tests when using this speech-rate conversion. The subjects' temporal resolution showed a correlation with the observed improvement, with a coefficient value of 0.859. The results suggest that this device provides support at the cognitive level of the auditory system.

1. INTRODUCTION

Hearing-impaired listeners, especially elderly people, often have difficulty in comprehending fast speech. Some studies have suggested that temporal processing factors other than peripheral hearing loss contribute to this difficulty [1]. However, conventional analog hearing aids are not suitable for the temporal manipulation of speech signals. Although the latest digital hearing aids employ digital technology, they use it mainly for fine fitting of the frequency characteristics, noise reduction, or acoustic feedback cancellation [2]. This paper presents a new approach to hearing aid design based on the time scaling of speech signals without changes in the pitch.

Several studies have suggested that time scaling of speech signals for hearing-impaired listeners is highly effective. However, the efficiency of digital signal processing for time scaling has not been demonstrated [3]. This may be because of extra distortion caused by the uniform processing algorithm for the entire speech signal. Our new device uses a dynamic processing algorithm to overcome this problem.

Recently, a speech-rate conversion system especially designed for television sets has been proposed [4]. This system provides very high quality slow speech. Our portable device can be used with an internal microphone and an external input terminal. This feature allows users to listen to any speech sources.

2. METHODS

2.1 Speech processing methods

The digital speech-rate converter shown in Figure 1 includes a 16-MIPS Digital Signal Processor (DSP), a 16-Mbit semiconductor memory, and 14-bit 13.3-kHz analog-digital interfaces. The device can work with a 5V battery. The total size of the device is 100mm x 180 mm x 20 mm.

The speech signal processing flow is illustrated in Figure 2. The time-scaling procedure expands the time scale of the input speech using two 48msec-long input frame-buffers alternately. The output ring-buffer holds the expanded speech signal. Since

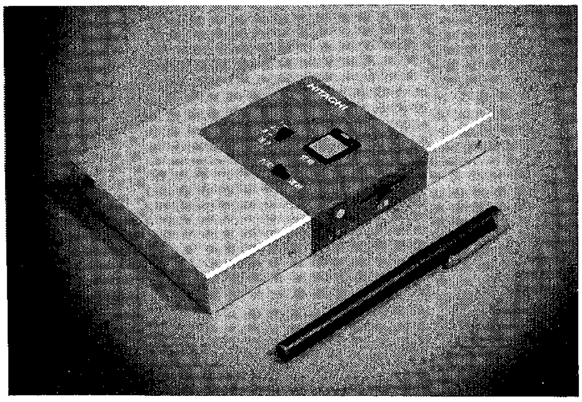

Figure 1. Digital speech-rate converter

the output buffer is circularly addressed, it keeps only data corresponding to the time difference between input and expanded output speech. The time-scaling procedure consists of two sub procedures, the power judgment and the wave expansion.

The operations of the sub-procedures are illustrated in Figure 3. The power judgment procedure calculates the power of each 48msec-long frame and makes comparison with the threshold value, **Th**. The wave expansion procedure processes only the duration of time whose power is larger than **Th**. Data in the frame whose power is smaller than **Th** are transferred to the output ring-buffer directly. Therefore, the procedures mainly manipulate voiced parts of the input speech signal. This operation avoids extra distortion on small power consonants and semivowels in speech. In addition, to shorten the delay between input and output speech, the power judgment procedure eliminates silences longer than 1 second by making comparisons with another threshold value, **To**. These two threshold values are often modified dynamically depending on the maximum and minimum frame power in the latest few seconds.

As illustrated in Figure 3, the wave expansion procedure consists of 4 parts, (a) pitch length detection based on the auto-correlation method, (b) forward wave transfer, (c) insertion of reproduced pitch pattern, and (d) rear wave transfer. The pitch pattern reproduction procedure produces a 2-pitch-long wave by using triangular weighting windows. The length of the windows is always adjusted to the pitch length detected just before the reproduction procedure. By this frequent adjustment, distortions caused by the difference between actual pitch length and the window length is minimized. Changes in the number of pitches **n,** in the rear wave transfer procedure permits three different expansion rates of 1.25, 1.33, and 1.50.

2.2 Methods of evaluation

To evaluate the efficiency of the proposed speech-rate conversion, speech recognition performances of hearing-impaired listeners with mild-to-moderate sensorineural hearing losses were investigated.

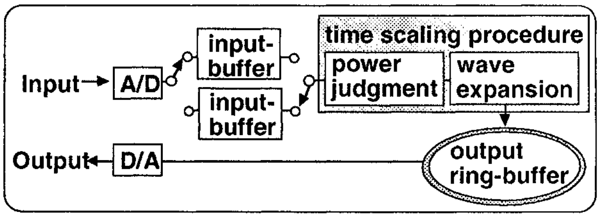

Figure 2. Speech signal processing flow.

