INFORMATICA, 2003, Vol. 14, No. 3, 349–356

New Implementation Scheme for the Restoration of Voiced Speech Signals [*]

Šarūnas PAULIKAS
Department of Telecommunication Engineering, Electronics Faculty
Vilnius Gediminas Technical University
Aušros Vartų 7a, 2600 Vilnius, Lithuania
e-mail: sarunas@el.vtu.lt

Dalius NAVAKAUSKAS
Department of Radioelectronics, Electronics Faculty
Vilnius Gediminas Technical University
Naugarduko 41, 2006 Vilnius, Lithuania
e-mail: dalius@el.vtu.lt

Received: May 2003

Abstract. Recently iterative procedure for the restoration of speech signals when prosodic elements: stress and accent, of comparatively long duration are missing was developed. Alternatively, it could be cast in a signal generation framework. Basing on that view the paper presents the efficient implementation scheme for the restoration of voiced speech signals. It enjoys parallel order of multirate processing utilizing interpolation and decimation filters parameterized by specific to problem coefficients. Presented simulation results confirm the feasibility of developed implementation.

Key words: speech signal processing, digital signal processing, multirate signal processing, polynomial approximation.

1. Introduction

Modern restoration techniques of speech signals employ Linear Prediction (Vaseghi, 2000), Hidden Markov Models (Vaseghi and Milner, 1993), Artificial Neural Networks (Czyzewski, 1997), various Bayesian technique (Godsill and Rayner, 1995), to name a few. However all methods fail when sufficiently long segments of speech signals constituting essential information are lost. In order to restore these segments additional a priori information must be available and easy includable into restoration procedure.

This paper continues started in (Paulikas and Navakauskas, 1998) investigation of a problem of restoration of homographs – words that meaning depends on the place of stress. There we assumed that big segments of signal composed of spoken words are completely destroyed by some kind of noise or are just lost. Also we supposed that there were

[*] For the support authors would like to thank prof. L. Ljung (Linköping University, Sweden).

homographs among other words and lost segments coincide with places of accent. Thus restoration of such lost segments was possible only when a priori information about existence of accent (extracted from the context of the whole sentence) and description (model) of accent was known (Paulikas, 1999). Based on that restoration technique that could be viewed as a weighted forward and backward processing of non-uniformly sampled and weighted speech signal was developed.

Here we propose alternative voiced speech signal restoration procedure that uses signal generation approach. It employs efficient implementation scheme based on parallel order multirate signal processing. That scheme utilizes interpolation and decimation filters parameterized by specific to problem coefficients (Paulikas and Navakauskas, 2003).

The paper is organized as follows: we start in Section 2 with brief overview of developed iterative speech restoration method that is based on accent model. In Section 3 we propose alternative efficient multirate implementation for speech signal restoration. Experimental study results of comparison of both restoration procedures we present in Section 4. Conclusions are given at the end.

2. Restoration of Voiced Speech Signal

For clearness of the presentation in Section 3, here we briefly overview iterative speech restoration method that uses characteristics and model of accent.

2.1. *Employed Characteristics*

Let us examine two words spoken by male speaker of 450 ms duration, recorded with sampling rate of 44.1 kHz (waveforms of corresponding word's syllables with accent are shown in Fig. 1(a)). These words are typical representatives of homographs (Barauskaitė *et al.*, 1995), as they are spelled identically, however their meaning in lithuanian is different – "chisel" and "guilty". Main difference between these words is in their accent – the first one has falling, while another has raising accent (Golovinas, 1982).

In Fig. 1(b) fundamental frequency (calculated by modified autocorrelation method with clipping (Dubnowski *et al.*, 1976), employing 30 ms window length and $1/3$ overlapping) and normalized intensity (calculated as a normalized power for each period of fundamental frequency) characteristics for the both accents are shown. More precisely, normalized intensity is expressed as

$$I_0 = \frac{1}{I_{0\,\max}} \sum_n |s_v(n)|. \tag{1}$$

Here $s_v(n)$ is a speech signal at a n time instance, summation is taken only for a particular period T_0, and $I_{0\,\max}$ is a maximum intensity value of the syllable.

