ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 *2006. Nr. 5(69)*
ELEKTRONIKA IR ELEKTROTECHNIKA

TELECOMMUNICATION ENGINEERING
T180
TELEKOMUNIKACIJŲ INŽINERIJA

Quality Estimation of the Interpolation of Missing Speech Signal Segments

Š. Paulikas

Department of Telecommunications Engineering, Vilnius Gediminas Technical University,
Naugarduko str. 41-210, LT-03227 Vilnius, Lithuania, phone: +370 5 2744977; e-mail: sarunas.paulikas@el.vtu.lt

Introduction

In modern voice communications systems transmission errors, noise, distortion and losses due to low bit-rate coding and packet transmission corrupts received speech signal. Modern restoration techniques of speech signals employ Linear Prediction [1], Hidden Markov Models [2], Artificial Neural Networks [2], various Bayesian technique [3], to name a few. However, quality of the restored speech signal and usefulness of restoration technique is usually expressed as improved signal to noise ratio (SNR) or reduces mean square error (MSE) between original and restored speech signals. But just how do they relate to human perception?

This paper continues started in [4, 5] investigation of a problem of restoration of speech signal, when sufficiently long segments of speech signals constituting essential information are lost. It was shown that proposed restoration technique improves quality of restored speech signal in MSE sense, however, had problem in a junction of original and restored speech segments that was overcame by low-pass filtering of signal samples in the vicinity of a junction.

Here new method for approximation of speech signal characteristics will be proposed and employed in speech signal restoration algorithm. Also, quality of the restored signal will be estimated using perceptual evaluation of speech quality (PESQ) algorithm according to ITU recommendation P.862.

Approximation of Characteristics

Let us examine first syllable of Lithuanian word "*káltas*" spoken by male speaker of 450 ms duration, recorded with sampling rate of 8 kHz. Corresponding characteristics of the period of fundamental frequency (calculated by modified autocorrelation method with clipping) and intensity (calculated as a normalized power for each period of fundamental frequency) are shown by circles in Fig. 1 (a) and (b) respectively.

Suppose that the speech signal segment approximately from 0.12 s until 0.28 s is missing. In such time interval non-linearity of characteristics became to reveal (see Fig. 1). To obtain missing values of period of fundamental frequency and intensity characteristics the third order MSE approximation over anchor points (shown by crosses in Fig. 1) is proposed.

The employment of third order polynomial lets us

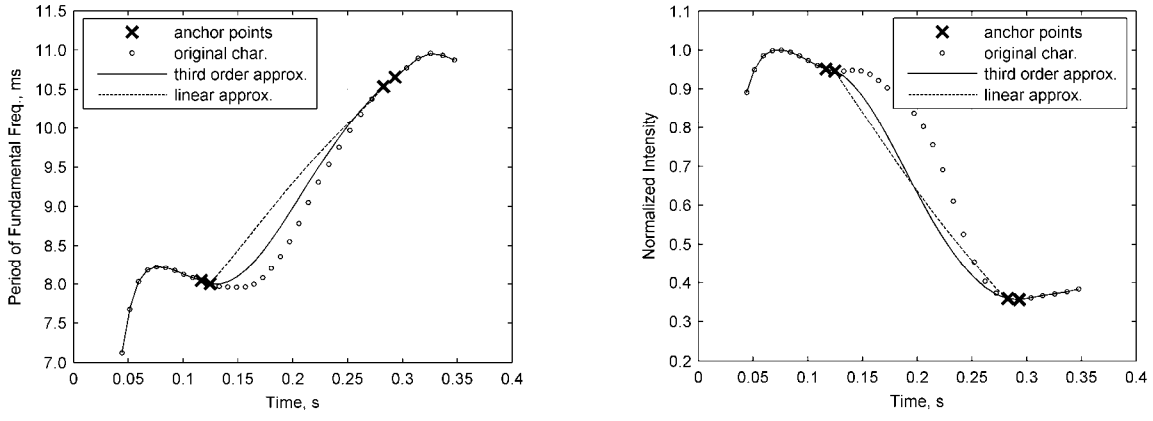

Fig. 1. Example of approximation of period of fundamental frequency and intensity characteristics of speech signal

Table 1. Approximation errors of period of fundamental frequency and intensity characteristics

	Third order approximation		Linear approximation	
	MSE	MAX	MSE	MAX
Intensity	0.0075	0.20	0.0078	0.19
Period of fund. freq.	1.7	3.04	7.7	6.15

have smooth transition between known and approximated parts of curves and preserve nature of characteristics course. In order to compare proposed method with ancestors which generally used linear interpolation [6], linear approximation of period of fundamental frequency and intensity characteristics is shown by dashed lines, too.

Approximation errors for example given in Fig. 1 are presented in Table 1. However values of these errors will depend not only on duration of approximated segment but also on its position. To evaluate relation of approximation errors with quality of restored speech signal the restoration experiment will be performed further.

