VA RESEARCH & DEVELOPMENT

JRRD
Journal of Rehabilitation Research & Development
Vol. 36 No. 3, July 1999

Multiresolutional modification of speech signals for listeners with hearing impairment

Osman Erogul, MSc, PhD and Irfan Karagöz, MSc, PhD

Biomedical and Clinical Engineering Center, Gülhane Military Medical Academy, Ankara, Turkey

Abstract--A new method, the multiresolutional modification algorithm (MMA), which modifies speech signals in frequency and time domains, is proposed for listeners with hearing impairment and early clinical results are reported. Unlike other methods, this algorithm modifies the wavelet coefficients of the speech signal in order to obtain a modified version of the original signal instead of modifying the speech waveform itself. The speech signal is first divided into subbands using an 11-level quadrature mirror filter (QMF) bank. These subbands are then modified using the modification algorithm. Finally, the inverse Wavelet transform is applied to these modified subband coefficients in order to reconstruct a modified version of the input signal. The efficacy of the MMA was evaluated using subjects with hearing impairment and subjects with no such impairment. Listening tests showed that the proposed algorithm increases the quality and intelligibility of the modified speech over the well-known modification algorithms.

Key words: *hearing impairment, modification, QMF, speech, wavelet.*

INTRODUCTION

Apart from pitch, loudness, and timbre, information in the world of sound is also characterized by more or less sudden temporary changes. For someone with hearing impairment these variations do not fit in his or her dynamic frequency range; therefore, there is lack of certain parts of the information. Listeners with hearing impairment, especially elderly people, often have

difficulty in comprehending fast speech. In addition, they may have hearing loss at certain frequencies. In this situation, hearing aids can offer certain improvement of the speech intelligibility (1,2).

However, conventional analog hearing aids are not suitable for the processing of speech signals, because they are designed only to apply frequency-selective amplification. Modern hearing aids permit the adjustment of a number of electroacoustic parameters, among them frequency response, saturation sound pressure level, various aspects of compression, fine tuning frequency characteristics, reducing noise, and canceling acoustic feedback (3). Digital hearing aids offer many advantages over conventional aids; such as signal processing capabilities that are superior to those of a conventional analog hearing aid, and methods of signal processing and control that are unique to digital systems and can not be implemented in conventional analog hearing aids (4). Although the latest hearing aids employ digital technology, they do not use it mainly for temporal and frequency modifications concurrently.

Several studies have been suggested to solve this problem (3). Slowing the speech speed is a common technique for helping listeners with hearing impairment comprehended more easily. Recent improvements in the speed of digital signal processors have enabled the time-scale modification (TSM) of speech signals in real time. Nakamura et al. (5) proposed a real-time speech rate conversion system especially designed for television sets. The portable digital speech-rate converter for hearing impairment developed by Nejime et al., which slows speech without changing the pitch, uses only a temporal time-scale modification (TSM) algorithm (3).

In this paper, a new algorithm that modifies the speech signals both in the time and frequency domains is proposed and the preliminary clinical results are reported. In addition to slowing the speech, using different normalization and modification factors for some frequency bands, frequency components of these bands can be shifted to different frequency bands in order to obtain a more intelligible speech signal for the listeners.

The efficacy of the multiresolutional modification algorithm (MMA) was evaluated using subjects with and without hearing impairment. Listening tests showed that the proposed algorithm increases the quality and intelligibility of the modified speech over the well-known modification algorithms. Therefore, it can be used for the implementation of a digital device that can aid listeners with hearing loss.

Several methods in both the time and frequency domains have been proposed for the time-and-frequency-scale modification of speech waveforms. One class of methods widely used in the application of TSM is based on a sinusoidal representation of speech (6,7). Another approach manipulates an excitation obtained by deconvolving the original speech with a vocal tract spectral envelope estimate (8). These systems operate entirely in the frequency domain.

On the other hand, there are many time domain algorithms for the TSM of speech signals; one example is the synchronized overlap-and-add procedure (SOLA) proposed by Roucos and Wilgus (9). A modified version of SOLA was proposed by Verhelst and Roelands (10). It is called the overlap-and-add technique based on waveform similarity (WSOLA).

The major difficulty in designing a transformation system based on short time Fourier transform (STFT) results from the uncertainty principle; that is, the analysis window can not be arbitrarily short in time and in frequency (6). Unfortunately, the time and frequency resolution is the same over the time and frequency plane. Furthermore, the vocal tract system is modeled as almost stationary for analysis segments under the assumptions in reference (6).

