

Application of Linear Prediction Coefficients Interpolation in Speech Signal Coding

Š. Paulikas, R. Karpavičius

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.vgtu.lt, rimantask@email.com

Introduction

Linear Prediction Coefficients (LPC) are extensively used in many fields of digital speech signal processing. Modern voice communication systems employ LPC for coding and decoding speech signals in order to reduce data bit rate and save bandwidth. Incorporating LPC interpolation techniques in voice codecs higher data compression rates could be achieved. However, transmission errors, noise, distortion and losses due to low bit-rate coding and packet transmission corrupt received data and thus received speech signal quality reduces. LPC interpolation in this case could help to recover lost or corrupted speech segments and improve decoded speech signal quality.

Interpolation of LPC is not new idea [1], however application of different interpolation methods dedicated to restoration of speech signal is not widely analyzed. Most recent LPC interpolation methods employ only linear interpolation and usually are limited to duration of few frames (~20 ms) interval.

This paper continues study [2] of investigation of a problem of restoration of speech signal, when sufficiently long segments of voiced speech signals are lost. Here we target problem of restoration of speech signal in packet communication systems that employs LPC speech coding techniques. We will investigate the dependence of quality of restored speech signal on duration of interpolation using different LPC interpolation approaches. The speech signal quality will be estimated using objective (spectral distortions, number of unstable filters) and subjective (mean opinion score (MOS), that encounter the perceptual abilities of a human) measures.

Interpolation of LPC

Most voice codecs uses linear prediction (LP) to model short time spectrum of speech signal. Usually LPC are calculated for 10÷30 ms duration speech signal segments (frames) [3]. However, if during such time interval the spectral characteristics of speech signal will vary, the LPC coefficients between adjacent frames will

change abruptly. Consequently, the restored speech signal will be significantly distorted. Such abrupt change in LPC values could be avoided by calculating LPC more often – for each 5 ms interval (sub-frame) [4]. But this will increase the data bit rate of speech codec. In order to make transition of LPC coefficients more continuous usually linear interpolation of LPC between sub-frames is employed [3].

Interpolation is possible not only for LPC, but also for other their representations: reflection coefficients, cepstrum coefficients, autocorrelation coefficients, linear spectral pair coefficients, etc. Each representation method models the same spectral envelope of speech signal; however, their interpolation produces different results [5, 6]. Results of [5, 6] shows, that using LPC interpolation some times produces unstable LP synthesis filters. To avoid that interpolated LPC should be additionally processed. Most suitable for interpolation are linear spectral pair coefficients that lead to always stable speech synthesis filters. However, representing LPC by linear spectral pairs also requires additional computations, thus some codecs still uses LPC interpolation. Also, these results were obtained for interpolation lengths of 10÷30 ms speech signal frames and only linear interpolation were used. In our research we extend the interpolation interval to duration of several frames and examine the performance of different interpolation methods.

Interpolation Methods

In this study for the interpolation of LPC coefficients we employ interpolation using square, cubic, Hermite polynomials and splines. Obtained results will be compared with results of linear interpolation.

The interpolation polynomials are described using general Newton polynomial formula [7]:

$$L_n(x) = f(x_0) + f(x_0, x_1)(x - x_0) + f(x_0, x_1, x_2)(x - x_0)(x - x_1) + \dots + f(x_0, x_1, \dots, x_n)(x - x_0)(x - x_1) \cdots (x - x_{n-1}). \quad (1)$$

Here $\{x_i\}$ are interpolation reference points. The linear and square interpolation polynomials are special case of (2).

Splines is relatively simple interpolation method, when interpolation function is composed from low order (first, second or third) polynomials [5].

Interpolation using second order splines is defined as piece wise square function in each interval $[x_i, x_{i+1}]$:

$$S_{2i}(x) = a_i x^2 + b_i x + c_i, \quad i = 0, 1, \dots, N-1. \quad (2)$$

The coefficients are such that first derivate of $S_2(x)$ is continuous function. Similarly, the third order splines are piece wise third order polynomials

$$S_{3i}(x) = a_i x^3 + b_i x^2 + c_i x + d_i, \quad i = 0, 1, \dots, N-1 \quad (3)$$

and first and second derivates of $S_3(x)$ are continuous functions.

