Commun. Fac. Sci. Univ. Ank. Series A2-A3
V. 45. pp. 39- 54 (2000)

TIME-SCALE MODIFICATION OF SPEECH SIGNALS BASED ON WAVELET TRANSFORM

O. EROĞUL[1] and Ö. TÜZÜNALP[2]

[1]*Gülhane Military Medical Academy, Biomedical and Clinical Engineering Centre, Ankara, Turkey*
[2]*Ankara University, Electronics Engineering Department, Ankara, Turkey*

(Received June 6,2000; Accepted Dec 22, 2000)

ABSTRACT

In this paper, a multiresolutional analysis/synthesis algorithm is introduced for the time-scale modification (TSM) of speech signals. Unlike most time domain methods, this algorithm modifies the wavelet coefficients of the speech signal instead of modifying the speech waveform itself. In this method the speech signal is first divided into its subbands in order to obtain more accurately localized temporal and frequency information for the TSM algorithm. These subbands are then modified using the waveform similarity overlap-add (WSOLA) method. Finally, the inverse wavelet transform is applied to these modified subband signals in order to reconstruct a time-scale modified version of the input signal. It has been shown that the multiresolutional time-scale modification (MTSM) of speech signals increases the intelligibility of the reconstructed speech while almost preserving its quality, over the well-known time-scale modification algorithms, namely the speech transformation system without pitch extraction (STWPE), sinusoidal analysis-synthesis model (SASM), and WSOLA. In order to assess the performance of the proposed MTSM algorithm, a novel evaluation procedure based on the subjective listening tests and statistical methods has been developed.

Keywords: Time-scale modification, quadrature mirror filter (QMF), subband, multiresolutional, speech.

1. INTRODUCTION

In many applications it is desirable to transform a speech waveform into a signal which is more useful than the original. For example, in time-scale modification speech can be sped up in order to compress the words spoken into an allocated time interval or to quickly scan a passage. As an application, full-duplex link can be achieved over a single-channel radio system if both ends of the link operate in what is known as Time Division Duplex Mode. In this mode the channel is allocated half of the time to transmit information in each of the two directions. That is, the radio channel is divided into time slots of T/2 seconds, with each end transmitting in alternate time slots stated simply when one end is transmitting, the opposite end is receiving the information and vice versa. In this way a single radio channel can support information flow in both directions resulting in a virtual full-duplex link (Serinken, Gagnon, and Eroğul, 1997). Alternatively, the articulation rate can be slowed down to make degraded speech more intelligible. For example, in phonocardiography, the heart sounds can be slowed down to improve physician's capability in recognition and discrimination of dissimilarities resulting from cardiac disorders (Eroğul, Karagöz, Bahadırlar, 1998).

Speech signals can also be slowed down to help the training of language learning impaired children (Eroğul, Karagöz, 1998). Numerous methods in both the frequency and time domains have been proposed for the modification of speech waveforms. The key requirement, however, is that qualities such as naturalness and intelligibility as well as speaker dependent features such as pitch and formant structure, be preserved.

One class of methods widely used in the application of time-scale modification is based on a sinusoidal representation of speech. The system developed by Portnoff (Portnoff, 1981), a refinement of the phase vocoder, represents each sine wave component by vocal cord excitation and vocal tract system contributions. Another approach manipulates an excitation obtained by deconvolving the original speech with a vocal tract spectral envelope estimate (Seneff, 1982). This system operates entirely in the frequency domain. Time expansion is achieved by doubling the unwrapped phase component of the spectrum. The other system, by Quatieri and McAulay (Quatieri and McAulay, 1986), is based on a sinusoidal representation that explicitly estimates the amplitude and phase of the vocal cord excitation and vocal tract system function contributions to each sine wave.

On the other hand, there are many time domain algorithms for the time-scale modification of speech signals. One example of a time domain algorithm is the synchronized overlap-and-add procedure (SOLA) proposed by Roucos and Wilgus (Roucos and Wilgus, 1985) which uses a modified overlap-and-add (OLA) procedure on the waveform. Another form of synchronization is obtained by applying a time domain pitch-synchronized OLA technique (TD-PSOLA) to the original waveform (Moulines and Chanpentier, 1990). With TD-PSOLA the OLA procedure is performed pitch-synchronously on the segments that are, accordingly, excised in a pitch synchronous way from an original signal. A modified version of TD-PSOLA called the overlap add technique, which is based on waveform similarity (WSOLA), was proposed by Verhelst and Roelands (Verhelst and Roelands, 1993). WSOLA ensures sufficient signal continuity at segment joints by requiring maximal similarity to the natural continuity that existed in the input signal.

Recently, the development of efficient algorithms for the time-scale and pitch modification of speech signals has been stressed. In particular, the use of the time domain SOLA algorithm in the development of these algorithms has been emphasized (George and Smith, 1997), (Yim and Pawate, 1996). While existing time domain algorithms are simpler than frequency domain algorithms, they do not address the problem of the frequency spectrum similarities in the speech signals, and therefore some of them produce poor results. Existing frequency domain approaches require very significant amounts of computation and may produce synthesized signals with a waveform that differs from the original one.