Despite these assumptions, the speech signals are inherently nonstationary and their characteristics can be best determined through the use of the Wavelet transform (11). In addition, rather than using the same analysis filter over the entire frequency spectrum, a tree-structured filter bank can be used to overcome the restrictions imposed by the STFT. Therefore, it has been proposed that the Wavelet analysis provides more accurately localized temporal and frequency information for speech signals and produces better results than the methods described above for the time-and-frequency-scale modification of speech signals.

For the Wavelet transform, a quadrature mirror filter (QMF) pair is called a *wavelet filter* and can be represented by a sequence of coefficients. These coefficients must satisfy certain conditions. Daubechies (12) derived a series of wavelet filters that satisfy these conditions and form an orthogonal basis. QMFs allow for the perfect reconstruction of a signal that has been passed through a QMF pair.

While two-band filter banks are convenient, subband applications generally require a resolution greater than that given by two-band systems alone. To address this issue, two-band filter banks were typically embedded in tree structures.

METHOD

Algorithm

In the proposed method, the speech signal is first divided into subbands using an 11-level QMF bank to obtain wavelet coefficients. These subbands are then modified using the WSOLA algorithm. Finally, the inverse Wavelet transform is applied to these modified subband coefficients in order to reconstruct a modified version of the input signal.

The wavelet coefficients of the input speech signal are obtained using an 11-level QMF bank based on Daubechies 4,...,20 filters. The high pass and low pass filters, *h(n)* and *g(n)*, respectively, are related by:

$$h(L - 1 - n) = (-1)^n g(n) \quad [1]$$

where *L* is the filter length. The synthesis filters *h'(n)* and *g'(n)* are identical to the analysis filters *h*

(n) and *g(n),* but are reversed in time.

Since the number of subbands is directly related to the input sequence length, $N=2^{LEVEL}$, the tree structure of the filter bank can be modified depending on the user's requirements. The speech signal is first divided into fixed length of frames. To process a waveform over successive frames, the filter bank is applied to the first frame of length 2048. To divide the entire signal into the subbands, the above procedure is repeated until the last frame is reached.

The WSOLA algorithm was used in the modification of the subbands. The WSOLA attempts to find a segment that will overlap-and-add with the previous segment, which lies within the prescribed tolerance interval around the synthesis instant. The position of the best segment is determined by finding the value that lies within a tolerance region around the analysis instant and maximizes the similarity measure. There are many similarity measures that can be applied to the search of maximum correlation, such as the cross-correlation coefficient, normalized cross-correlation coefficient, mean square error, and cross-absolute magnitude difference functions. The normalized cross-correlation coefficient was used in this study. The subband coefficients are then modified using the WSOLA algorithm. The expansion and compression factors of the modified subbands are equal to the frequency shifting factors. For shifting-up the frequency band in which the subject has hearing loss, WSOLA finds a segment in the subband being processed that will overlap-and-add with the previous segment to form a similar continuity of the original subband. After the subbands are squeezed by a modification factor, the lengths of subbands become shorter. Shift-down operation is achieved by repeating the speech segments excised from the previous ones within the tolerance interval in the same manner as the shift-up procedure, except that the subbands are dilated. The time alignment between successive windows with respect to the signal similarities removes the phase discontinuities. A Hann window with 50 percent overlap is used in the overlap-and-add process. Each signal segment in this process is weighted with one half of the windowing function for smooth transition from one segment to the next segment. The old signal segment is weighted with the falling portion of the windowing function while the new segment is weighted with the raising portion of the windowing function.

After the modification of the subbands, the signal contents in each subband do not change; however, their playing time becomes shorter or longer. Finally, the inverse Wavelet transform is applied to the modified coefficients produced in order to reconstruct a modified version y(*n*) of the input signal *x(n).* After synthesis, the duration of the synthesis signal is not the same as the input signal. However, the playing time of the synthesized signal can be made equal (shorter or longer) to that of the input signal by applying the synthesis signal to a digital-to-analog converter with an adjustable sampling rate. Another approach is to interpolate or decimate the synthesized signal such that the final signal has the same amount of data as that of the input signal. Both of these approaches were used in the development of this algorithm.

