ON THE STATISTICAL PROPERTIES OF LINE SPECTRUM PAIRS

J.S. Erkelens and P.M.T. Broersen

Delft University of Technology, Department of Applied Physics, P.O. Box 5046, 2600 GA Delft, The Netherlands
email: erkelens@duttncb.tn.tudelft.nl / broersen@duttncb.tn.tudelft.nl

ABSTRACT

In literature, the quantization properties of several representations of the LPC model have been studied. Until recently, best results have generally been obtained with the LSP frequencies. In scalar quantization schemes, the Immitance Spectrum Pairs (ISP's) [1] perform even slightly better. The good quantization performance of LSP and ISP can be attributed to their theoretical statistical properties: they are uncorrelated when estimated from stationary autoregressive processes, in contrast to the other representations. For small variations in the coefficients of any representation, the Spectral Distortion can be expressed as a weighted squared distortion measure. The optimal weighting matrix is the inverse of the covariance matrix of the coefficients. For LSP and ISP this matrix is a diagonal matrix and hence the best weighting factors are the inverses of the theoretical variances. The difference between LSP and ISP is due to their distributions in speech.

I. INTRODUCTION

Accurate quantization of the LPC model is of prime importance for the quality of low bitrate speech coders. The objective of spectral quantization is to achieve transparent quality, i.e. speech coded with quantization of the LPC model and speech coded without spectral quantization are indistinguishable in terms of subjective quality. For a comparison of quantization methods, an objective measure of quality is needed. This measure has to be relevant for speech. Paliwal and Atal [2] showed that transparent quantization is achieved if the average of the well-known Spectral Distortion measure is about 1 dB with not too many outliers. For quantization of the spectral model different representations can be used, such as the Reflection Coefficients (RC's), Log Area Ratios (LAR's), Arcsine of Reflection Coefficients (ASRC's) and Line Spectrum Pair frequencies (LSP frequencies) [3]-[5]. Stability is easily guaranteed with these representations. Scalar quantization (SQ) schemes need approximately 35-40 bits per frame for transparent quantization and vector quantization (VQ) schemes about 25-30 bits per frame. The large VQ codebook size associated with these numbers of bits urges the necessity of fast search measures and complexity reducing codebook structures, such as split VQ (SVQ) and multi-stage VQ.

An ideal search measure has three characteristics. Firstly, it can be computed very fast. Secondly, it ensures that the vector of minimum distance is chosen, even from constrained codebooks. Thirdly, it is a good approximation of Spectral Distortion.

A weighted Euclidean distance measure (WEDM) is a search measure that fulfils the first and second demand for SVQ. Whether it fulfils the third demand, depends on which representation and weighting factors are used.

In section II an approximation is derived for the multi-parameter spectral sensitivity of the Spectral Distortion in terms of the theoretical covariance matrix of the coefficients of any representation. This expression shows that the optimal weighting matrix in a weighted squared distance measure is the inverse of the theoretical covariance matrix of the coefficients of any representation. Hence, a WEDM is an approximation of Spectral Distortion if the theoretical covariance matrix is a diagonal matrix, i.e. the coefficients are uncorrelated.

Best quantization results are generally obtained with the LSP frequencies, both for SQ (Soong and Juang [6]) and SVQ (Paliwal and Atal [2]). An interesting new representation for quantization is formed by the Immitance Spectrum Pairs (ISP's) [1]. In SQ schemes, ISP performs even slightly better than LSP.

In section III we will show that the good quantization performance of LSP and ISP can be attributed to the theoretical statistical properties of these representations; they both have uncorrelated coefficients.

In section IV the performance of the theoretically derived weighting factors is compared to different weighting factors used in literature.

The improvement of ISP over LSP can be attributed to their distributions in speech, which is the subject of section V. Conclusions follow in section VI.

II. MULTI-PARAMETER SPECTRAL SENSITIVITIES

An approximation will be derived of the Spectral Distortion due to small disturbances in all coefficients of an arbitrary representation.