Restoration of Speech Signal

As it was stated in [5], the voiced speech signal can be viewed as quasi-periodical signal (black and gray solid lines in Fig. 2 (a)) s_v with varying intensity $I(n)$ and period $T(n)$:

$$s_v(I(n),T(n),n) = I(n)\cdot s_v(n-T(n)). \quad (1)$$

In the restoration of a missing segment of speech signal (black solid line in Fig. 2 (a)) we employ forward together with backward processing in time [6], more precisely

$$\hat{s}_v(n) = w^f(n)\hat{s}_v^f(n) + w^b(n)\hat{s}_v^b(n), \quad (2)$$

where used weighting function is of a form

$$w^{f/b}(n) = \frac{1}{2}\cdot\left[1 \pm \cos\left(\frac{n\cdot\pi}{n_2 - n_1}\right)\right]. \quad (3)$$

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

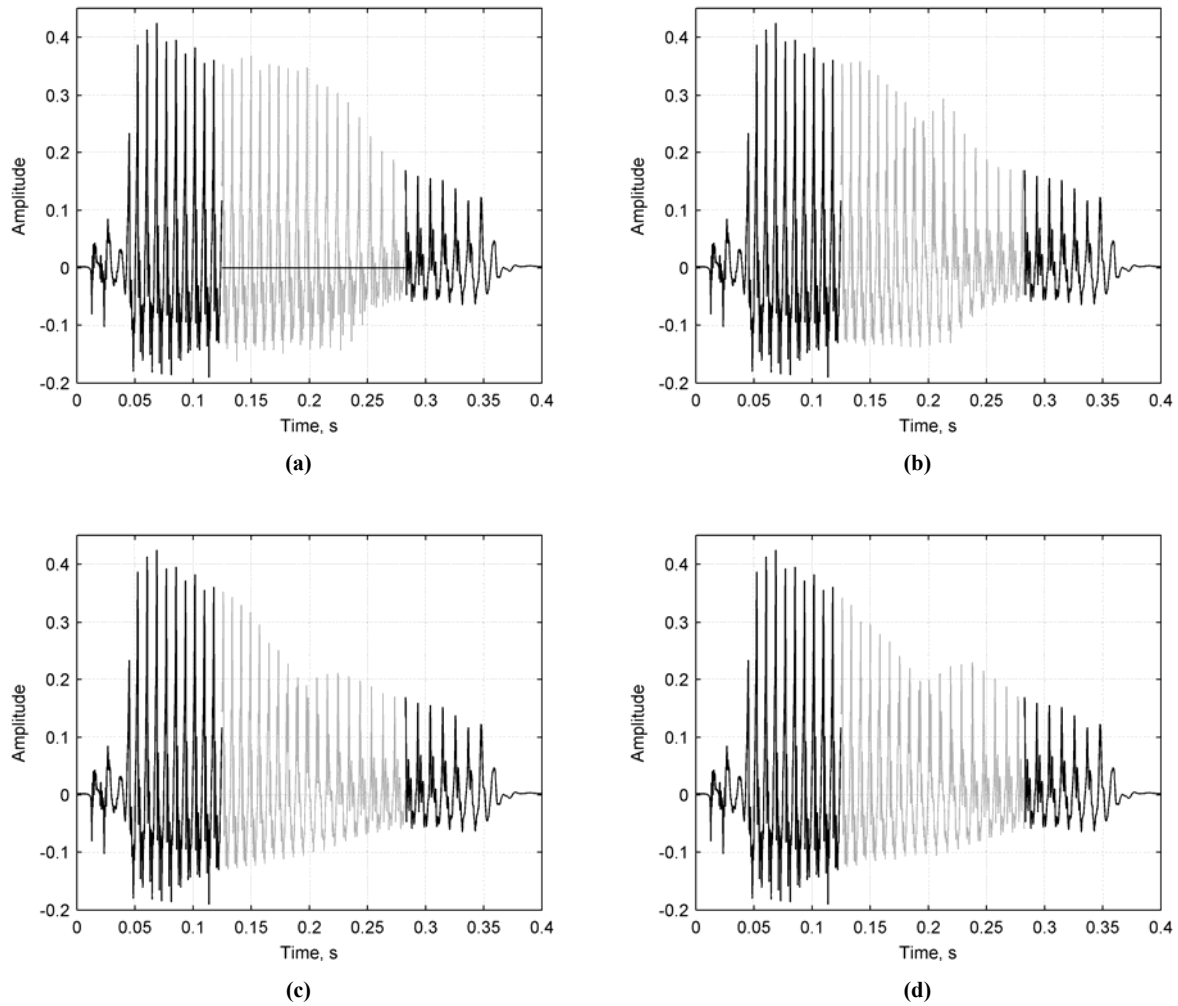

Fig. 2. Restoration example of the utterance of Lithuanian word "*káltas*": (a) shows distorted (black solid line) and undistorted (gray solid line) waveforms; (b), (c) and (d) presents restored waveforms (gray solid line) using original, approximated by third order and linear polynomials intensity and fundamental frequency characteristics respectively