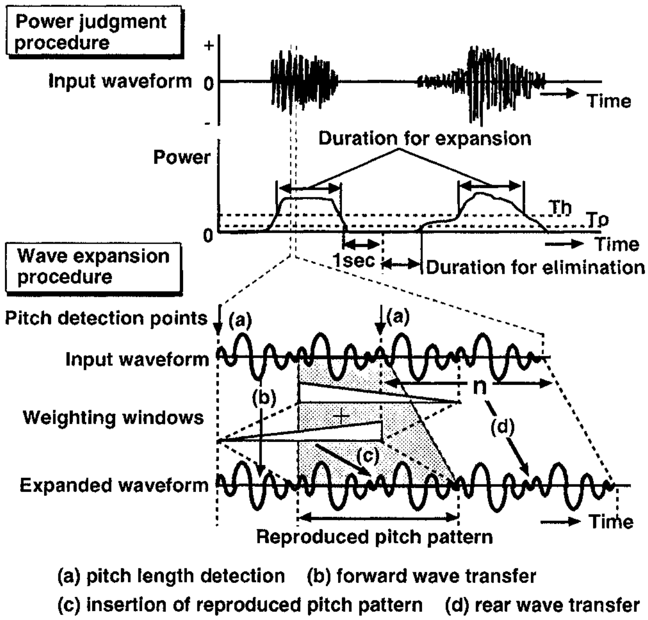

Figure 3. Time-scaling signal processing procedure.

The subjects' time gap detection ability was measured as their auditory temporal resolution. Two bursts of a single tone were repeated at several time gaps Δt; the smallest detectable **Δt** was taken to be the subjects' auditory temporal resolution. The frequency of the single tone was 1 kHz, the value of **Δt** for young normal listeners was about 13 msec.

Described below, are two kinds of speech recognition tests that were performed for this evaluation task.

<u>Test 1 : Word recognition test.</u> Fifty Japanese trisyllabic words converted by the speech-rate converter with the three expansion rates were randomly presented to 5 subjects. Each word had a clear meaning. Subjects listened to the words at a 20dB(SL) volume level using headphones, and repeated the presented words. Only when 3 syllables were perfectly repeated, was the answer taken as correct.

Test 2 : Sentence recognition test. Two sets of 20 Japanese sentences converted by the speech-rate converter with the three expansion rates were randomly presented to 10 elderly (age range 56-83 years) subjects. Each sentence contained 4 words. Although each word had a clear meaning, the sentences were made grammatically correct but meaningless in order to remove contextual effect. Subjects listened to the sentences at the most comfortable volume level using headphones, and repeated the presented sentences.

3. RESULTS

Figure 4 shows the original and the expanded speech signal wave form. The expansion rate was 1.50. As shown in Figure 4, the time scale of the vowel part is expanded by increasing the pitch patterns. However, the small powered consonant part maintains its original wave form. The distortion caused by the conversion is small enough for recognizing speech for normal hearing listeners. However, when the expansion rate was larger than 1.50, the converted speech sounds were too slow.

Table 1 . Correct answer rate at each expansion rate given by 5 subjects in Test 1.

Subject	Age	Δt (msec)	Correct Answer rate (%) for each expansion rate			
			Original	1.25	1.33	1.50
a	54	14	92	88	88	92
b	59	14	94	94	94	96
c	48	21	96	100	100	94
d	60	27	86	94	86	92
e	21	50	94	96	96	84

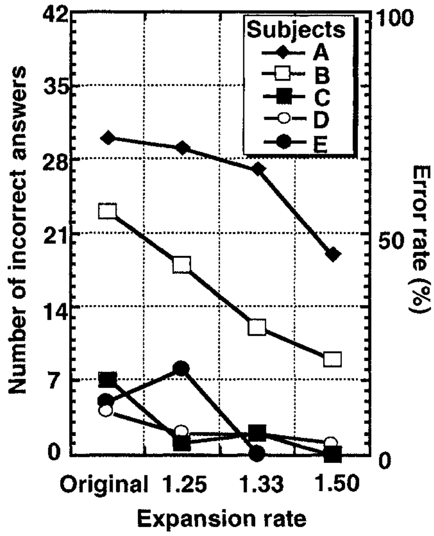

Figure 5. Number of incorrect answers given by 5 subjects at three expansion rates in Test 2.

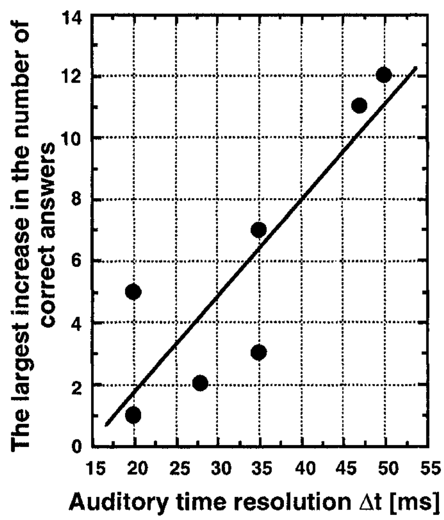

Figure 6. Largest increase in the number of correct answers as a function of auditory time resolution in Test 2.

