LPC Quantization and Interpolation in Coding for Speech Storage Applications

Carsten HOELPER, Astrid FRANKORT

Institute of Communication Systems and Data Processing, RWTH Aachen University,
Templergraben 55, 52056 Aachen, Germany

hoelper@ind.rwth-aachen.de, frankort@ind.rwth-aachen.de

Abstract. *In this paper, personalized quantization of the filter coefficients of Linear Predictive Coding (LPC) is studied. The study covers two aspects. On the one hand, a signal-adaptive algorithm which determines when to transmit a set of LPC coefficients is introduced. This algorithm allows a reduction of about 35% of the bit rate needed to code the LPC coefficients in speech storage applications – e.g. voice prompts in mobile terminals.*

The second part of this paper deals with the quantizing of LPC coefficients. Different approaches to quantization are compared in the context of speech coding for storage applications: Split Vector Quantization (SVQ), e.g. as used in the Adaptive Multi Rate Speech Codec (AMR) for GSM and UMTS, near optimum vector quantization with the LBG Algorithm (Linde, Buzo, Gray), and Lattice Quantization (LQ) with and without entropy coding.

Keywords

Linear Prediction, LSP Quantization, LPC Interpolation, Personal Speech Coding.

1. Introduction

Speech coding for storage applications (e.g. voice prompts in mobile terminals) has different limitations and demands compared to coding for telephone transmission. Real time operation is not an issue – the signal is encoded only once on a powerful workstation, pre-recorded speech is not degraded by unknown noise, only one speaker pronounces the whole text, and the complete speech signal is known at the time of encoding. These special circumstances allow a number of unconventional coding strategies.

This paper concentrates on calculation, quantization, and storage of the filter coefficients for Linear Predictive Coding (LPC). In section two, a new algorithm is presented that decides at which time instants a set of LPC coefficients needs to be stored. As all sets of stored coefficients are available at the time of decoding, no additional delay is introduced with this extended interpolation scheme. Section three comprises a survey of different quantization schemes that can be used to quantize the LPC coefficients. In section four the results are briefly summarized.

2. Signal Adaptive Interpolation of LPC Parameters

Although filtering in linear prediction is usually carried out on a 5 ms subframe basis, LPC coefficients are only transmitted every 20 ms and interpolated in between (e.g. in the AMR speech codec [1]).

As, in speech storage systems, the speech signal is known completely at the time of encoding and all parameters are present at the time of decoding, the coefficients may be stored in variable distances.

Most commonly in narrowband speech coding, an LPC filter of order $N_p = 10$ is studied. To eliminate block effects, windowing is applied. The window function $w(k)$ consists of half a Hamming window and a quarter cosine function:

$$w(k) = \begin{cases} 0.54 - 0.46\cos\left(\dfrac{2\pi k}{2L_1 - 1}\right), & k = 0,\ldots, L_1 - 1 \\ \cos\left(\dfrac{2\pi(k - L_1)}{4L_2 - 1}\right), & k = L_1,\ldots, L_1 + L_2 - 1 \end{cases} \quad (1)$$

with $L_1 = 200$ and $L_2 = 40$. Lookahead and lookback have a length of 5ms.

Since the emphasis of the window is on the last subframe of the actual speech frame, the transmitted LPC coefficients correspond to the fourth subframe. The coefficients that correspond to the remaining three subframes are linearly interpolated in the LSP domain (Line Spectral Pairs), as described in [3].

If plotted over time, the Line Spectral Pairs (LSP) will not show any discontinuities, which is important for interpolation. In voiced speech, they stay constant for even longer periods of time. Thus, the fixed pattern of calculated and transmitted / stored LSP vectors (called sampling points in this paper) every four subframes as in AMR may

be given up, and the distance between the sampling points can be adapted to the actual speech signal. The distance between two sampling points is limited by the perceptual quality of the output speech. Possible measures to predict the speech quality from the similarity of LSP vectors are

- the cepstral distance (d_{cep}) of two sets of coefficients as presented in [4]
- the short-time prediction gain (*STPG*) which measures the quality of the linear prediction by comparing the residual with the original signal.
- a weighted distance measure (E_{LSP}) as it is used for quantization in the AMR codec [1]:

$$E_{LSP} = \sum_{i=1}^{10} \left(w_i LSP_i^{AMR} - w_i LSP_i^{new} \right) \tag{2}$$

$$w_i = \begin{cases} 3.347 - \dfrac{1.547 d_i}{450}, & d_i < 450 \\ 1.8 - \dfrac{0.8(d_i - 450)}{1050}, & else \end{cases} \tag{3}$$

Measurement of similarity can be performed either between the LSPs of subsequent subframes or towards a reference coder with fixed LSP calculation pattern. Both methods have been analyzed and implemented. The latter proved to be more suitable in practice.

