

DIRECT LINEAR CONVERSION OF LSP PARAMETERS FOR PERCEPTUAL CONTROL IN SPEECH AND AUDIO CODING

R. Sugiura¹, Y. Kamamoto², N. Harada², H. Kameoka², T. Moriya²

¹Graduate School of Information Science and Technology, the University of Tokyo

²NTT Communication Science Labs., Nippon Telegraph and Telephone Corp.

ABSTRACT

We have devised a direct and simple scheme for linear conversion of line spectrum pairs (LSP) with low computational complexity aiming at weighting or inverse weighting spectral envelopes for noise control in speech and audio coders. Using optimally prepared coefficients, we can perform the conversion directly in the LSP domain, which ensures low computational costs and also simplifies the check or the modification of unstable parameters. We show that this method performs the same as the weighting in the linear prediction coding domain but with lower complexity in a low-bit-rate situation. The devised method is therefore expected to be useful for low-bit-rate speech and audio coders for mobile communications.

Index Terms— audio coding, signal processing, LSP, linear approximation, TCX

1. INTRODUCTION

The comfortability of voice communication tools, such as mobile phones greatly depends on the speech coder used in them. Although the technology in speech coders has been progressing for years, there is still a demand for higher quality not only in speech but also in the other audio signals such as music.

3GPP Extended Adaptive Multi-Rate WideBand (AMR-WB+) and MPEG-D Unified Speech and Audio Coding (USAC) [1, 2] are known as the state-of-the-art speech and audio coders. Both coders have at least two different modes and use the one that suits the input signals. Voice signals are coded in the time domain by Algebraic Code Exited Linear Prediction (ACELP), and the other audio signals are coded in the frequency domain by Transform Coded eXcitation (TCX). The goal of this work is to design a high-quality low-bit-rate speech and audio coder based on their ideas with lower delay aiming at its use in mobile phones. In the context of mobile communications, the coder is also required to have a low computational cost since there is a limit in the cost for real-time processing. Lowering the computational complexity of the process in the coder allows the coder to meet the limit of the cost when the coder is too complex or makes a room for additional processes when the cost is low enough, which leads to higher quality. This paper discusses saving computational costs by focusing on the following conventional process.

Both AMR-WB+ and USAC, mentioned above, quantize and code linear prediction coding (LPC) parameters, which

are used for constructing whitening filters or for representing spectral envelopes. Weighted parameters are also used in order to shape quantization noise perceptually, and this perceptual weighting is often performed in the LPC domain since it requires only a single parameter for tuning the weights. However, in the coders, these parameters are quantized and coded in the form of line spectrum pairs (LSP) [4], which is robust for quantization or interpolation and also makes it much easier to guarantee the stability of the parameters. Since the conversion between LPC and LSP parameters costs complexity, it is reasonable to perform the weighting directly in the LSP domain. Furthermore, by weighting in the LSP domain, the useful properties of LSP are available in the process. Here, we present a low-complexity LSP conversion scheme based on simple linear operations with trained coefficients that approximate either the perceptual weighting or the inverse weighting. This method of conversion enables several types of coders to reduce the computational costs without degrading quality.

In section 2, we explain the perceptual weighting used in the coder. In section 3, the properties of LSP are described. Then, in section 4, the method of approximating the weighting in the LSP domain is introduced. Finally, in section 5, the applications of this conversion are shown and the conversion is evaluated in those contexts.

2. PERCEPTUAL WEIGHTING IN THE CODER

First of all, we explain perceptual weighting using the TCX coder [3] as an example. The TCX coder quantizes and codes two kinds of information: the LPC parameters explained above and residual spectra, the input frequency spectra divided by their envelopes.

The weighted LPC parameters are used to calculate the envelopes divided into the spectra. Normally, envelopes can be computed from the frequency response of an all-pole filter with the LPC parameters:

$$H(z) = 1/(1 + \sum_{n=1}^p a_n z^{-n}) \quad (1)$$

where a_n is the LPC parameter of n th order. To control quantization noise in the spectra perceptually, the weighted parameters are applied to the filter as

$$\tilde{H}(z) = 1/(1 + \sum_{n=1}^p \gamma^n a_n z^{-n}), \quad (0 < \gamma < 1) \quad (2)$$

and this weighting smooths the envelopes. The residual spec-

tra, which are divided by the weighted envelopes, are scalar quantized, followed by entropy coding. Here, the weighted envelopes shape the quantization noise into approximately $\tilde{H}(z)/H(z)$, eventually causing smaller distortion in the spectral peaks, which is more important than the spectral valleys for human perception. It is experimentally known that the perceptual weighting works well when 0.92 is chosen for the value of the parameter γ .