These two characteristics together with accent duration are main in the description of accent (Pakerys, 1986). Because of fact that accent duration must be tailored to a real speech signal (Paulikas, 1999) it is not included into developed accent model to be described next.

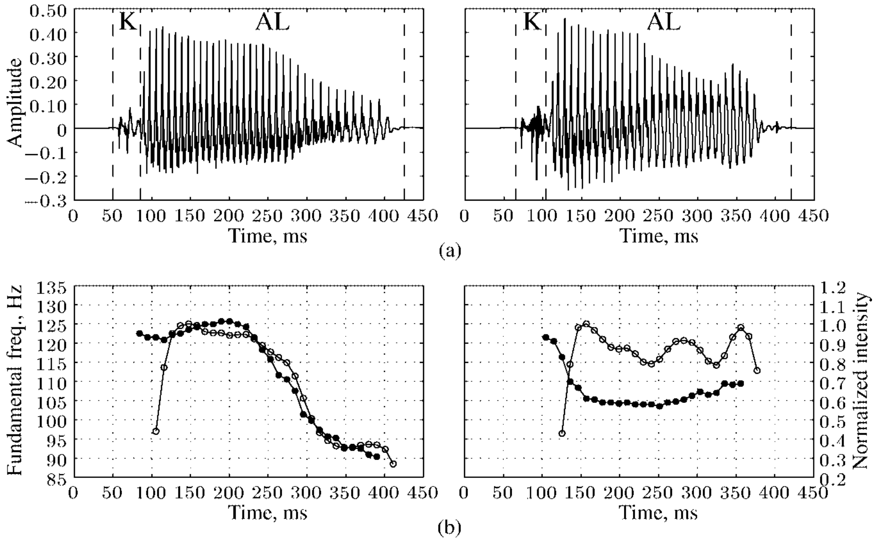

Fig. 1. Data for two syllables with accent that distinguish between two homographs: Lithuanian words meaning “chisel” (left column) and “guilty” (right column). (a) – waveforms of syllables spoken by male speaker and recorded with sampling rate of 44.1 kHz; (b) – composite plots of fundamental frequency (stars) and intensity (circles) characteristics.

2.2. *Utilized Model*

In order to describe accent in terms of signal characteristics let us implicitly express voiced speech signal s_v as a periodic signal with constant period T_0 and intensity I_0

$$s_v\left(I_0, T_0, n\right) = I_0 \cdot s_v\left(n - T_0\right). \tag{2}$$

This definition of speech signal does not formalize speech signal in detail, however enables to control it through chosen speech characteristics. That is very suitable in non-stationary signal case, when some speech signal characteristics could be determined easier than the speech signal model (Botinis *et al.*, 2001). Taking into account that period of fundamental frequency $T(n)$ and intensity $I(n)$ are time varying we re-write previous equation ending in general accent model

$$s_v\big(I(n), T(n), n\big) = I(n) \cdot s_v\big(n - T(n)\big). \tag{3}$$

Now, accent model development squeezes into modeling of fundamental frequency and intensity characteristics (Paulikas and Navakauskas, 1998).

2.3. *Restoration Procedure*

As was mentioned earlier, we assume that in a place of homograph speech signal is completely destroyed. However, there exists a priori information (extracted, e.g., from context) about its type and corresponding intensity and fundamental frequency characteristics. In the restoration of homograph we employ forward together with backward processing in time (Etter, 1996), more precisely

$$\hat{s}_v(n) = w^f(n)\hat{s}_v^f(n) + w^b(n)\hat{s}_v^b(n), \tag{4}$$

where used weighting function is of a form

$$w^{f/b}(n) = \frac{1}{2} \cdot \left[1 \pm \cos\left(\frac{n \cdot \pi}{n_2 - n_1}\right)\right]. \tag{5}$$

Here $n \in [n_1, n_2]$, n_1 and n_2 are limits of the restoration, subscripts f/b indicate processing direction.