In the above procedure, the playing time adjustment or the interpolation/decimation can be made between the subband modification and inverse Wavelet transform stages. In other words, after the modification of each subband by the WSOLA algorithm, the playing time of each subband can be made equal (shorter or longer) to that of the original subband using a digital-to-analog converter with an adjustable sampling rate or by decimating/interpolating each subband. In this manner, each modified subband will have the same amount of data as the unmodified version. Later, the

inverse Wavelet transform can be applied to these modified coefficients produced in order to reconstruct a frequency-scale-modified version of the input signal. After the frequency-scale modification, speech signals were time-scale modified by a factor of 1.5 using the same algorithm. A block diagram illustrating the overall approach is given in **Figure 1.**

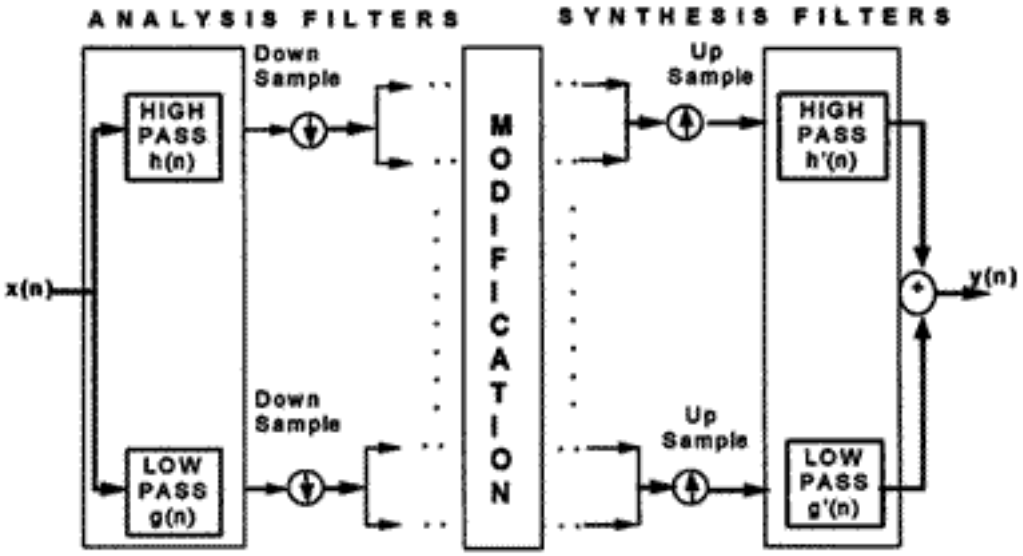


Figure 1.
Block diagram of the multiresolutional modification algorithm.

RESULTS AND DISCUSSION

In order to show how a multiresolutional analysis technique can be used in the modification of speech signals, the above algorithm has been applied to speech signals from both genders and its performance evaluated through a series of subjective listening tests using subjects with and without hearing impairment.

Signal Modification

Subbands of the original signal were obtained by using DAUB 4 filters. The time and frequency scales of the speech samples were modified by various modification factors in order to evaluate the performance of the multiresolutional modification algorithm (MMA). For unimpaired subjects, the speech signals are first modified by a factor of 0.5 and then this modified signal is again changed by a factor 2 in order to reconstruct the original signal. Frequency shaping was not used for these subjects. However, both the time-and-frequency-scale modifications were used for subjects with hearing impairment. **Figures 2a,** and **2b,** respectively, show the original signal and its modified version for a subject with hearing impairment using the MMA approach.