The other problems often encountered with some of the time-scale modification algorithms are the artifacts associated with processing the entire spectral band. When sinusoidal based techniques are applied, these artifacts often results in musical noise artifacts that appear due to the correlations being introduced in the high frequency regions (Quatieri and McAulay, 1992). Also, music signals tend to be very complex, containing sharp attacks or transitions. To solve this problem, the deterministic part and stochastic part of the input signal are processed separately instead of processing entire signal. For example, to make the time-scaled signals free from the 'buzzy' quality, the deterministic part and stochastic part are treated differentially (Laroche, Stylianou, and Moulines, 1993). To preserve edges, relative distances between the edges and stationary components (noise), the signal was represented as the sum of sinusoidal components and a residual (edges and noise) by Hamdy et.al. 1997. In their approach the decomposition was computed via a combined harmonic and wavelet representation. Time scaling was performed on the harmonic components and residual components separately. They analyzed the residual with a wavelet transform and performed the time-scale modifications in the wavelet domain. In our approach, entire signal was analyzed with wavelet transform and time-scale modification was performed on the subbands.

The major difficulty in designing a speech-rate change system based on short time Fourier transform (STFT) results from the uncertainty principle, i.e., the analysis window can not be arbitrarily short in time and in frequency (Portnoff, 1981). 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 the duration of its memory under the assumptions in (Portnoff, 1981).

Despite these assumptions, speech signals under consideration are inherently non-stationary, and the wavelet transform can be used effectively in TSM applications (Eroğul, 1997) since (a) decomposition of a signal spectrum into its subbands allows to operate on different resolution levels, where the degree of nonstationarity is reduced (Krim, Pesquet and Drouiche, 1993), (Sankur, Güler and Kahya, 1996), (b) observation of a signal in the subbands is advantageous especially when one would like to take into account the signal characteristics which may affect little the total spectrum, but may have more bearing on one of the band spectra. For example, in our particular application, the modification of each signal subband via WSOLA is useful since different "useful" signal information is accessed and zoomed in each subband. In other words, WSOLA takes into account the correlation in between the signal frames. Hence, each subband signal carries more refined information for better operation of WSOLA compared to the original full-band signal. 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. It has therefore been proposed that the Wavelet analysis provides more accurately localized temporal and frequency information for non-stationary speech signals. The pre-processing of the signals by using the wavelet analysis produces better results than the other well-known methods for the

TSM of speech signals. The subband approach introduced by Quatieri et al., 1995, to expand transient signals combines Wavelet Transform and Fourier Transform, but requires complex mathematical calculations such as "birth and death" tracking of frequency components, phase unwrapping, phase corrections among subbands, etc.

In this work, a novel TSM method based on the multiresolution decomposition of speech signals is proposed. It is shown that a combination of the time domain approach, WSOLA, and the wavelet transform can be used to construct a high performance and mathematically inexpensive multiresolutional time-scale modification (MTSM) algorithm in which wavelet coefficients are modified using the WSOLA algorithm. The proposed method will be called multiresolutional time-scale modification (MTSM) in the sequel. Performance of the MTSM algorithm is assessed via a novel test method and procedure based on the quantitative evaluation of human observers' subjective judgements.

II. FRAMEWORK

A. Wavelet Basis

A signal can be represented in many different forms. The optimal representation of the signal must be defined in terms of the specific problem being considered. For our particular application, the intention is to represent the signal in such a way as to maximize the time and frequency resolution so that time-scale modification algorithms will be able to modify the signal in a better manner.

The continuous wavelet transform (CWT) of the input signal, *x(t)*, is defined as (Rioul and Vetterli, 1991):

$$\mathrm{CWT}_{\mathrm{x}}(\tau,a) = \int \mathrm{x(t)h}^{*}_{a,\tau}(\mathrm{t})\,\mathrm{dt}$$

where $h^{*}_{a,\tau}(t)$ is the complex conjugate of the wavelet, $h_{a,\tau}(t)$. The wavelets themselves are scaled and translated versions of the basic wavelet prototype *h(t)*, called the *mother wavelet*, and are given by:

$$\mathrm{h}_{a,\tau}(\mathrm{t}) = \frac{1}{\sqrt{a}}\mathrm{h}(\frac{\mathrm{t}-\tau}{a}) \qquad \text{where } a>0 \text{ is the } \textit{scaling factor}.$$

Reconstruction of the original signal can be accomplished by summing up all the orthogonal projections of the signal onto the wavelets through the use of the inverse Wavelet transform, which is given by:

$$\mathrm{x(t)} = \mathrm{c}\iint_{a>0} \mathrm{CWTx}(\tau,a)\mathrm{h}_{a,\tau}(\mathrm{t})\frac{\mathrm{d}a\mathrm{d}\tau}{a^2}$$ where c is a constant that depends only on *h(t)*.

To remove the redundancy from the continuous wavelet transform, discrete values for the scale and translation parameters can be used and the wavelet basis functions can be implemented as a finite impulse response (*FIR*) filter or an infinite impulse response (*IIR*) filter depending on the particular properties required. For the Wavelet transform, a QMF pair is called a *wavelet filter* and can be represented by a sequence of coefficients. These coefficients must satisfy certain conditions (Wickerhauser, 1992). Daubechies derived a series of filters that satisfy these conditions and form an orthogonal basis. Quadrature mirror filters allow for the perfect reconstruction of a signal which has been passed through a QMF pair (Daubechies, 1988).