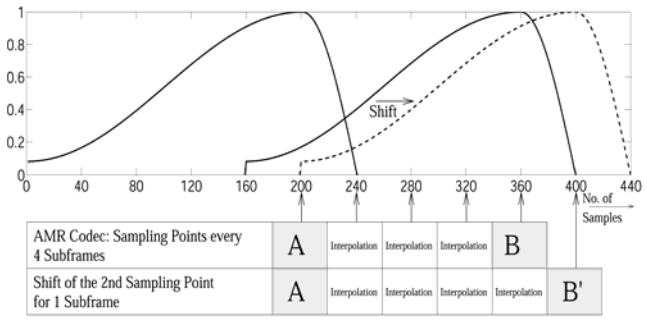

Fig. 1. AMR Sampling Points and Interpolated Subframes.

The LSP sets of the new interpolation scheme are compared to those of the AMR codec, mode 7.95 kbit/s, to determine the maximum possible distance between two sampling points,. The algorithm is designed as follows: The very first sampling point *A* (see Figure 1) is identical to the first LSP set of the AMR codec. Starting from here, a few frames are analyzed, using the AMR routines which calculate a set of LSPs every 4th subframe, and which interpolate three LSP sets in between. The resulting LSP vectors are memorized, whereupon sampling point *B* is shifted by one subframe (which makes *B'* the exactly calculated LSP set), thus 4 LSP sets now have to be interpolated in between. After this, the interval *AB'* is compared to the reference (AMR), using one of the three measures mentioned above. If the termination criterion of the measure has not yet been fulfilled, B' is shifted by one more subframe to become B'', and so on.

Reaching the termination criterion, the current interval AB^n is no longer similar enough to the reference, and B^{n-1} will be stored and used as starting point for the next iteration. This is repeated until the end of the speech signal is reached.

As the algorithm starts with a minimum distance of four subframes between the sampling points, it will never deliver more sampling points than the AMR codec does.

The quality of this algorithm was determined by speech synthesis using the reduced number of sampling points. As the aural impression cannot be measured objectively, the overall prediction gain is used as a quality measure. The overall prediction gain is the mean prediction gains over all subframes. It adequately resembles the aural impression and rises with increasing quality. If LSPs are calculated for each subframe explicitly, which means none of them is interpolated, the overall prediction gain is about 16.2dB, depending on the processed speech material. Comparing all three measures, it turns out, that the overall prediction gain using the short term prediction gain algorithm is the highest for the same reduction rates. Informal listening tests have proved these results. Figure 2 shows an exemplary progression of the overall prediction gain versus the reduction of sampling points. Reductions of up to 35% will not significantly lower the perceptual speech quality.

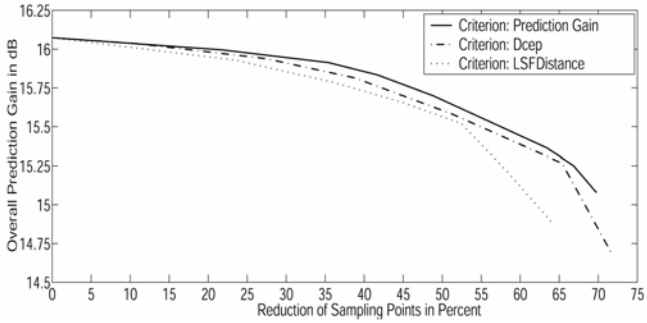

Fig. 2. Overall Prediction Gain in dB versus Percentile Reduction of Sampling Points.

3. Quantization of LSP Vectors

The LSP vectors of each sampling point have to be quantized for storage. While quantization should be accomplished with minimum bit rate, the distortion introduced by this quantization should remain as small as possible. A common criterion is that the average spectral distortion must be less than 1dB, while no more than 2% of the quantized LSP vectors may exceed 2dB spectral distortion [5]. In speech storage systems, not only the bit rate required per set of LSPs is crucial, but also the amount of memory needed for quantization tables. Depending on the length of the actual speech signal, one or the other part is dominant for the gross bit rate.

3.1 Split Vector Quantization

The first method to be evaluated was the SVQ scheme of the AMR speech codec, mode 7.95 kbit/s. The LSP vector of dimension $N = 10$ is divided into three subvectors of dimensions $N_1 = 3$, $N_2 = 3$, and $N_3 = 4$. Each subvector is quantized with 9 bit according to three fixed quantization tables.

This results in a bit rate of 27 bit per LSP vector plus one single quantization table (independent of the speech signal) of size $(3+3+4) \cdot 2^9 \cdot 16$ bit = 81 920 bit.