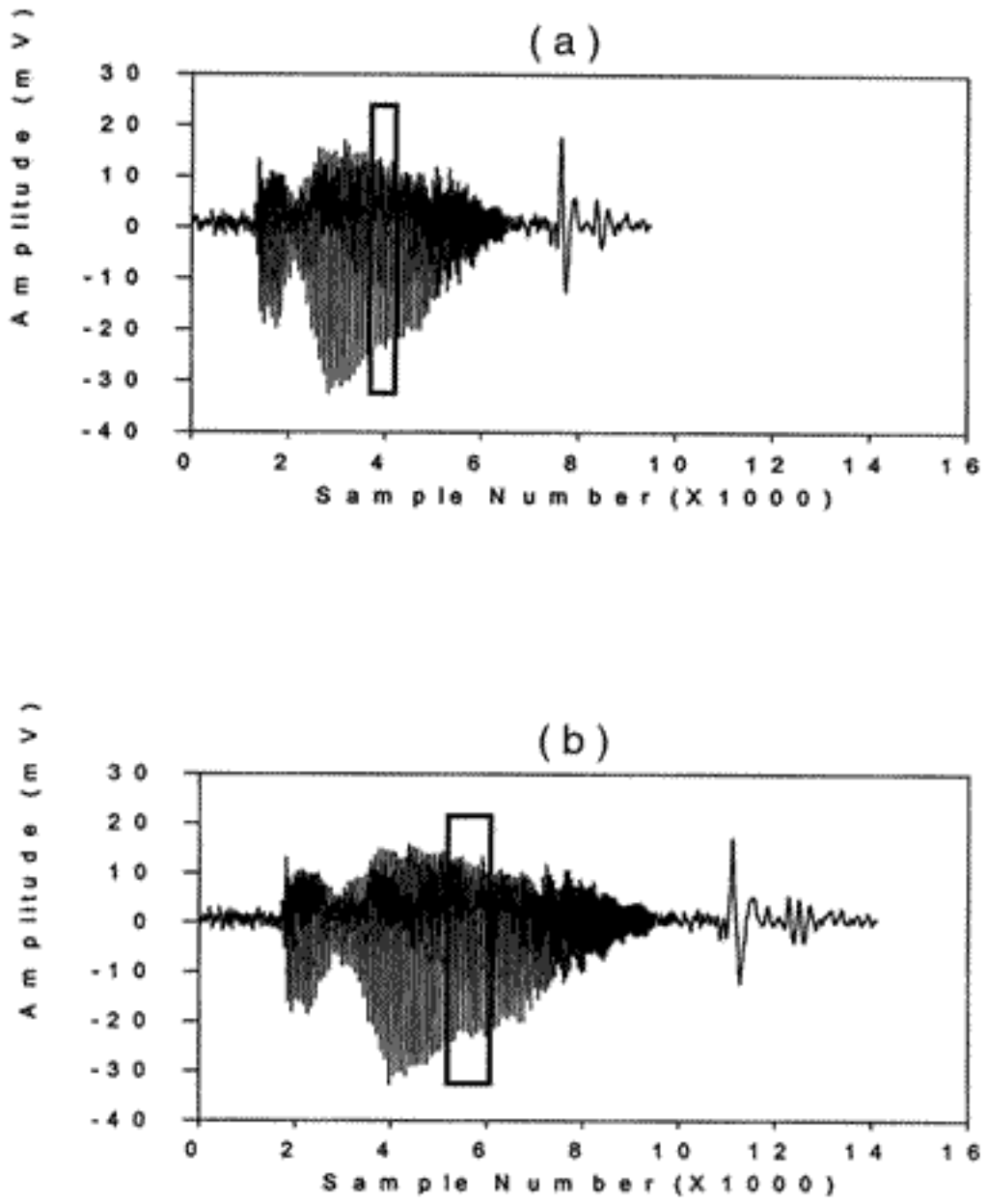
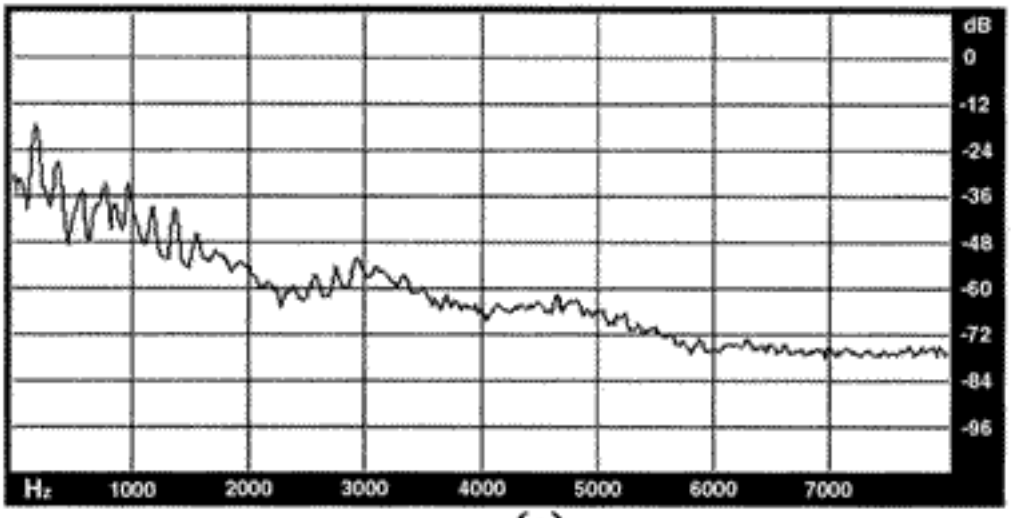
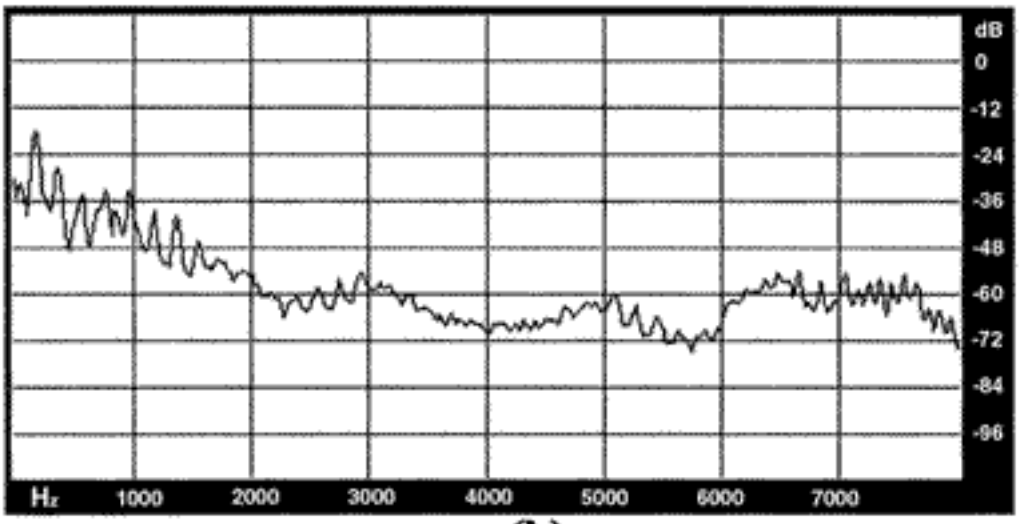