One more widely used in practice interpolation method employs piece wise cubic Hermite polynomials [8]:

$$H(x) = \begin{cases} H_1(x), & x_1 \leq x \leq x_2, \\ H_2(x), & x_2 \leq x \leq x_3, \\ \vdots & \vdots \\ H_{n-1}(x), & x_{n-1} \leq x \leq x_n. \end{cases} \quad (4)$$

Here polynomials $H_i(x)$ are:

$$H_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad (5)$$

Here $a_i = y_i$, $b_i = y'_i$,

$$c_i = \frac{1}{\delta} \left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i} - y'_i \right),$$

$$d_i = \frac{1}{\delta^2} \left(\frac{1}{2} (y'_i + y'_{i+1}) - \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right)$$

and

$$\delta = x_2 - x_1.$$

Quality Evaluation

To compare performance of proposed interpolation techniques we evaluated objective quality of restored speech signal using LPC interpolation. The quality of speech signal is evaluated using objective and subjective measures. The objective measures estimates quality of speech signal by comparing its properties with non-distorted reference signal, and do not takes into account the nature of speech and perceptual abilities of a human listener. More common objective measures are signal to noise ratio (SNR) and spectral distortions of restored signal.

Logarithmic spectral distortions for the duration of interpolation interval could be calculated as follows [9]:

$$D = \sqrt{\frac{1}{n_1 - n_0} \sum_{k=n_0}^{n_1-1} \left[10 \log_{10} \left(\frac{S(e^{j2\pi k/N})}{\hat{S}(e^{j2\pi k/N})} \right) \right]^2}. \quad (6)$$

Here $S(e^{j2\pi k/N})$ and $\hat{S}(e^{j2\pi k/N})$ – original and restored speech signal power spectrum, respectively, n_0 and n_1 - limits of interpolation interval, and $N = n_1 - n_0$.

The interpolation error for LPC could be estimated as ratio of signal and noise energies:

$$SNR = 10 \log_{10} \frac{\sum_{n=-\infty}^{\infty} s[n]^2}{\sum_{n=-\infty}^{\infty} (s[n] - \hat{s}[n])^2}, \quad (7)$$

here $s[n]$ – original and $\hat{s}[n]$ – restored signals.

Contrarily, the subjective quality measures of speech signal encounter the perceptual abilities of a human. The subjective quality of speech usually is evaluated by group of listeners performing listening test. As result the mean opinion score (MOS) (numbers from 1 - worst until 5 - excellent) of speech quality is obtained. However, such experiments usually are expensive and take a long time to setup.

To overcome that the quality of restored speech signal could be 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 [10]. 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. Statistics however prove that the best average result one can generally expect from a listening test is not 5.0 ("excellent") but 4.5.

Experimental Study

For the experiments we used voice records from NOIZEUS data base [11]. The discretization frequency of voice record is 8000 Hz and quantization - 16 bits. The LPC interpolation were done on voiced speech with interpolation intervals and sounds of examined words summarized in Table 1.

During experiment signals were segmented into frames of 20 ms, and place and number of interpolated segments were changed. In total in the experiments were used 15 different words, 12 of them were spoken by male and 3 - by female speakers.

The interpolation was performed using six interpolation techniques discussed previously: linear, square, cubic and Hermite polynomials and second and third order splines. So, during experiments totally 2064 interpolations were performed.

Table 1. Words and sounds used in experiment

Sound	Word	Max. interpolation interval, ms
b	bright	100
	best	40
d	read	40
	seldom	60
g	green	80
	good	60
dž	joy	120
	edges	120
h	hands	80
em	items	180
en	hands	180
il	skill	160
im	simplest	120
in	thin	180
un	round	120

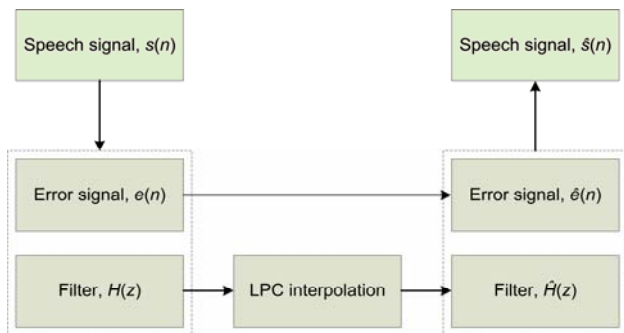


Fig. 1. Experimentation scheme

Experimentation was done using scheme shown in Fig. 1, and experiment averaged results (except Fig. 2 that presents total numbers) are depicted in Figs. 2 – 5. In order to indicate dispersion of obtained results, the standard deviations were calculated and lower limits (worst situation) also are shown in Figs. 3 - 5. In order to clearly present overlapping standard deviations, all curves (except linear approximation) are shifted in X -axis around particular integer number of interpolated LPC sets.