This method of weighting is widely used since it needs only one parameter γ and p operations of multiplication.

3. PROPERTIES OF LSP

LSP parameters have roughly two intuitive properties, which makes them worth using in audio coders. One is about the stability of the parameters [5]. The all pole filter eq. (1) is stable if and only if it can be written with real value LSP parameters, $\omega_1, \dots, \omega_p$, when $p = 2m - 1$ as

$$H(z)^{-1} = (1 - z^{-2}) \prod_{i=1}^{m-1} (1 - 2 \cos(\omega_{2i})z^{-1} + z^{-2}) + \prod_{i=1}^m (1 - 2 \cos(\omega_{2i-1})z^{-1} + z^{-2}) \quad (3)$$

and when $p = 2m$ as

$$H(z)^{-1} = (1 - z^{-1}) \prod_{i=1}^{m-1} (1 - 2 \cos(\omega_{2i})z^{-1} + z^{-2}) + (1 + z^{-1}) \prod_{i=1}^m (1 - 2 \cos(\omega_{2i-1})z^{-1} + z^{-2}) \quad (4)$$

with $0 < \omega_1 < \omega_2 < \dots < \omega_p < \pi$ satisfied. This property makes it easy to not only detect the stability of the parameters but also to modify the parameters to stabilize them if they become unstable by quantization or other transformations.

The other property of LSP is the relation between the parameters and the envelopes represented by them. The arrangement of the parameters influences the steepness and smoothness of the envelopes. Fig. 1 shows an envelope and its LSP parameters. The envelope gets steeper when the parameters are closely neighboring each other, and vice versa. For this reason, an intuitive process can be taken, such as the simple interpolation in an ACELP coder, which interpolates linearly the LSP parameters in subframes from those in current and past frames.

4. PERCEPTUAL WEIGHTING IN LSP DOMAIN

4.1. Linear conversion model

To approximate the perceptual weighting in the LSP domain, several assumptions should be made by considering the following facts.

First, the weights decrease the steepness of the envelopes, thus the weighted LSP parameters get farther away from each other, ending up in even intervals where $\gamma = 0$, in which the envelopes become completely flat. Fig. 2 shows an example of the change in LSP parameters when the corresponding LPC parameters are weighted by γ . It can be seen that in

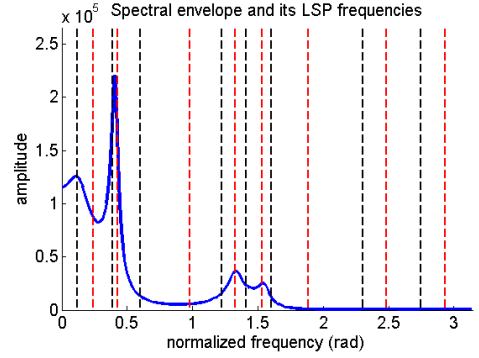


Fig. 1. LSP parameters and the envelope represented by them in normalized frequency axis. The blue solid line indicates the spectral envelope. The red and black vertical dashed lines indicate the LSP parameters for the respective order.

$\gamma = 1$, which indicates the unweighted parameters, the LSP parameters are arranged unevenly, causing peaks and valleys in the envelope. This uneven arrangement of LSP parameters gradually becomes even as γ gets smaller.

Second, the change in the parameters due to γ is large when the parameters are far from their corresponding frequencies in $\gamma = 0$. Fig. 3 displays the gradients of the respective LSP parameters by their displacements. The gradients are given by

$$\frac{\omega_i(0.92) - \omega_i(1)}{0.92 - 1} \quad (5)$$

where $\omega_i(\gamma)$ indicates the i -th LSP parameter weighted by γ , and the displacements are defined as

$$\omega_i(1) - \omega_i(0) = \omega_i(1) - \frac{\pi}{p+1}i. \quad (6)$$

This figure shows that the gradients and the displacements have a positive correlation, proving the possibility of linear approximation of the weighting.