The Spectral Distortion (SD) measure between two autoregressive models is defined as:

$$SD^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left| \log \left| \frac{\hat{A}}{A} \right|^2 \right|^2 d\omega \quad . \tag{1}$$

A and $\hat{A}$ are polynomials in z, with LPC parameters as coefficients. A first order approximation to the multi-parameter spectral sensitivity of SD with respect to all parameters is:

$$SD^2 \approx \frac{2}{2\pi}\int_{-\pi}^{\pi} \left| \frac{\Delta A}{A} \right|^2 d\omega \quad . \tag{2}$$

ΔA is the difference between the original polynomial and the disturbed polynomial. Eq.(2) can be written in the time domain as:

$$SD^2 \approx 2\ \Delta\alpha^T R_p \Delta\alpha \quad , \tag{3}$$

where $\Delta\alpha$ is the vector of differences between LPC parameters of the disturbed and original model and R_p is the (p×p) covariance matrix of the autoregressive process described by the polynomial A, normalized with respect to the innovation variance σ^2.

The theoretical asymptotical covariance matrix C_α of estimated LPC parameters follows from Maximum Likelihood theory and is given by

$$C_\alpha = \frac{1}{N} R_p^{-1} \quad , \tag{4}$$

where N is the number of observations used to estimate the parameters. The theoretical covariance matrix C_λ of any other representation of the LPC model can be found from the covariance matrix of the LPC parameters and from the matrix of partial derivatives A of the representation with respect to the LPC parameters:

$$C_\lambda = A\ C_\alpha\ A^T \tag{5}$$

The (i,j)-th element of A is the partial derivative of the i-th coefficient λ_i with respect to the j-th LPC parameter.

For any representation λ, the inverse of the "λ to LPC derivative matrix" A is the "LPC to λ derivative matrix" B, because the LPC parameters and the λ-coefficients are one-to-one related. Hence, the inverse of the covariance matrix of the other representation can be written as:

$$C_\lambda^{-1} = N\ B^T\ R_p\ B \tag{6}$$

For small disturbances $\Delta\lambda_i$ in the coefficients λ_i, $\Delta\alpha$ in eq.(3) is approximately equal to B $\Delta\lambda$, where B is the matrix of partial derivatives of the LPC parameters with respect to the λ-coefficients. Together with eq.(6) it follows from eq.(3) that

$$SD^2 \approx \frac{2}{N} \Delta\lambda^T\ C_\lambda^{-1} \Delta\lambda \tag{7}$$

This formula shows that the Spectral Distortion due to small disturbances in the coefficients of any representation can be approximated by a weighted squared distance measure in the coefficients of that particular representation, the weighting matrix being a constant times the inverse of the theoretical covariance matrix of the coefficients. It can be interpreted as a relation between the deterministic properties of LPC representations and their statistical properties. For example, coefficients with a high spectral sensitivity have a small variance. Further, strong coefficient couplings cause large correlations between the estimated coefficients. The single parameter spectral sensitivities are the diagonal elements of C_λ^{-1}.

In SVQ, the vector is split into two or more parts and these parts are quantized independently using VQ [2]. A WEDM is a separable measure, hence a minimum for each part of the vectors gives a minimum for the total vector. Eq.(7) reduces to a WEDM with the inverses of the variances as weights, if the coefficients of the representation are uncorrelated when estimated from stationary autoregressive processes. In the next section we will show that the LSP frequencies and the ISP's satisfy this property.

III. STATISTICAL PROPERTIES OF LSP AND ISP

The statistical properties of representations of the LPC model are important for quantization. The theoretical statistics of most representations are known. For example, Kay and Makhoul [7] have investigated them for RC. The statistical properties of LAR and ASRC can be easily derived from those of reflection coefficients. So far, not much is known about the statistical properties of LSP frequencies. In this section we will show that LSP frequencies and ISP coefficients are uncorrelated when estimated from stationary autoregressive processes.

Autoregressive estimation leads to a p-th order model with

$$A_p(z) = \sum_{i=0}^{p} \alpha_i z^{-i}, \quad \alpha_0 = 1 \tag{8}$$

The polynomial for the order p+1 follows from the polynomial of order p and the (p+1)-th reflection coefficient κ_{p+1} with the Levinson recursion:

$$A_{p+1}(z) = A_p(z) + \kappa_{p+1} z^{-(p+1)} A_p(z^{-1}) \tag{9}$$

The odd and even LSP polynomials P(z) and Q(z) are formed by setting the (p+1)-th reflection coefficient to -1 or +1 respectively in eq. (9) and can be written as:

$$\begin{aligned} P(z) &= (1 - z^{-1}) \prod_{i=1}^{p/2} (1 + 2\gamma_{2i-1}\ z^{-1} + z^{-2}) \quad , \\ Q(z) &= (1 + z^{-1}) \prod_{i=1}^{p/2} (1 + 2\gamma_{2i}\ z^{-1} + z^{-2}) \end{aligned} \tag{10}$$

These expressions are given here for even order LPC models. We only consider this case here for ease of notation, without loss of generality. The LSP coefficients γ_i are related to the LSP frequencies ω_i as $\gamma_i = -\cos(\omega_i)$.