Signal restoration carried out from different directions inheritable is the same, the main difference being in the time direction, i.e., indexes. Thus further in this section we will discuss processing of signal only in the forward direction. It is unnecessary to take into account variation of intensity and fundamental frequency at each time step (Paulikas, 2001), thus we use only their variation with period of fundamental frequency, in the following expressions introducing new variable k as index of period of fundamental frequency. Therefore, expression (3) of the restored voiced speech signal could be re-written by

$$\hat{s}_v^f(n) = \frac{I(k)}{I(k-1)}\hat{s}_v^f\left(\left\lfloor\frac{T(k-1)\big(n-T(k-1)\big)}{T(k)}\right\rfloor\right). \tag{6}$$

Now intensity and fundamental frequency characteristics do not depend on time index and expression is valid for particular speech signal period. Note, that in the calculation of time indexes ratio of periods must be integer number that is why operation of rounding to minus infinity, $\lfloor \cdot \rfloor$ is used here. Expression (6) could be used recursively in the restoration of accent in speech signal.

3. Multirate Implementation

Direct implementation (Paulikas and Navakauskas, 1998) of speech signal restoration using Eqs. 4–6 has several drawbacks: error accumulation (because of recursive order of processing) and long duration (because of iterative calculations). To avoid mentioned effects, here we propose to use only two reference periods of signal that are directly connected to the segment of speech to be restored. Such restoration could be viewed as parallel signal generation problem when restored signal periods are obtained changing parameters of reference signal period. Efficiency could be gained employing multirate processing technique for re-sampling of reference signal (Section 3.6 in Oppenheim and Schafer, 1989).

General sampling rate conversion by factor M/L (here M and L are integers) is expressed by

$$\hat{s}_v(m)=\sum_{n=n_1}^{n_2}\left[h\big(nL+(\!(mM)\!)_L\big)s_v\left(\left\lfloor\frac{mM}{L}\right\rfloor-n\right)\right]. \tag{7}$$

Here m indexes time, $\hat{s}_v(m)$ is re-sampled by factor M/L signal $s_v(n)$, $h(\,\cdot\,)$ is impulse response of low-pass FIR filter with gain equal to L and cutoff frequency equal to $\min(\pi/L, \pi/M)$, and $(\!(q)\!)_L = q - L\,\lfloor q/L \rfloor$ denotes q modulo L. In our forward processing case M and L are equal to the number of samples in reference period on the left $T(0)$ and period to be restored $T(k)$, correspondingly.

Taking into account (6), last equation could be re-written as follows

$$\hat{s}_v^f(m) = G^f(k))\sum_{n=n_1}^{n_2}\left[h^f\big(nL(k)+(\!(mM)\!)_{L(k)}\big)s_v\left(\left\lfloor\frac{mM}{L(k)}\right\rfloor-n\right)\right], \tag{8}$$

here k indexes periods of fundamental frequency, $G^f(k) = I(k)/I(0)$ is a gain of filter $h^f(\,\cdot\,)$, $I(0)$ and $I(k)$ are intensities of reference period and period to be restored, correspondingly. Notice that in (8) only parameter L and filters gain are dependent on k. Moreover, the equation can be viewed as cascade connections of interpolation and decimation procedures

$$\hat{s}_v^f(l) = G^f(k)\sum_{n=n_1}^{n_2}\left[h_L^f\big(nL+(\!(m)\!)_{L(k)}\big)s_v\left(\left\lfloor\frac{l}{L(k)}\right\rfloor-n\right)\right], \tag{9a}$$

$$\hat{s}_v(m) = \sum_{l=n_1}^{n_2} h_M^f\,(Mm-l)\,\hat{s}_v^f\,(l)\,. \tag{9b}$$

Here h_L^f is a lowpass FIR filter with gain equal to L and cutoff frequency equal to π/L, while h_M^f is a lowpass FIR filter with gain equal to 1 and cutoff frequency equal to π/M. From (9) follows that decimation has constant parameters for all periods of speech signal to be restored.

Notice that signal restoration could be carried out in both forward and backward in time directions simultaneously, moreover processing of all periods to be restored could be done in parallel.

Taking that view, incorporation of weighting function in (5) directly into a corresponding gain terms $G^f(k)$ could be beneficial. In order to proceed, let us re-write equation of weighting function switching indexes from time samples to periods as follows

$$w^{f/b}(k) = \frac{1}{2}\cdot\left[1 \pm \cos\left(\frac{k\cdot\pi}{K-1}\right)\right], \tag{10}$$

here K is a total number of periods to be restored. Now modified expression for gain in (9a) becomes $G^f(k) = w^f(k)I(k)/I(0)$.