Two-band filter banks are convenient, but 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. By cascading the two-band filter banks such that each band is split in two successively, an infinity of decompositions can be realized.

B. Time-Scale Modification

Four different TSM algorithms namely the MTSM, STWPE, SASM and WSOLA, were used in the time-scale modification of speech signals. The first algorithm is the proposed MTSM method which operates in the subbands of the full-band signal rather than the full-band signal itself. In other words, in MTSM, each of the subband signals are first modified using WSOLA and then they are synthesized. The MTSM algorithm is detailed in Section 3. On the other extreme, WSOLA, STWPE and SASM directly operate on the original full-band signal. These three methods were used to assess the performance of the proposed MTSM algorithm.

In this work, Waveform Similarity Overlap-and-Add (WSOLA) technique is used both in the MTSM method and in the TSM of the original full-band signal. WSOLA seeks to find a segment that will overlap-add with the previous segment which lies within the prescribed tolerance interval around the synthesis instant. The position of the best segment m is determined by finding the value $\Delta_{optimum}$ lying within a tolerance region $[-\Delta_{\max}.. \Delta_{\max}]$ around the analysis instant that maximizes the cross-correlation coefficients between the previous segment and the segment under consideration.

The second TSM algorithm, the speech transformation system without pitch extraction (STWPE) developed by Seneff (Seneff, 1982), is capable of independent manipulation of the fundamental frequency and spectral envelope of a speech waveform. The system developed by Seneff deconvolves the original speech with the spectral envelope estimate to obtain a model for the excitation. Hence, explicit pitch extraction is not required. To alter only the temporal characteristics the phase spectrum must be modified. Seneff's method is essentially equivalent to that used in a standard phase vocoder.

The third approach used in this study utilizes a sinusoidal analysis-synthesis model (SASM) that is based on the amplitudes, frequencies, and phases of the component sine waves (Quatieri and McAulay, 1986). The reconstruction requires the manipulation of functions which describe the time evolution of the vocal cord excitation and vocal tract system contributions of the amplitude and phase of each sine wave component. The parameters used in these functions are estimated from the STFT using a simple peak-picking algorithm. Rapid changes in the highly resolved spectral components are tracked using the concept of "birth" and "death" of the underlying sine waves. In the sine wave model for time-scale modification the events which are time-scaled are the system amplitudes and phases, and the excitation amplitudes and frequencies of each underlying sine wave.

III. MTSM METHOD

The proposed multiresolutional time-scale modification (MTSM) method is based on the use of the wavelet transform in association with the conventional WSOLA algorithm for TSM of speech signals. The steps of the method are given and detailed below.

(i) The original fullband speech signal is first decomposed into its subband components via the wavelet transform,

(ii) TSM of each of the subbands is then obtained using the WSOLA algorithm,

(iii) Finally, an output signal is synthesized from the modified subband signals.

The wavelet coefficients of the input speech signal are obtained using an 11-level Quadrature Mirror Filter Bank (QMF) based on Daubechies 4,...,20 filters (DAUB4, ..., DAUB20) (Daubechies, 1992). As the filters increase in length, they also increase in smoothness. A longer filter is suitable for representing low frequency signals, while a short filter would be ideal for high frequency applications. The high pass and low pass filters (*h(n)* and *g(n)*, respectively are related by:

$$h(L-1-n) = (-1)^n g(n)$$

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.

A tree structured filter bank was used to perform the transform. Taking as an example the decomposition of the input speech signal onto the basic wavelet packets, the high-pass filter of the QMF pair was first applied to the signal and the result was decimated by 2. Then the low-pass filter was applied to and the result was decimated by 2. What this has done is to split the original signal into two parts, high-pass part and a low-pass part, each has half of the length of the original signal. The above procedure produces the first level wavelet packet transform. Decomposing the signal onto the general wavelet packet requires to apply the QMF pair to all of the high-pass and low-pass results of the previously calculated level as shown in Figure 1. The filtering is continued until the output of each filter is a unit length sample. For a signal with a length of 2048 samples, N, there would be a total of 11 levels of filters ($N = 2^{LEVEL}$). The entire process produces a large library of wavelet coefficients. The proposed algorithm is capable of selecting a number of different bases, such as wavelet basis, user selected level basis, threshold criteria basis and minimal entropy criterion basis. The easiest of these is to select "*a user selected level basis*" which simply takes the coefficient from the selected level and writes them to a data file. To process a waveform over successive frames, the filterbank is applied to the first frame of length 2048. To divide entire signal into the subbands the above procedure is repeated until the last frame is reached. These subbands are then modified using the WSOLA time-scale modification algorithm. Finally, the inverse Wavelet transform is applied to the modified coefficients produced in order to reconstruct a time-scale modified version *y(n)* of the input signal *x(n)*. A block diagram illustrating the overall approach is given in Figure 1.