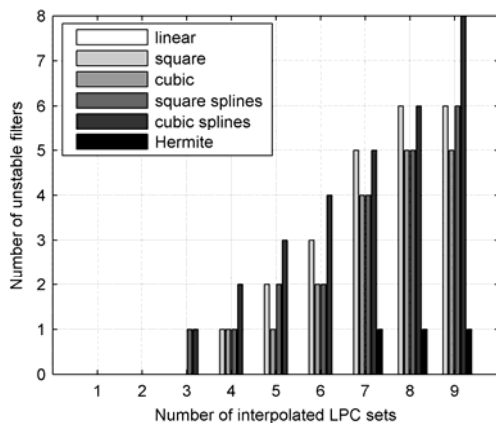


Fig. 2. Dependence of number of unstable filters on duration of interpolated interval

Fig. 2 shows number of unstable filters obtained after interpolation of LPC's. It should be mentioned that only linear interpolation guarantees filters stability. All other interpolations develop unstable filters. However, Hermite

polynomials perform quite well; while interpolation length is less than six LPC sets all filters were stable. Worst results were obtained by using third order splines.

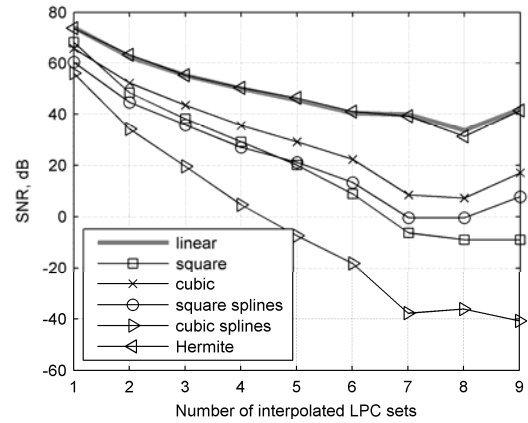


Fig. 3. Average signal to noise ratio dependence on duration of interpolated interval

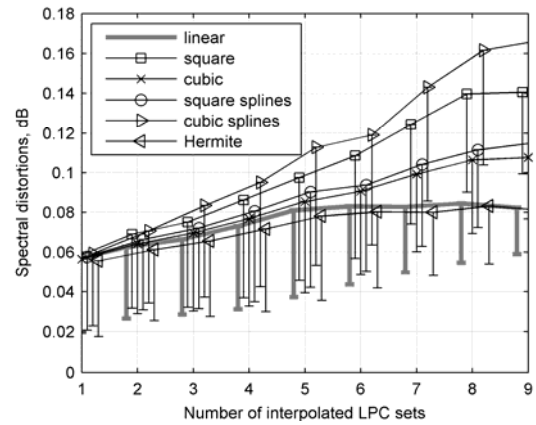


Fig. 4. Average spectral distortions and its standard deviation dependence on duration of interpolated interval

Figs. 3 and 4 show averaged SNR and averaged spectral distortions of restored signals after LPC interpolation. In both cases, the linear and Hermite polynomials give best results, i.e. they produce minimum distortions. Maximal distortions were obtained when interpolation is done using third order splines.

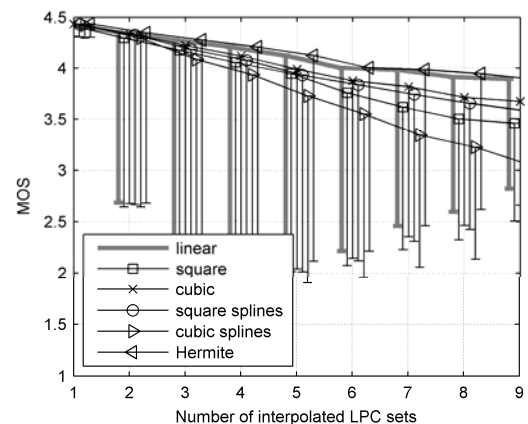


Fig. 5. Average MOS scores by PESQ and its standard deviations dependence on duration of interpolated interval

Fig. 5 shows obtained MOS scores using PESQ algorithm. Again, it can be seen that linear and Hermite polynomials perform best, and third order splines produces worst MOS scores. However, standard deviations of MOS results are greater than standard deviations of SNR and spectral distortions.

Conclusions

- Presented study shows that direct interpolation of LPC can be applied for synthesis of comparatively long, several frames duration voiced speech segments.
- According to experimental results the most suitable techniques for interpolation of LPC are using linear and Hermite polynomials.
- Interpolation using higher order polynomials and cubic and third order splines without extra post processing, produces relatively large alternations in obtained LPC that led to big number of unstable filters and other distortions.