Figure 2.
(a) Original speech signal. (b) Modified version of original signal (frequency shifted and time scale modified by a factor of 1.5).

Figures 3a, and **3b** give the spectra of the small portion of the signals in the rectangular sections in each of **Figures 2a,** and **2b.** Note that the original signal in **Figure 2a** is slowed by a factor of 1.50 in order to obtain the modified version (**Figure 2b**). Also note that the frequency spectra between 4-5 kHz and 6-8 kHz in **Figure 3a** are modified and amplified to increase the intelligibility while the others remain unchanged as seen in **Figure 3b.**



(a)



(b)

Figure 3.
(a) Spectrum of the small portion of the original signal in the rectangle in **Figure 2a.** (b) Spectrum of the small portion of the modified signal in the rectangle in **Figure 2b.**

Later, the modified signals were recorded on audiotapes. The speech samples used in the experiments were recorded under quiet conditions.

Evaluation Tests and Results for Unimpaired Subjects

Three different test procedures were used in order to evaluate the performance of the MMA (13). To evaluate the intelligibility of the reconstructed speech, the diagnostic rhyme test (DRT) was used in the first test. To assess the speech quality, the mean opinion score (MOS) test was used while the degradation mean opinion score (DMOS) test was used to measure degradation in the quality of the reconstructed speech with respect to the reference.

The DRT uses a corpus of words: 232 words in 116 rhyming pairs (13). Six different filesets were used. Data collected for the different filesets are essentially replications. Speech samples in

the filesets were obtained from three males and three females. In all cases, the sampling frequency was 16 kHz. Speech sample durations are in between 5 minutes 41 seconds and 5 minutes 55 seconds.

Two modification algorithms, the speech transformation system without pitch extraction (STWPE), introduced by Seneff (8), and the sinusoidal analysis-synthesis model (SASM), developed by Quatieri and McAulay (7), were compared to the MMA to evaluate its performance.

Source speech samples for the MOS and the DMOS tests were obtained from the TIMIT acoustic-phonetic continuous speech corpus (14). On the other hand, the MOS and the DMOS use Harward-type sentences having sampling frequencies of 16 kHz. Two sentences, one spoken by a male and the other by a female, are used in each sample separated by a short silence. Sample durations are between 6 and 9 seconds. In the DMOS test, a reference sample, processed through the MMA algorithm, is presented first on each trial followed by the identical sample processed through the other modification algorithms.