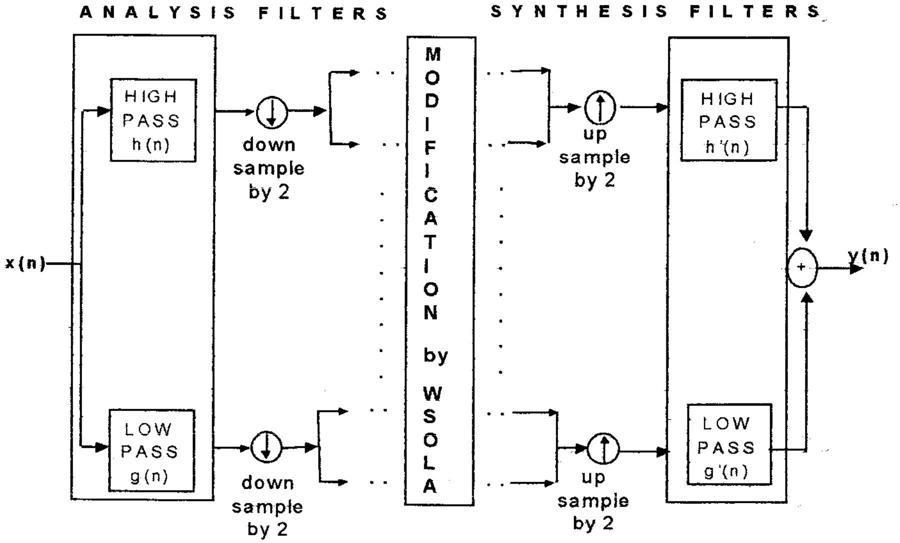


Fig. 1. Block diagram of the multiresolutional time-scale modification algorithm.

As mentioned in Section 2, the performance of the proposed MTSM method was assessed with respect to three other single-band TSM algorithms, namely, the STWPE, SASM, and WSOLA. Note that, in the proposed MTSM method, the time-scale modification is performed on the subband signals using the WSOLA, while STWPE, SASM, and WSOLA algorithms directly operate on the fullband signal.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In order to show how a multiresolutional analysis technique can be used in the TSM 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.

As an example the word /Bob/, recorded at 8 kHz sampling rate and spoken by a male, was used as the original signal and is shown in Figure 2.

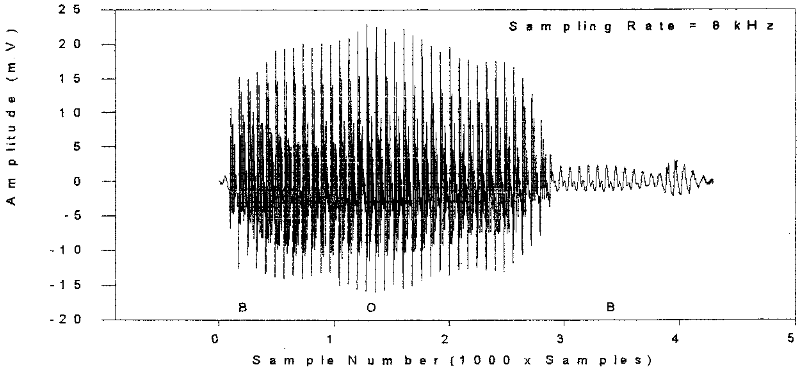
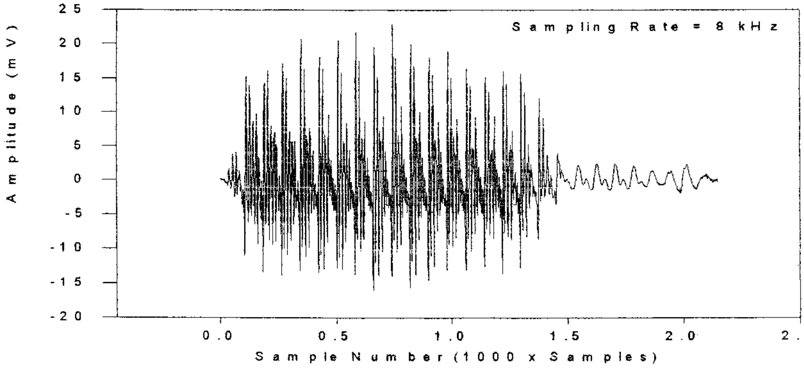
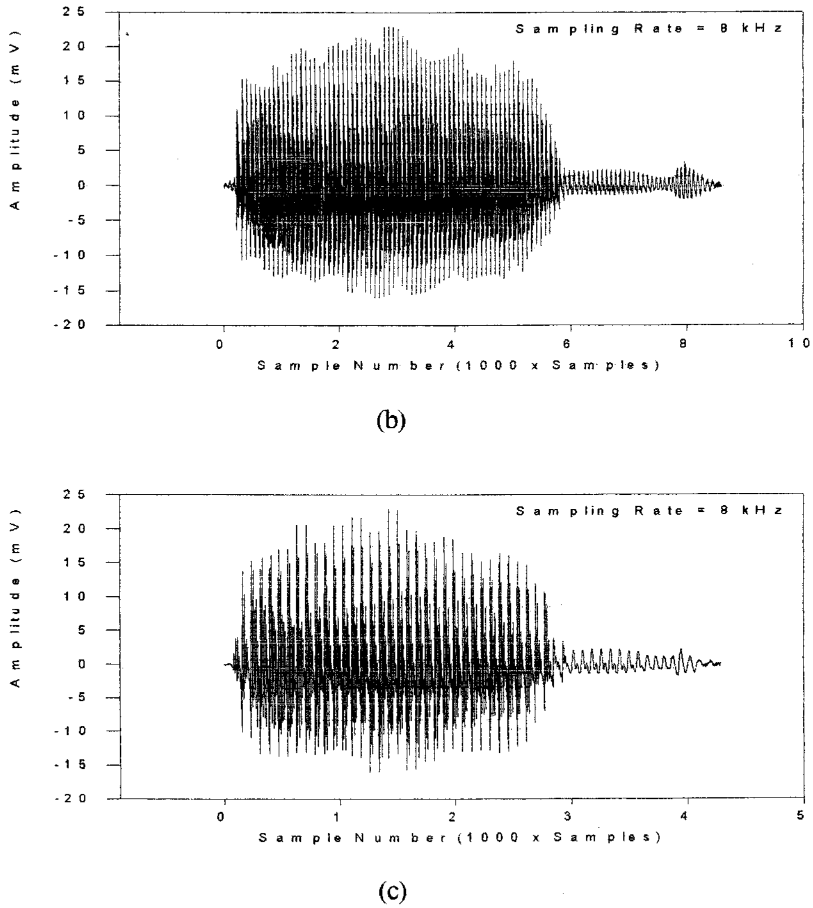