3.2 Source Adaptive Vector Quantization with the LBG Algorithm

As illustrated in section one, in speech storage applications the complete signal is known, and spoken by only one single speaker. This allows useing a trained LBG vector quantizer [2] of full order $N = 10$ to find the minimum bit rate needed for quantization of each set of LSPs. Simulations with speech material in German, French, and English indicate that the quantization of the LSP sets of a single speaker requires about 14 bit. If voiced speech and unvoiced speech are handled separately as suggested in [6], the LSP vectors of a single person can be quantized with 12 bit for voiced speech and 9 bit for unvoiced speech. In the latter case, the distortion criterion is an average spectral distortion below 2dB for unvoiced speech [6].

The penalty for achieving optimum bit rates for the quantization of each LSP set are huge quantization tables of $2^{14} \cdot 10 \cdot 16$ bit = 2 621 440 bit for the general scheme or $(2^9 + 2^{12}) \cdot 10 \cdot 16$ bit = 737 280 bit for the voicing specific quantization.

3.3 Lattice Quantization

To avoid the huge tables in case of short speech signals, Lattice Quantization was studied. A lattice is the set of all points l

$$\Lambda = \left\{ l \in \mathbf{R}^n : l = \sum_{i=1}^{n} v_i a_i \quad v_1, v_2, ..., v_i \in \mathbf{Z} \right\} \tag{4}$$

which can be expressed as a linear combination of integer multiples of the basis vectors $a_1, a_2, \ldots, a_n \in \boldsymbol{R}^n$. Different possible lattices of order $N = 2$ are shown in Figure 3.

Vector quantizers which use such an equally distributed n-dimensional lattice of code vectors are called lattice quantizers. As the position of each code vector can be expressed analytically, no codebook needs to be stored. The main drawback compared to source adaptive vector quantizers is, that the distribution of code vectors does not match the distribution of the LSP vectors, which results in a higher bit rate for each vector to be quantized. For further reference to lattice quantizers the reader is referred to [7].

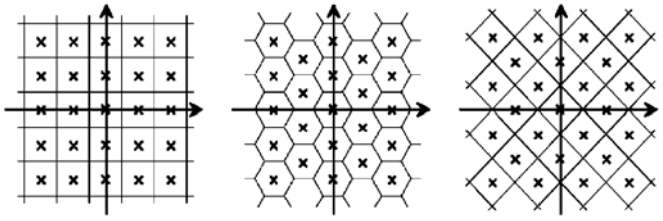

Fig. 3 Lattices Z_2, A_2 and D_2.

For quantization of the 10-dimensional LSP vectors, the D_{10}^+ lattice was chosen, which yields a low distortion with a simple quantization and index assignment scheme based on iso-norm, absolute leader, and signed leader [7].

Using the D_{10}^+ lattice for quantization of LSP vectors of one single speaker, a bit rate of 50 bit per vector is necessary to keep the average spectral distortion below 1dB for voiced speech, and 39 bit to keep the average distortion of unvoiced speech below 2dB. Due to the regular structure of the lattice vector quantizer, no advantage can be taken from the fact that the speaker and the complete signal are known at the time of encoding.

3.4 Lattice Quantization with Entropy Coded Indices

Lattice quantization as described in section 3.3 is not adapted to the distribution of the source and assigns a fixed number of bits to each possible index of the lattice. However, since some codevectors will be used more often than others (usually most codevectors of the regular lattice grid will not be used at all by one single speaker), entropy coding can reduce the average bit rate per coded LSP set at the expense of a small entropy coding table.

Using the knowledge of the complete speech signal for Huffman compression [8] of the indices of the lattice vector quantizer, the average bit rate for voiced LSP vectors is about 40 bit per vector, for unvoiced speech it is about 30 bit per vector. The coding table for the corresponding Huffman code must be taken into account with approximately 3500 bit.

4. Results

The LSP vectors that have to be stored after the application of the signal adaptive interpolation can be quantized in different ways, depending on the length of the speech signal.

Figure 4 shows the mean spectral distortion of a trained vector quantizer (left), a lattice quantizer (right), entropy coded lattice quantizers (middle) and the AMR split vector quantizer scheme (circles) for the complete speech signal (all), as well as voiced parts and unvoiced parts separately, for a speech signal of 80 minutes length, spoken by a male German.