By expanding the polynomials P(z) and Q(z) and comparing the coefficients of the powers of z with those of eq.(9) for κ_{p+1} equal to -1 and +1 respectively, a set of equations is obtained connecting the LSP frequencies ω_i to the LPC parameters α_i. From these relations it can be readily seen that the LSP frequencies are functions of parameters β_i, which themselves are differences and sums of the LPC parameters:

$$\begin{aligned} \beta_i &= \alpha_i - \alpha_{p+1-i} \\ \beta_{i+p/2} &= \alpha_i + \alpha_{p+1-i}, \quad i = 1, \ldots, p/2 \end{aligned} \tag{11}$$

The odd LSP frequencies are functions of the differences of LPC parameters and the even LSP frequencies of the sums of LPC parameters only. Therefore, the following symmetry relations hold for the elements of the LPC to LSP derivative matrix:

$$\begin{aligned} \frac{\partial\alpha_i}{\partial\omega_j} &= -\frac{\partial\alpha_{p+1-i}}{\partial\omega_j}, \quad j \text{ odd} \\ \frac{\partial\alpha_i}{\partial\omega_j} &= +\frac{\partial\alpha_{p+1-i}}{\partial\omega_j}, \quad j \text{ even} \end{aligned} \tag{12}$$

The covariance matrix R_p of an autoregressive process has a persymmetric Toeplitz structure, so R(i,j)=R(j,i)=R(p+1-i,j)=R(p+1-j,i). Using this Toeplitz structure and the symmetry relations of eq.(12) it follows from eq.(6) that the even and odd LSP frequencies are mutually uncorrelated.

Analytical computation of the inverse of the covariance matrix of the LSP frequencies for process orders up to p=6 with eq.(6) showed that *all* LSP frequencies are uncorrelated for these orders, i.e. C_ω is a diagonal matrix. For higher orders, analytical computation of the covariance matrix is rather straightforward but very tedious. However, numerical computation of the covariance matrix of LSP frequencies showed that LSP frequencies are also uncorrelated for higher order processes. For a given process, the elements of the LPC to LSP derivative matrix **B** can be found numerically. Next, the inverse of the covariance matrix of the LSP frequencies can be computed with eq.(6). This procedure has been followed for many different process orders and process parameters. For every process considered, the LSP frequencies turned out to be uncorrelated.

The Immitance Spectrum Pairs (ISP's) are a new representation for quantization (Peller and Bistritz [1]). They consist of p-1 frequency parameters and (a transformation of) the last reflection coefficient of the model. They were shown to perform slightly better than LSP frequencies in SQ. Chan [8] has shown that the set of ISP's are actually the LSP's of the model of one order lower plus the last reflection coefficient. However, the theoretical variances and covariances of the ISP frequency parameters are not equal to those of the LSP frequencies of a process of one order lower, because the covariance matrix of the parameters of that lower order model depends on the last reflection coefficient. In the case of LSP frequencies, no higher order reflection coefficients exist.

We investigated the statistical properties of ISP's and found that they are uncorrelated, just like the LSP frequencies. We used symmetry relations to show theoretically that odd and even ISP frequency parameters are mutually uncorrelated. Numerical differentiation showed that all ISP frequency parameters are uncorrelated.

Because the LSP's and ISP's are uncorrelated, the Taylor approximation of SD in eq.(7) reduces for these representations to a WEDM. The weighting factors in this measure are the inverses of the theoretical variances v_i of the coefficients:

$$SD^2 \approx \frac{2}{N}\sum_{i=1}^{p} \frac{\Delta\omega_i^2}{v_i} \qquad (13)$$

The variances can be found from eq.(6), either analytically or numerically. The variances are smallest for LSP frequencies corresponding to a formant. The smaller the bandwidth, the smaller the variances. Use of a WEDM with the inverses of the variances as weighting factors results in accurate quantization of formants.

