A NEW GENERAL DISTANCE MEASURE FOR QUANTIZATION OF LSF AND THEIR TRANSFORMED COEFFICIENTS

Hai Le Vu and László Lois

Department of Telecommunications
Technical University of Budapest
Sztoczek 2, 1111 Budapest, Hungary
Email: hai@hit.bme.hu

ABSTRACT

In this paper, we have developed a new general distance measure that not only can be used in a vector quantization (VQ) of the line spectrum frequency (LSF) parameters but performes well in the LSF transformed domain. The new distance is based on the spectral sensitivity of LSF and their transformed coefficients. In addition, the fix scaling factor is used to decrease the sensitivity of spectral error at higher frequencies.

Experimental results have shown that the proposed distance measure leads to as good as or better perfomance of VQ compared to other methods in the field of LSF coding. The use of this distance as the weighting function of the LSFs' transformed parameters is also suggested.

1. INTRODUCTION

Vector quantization of the LSF parameters is an efficient method for minimizing the number of bits needed to represent the speech spectrum in low-rate speech coding system. The aim of speech coding is to obtain a synthetized speech signal perceptually as closed as possible to the original signal. As the human ear is very sensitive to distortions of the spectrum, spectral distortion has been found to allow the best subjective evaluation of LSF encoding quality.

During the designing of a vector quantizer, one needs to find the codeword that minimizes spectral distortion in its region. Because of its high computational complexity, a weighted Euclidean distance is used in most recent research for VQ in a real-time speech coder [1]. However, for small distances Gardner shown that the spectral quantization distortion approaches a simple quadratically weighted Euclidean error, where the weighting matrix is a sensitivity matrix that is an extension of the concept of the scalar sensitivity [2].

In this paper, based on the Gardner form we define the spectral sensitivity with respect to LSF and to their transformed coefficients. We then introduce the new general squared weighting function by modify the sensitivity matrix in order to make use of a property of human ear. The aim of this paper is to study different distance measures and develop the new one that can be used in the LSF and their transformed domain as well.

2. WEIGHTED LSF DISTANCE MEASURES

Selection of a proper distortion measure is the most important issue in the design and operation of a VQ. This measure should combine two features: it must be able to predict subjective tests accurately and, secondly, it should be computationally efficient. One of the best available suggestions for a measure of spectral distortion is the log spectral distortion (LSD) in dB, defined by

$$LSD^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[10 \log_{10} P(\omega) - 10 \log_{10} \widehat{P}(\omega)\right]^2 d\omega \tag{1}$$

where ω is the radian frequency, and $P(\omega)$ and $\widehat{P}(\omega)$ are the original power spectra of the nth speech frame and its quantized version. Experimental results from several studies indicate that whenever the LSD value drops below 1 dB, the quantizer has introduced distortion that is close to negligible. As of now, in practice, the spectral distortion is often replaced by a weighted squared Euclidean distance (WED) both in the designing and coding phases of VQ. A WED, d, has the form

$$d^2\left(\mathbf{f}, \hat{\mathbf{f}}\right) = \left(\mathbf{f} - \hat{\mathbf{f}}\right)^T \mathbf{W} \left(\mathbf{f} - \hat{\mathbf{f}}\right) \tag{2}$$

where $\mathbf{f}$ and $\hat{\mathbf{f}}$ are column vectors of the original and quantized LSF vectors, and $\mathbf{W}$ is a diagonal weighting matrix which may depend on $\mathbf{f}$. The task to design the code-books and the quantizer algorithm of VQ, is to find a weighting function that gives the best approximation of the spectral distortion. In relevant research, we found many suggestions for the weights, all determined in an empirical way. Here follow we give some examples of these weights for comparisons purposes

- The LPC spectral Weights (LPCW) [1]

$$w_i = \left\{ c_i \left[P(f_i) \right]^{0.15} \right\}^2, \ 1 \le i \le 10 \tag{3}$$

where $P(f_i)$ is the unquantized linear prediction power spectrum at the frequency of the i^{th} LSF component. The c_i gives more weight to the lower LSFs than to the higher LSFs and are fixed

$$c_i = \begin{cases} 1, & 1 \le i \le 8 \\ 0.8, & i = 9 \\ 0.4, & i = 10 \end{cases} \tag{4}$$

- The Invers Harmonic Mean Weights (IHMW) [3]

$$w_i = \frac{1}{\omega_i - \omega_{i-1}} + \frac{1}{\omega_{i+1} - \omega_i} \tag{5}$$

where ω_i is the i^{th} LSF parameter (in radian), and $\omega_0 = 0$, $\omega_{11} = \pi$.