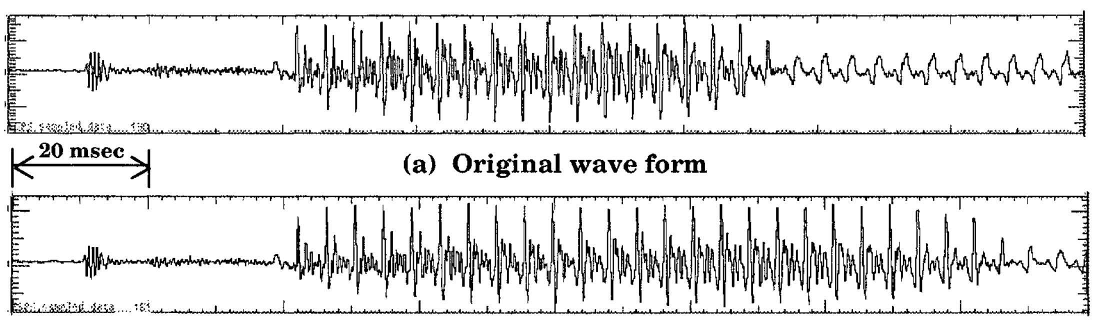

Figure 4. Original and Expanded wave forms

Table 1 shows the correct answer rate for Test 1. The correct answer rate did not differ much at each expansion rate for all of the 5 subjects. Trisyllabic words were too short to demonstrate the effect of the speech-rate conversion.

Figure 5 shows the number of incorrect words at each expansion rate for the 5 subjects who showed a clear improvement in Test 2. Subjects who performed poorly at the original speed exhibited a significant improvement when using the speech-rate conversion.

Figure 6 shows the relationship between the subjects' auditory temporal resolution and the largest improvement with the speech-rate conversion for the 7 subjects who exhibited an improvement in Test 2. The correlation coefficient of 0.859 was large enough to demonstrate the greater effectiveness of speech-rate conversion for the subjects with poor temporal resolution.

4. DISCUSSION

From Table 1, the speech-rate conversion has no effect for recognizing short speech sources. However, Figure 5 shows that it is effective when recognizing relatively long speech sources. This suggests that the speech-rate conversion works effectively on the memory function in the human auditory system. Although only a few parts of the auditory cognitive mechanism have been clarified, it has been widely accepted that during speech recognition the auditory system integrates acoustical impressions into a linguistic unit in the short term memory. This integration process might use the extra time interval provided by the speech-rate conversion effectively. From these results, this device might be suitable for aiding elderly hearing-impaired listeners with declining short term memory function.

On the other hand, Figure 6 shows that the speech-rate converter can overcome deterioration of auditory temporal resolution. It is certain that the measured resolution values indicate temporal resolution of the peripheral level of the auditory system. However, the measured values might also be related to the temporal ability of the cognitive level.

From this point of view, it might be possible for this device to be applied to the alleviation of other temporal cognitive disabilities. One of the promising applications is rehabilitation from aphasia. Another practical application is the aid for foreign language recognition. This device might also be useful when listening to unfamiliar languages.

Conventional hearing aids have improved the lowered peripheral functions of the hearing directly. In contrast, the speech-rate conversion works indirectly by supporting cognitive function. Such effective utilization of remaining function provides a natural form of aid for disability.

5. CONCLUSIONS

A new form of digital hearing aid that reduces speech speed in real time is proposed. Dynamic signal processing is used to achieve low distortion. Improvement in word recognition performance for short sentences was achieved when using this speech-rate conversion by 7 out of 10 elderly hearing-impaired subjects. The temporal performance of the subjects was strongly correlated to the improvement. The tests suggested that the speech-rate conversion affects not only the peripheral level but the cognitive level of the auditory system as well. Consequently, unlike any conventional hearing aid, the device can aid the hearing process at the cognitive level.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to T. Hokimoto for his support in the evaluation tests.

REFERENCES

[1] S. Gordon-Salant, et al, "Temporal Factors and Speech Recognition Performance in Young and Elderly Listeners," J. of Speech & Hearing Res., Vol. 36, 1276-1285, Dec. 1993.

[2] A. M. Engebretson, "Benefits of Digital Hearing Aids", IEEE EMBS Magazine, Vol. 13, 238-248, May. 1994.

[3] M. A. Picheny, et al, "Speaking Clearly for the Hard of Hearing III," J. of Speech & Hearing Res., Vol. 32, 600-603, Sep. 1989.

[4] A. Nakamura, et al, "Real Time Speech Rate Converting System for Elderly People," Proc. of IEEE ICASSP-94, Vol. II, 225-228, May. 1994.