Twenty-eight subjects were accessed for this study. Listeners were drawn from the Communications Research Centre (CRC), Ottawa, Canada, and from the ordinary people. Many of these subjects drawn from the CRC are familiar with the testing procedure, are sophisticated technically, and are knowledgeable about speech technology. On the other hand, subjects drawn from the general public have had no experience with speech evaluation.

Subjects were told that the samples they were going to listen to had been processed by a different TSM algorithm, and were given a simple general description of what TSM is used for. They were then told that they should listen carefully to the samples, and try to make distinctions between them in their choice of ratings. Printed instructions were given to the listeners. All subjects judged each of the modification algorithms with different speakers.

Statistical analyses were performed with the SPSS Ver. 7.0, 1997. One-way ANOVA technique was used for analysis. All statistical tests were evaluated at the 0.01, 0.05, and 0.1 levels of significance. In the DRT, it was found that the MTSM algorithm increases the intelligibility of the reconstructed speech over the other two test algorithms ($p=0.09<0.1$). The mean and standard deviations of each algorithm's score for the DRT are given in **Table 1.**

Table 1.

The DRT test results.

Test	**DRT**		
Algorithms	MMA	STWPE	SASM
Std. Dev. (σ)	0.69	0.62	0.74
Average (%)	93	92	90

DRT = diagnostic rhyme test; MMA = multiresolutional modification algorithm; STWPE = speech transformation system without pitch extraction; SASM = sinusoidal analysis-synthesis model.

The MOS test demonstrated that the MMA algorithm preserves the quality of modified speech over the other test algorithms ($p=0.000<0.01$). Statistical analysis of speech quality also demonstrated that there is no significant difference between the STWPE and the SASM algorithm according to 0.05 criteria ($p=0.52>0.05$).

The DMOS test also demonstrated that the MMA algorithm distorts the speech less than the others do ($p=0.000<0.01$). Statistical analysis also indicated that the SASM preserves the speech quality better than the STWPE algorithm ($p=0.013<0.05$). Results of the MOS and the DMOS tests are given in **Table 2.**

Table 2.

The MOS and the DMOS test results.

Test	**MOS**			**DMOS**	
Algorithms	MMA	STWPE	SASM	STWPE	SASM
Std. Dev. (σ)	0.69	0.62	0.74	0.71	0.66
Average (%)	3.76	1.84	2.55	1.89	2.72

Full score for tests: from 1 = poor to 5 = excellent; MOS = mean opinion score; DMOS = degradation mean opinion score; MMA = multiresolutional modification algorithm; STWPE = speech transformation system without pitch extraction; SASM = sinusoidal analysis-synthesis model.

Results of the subjective evaluation test show that the MMA algorithm increases the quality and intelligibility of the modified speech over the well-known algorithms. Therefore, the MMA algorithm was chosen for the modification of speech signals used for listeners with hearing impairment.

Evaluation Tests and Results for Subjects with Hearing Impairment

To evaluate the effectiveness of the proposed algorithm, speech recognition by listeners with sensorineural hearing loss or acoustic trauma was investigated using recorded speech materials.

The principal objective of speech-rate conversion is to give listeners extra time to recognize fast speech. The frequency shaping was also used for this test for the subjects having hearing loss.

Twenty-six subjects (24 males and 2 females) were given this test. Their ages ranged from 20 to 72 years. Their hearing levels and the shapes of their audiograms were not uniform. Both right and left ear audiograms indicating hearing loss above 3-4 kHz were included in this investigation. Seven subjects were 21-22-year-old males having high frequency hearing loss. Nine of the subjects had acoustic trauma. The remainder also had high frequency hearing loss. None of the subjects had any experience with hearing aids.

Standard 150 one-syllable and 165 three-syllable Turkish words, used in the hearing tests in the audiology department, were presented to these subjects. They were first asked to repeat the words immediately after they were presented. The number of incorrectly repeated or not repeated words was assumed as a false answer. Later, the same words were processed through the MMA algorithm, and they were again presented to the subjects, who were asked to repeat the words. Again, the number of incorrectly repeated or not repeated words was assumed as a false answer.