- The Local Spectral Approximation Weights (LSAW) [4]

$$\begin{aligned} w_i &= c_i \left\{ \left[1 - \cos \pi (f_i - f_{i-1}) \right]^{-0.8} \right. \\ &\quad \left. + \left[1 - \cos \pi (f_{i+1} - f_i) \right]^{-0.8} \right\} \\ i &= 1, 2, \ldots, 10 \end{aligned} \tag{6}$$

where f_i is the i^{th} LSF parameter (in Hz), and $f_0 = 0$, $f_{11} = F_s/2$. F_s is the sampling frequency in Hz. The value of c_i's are the same as those used in the LPCW weights.

3. SPECTRAL SENSITIVITY OF LSFS AND THEIR TRANSFORMED PARAMETERS

While use of a WED (rather than spectral distortion) for VQ design and coding, the spectral distortion is not minimized in general. However, Gardner has found that the spectral distortion, LSD, given by Equ. 1, is equal to the WED (Equ. 2) for small distances, i.e.

$$LSD^2 = \left(\mathbf{f} - \hat{\mathbf{f}} \right)^T \mathbf{D} \left(\mathbf{f} - \hat{\mathbf{f}} \right) \tag{7}$$

when the cubic and higher terms in a Taylor series expansion can be neglected [2]. The diagonal weighting matrix $\mathbf{D}$ in Equ. 7 is called sensitivity matrix that depend on LSF parameters, but their calculation is simple enough for real-time coder.

3.1. Spectral sensitivity of LSF parameters

Let S denote the log power spectrum of a given LPC filter, then the spectral sensitivity with respect to an LSF frequency, ω_i, is defined as [5]

$$SEN_i(\boldsymbol{\omega}) = \int_{-0.5}^{0.5} \left| \frac{\partial S}{\partial \omega_i} \right|^2 d\omega = \frac{1}{2\pi \Delta\omega_i^2} \int_{-\pi}^{\pi} |\Delta S|^2 \, d\omega \tag{8}$$

To evaluate the spectral sensitivity, we modify only one of ten components of the vector LSF in order to get the distorted vector while the other nine parameters of the two vectors were unchanged. Using the expression in Equ. 7, the LSF spectral sensitivity in Equ. 8 can be computed in a closed-form as follow

$$\begin{aligned} SEN_i(\boldsymbol{\omega}) & \\ &= \frac{1}{2\pi\Delta\omega_i^2} \int_{-\pi}^{\pi} \left[10 \log_{10} \left(P(\omega)/\widehat{P}(\omega) \right) \right]^2 d\omega \\ &= \left. \frac{LSD^2(\boldsymbol{\omega}, \hat{\boldsymbol{\omega}})}{\Delta\omega_i^2} \right|_{\hat{\omega}_i = \omega_i + \Delta\omega_i} = D_i \end{aligned} \tag{9}$$

where D_i is i^{th} element of the diagonal sensitivity matrix, $\mathbf{D}$, in Equ. 7.

3.2. Spectral sensitivity of LSFs' transformed coefficients

There are several one-to-one vector functions that transform the vector of LSF into another vector, e.g., difference LSF frequencies [5], Karhunen-Loeve (KL) coefficients [6], etc. It can be shown that the spectral distortion with respect to the transformed coefficients is given by

$$LSD^2 = (\mathbf{u} - \hat{\mathbf{u}})^T \, \mathbf{T}^{-1} \mathbf{D} \left(\mathbf{T}^{-1} \right)^T (\mathbf{u} - \hat{\mathbf{u}}) \tag{10}$$

where $\mathbf{T}$ is the linear transform matrix, $\mathbf{u}$ is the transformed vector of LSF frequencies. In the above expression the weighting matrix is $\mathbf{T}^{-1}\mathbf{D}\left(\mathbf{T}^{-1}\right)^T$ and generally is not diagonal. In this paper we only consider the KL transformed coefficients that was proposed in our previous study of optimal transformation of LSF parameters. In such a case the KL transform matrix is defined as $\mathbf{T} = [\mathbf{b}_1 \mathbf{b}_2 \ldots \mathbf{b}_{10}]^T$, where $\mathbf{b}_i$ is the eigenvector of the autocorrelation matrix of the LSF parameters. Similar to LSF parameter, we can get the spectral sensitivity of KL coefficients as follow [7]