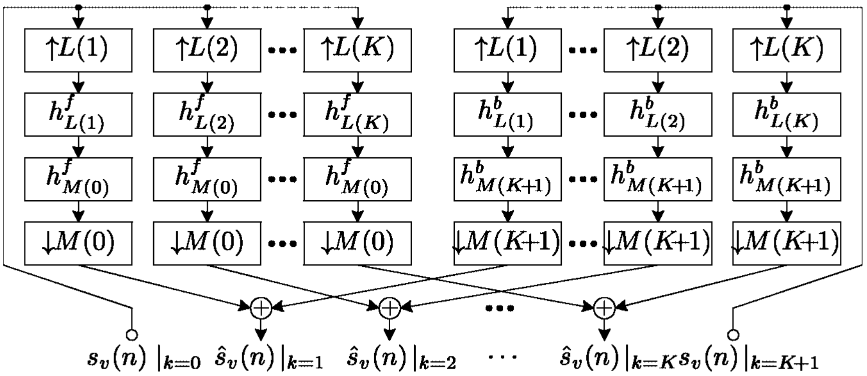

Fig. 2. Multirate implementation scheme for restoration of voiced speech signals $s_v(n)$. Notations: $\downarrow M$ – decimator, $\uparrow L$ – interpolator, h – filter, $k \in [0, K{+}1]$ – index of period of fundamental frequency.

Developed implementation scheme for restoration of voiced speech signals is presented in Fig. 2.

4. Experimental Study

For the following comparison of efficiency of implementations, let us to introduce mean error characteristics.

Total mean sum square error (MSE) of restoration length K of fundamental frequency periods invariant (up to $\Delta_{\max}$ periods shift) to restoration segment position is expressed (Paulikas and Navakauskas, 1998) by

$$e_{M,\,\Delta_{\max}} = \frac{1}{\Delta_{\max}} \sum_{\Delta=0}^{\Delta_{\max}} e_M(\Delta). \tag{11}$$

with bounds of it

$$e_M^{\min\,/\,\max} = \min_{\Delta\,\in[0,\,\Delta_{\max}]} / \max\; e_M(\Delta), \tag{12}$$

where $\Delta_{\max}$ is a maximum value of shift in periods, e_M is MSE of whole restoration with length of K fundamental frequency periods.

In Fig. 3 calculation results for total MSE and its limits for iterative restoration (shown in black color) using Eqs. 4–6 and multirate restoration implementation (shown in gray color) using (9a) and (9b) are shown. They summarize in total 17 different length restoration experiments carried out for each homograph.

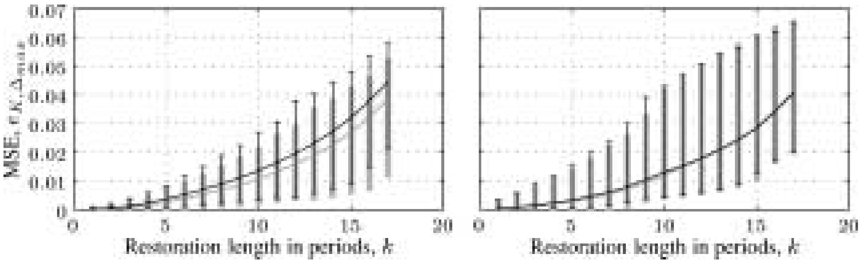

Fig. 3. Total MSE and its bounds for restoration of two syllables employing iterative (in black) and multirate (in gray) implementation.

5. Conclusions

The paper in general dealt with the restoration of comparatevely long duration voiced speech signals. We proposed implementation scheme that is based on the signal generation framework and is an alternative to previously developed iterative restoration procedure.

Presented efficient multirate implementation for restoration of voiced speech signals has following features:

- utilizes parallel processing order;
- possesses simplified non-iterative structure that employs ordinary interpolator and decimator filters;
- has comparable to iterative procedure total MSE performance.

Acknowledgements

Research was supported by The Swedish Institute New Visby project Ref. No. 2473/202 (381/T81) and The Royal Swedish Academy of Sciences.