Fig. 2. Original speech sample: The word /Bob/.

The proposed MTSM algorithm was then applied to the sample speech signal. In the application of the method, subbands of the original signal were obtained by using DAUB 10 filters. These subbands were then modified using the WSOLA algorithm. The inverse wavelet transform was then applied to these modified subbands in order to reconstruct a time-scale modified version of the input signal. Figure 3 shows the time-scaled and reconstructed versions of the original signal using the MTSM method.



(a)



(b)

(c)

Fig. 3. Multiresolutional time-scale modification of the original signal. Subbands were modified by using the WSOLA algorithm: (a) Compressed version (Ratio = 0.5) of Fig.2; (b) Expanded version (Ratio = 2) of Fig. 2; (c) Reconstructed version (Ratio = 0.5/2) of Fig. 2.

As mentioned above, WSOLA is used in order to obtain the TSM of the subband signals in the MTSM method, furthermore subbands can also be modified using various TSM algorithms found in literature. Figures 4 and 5 illustrate the time-scaled and reconstructed versions of the original signal in which the subbands are modified by using STWPE and SASM algorithms, respectively.

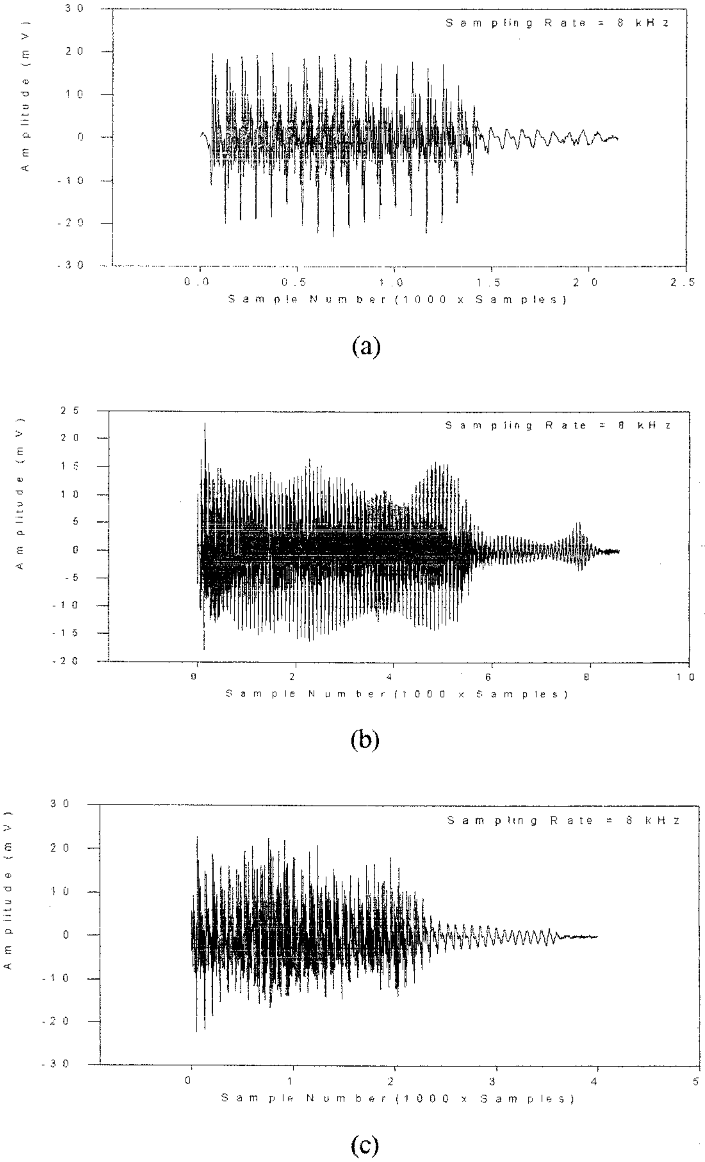
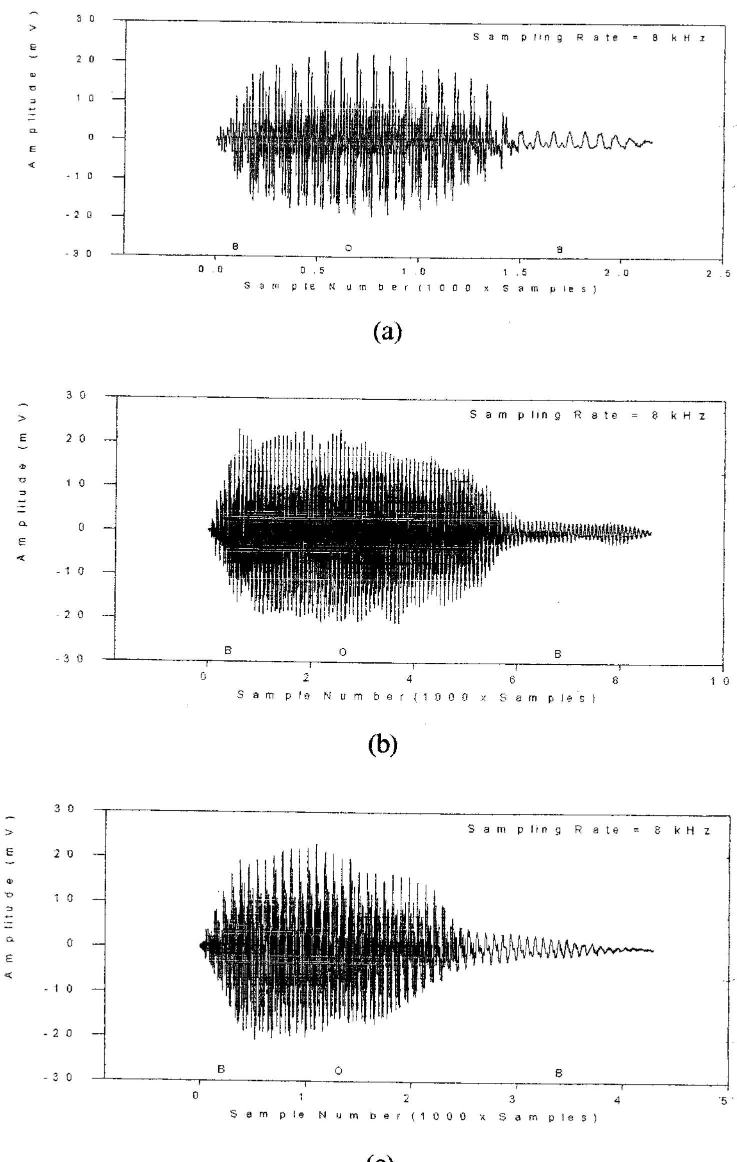


Fig. 4. Time-scale modification of the original signal. Subbands were modified by using the STWPE algorithm: (a) Compressed version (Ratio = 0.5) of Fig.2; (b) Expanded version (Ratio = 2) of Fig. 2; (c) Reconstructed version (Ratio = 0.5/2) of Fig. 2.



(a)

(b)

(c)

Fig. 5. Time-scale modification of the original signal. Subbands were modified by using the SASM algorithm: (a) Compressed version (Ratio = 0.5) of Fig.2; (b) Expanded version (Ratio = 2) of Fig. 2; (c) Reconstructed version (Ratio = 0.5/2) of Fig. 2.