Signal restoration carried out from different directions inheritable is the same, the main difference being in the time direction, i.e. indexes. Also, it is unnecessary to take into account variation of intensity and fundamental frequency at each time step [4], 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 (1) of the restored voiced speech signal could be re-written by

$$\hat{s}_v^f(n) = \frac{I(k)}{I(k-1)} \hat{s}_v^f\left(\left\lfloor \frac{T(k-1)(n-T(k-1))}{T(k)} \right\rfloor \right). \quad (4)$$

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, $\lfloor \bullet \rfloor$ is used here. Expression (4) could be used recursively in the restoration of accent in speech signal.

A restoration example of the utterance of Lithuanian word "káltas" using described method is shown in Fig. 2. Here (a) shows distorted (black solid line) and original (gray line) waveforms; (b) – restored using original, (c) and (d) – restored using approximated by third order and linear polynomial characteristics (see Fig. 1).

From restoration example shown in Fig. 2 can be seen that in all three approximation cases the lost speech signal segment was restored. However to judge quality of restoration from depicted waveforms is difficult or perhaps impossible.

Quality Evaluation

The quality of restored segment of speech signal will depend not only on distortion duration but also on place of missing segment. In order to evaluate quality of restored word using proposed method, restoration experiments were performed. Restoration was done on artificially distorted utterance of Lithuanian word "*káltas*" varying the place and duration of the distortion. To be more precise, the duration of distorted segment is varied from length of the 1 till the 25 periods of fundamental frequency, while the beginning of distorted segment varies approximately from 0.07 s until 0.32 s.

In restoration algorithm proposed new approximation of intensity and period of fundamental frequency characteristics method was used. In order to judge proposed approximation method a restoration using linear approximation of employed speech signal characteristics was performed, too. To show limits, i.e. best restoration possible, of described restoration technique a restoration with original characteristics also was done. In total 1405 restorations were performed.

The quality of restored speech signal was evaluated using PESQ algorithm that is able to predict subjective quality of speech signal with good correlation in a very wide range of conditions, which may include coding distortions, errors, noise, filtering, delay, and variable delay [7, 8]. A PESQ result is objective Mean Opinion Score (MOS), which is mapped to the MOS scale that directly expresses the voice quality. The PESQ MOS as defined by the ITU recommendation P.862 ranges from 1.0 (worst) up to 4.5 (best). The ITU scale ranges up to 5.0, but PESQ simulates a listening test and is optimized to reproduce the average result of all listeners (MOS stands for Mean Opinion Score). Statistics however prove that the best average result one can generally expect from a listening test is not 5.0 ("excellent") but 4.5.

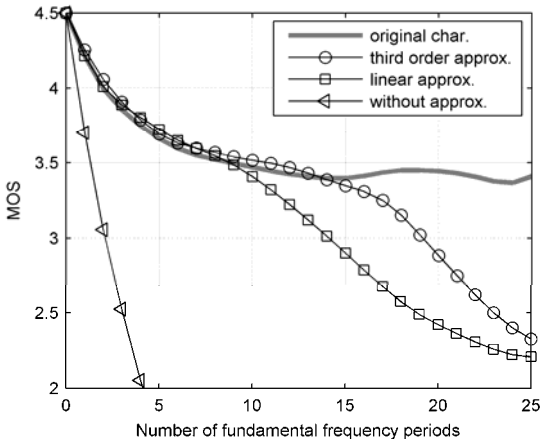

Fig. 3. Results of quality evaluation of restored utterance of Lithuanian word "*káltas*"

Experimentation results were summarized and mean values of MOS are depicted in Fig. 3. As a reference point was taken $\mathrm{MOS} = 4.5$ obtained by passing to PESQ algorithm undistorted word "*káltas*".

It can be seen that proposed restoration technique using third order approximation of employed speech signal characteristics outperforms linear interpolation starting from ninth period of fundamental frequency where starts to reveal non-linearity of intensity and fundamental frequency characteristics.

The fundamental limit of proposed method is depicted by solid gray line in Fig. 2. In this case speech signal restoration algorithm uses original intensity and fundamental frequency characteristics.

Conclusions

- Proposed approximation method of speech signal characteristics increases restoration accuracy in the cases when speech signal is corrupted by long term impulsive noise or gaps.
- The employment of third order polynomial gives smooth transition between known and approximated parts of curves and preserves nature of characteristics course, thus bypassing low-pass filtering of signal samples in the vicinity of a junction and simplifying restoration procedure.
- By the experimental study was shown that the approximation of period of fundamental frequency and intensity characteristics using the third order polynomial comparing with the linear approximation improves quality of restored speech signal in case where intensity and fundamental frequency characteristics becomes non-liner.