$$SEN_i(\boldsymbol{\kappa}) = \left. \frac{LSD\left(\boldsymbol{\kappa}, \hat{\boldsymbol{\kappa}}\right)}{\Delta\kappa_i^2} \right|_{\hat{\kappa}_i = \kappa_i + \Delta\kappa_i} = \sum_{j=1}^{10} D_j \left(\mathbf{b}_i(j) \right)^2 \tag{11}$$

where D_j is the spectral sensitivity of the j^{th} LSF frequency and $\mathbf{b}_i(j)$ is the j^{th} component of vector $\mathbf{b}_i$. So far, we have derived in closed-form the spectral sensitivity of LSF and their KL transformed coefficients as the function of element of the LSF diagonal sensitivity matrix ($\mathbf{D}$). The proposed parameter spectral sensitivity can be efficiently used as a weighting function in the VQ distance measure, thought this formulation is only valid when the quantizer's average distortion is sufficiently small. However, quantizers for speech coders usually meet this requirement.

4. GENERAL SQUARED WEIGHTING FUNCTION

In what follows, we introduce the new general weighting function (based on the above derived parameter spectral sensitivity in Section 3). This weighting can be used both in coding the LSF frequencies and their transformed parameters as well. Since the human ear can not resolve differences at high frequencies as accurately as at low frequencies, we also use the scaling factor c_i defined by Paliwal (Equ. 4) to make use of this property of human ear. For LSF vector we define the general weighting (GW) function as

$$w_i = c_i \cdot SEN_i(LSF) \tag{12}$$

and for KL transformed coefficients the GW function is defined as (from Equ. 11)

$$w_i = \sum_{j=1}^{10} c_j \cdot SEN_j(LSF) \cdot (\mathbf{b}_i(j))^2 \tag{13}$$

where SEN_i is the spectral sensitivity of LSF parameters given in Equ. 9.

We show here in Fig. 1 the normalized value of different weighting functions which are mentioned in Section 2 and the new general weights proposed in this section for a typical speech vowel. It can be seen here that the new distance measure (GW) allows for quantization of LSFs in the formant regions better than those in the non-formant regions. Also, the distance measure gives more weight to the formant associated with the more closed LSF pair than to those associated with the less closed LSFs; the LSFs corresponding to the valleyes in the LPC power spectrum get the least weight.

5. EXPERIMENTS AND RESULTS

In this section, the LSF quantization performance of the Multi Stage VQ (MSVQ) and Split VQ (SVQ) (in terms of average spectral distortion and number of outliers) is studied and results are reported at using differrent weighting function.

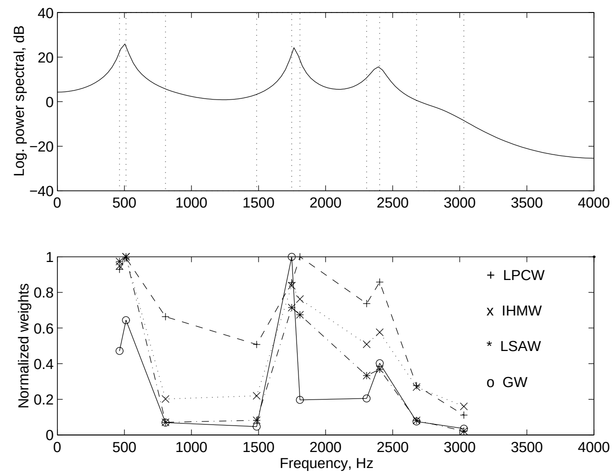

Figure 1:
(a) LPC power spectral and associated LSFs for a typical speech vowel
(b) LPCW, IHMW, LSAW and the new General Weights (GW) for LSF parameters

The speech data base was used in this study consist of 24.000 frames of male and female speech data taken from 18 speakers (11 male and 7 female). Six short sentences were recorded for each speaker. The sampling rate was 8 kHz, each frame was 20 ms long and 10 order LPC analysis was employed. About 21.000 frames of speech is used for training, and the last 3.000 frames of speech (different from those used for traning) is used for testing.

In order to see the effect of weighting, we study the performance of the MSVQ and SVQ using the unweighted Euclidean distance measure (UNW) and the WED with different weighting functions as follows: LPCW (Equ. 3); IHMW (Equ. 5); LSAW (Equ. 6); GW for LSFs (Equ. 12); GW for KLs (Equ. 13).

The VQ is designed by using LBG algorithm on the training data, and its performance is evaluated both from the training and the test data.