Careful examination of Figure 2, original signal, and 3(c), 4(c), 5(c), reconstructed signals, shows that the MTSM algorithm, in which subbands are modified by using the WSOLA algorithm, Figure 3(c), preserves the envelope of the modified signal best. Furthermore, as will be demonstrated, subjective listening tests applied to original full-band speech signals prove that the performance of WSOLA is considerably better than the STWPE and SASM algorithms. Therefore, WSOLA seems to be the best choice to be applied to subband signals. In conclusion, the MTSM method consists of decomposition of the original fullband signal into its subbands using the wavelet transform, the TSM of subband signals using WSOLA, and reconstruction via the inverse wavelet transform.

Test Procedure

Three different test procedures were used in order to evaluate the performance of the MTSM algorithm. 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 a reference (Papamichalis, 1987).

The DRT works as follows: It uses a corpus of words, 232 words in 116 rhyming pairs. In a given instance, one word of the pair is presented and the listener is asked to determine which word was spoken. The two words of each pair, for instance "Bob", "Gob", differ only in one attribute of the first consonant. So, a correct response from the listener indicates that the speech processing system under examination preserves that attribute. Source speech samples for the DRT were obtained from three male and three female talkers (CRC, 1997). In all cases the sampling frequency was 8kHz. Six different filesets, containing the same words but spoken different talkers were used. A fileset corresponds to a speaker. Speech sample durations in filesets are between 5 minutes 41 seconds and 5 minutes 55 seconds.