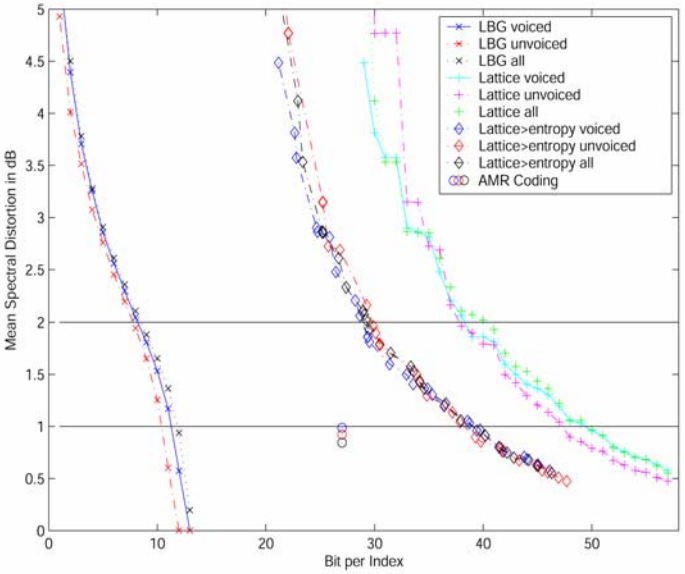

Fig. 4. Mean Spectral Distortion Versus Bit Rate for AMR-SVQ, VQ and Lattice VQ.

Figure 5 plots the expected gross bit rates for the quantizers that were studied over time with fixed interpolation over three subframes. Using our adaptive interpolation scheme the time axis would be stretched by a factor of 1.5.

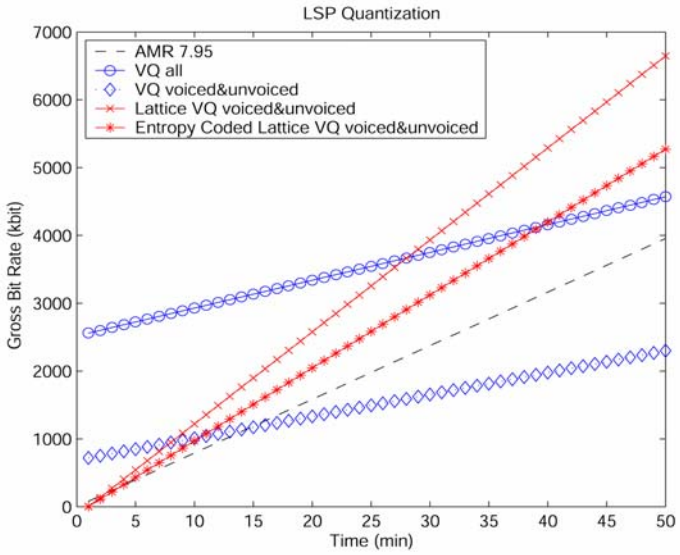

Fig. 5. Gross Bit Rate Versus Time.

It can be seen that for speech signals of more than 15 minutes the voicing specific LBG quantizer offers the best performance, while for signals of less than 3.5 minutes an entropy coded Lattice VQ scheme should be used.

5. Summary

It was shown, that with a signal adaptive interpolation scheme the number of LPC coefficients to be stored can be reduced by about a third, if the knowledge of the complete speech signal is exploited.

For quantization of the LSP vectors that need to be stored, different approaches have been compared with respect to the required amount of bits for the quantization tables and for each codebook index. Means of utilizing the fact that one single speaker will generally not drastically change his voice have been considered for this.

Finally, a guideline which quantization scheme to prefer for which speech signal has been given.

Acknowledgements

The authors would like to thank the head of the Institute of Communication Systems and Data Processing, Prof. P. Vary. The underlying work of this paper is the result of a research project which has been carried out at his institute.

References

[1] EUROPEAN TELECOMMUNICATION STANDARD Digital Cellular Telecommunications System; Adaptive Multi Rate (AMR) Speech Transcoding, GSM 06.60.AMR, 1998.

[2] LINDE, Y., BUZO, A., GRAY, A.M. An algorithm for vector quantizer design. *Itoc Journal*, 1980, vol. 28.

[3] ITAKURA, F. Line spectrum representation of linear predictive coefficients of speech signals. *J. Acoust. Soc. Am.*, 1975.

[4] PAULUS, J. Codierung breitbandiger Sprachsignale mit niedriger Datenrate, *Aachener Beiträge zu Digitalen Nachrichtensystemen*, 1997.

[5] PALIWAL, K.K., ATAL, B.S. Efficient vector quantization of LPC parameters at 24 Bits/Frame. *IEEE Transactions on Speech and Audio Processing*, 1993, vol.1, no. 2, p. 3 - 14.

[6] HAGEN, R., PAKSOY, E., GERSHO, A. Voicing-specific LPC quantization for variable-rate speech coding, *IEEE Transactions on Speech and Audio Processing*, 1999, vol. 7, no. 5, p485 – 493.

[7] CONWAY,J.H., SLOANE, N.J.A. Sphere packings, lattices and groups, *Springer Verlag*, New York, 1988

[8] Huffman, D.A. A Method for the construction of minimum redundancy codes, *Proceedings of IRE*, 1952, vol. 20, no. 9, p.1098 -1101.