References

1. **Erkelens J. S., Broersen P. M. T.** LPC Interpolation by Approximation of the Sample Autocorrelation Function // . IEEE Trans. On Audio and Speech Processing. – 1998. – Vol. 6, No. 6. – P. 569–573.
2. **Paulikas Š.** Quality Estimation of the Interpolation of Missing Speech Signal Segments // Electronics and Electrical Engineering. – Kaunas: Technologija, 2006. – No. 5(69). – P. 63–66.
3. **Islam T., Kabal P.** Partial-energy weighted interpolation of linear prediction coefficients. Electrical & Computer Engineering McGill University, Montreal, Canada. Proc. IEEE Workshop Speech Coding (Delavan, WI). – Sept. 2000. – P. 105–107.
4. **Paliwal K. K., Kleijn W. B.** Quantization of LPC Parameters // Speech Coding and Synthesis. – Elsevier, Amsterdam, 1995. – P. 443–466.
5. **Umezaki T., Itakura F.** Analysis of time fluctuating characteristics of Linear predictive coefficients // Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing. –Tokyo. – 1986. – P. 1257–1260.
6. **Paliwal K. K.** Interpolation properties of linear prediction parametric representations // Proc. European Conf. on Speech Commun. and Technology. – Madrid. – 1995.
7. **Čiegis R., Būda V.** Skaičiuojamoji matematika. – Vilnius: TEV, 1997. – 240 p.
8. **Fritsch F. N., Carlson R. E.** Monotone Piecewise Cubic Interpolation // SIAM Journal of Numerical Analysis. – No. 2(17). – 1980. – P. 238–246.
9. **Pereira W.** Modifying LPC Parameter Dynamics to Improve Speech Coder Efficiency, Master's thesis, McGill University, Montreal, Kanada. – 2001.
10. **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. 755–778.
11. **NOIZEUS:** A noisy speech corpus for evaluation of speech enhancement algorithms. Accessed on internet at <http://www.utdallas.edu/~loizou/speech/noizeus/>.

Submitted for publication 2006 02 28

Š. Paulikas, R. Karpavičius. Application of Linear Prediction Coefficients Interpolation in Speech Signal Coding // Electronics and Electrical Engineering. – Kaunas: Technologija, 2007. – No. 8(80). – P. 39–42.

Several interpolation techniques of linear prediction coefficients (LPC) for speech signal coding were experimentally analyzed. The obtained results show that direct interpolation of LPC can be applied for coding and decoding of significantly long segments of voiced speech. The best performance in sense of least number of unstable filters, smallest spectral distortions and greatest MOS score, was achieved when for interpolation were used linear and Hermite polynomials. Higher order interpolation polynomials such as cubic or third order splines, produced significant alternations in obtained LPC that led to worst results of the quality of the decoded speech. Also higher order polynomials without additional post processing of interpolated LPC produced more unstable filters. Ill. 5, bibl. 11 (in English; summaries in English, Russian, Lithuanian).

Ш. Пауликas, Р. Карпавичюс. Применение интерполяции коэффициентов линейного прогноза для кодировки речевого сигнала // Электроника и электротехника. – Каунас: Технология, 2007. – № 8(80). – С. 39–42.

Экспериментально было проанализировано несколько методик интерполяции коэффициентов линейного прогноза (КЛП) для кодировки речевого сигнала. Полученные результаты показывают, что прямую интерполяцию КЛП можно использовать для синтеза сравнительно длинных сегментов речи. Лучшие результаты по наименьшему числу неустойчивых фильтров, минимуму спектральных искажений и высшего MOS были достигнуты при интерполяции линейными и Hermite полиномами. Интерполяция полиномами более высокого порядка, такими как кубические полиномы или сплайны, вызывает значительные вариации в полученных КЛП, что привело к самым плохим результатам качества декодированной речи. Также они без дополнительной обработки результатов интерполяции КЛП произвели больше неустойчивых фильтров. Ил. 5, библи. 11 (на английском языке; рефераты на английском, русском и литовском яз.).

Š. Paulikas, R. Karpavičius. Tiesinės prognozės koeficientų interpoliavimo taikymas kalbos signalui koduoti // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2007. – Nr. 8(80). – P. 39–42.

Eksperimentiškai tiriami įvairūs tiesinės prognozės koeficientų (TPK) interpoliavimo būdai, taikomi koduojant kalbos signalą. Eksperimento rezultatai rodo, kad tiesioginis TPK interpoliavimas gali būti taikomas gana ilgiems kalbos signalo segmentams koduoti ir dekoduoti. Pagal mažiausią sukurtų nestabilių filtrų skaičių, mažiausius spektrinius iškreipius ir gautus didžiausius MOS įverčius tam geriausiai tinka tiesinis TPK interpoliavimas arba interpoliavimas Hermite'o daugianariais. Interpoliavimas aukštesnio laipsnio daugianariais bei antrojo ir trečiojo laipsnio splineais sukelia didokius TPK svyravimus, o tai blogina sintezuotos kalbos kokybę. Be papildomo interpoliavimo rezultatų apdorojimo šie daugianariai sukuria ir gana daug nestabilių filtrų. Il. 5, bibl. 11 (anglų kalba; santraukos anglų, rusų, lietuvių k.).