In MOS procedure, one sentence is presented on each trial, and the listener is asked to rate the sample according to the absolute scale, ranging between 1 and 5. The quality scale ranges from "bad" for grade 1 to "excellent" for grade 5. The drawback of this procedure is that, ceiling and floor effects may obscure real differences in performance. To overcome this limitation, the degradation mean opinion score test was also used.

In DMOS procedure two samples are presented on each trial, a reference sentence and a test sentence. Listeners are asked to rate the quality of the second sample relative to the quality of the first. The quality scale ranges from "very much poòr quality" for grade 1 to "the same or better quality" for grade 5.

Source speech samples for the MOS and the DMOS tests were obtained from six males and six females. These sources had sampling frequencies of 16kHz (TIMIT, 1990). The MOS and the DMOS use Harward type-sentences. 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.

Twenty-eight subjects were accessed to 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 and they are sophisticated technically and know much about speech technology. Subjects drawn from ordinary people 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.

To emphasize the effects of algorithms on speech signals, speech samples were first compressed by a factor of 0.5, and then expanded by a factor of 2 in order to recover the original signal. In order to evaluate the performances of algorithms and to assess the performance of the proposed MTSM algorithm, modified versions of reconstructed speech_signals using the MTSM, and original single band speech signals via STWPE, SASM and WSOLA_algorithms were recorded on audio tapes. All of the recordings used in the subjective listening tests were taken under quiet conditions.

In the DMOS test a reference sample, processed through the WSOLA algorithm was presented first on each trial followed by the identical sample processed through the other three algorithms (MTSM, STWPE, SASM). All subjects judged each of the TSM algorithms with different speakers.

Statistical Analysis and Results

One-way ANOVA technique was used for statistical 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 three test algorithms ($p=0.09<0.1$). The mean and standard deviations of each algorithm's score for the DRT are given in Table-I.

TABLE I
THE DRT TEST RESULTS

Test	**D R T**			
Algorithms	MTSM	STWPE	SASM	WSOLA
Std. Dev.(σ)	0.69	0.62	0.74	0.84
Average (%)	93	92	90	92

The MOS test demonstrated that the MTSM algorithm preserves the quality of modified speech over the other STPWE and SASM 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). In MOS test, on the other hand, performance of WSOLA algorithm was found to be slightly better than that of the MTSM algorithm, i. e. means of WSOLA and MTSM are 3.97 and 3.76, respectively (see Table II). But, the statistical analysis of the result showed that the difference is not significant according to 0.05 criteria (p=0.395>0.05).

TABLE II
THE MOS AND THE DMOS TESTS RESULTS

Test	**MOS**				**DMOS**		
Algorithms	MTSM	STWPE	SASM	WSOLA	MTSM	STWPE	SASM
Std. Dev.(σ)	0.69	0.62	0.74	0.84	0.85	0.71	0.66
Average	3.76	1.84	2.55	3.97	4.07	1.89	2.72

The DMOS test also demonstrated that the MTSM algorithm preserves the quality of the modified speech better than the STWPE and SASM do (p=0.000<0.01). Statistical analysis also indicated that the SASM preserves the speech quality better than the STWPE algorithm does (p=0.013<0.05).

Subjective evaluation test results show that the MTSM increases the intelligibility of the reconstructed speech signals while almost preserving the quality of the modified signal.

V. CONCLUSIONS

There are two methods of time-scale modification which can be used in the subband approach. One works in the time domain, i.e., compresses or expands the wavelet coefficients in the time domain. The other works in the frequency domain, i.e., the modification of subband signals in the frequency domain. The WSOLA, and the SASM and STWPE algorithms were used as examples of the former and latter methods, respectively.

The application of the Wavelet transform to the analysis, modification and synthesis of non-stationary speech signals suggests that this approach can be used to increase the intelligibility of time-scale modified speech with the desired time-scale modification while preserving the pitch and formant structure of the original signal. Such an algorithm was developed and was demonstrated to be also capable of producing high quality rate-changed speech. It was also shown that it is possible to use this approach in conjunction with various TSM algorithms found in literature. The same approach can also be extended to the pitch modification of speech signals. Another approach might be to modify each subband by a different TSM algorithm rather than modifying all of the subbands via a single TSM algorithm. For example, first subband signal might be modified using the STPWE algorithm, while the second one is modified with the SASM, and so on. Such a treatment might improve the performance of the proposed MTSM method.

ACKNOWLEDGEMENTS

Authors wish to thank Dr. Nur Serinken of the Communications Research Center (CRC), Ottawa, Canada for his support in this study.

REFERENCES

[1] CRC, (1997) Speech Corpus Used for the Diagnostic Rhythm Test (DRT), Communications Research Centre (CRC), Ottawa, Canada.

[2] Daubechies, I. (1988), "Orthonormal Bases of Compactly Supported Wavelets", *Comm. in Pure and Applied Math.*, Vol. 41, No. 7, pp. 909-996.

[3] Daubechies, I. (1992), "Ten Lectures on Wavelets", Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, pp 167-213.

[4] Eroğul O., and Karagöz İ. (1997), "Time-Scale Modification of Speech Signals for Language-Learning Impaired Children", *IEEE 2nd Int. Biomedical Engineering Days*, pp 33-35.