The localized spectral sensitivity of the LSP frequencies has always been used as a heuristic justification for the use of a WEDM. However, such a measure can also be used for the ISP's, although the spectral sensitivity of the last ISP coefficient, which is a transformation of the last reflection coefficient of the model, is *certainly not* localized. The true justification for the use of a WEDM is that the theoretical covariance matrices of LSP's and ISP's are diagonal.

IV. COMPARISON OF WEIGHTING FACTORS

In this section we will compare the weighting factors following from eq.(7) with some others used in literature. The weighting matrix in eq.(7) is a diagonal matrix for the LSP's and ISP's. For VQ this is very useful because one can now use a fast search measure that is a linearized approximation of the Spectral Distortion. For the reflection coefficient based representations, the weighting matrix is not a diagonal matrix. The off-diagonal elements can be very important because the correlations between the coefficients can be large for these representations (Erkelens and Broersen [9]). Therefore, a WEDM on the RC based representations doesn't approximate the Spectral Distortion as accurately as eq.(7). To illustrate this, an experiment was performed to compare seven different distortion measures. From a segment of speech data, an LPC model has been estimated. The LPC power spectrum and the LSP frequencies of this model are shown in Figure 1. With this model as reference, a simulation study has been made, by adding to the Log Area Ratios of the original model a vector of Gaussian numbers, scaled in such a way that the Spectral Distortion was 1dB with respect to the original model. We computed the seven different distortion measures 25000 times between the distorted model and the original model. A perfect measure would always give exactly the same result. Not all measures had the same average value. For a comparison, they were properly scaled. Now, the distance measure with lowest variance for all vectors generated from the model, is the best approximation of the Spectral Distortion. The measures we compared are:

- (a) / (b) / (c) the approximation of the Spectral Distortion of eq.(7) for LSP, ISP and LAR,
- (d) a WEDM for LAR, using as weighting factors the single parameter spectral sensitivities, i.e. only the diagonal elements of C_{LAR}^{-1},
- (e) / (f) / (g) WEDM for LSP using the heuristic weights in [2], [10] and [11].

The standard deviations of the measures were:
(a) 0.02 (b) 0.03 (c) 0.04 (d) 0.26 (e) 0.15 (f) 0.08 (g) 0.07

respectively. The WEDM for LAR has worst performance, because the off-diagonal elements of C_{LAR}^{-1} are neglected.

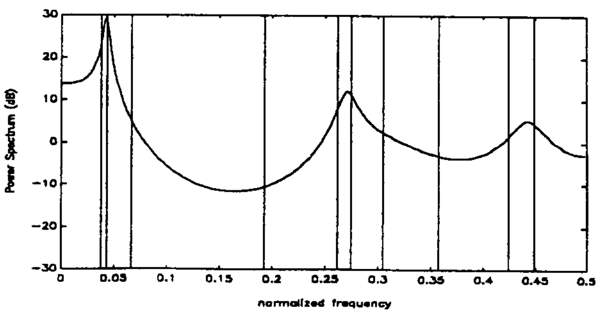

Figure 1. Power Spectrum and LSP frequencies of an LPC model for a voiced sound.

Models from other speech data give different values for the standard deviations. However, several different examples were considered and in all examples the measures based on eq.(7) were the best approximation of the Spectral Distortion.
The theoretical properties of the RC based representations are not suitable for SVQ. ISP and LSP have optimal theoretical properties for SVQ. The difference in performance between ISP and LSP is due to their distributions in practice.