In the case of MSVQ, the codewords of the second and further stages represent the quantization error of the previous stages that do not have individual weights. Therefore the effect of different distance measures can be investigated on the first stage only. The average spectral distortions (SD) of the first stage MSVQ using the codebook of 1024 codewords as a function of different distance measure are shown in Table 1. It can be seen from these results that the new weighting function improves the VQ performance for both LSF parameters and their KL transformed coefficients.

In the case of SVQ, we study the performance of the 22 bits/frame split VQ using the above described distance measures. The LSF vector is splited here into three parts: the first part has the first four LSFs, the second part has the next three LSFs and the last part has the remaining LSFs. The bitallocation vector is [8,7,7] for the LSF vectors and [12,6,4] for KL coefficients.

Tables 2 shows the average and outlier spectral distortion performance of the SVQ quantizers using different weighting function. It can be seen that the new weighting function introduced better performance than the other distance measures with all the respect to AVSD and the percentages of outliers. Exceptionally in the case of test data, we can find that SVQ for KL using the proposed GW has the same average distortion as the unweighted SVQ, whereas the percentage of outlier of above 4 dB is significantly reduced.

6. CONCLUSIONS

In this paper, we have derived in closed-form the spectral sensitivity with respect to LSF frequencies and to their linear transformed coefficients. Using this spectral sensitivity, we have proposed the new general weighting function for LSF parameters and their KL transformed coefficients.

The performance of the fisrt stage MSVQ using 1024 vectors codebook and the 22 bits/frame SVQ using the unweighted- and weighted Euclidean distance measure with different weighting functions are studied.

Our experimental results have shown that with respect to other conventional weights, this distance measure introduce better perfomance for both MSVQ and SVQ scheme.

7. REFERENCES

[1] K.K. Paliwal and B.S. Antal, "Efficient Vector Quantization of LPC Parameters at 24 bits/frame," in *Proc. ICASSP*, 1991, pp. 661-664

[2] W.R. Gardner and B.D. Rao, "Theoretical Analysis of the High-Rate Vector Quantization of LPC Parameters," *IEEE Trans. Speech Audio Processing*, Vol.3, No.5, Sep. 1995, pp. 367-381

[3] R. Laroia, N. Phamdo and N. Farvardin, "Robust and Efficient Quantization of Speech LSP Parameters Using Structured Vector Quantizers," in *Proc. ICASSP*, 1991, pp. 641-644

[4] D. Chang, S. Ann and C.W. Lee, "A Classified Vector Quantization of LSF Parameters," *Signal Processing*, Vol.59, No.3, Jun. 1997, pp. 267-273

[5] F.K. Soong and B.H. Juang, "Optimal Quantization of LSP Parameters," in *Proc. ICASSP*, 1988, pp. 394-397

[6] H.L. Vu and L. Lois, "Optimal Transformation of LSP Parameters Using Neural Network," in *Proc. ICASSP*, 1997, pp. 1339-1342

[7] H.L. Vu and L. Lois, "Spectral Sensitivity of LSP Parameters and Their Transformed Coefficients," in *Proc. European Conf. Speech Comm. Technology, EUROSPEECH*, 1997, pp. 1251-1254

	Weights	Average SD (dB)	
		Train	Test
LSF	UNW	2.22	2.55
	LPCW	2.19	2.51
	IHMW	2.13	2.46
	LSAW	2.18	2.51
	GW	2.11	2.46
KL	UNW	2.19	2.53
	GW	2.11	2.46

Table 1: Spectral distortion of the fisrt stage MSVQ (10 bits/frame) for different weighting functions

LSF Parameters						
Weights	Train			Test		
	SD (dB)	Outliers (%) 2-4dB	Outliers (%) > 4dB	SD (dB)	Outliers (%) 2-4dB	Outliers (%) > 4dB
UNW	1.46	12.91	0.03	1.51	14.66	0.11
LPCW	1.41	10.19	0.02	1.48	13.51	0.11
IHMW	1.36	7.14	0.00	1.42	9.55	0.08
LSAW	1.39	9.11	0.02	1.45	11.91	0.05
GW	1.34	6.00	0.00	1.40	8.39	0.02

KL Parameters						
Weights	Train			Test		
	SD (dB)	Outliers (%) 2-4dB	Outliers (%) > 4dB	SD (dB)	Outliers (%) 2-4dB	Outliers (%) > 4dB
UNW	1.36	8.81	0.00	1.50	17.35	0.10
GW	1.35	8.67	0.00	1.50	17.47	0.06

Table 2: Performance of 22 bits/frame SVQ of LSF and KL parameters for different weighting functions