[5] Eroğul, O. (1997), "Multiresolutional Time-Scale and Pitch Modifications of Speech Signals based on Wavelet Transform (in Turkish)", *Ph. D. Thesis*, Ankara University.

[6] Eroğul, O., Karagöz, İ. and Bahadırlar Y. (1998), "A New Approach for Diagnosis of Cardiac Disorders: Time-Scale Modification of Phonocardiographic Signals using Wavelet Transform", *Biomedical Engineering Bulletin, Boğaziçi University Biomedical Engineering Institute*, Turkey, accepted for publication.

[7] George, E.B. and Smith J.T. (1997), "Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model", *IEEE Trans. on Speech and Audio Process.*, Vol. 5, No. 5, pp. 389-406.

[8] Hamdy, K. N., Tewfik, A. H., Chen, T. and Takagi S. (1997), "Time-Scale Modification of Audio Signals with Combined Harmonic and Wavelet Representations", *IEEE Int. Conf. Acoust., Speech, Signal Process.*, ICASSP-97, Munich, Germany, Vol. 1, pp. 439-442.

[9] Krim, H., Pesquet, J.C. and Drouchie K. (1993), "Tracking Nonstationarities with a Wavelet Transform", in Proc. *IEEE-ICASSP*, pp. 245-248.

[10] Laroche, J., Stylianou, Y. and Moulines E. (1993), "HNS: Speech Modification Based on a Harmonic + Noise Model", *IEEE Int. Conf. Acoust., Speech, Signal Process.*, ICASSP-93, Minneapolis, MN, Vol. 2, pp. 550-553.

[11] Moulines, E. and Charpentier F. (1990), "Pitch-Synchronous Waveform Processing Techniques for text-to-Speech Synthesis Using Diphones", *Speech Communication,* Vol. 9 (5/6), pp. 453-467.

[12] NIST Speech Acoustic-Phonetic Continuous Speech Corpus, DARPA, TIMIT Disc 1-1.1 (1990), U.S. Department of Commerce, National Institute of Standards and Technology, Gaithersburg, MD.

[13] Papamichalis, P. E. (1987), "Practical Approaches to Speech Coding", N.J., pp. 177-198, Englewood Cliffs, Prentice-Hall Inc.

[14] Portnoff, M.R. (1981), "Time-Scale Modification of Speech Based on Short-Time Fourier Analysis", *IEEE Trans.Acoust., Speech, Signal Process.*, Vol. ASSP-29, No. 3, pp. 374-390.

[15] Quatieri, T. F. and McAulay R. J. (1992), "Shape Invariant Time-Scale and Pitch Modification of Speech", *IEEE Trans. Signal Process.*, Vol. 40, No. 3, pp. 497-510.

[16] Quatieri, T. F., and McAulay R. J. (1986), "Speech Transformation Based on a Sinusoidal Representation", Technical Report, Lincoln Laboratory MIT, Lexington, Massachusetts, pp. 1-56.

[17] Quatieri, T.F., Dunn, R.B. and Hanna T.E. (1995), "A Subband Approach to Time-Scale Expansion of Complex Acoustic Signals", *IEEE Trans. On Speech and Audio Process.*, Vol.3, No.6, pp. 515-519.

[18] Rioul, O. and Vetterli M. (1991), "Wavelets and Signal Processing", *IEEE Signal Processing Magazine*, Vol. 8, No. 4, pp. 14-38.

[19] Roucos, S. and Wilgus A. M. (1985), "High Quality Time-Scale Modification for Speech", *IEEE Int. Conf. Acoust.,Speech, Signal Process.*, ICASSP-85, pp. 493-496.

[20] Sankur, B., Güler E.Ç. and Kahya Y.P. (1996), "Multiresolution Transient Extraction Applied to Respiratory Crackles", *Computers in Biology and Medicine*, Vol. 26, No. 1, pp. 25-39

[21] Seneff, S. (1982), "System to Independently Modify Excitation and/or Spectrum of Speech Waveform Without Explicit Pitch Extraction", *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. ASSP-30, No. 4, pp. 566-578.

[22] Serinken, N., Gagnon, B. and Eroğul O. (1997), "Full-Duplex Speech for HF Radio Systems", *IEE HF Radio Systems and Techniques, Conf. Pub. No. 411*, pp 281-284.

[23] Verhelst, W. and Roelands M. (1993), "An Overlap-add Technique Based on Waveform Similarity (WSOLA) For High Quality Time-Scale Modification of Speech.", *IEEE Int. Conf. Acoust., Speech, Signal Process.*, ICASSP-93, Vol. 2, pp. 554-557.

[24] Wickerhauser, M.V. (1992), "Acoustic Signal Compression with Wavelet Packets", in Wavelets-A Tutorial in Theory and Applications, C.K. Chui (ed.), pp. 679-700, Academic Press Inc.

[25] Yim, S. and Pawate B.I. (1996), "Computationally Efficient Algorithm for Time Scale Modification (GLS-TSM) ", *IEEE Int. Conf. Acoust., Speech, Signal Process.*, ICASSP-96, Vol. 2, Atlanta, Georgia, pp. 1009-1012.