References

1. **Vaseghi S. V.** Advanced Signal Processing and Digital Noise Reduction // John Wiley & Sons Ltd. 2nd edition, 2000.
2. **Czyzewski A.** Learning algorithms for audio signal enhancement: part 1 neural network implementation for the removal of impulse distortions // J. Audio Eng. Soc. – 1997. – Vol. 45. – P. 815 – 31.
3. **Godsill S. J., Rayner P. J. W. A** Bayesian approach to the restoration of degraded audio signals // IEEE Trans. on Speech and Audio Processing. – 1995. – Vol. 3. – No. 4. – P. 267 – 278.
4. **Paulikas Š.** Application of Methods of Digital Electronics in Restoration of Accent in Speech Signals // Electronics and Electrical Engineering. – 2000. – Nr. 6(29). – P. 54 – 59.
5. **Paulikas Š., Navakauskas D.** Restoration of voiced speech signals preserving prosodic features // Speech Communication.– 2005. – Vol. 47. – Issue. 4. – P. 457 – 468.
6. **Etter W.** Restoration of a Discrete-Time Signal Segment by Interpolation Based on the Left-Sided and Right-Sided Autoregressive Parameters // IEEE Trans. on Signal Processing. – 1996. – Vol. 44, No. 5. – P. 1124–1135.
7. **Beerends J. G., Rix A. W., Hollier M. P., Hekstra A. P.** Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for End-to-End Speech Quality Assessment, Part I – Time-Delay Compensation // J. Audio Eng. Soc. – 2002. – Vol. 50. – No. 10. – P. 755–764.
8. **Beerends J. G., Rix A. W., Hollier M. P., Hekstra A. P.** Perceptual Evaluation of Speech Quality (PESQ) The New ITU Standard for End-to-End Speech Quality Assessment, Part II – Psychoacoustic Model // J. Audio Eng. Soc. – 2002. – Vol. 50. – No. 10. – P. 765–778.

Submitted for publication 2006 02 28

Š. Paulikas. Quality Estimation of the Interpolation of Missing Speech Signal Segments // Electronics and Electrical Engineering. – Kaunas: Technologija, 2006. – No. 5(69). – P. 63–66.

New method for approximation of intensity and period of fundamental frequency characteristics of speech signals is proposed and incorporated in algorithm of restoration of speech signal distorted by lengthy impulsive noise or gap. The suitability of presented restoration technique is judged by evaluating quality of restored speech signal. The quality is estimated using perceptual evaluation of speech quality (PESQ) algorithm that is based on ITU recommendation P.862. Shown that proposed method is more accurate in restoration of distorted speech segment where characteristics of speech signal have non-linear change in time. Ill. 3, bibl. 8 (in English; summaries in English, Russian and Lithuanian).

Ш. Пауликас. Качество инерполяции потеряных сегментов речевого сигнала // Электроника и электротехника. – Каунас: Технология, 2001. – № 5(69). – С. 63–66.

Предлагается новый метод для аппроксимации характеристик интенсивности и периода фундаментального тона речевого сигнала на котором основана техника восстановления речевого сигнала изкаженного длительным импульсным шумом. Годность представленого метода измерена качеством восстановленного речевого сигнала, которая найдена используя алгоритм для измерения понятливости речи (PESQ) по ITU рекомендации P.862. Экспериментально показано, что предложен метод точнее воспроизводит изкаженный сегмент речевого сигнала, в котором характеристики речевого сигнала имеют нелинейное изменение во времени. Ил. 3, библ. 8 (на английском языке, рефераты на английском, русском и литовском яз.).

Š. Paulikas. Prarastų kalbos signalo segmentų interpoliavimo kokybės tyrimas // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2006. – Nr. 5(69). – P. 63–66.

Pristatomas naujas kalbos signalo intensyvumo ir pagrindinio tono periodo charakteristikų aproksimavimo metodas, kuris toliau pritaikomas atkuriant kalbos signalą sugadintą ilgai trunkančio impulsinio triukšmo ar segmento praradimo. Pasiūlytos atkūrimo technikos tinkamumas vertinamas atkurto kalbos signalo kokybe. Kalbos kokybė nustatoma taikant kalbos suprantamumo įverčio (PESQ) radimo algoritmą, kurį reglamentuoja ITU P.862 rekomendacija. Eksperimentiškai įrodoma, kad pasiūlytu metodu tiksliau atkuriami prarasti kalbos signalo segmentai, kuriuose kalbos charakteristikos laikui bėgant kinta netiesiškai. Il. 3, bibl. 8 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).