References

Barauskaitė, J., G. Čepaitienė, D. Mikulenienė, J. Pabrėža and R. Petkevičienė (1995). *Lithuanian Language 1 (Lexicology, Phonetics, Accentology, Dialectology, Orthogrophy).* Science and Encyclopedia, Vilnius (in Lithuanian).

Botinis, A., B. Granström and B. Möbius (2001). Developments and paradigms in intonation research. *Speech Communication*, **33**(4), 263–296.

Czyzewski, A. (1997). Learning algorithms for audio signal enhancement. Part 1. Neural network implementation for the removal of impulse distortions. *Journal of the Audio Engineering Society*, **45**(10), 815–831.

Dubnowski, J. J., R.W. Schafer and L.R. Rabiner (1976). Real-time digital hardware pitch detector. *IEEE Transactions on Acoustic, Speech, and Signal Procesing*, **24**, 2–8.

Etter, W. (1996). Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters. *IEEE Transactions on Signal Processing*, **44**(5), 1124–1135.

Godsill, S. J., and P.J.W. Rayner (1995). A Bayesian approach to the restoration of degraded audio signals. *IEEE Transactions on Speech and Audio Processing*, **3**(4), 267–278.

Golovinas, B. (1982). *Introduction to Linguistics*. Mokslas, Vilnius (in Lithuanian).

Oppenheim, A. V., and R. Schafer (1989). *Discrete-Time Signal Processing*. Prentice Hall.

Pakerys, A. (1986). *Phonetics of Common Lithuanian Language*. Mokslas, Vilnius (in Lithuanian).

Paulikas, Š. (1999). *Restoration of Noise Distorted Accents Audio Records*. Doctoral dissertation, Vilnius Gediminas Technical University (in Lithuanian).

Paulikas, Š. (2001). Application of methods of digital electronics in frequency domain restoration of accent in speech signals. *Elektronika ir elektrotechnika*, **5**, 32–35.

Paulikas, Š., and D. Navakauskas (1998). Restoration of localized pitch and intensity variations of speech signals. In *Proceedings of the 1st International Conference Digital Signal Processing and its Applications*, vol. 1. Moscow, Russia. pp. 130–136.

Paulikas, Š., and D. Navakauskas (2003). Multirate implementation scheme for restoration of voiced speech signals. *Technical Report LiTH-ISY-R-2508*, Division of Automatic Control, Department of Electrical Engineering, Linköpings universitet, SE-581 83 Linköping, Sweden.

Vaseghi, S. V. (2000). *Advanced Signal Processing and Digital Noise Reduction*. John Wiley & Sons Ltd., England, 2 edition.

Vaseghi, S. V., and B.P. Milner (1993). Noise adaptive hidden Markov models based on Wiener filters. *Eurospeech*, **2**, 1023–1026.

Š. Paulikas born 1969 in Vilnius, Lithuania. Received engineers diploma with honor (1992), MSc degree (1994) and PhD degree (1999). Presently working as associated professor at Telecommunication Engineering Department of Vilnius Gediminas Technical University. Research interests are digital and speech signal processing.

D. Navakauskas is an associate professor at Radioelectronics Department of Vilnius Gediminas Technical University. He received honor diploma of radioelectronics engineer in 1992, MSc in electronics degree in 1994, doctor of electrical and electronical engineering degree in 1999, all at Vilnius Gediminas Technical University. His main research interests include artificial neural networks, speech signal processing and nonlinear signal processing.

Multidažninė schema skardiesiems kalbos signalams restauruoti

Šarūnas PAULIKAS, Dalius NAVAKAUSKAS

Šiame straipsnyje į anksčiau sukurtą iteracinę kalbos signalo restauravimo procedūrą, kai palyginti ilgos trukmės prozodiniai elementai tokie kaip kirtis ir priegaidė yra sugadinti, yra pažvelgiama iš signalo generavimo pusės. Remiantis šiuo požiūriu, čia aprašoma efektyvi skardaus kalbos signalo restauravimo realizacija. Ji remiasi lygiaigrečiu signalo apdorojimu naudojant interpoliacijos ir decimacijos filtrus. Pateikti eksperimentinio tyrimo rezultatai patvirtina pasiūlytos realizacijos tinkamumą signalų restauravimui.