V. EXPERIMENTAL DISTRIBUTIONS OF LSP AND ISP

The superiority of weighting based on eq.(7) over other weighting factors has been verified experimentally in the last section.
Single stage VQ is able to take fully into account all dependencies between coefficients of a representation. SQ and SVQ are suboptimal due to the constrained location of the vectors and can have reduced performance if the search measure does not ensure that the vector of smallest distance is selected. For LSP and ISP a WEDM can be used, so here the decrease in quality of SQ and SVQ as compared with that of single stage VQ is due to the constraints posed on the locations of the codebook vectors. Therefore it must be that the experimental distributions of ISP coefficients in speech are more suitable for SVQ than the distributions of the LSP frequencies. This was first noted by Chan [8]. He showed that the set of ISP's are actually the LSP frequencies of a model of one order lower plus a transformation of the last reflection coefficient. This explains why ISP's perform better in SQ than LSP frequencies. Every LSP frequency needs approximately 4 bits. The last reflection coefficient of a model is generally small and needs only about 2 bits. This accounts for the gain in performance of ISP's over LSP frequencies. Chan used this to propose a mixed LSP/RC representation for quantization, which resulted in slightly increased performance over 'pure' LSP and RC, both in SQ and SVQ. The LSP frequencies computed from the first p-k, k>1 reflection coefficients of model are not uncorrelated. An experimental advantage of a mixed LSP/RC representation over LSP exists because the last reflection coefficients are generally small and need fewer bits than LSP frequencies, but a theoretical disadvantage exists because correlations are neglected.
The quality loss due to the constrained codebook structure is less for ISP than it is for LSP. Other representations may have experimental distributions more suitable for SQ and SVQ, but are not suitable for use with a WEDM.
We think that for SVQ only representations should be used, which have optimal theoretical properties for this constrained codebook structure, i.e. are uncorrelated. More representations satisfy this property. For example, the Karhunen-Loeve transform does, but stability may become a problem. An experimental evaluation is necessary to choose between uncorrelated representations, because the amount of quality loss in constrained VQ is influenced by the experimental distributions of representations.

VI. CONCLUSIONS

A weighted squared distortion measure is a Taylor approximation of the Spectral Distortion. The optimal weighting matrix in this measure is the inverse of the theoretical covariance matrix of the coefficients.
LSP frequencies and ISP's have optimal theoretical statistical properties for SQ and SVQ, because they are uncorrelated. For uncorrelated representations a weighted Euclidean distance measure with the inverses of the variances as weights is an approximation of Spectral Distortion.
The theoretical statistical properties are not the only important factors for SQ and SVQ. Also of great importance are the experimental distributions, i.e. the dependencies and variations of the coefficients in speech. Because of its experimental properties, ISP is a better representation for SQ and SVQ than LSP.
Representations with optimal theoretical properties do not necessarily give the same performance, because the amount of quality loss in SVQ depends on the distributions of a representation in speech. It has to be found out experimentally which of these representations is the best, because the distributions in speech are not fully known.

ACKNOWLEDGEMENT

This research is supported by the Technology Foundation under project no. DTN11.2436.

REFERENCES

[1] S. Peller and Y. Bistritz, "Performance of Immitance Spectral Pairs (ISP) in LPC Scalar Quantizer Schemes", Signal Processing VII, pp. 403-406, 1994.

[2] K.K. Paliwal and B.S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame", IEEE Trans. on SAP, vol. 1, no. 1, pp. 3-14, Jan. 1993.

[3] F. Itakura, "Line Spectral Representation of Linear Predictive Coefficients of Speech Signals", Journ. Acoust. Soc. Amer., vol. 57, Suppl. no. 1, S35, 1975.

[4] V.R. Viswanathan and J. Makhoul, "Quantization Properties of Transmission Parameters in Linear Predictive Systems", IEEE Trans. on ASSP, vol. ASSP-23, pp. 309-321, June 1975.

[5] A.H. Gray and J.D. Markel, "Quantization and Bit Allocation in Speech Processing", IEEE Trans. on ASSP, vol. ASSP-24, pp. 459-473, 1976.

[6] F.K. Soong and B.H. Juang, "Optimal Quantization of LSP Parameters", IEEE Trans. on SAP, vol. 1, no. 1, Jan. 1993

[7] S. Kay and J. Makhoul, "On the Statistics of the Estimated Reflection Coefficients of an Autoregressive Process", IEEE Trans. on ASSP, vol. ASSP-31, pp. 1447-1455, Dec. 1983.

[8] C.F. Chan, "Efficient Quantization of LPC Parameters Using a Mixed LSP/PARCOR Representation", Signal Processing VII, pp. 939-942, 1994.

[9] J.S. Erkelens and P.M.T. Broersen, "Interpolation of Autoregressive Processes at Discontinuities: Application to LPC based Speech Coding", Signal Processing VII, pp. 935-938, 1994.

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

[11] F.F. Tzeng, "Analysis-by-Synthesis Linear Predictive Speech Coding at 2.4 Kbit/s", Proc. Globecom, pp. 1253-1257, 1989.