Statistical analysis was performed with the SPSS Ver. 5.0, 1997. The paired samples t test was used to test whether there is a statistically significant difference between the means of false answers of two groups. H_0 hypothesis was rejected in 0.005 significance level for one-syllable words; that is, two group means (intelligibility of the normal words and the modified words) were found to be statistically different in 0.005 level. The statistical results for one-syllable words are given in **Table 3.**

Table 3.

T tests for paired samples.

Variable	# of Pairs	2-tail Corr	Sig	Mean	SD	SE of Mean
MODIFIED				38.1538	30.735	6.028
	26	0.975	0.000			
NORMAL				43.8846	32.492	6.372

Paired Differences

Mean	SD	SE of Mean	t-value	df	2-tail Sig
-5.7308	7.280	1.428	-401	25	0.000

95% CI (-8.672, -2.789)

Corr = correlation; Sig= significance; SD = standard deviation; SE = standard error; df = difference.

On the other hand, there was no statistical difference between the means of two groups for three-syllable words. If the speech stimuli are as long as three syllables, the listeners can simply guess

the words from the complete acoustics impression. In contrast, if the speech stimuli are as short as one phoneme, slowing the speech speed should be effective because it gives extra time during the utterance. Frequency shifting provides the listeners' high frequency components with large amplitudes.

CONCLUSIONS

Multiresolutional analysis, modification, and synthesis of speech signals suggest that this new approach can be used to increase the quality and intelligibility of the modified speech with the desired time and frequency-scale modification. The preliminary clinical test results also suggest that the MMA can be used to overcome the deterioration of peripheral ability. It is possible to use this approach in conjunction with the various modification algorithms found in the literature for the modification of each subband by different modification algorithms to obtain the best results. Recent improvements in the speed of digital signal processors have enabled the time and frequency-scale modification of speech signal in real time. Therefore, this algorithm can be used in the implementation of digital hearing instruments regarded as having a function complementary to that of conventional hearing aids. TSM part of the proposed MMA algorithm, without frequency domain modification, can also be used for training children with language-learning impairment (15).

ACKNOWLEDGMENTS

The authors thank Dr. Nur Serinken of the Communications Research Centre (CRC), Ottawa, Canada for his support in the development of this algorithm.

REFERENCES

1. Engebretson AM. Benefits of digital hearing aids. IEEE Med Biol Mag 1994;13-238.
2. Van Tasell DJ. Hearing loss, speech, and hearing aids. J Speech Hear Res 1993;36(2): 228.
3. Nejime Y, Aritsuka T, Imamura T, Ifukube T, Matsushima J. A portable digital speech-rate converter for hearing impairment. IEEE Trans Rehabil Eng 1996;4-73.
4. Lewitt H, Neuman A, Sullivan J. Studies with digital hearing aids. Acta Otolaryngol Suppl (Stockh) 1990;496-57.
5. Nakamura A, et al. Real time speech rate converting system for elderly people. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Adelaide, Australia; 1994;2-225.
6. Portnoff MR. Time-scale modification of speech based on short-time Fourier analysis.

IEEE Trans Acoust, Speech, Signal Process 1981; 29:3-374.
7. Quatieri TF, McAulay RJ. Speech transformation based on a sinusoidal representation. Technical Report, Lexington, MA: Lincoln Laboratory MIT; 1986.
8. Seneff S. System to independently modify excitation and/or spectrum of speech waveform without explicit pitch extraction. IEEE Trans Acoust Speech Signal Process. 1982;30:4-566.
9. Roucos S., Wilgus AM. High quality time-scale modification for speech. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-85; 1985; 493.
10. Verhelst W, Roelands M. An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-93 1993; 2-554.
11. Rioul O, Vetterli M. Wavelets and signal processing. IEEE Signal Process Mag 1991;8:4-14.
12. Daubechies I. Orthonormal bases of compactly supported wavelets. Comm Pure Appl Math 1988;41:7-909.
13. Papamichalis PE. Practical approaches to speech coding. Englewood Cliffs, NJ: Prentice-Hall Inc.; 1987. p.177-98.
14. NIST speech acoustic-phonetic continuous speech corpus, DARPA, TIMIT Disc 1-1.1. U. S. Dept. of Commerce, National Institute of Standards and Technology, Gaithersburg, MD; 1990; Oct.
15. Erogul O, Karagoz I. Time-scale modification of speech signals for language-learning impaired children. IEEE 2nd International Biomedical Engineering Days, IBED-98, 1998; 33.

Contents

Back to Top