Third the shapes of the envelopes are mainly influenced by combinations of only a few LSP parameters around the respective frequency.

With the considerations above, we make the following assumptions for the direct weighting in LSP domain:

1. It can be performed by linear operations.
2. Changes in LSP parameters depend only on their values and the values of the neighboring parameters.
3. Evenly arranged LSP parameters do not change no matter what γ is.

These assumptions lead to the conversion model:

$$\begin{bmatrix} \hat{\omega}_1(\gamma_2) \\ \vdots \\ \hat{\omega}_p(\gamma_2) \end{bmatrix} = K \begin{bmatrix} \omega_1(\gamma_1) - \frac{\pi}{p+1} \\ \vdots \\ \omega_p(\gamma_1) - \frac{\pi p}{p+1} \end{bmatrix} (\gamma_2 - \gamma_1) + \begin{bmatrix} \omega_1(\gamma_1) \\ \vdots \\ \omega_p(\gamma_1) \end{bmatrix} \quad (7)$$

where

$$K = \begin{bmatrix} x_1 & y_1 & & 0 \\ z_2 & x_2 & y_2 & \\ & \ddots & \ddots & \ddots \\ 0 & & & z_p & x_p \end{bmatrix}. \quad (8)$$

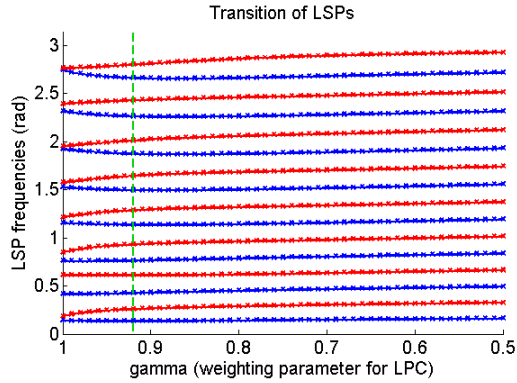


Fig. 2. The transition of 16th-order LSP parameters due to the weighting parameter γ . The leftmost group is the group of unweighted LSP and the green vertical dashed line indicates $\gamma = 0.92$, which is actually used for the weighting.

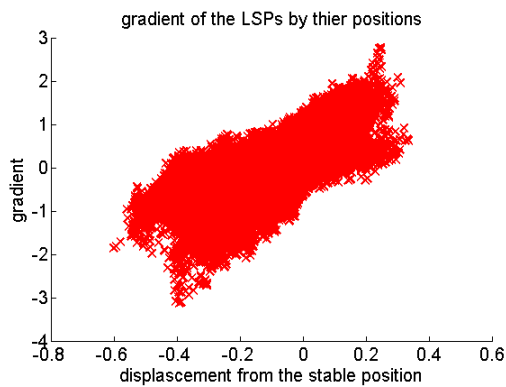


Fig. 3. The gradients in $\gamma \in [0.92 \ 1]$ at each LSP parameter position (total 15044 frames, 16th-order LSP).

When $\gamma_1 = 1$ and $\gamma_2 = 0.92$, this model approximates the weighting, and when $\gamma_1 = 0.92$ and $\gamma_2 = 1$, this model approximates the inverse. This conversion needs only $3p$ operations of multiplication, which is a minor complexity cost for the coders.

4.2. Optimizing the coefficients of the model

The coefficients for the model proposed above are decided in advance by means of an optimization for training data. This optimization can be represented in the following minimization problem, with LSP parameters $\omega_i(\gamma_1)$ and $\omega_i(\gamma_2)$ ($i = 1, \dots, p$) given from the data by weighting in the LPC domain:

$$\min_K \sum_n \|\omega^{(n)}(\gamma_2) - \hat{\omega}^{(n)}(\gamma_2)\|_2^2 \quad (9)$$

where K is the one in eq. (8), and

$$\begin{aligned} \omega^{(n)}(\gamma) &= (\omega_1^{(n)}(\gamma), \dots, \omega_p^{(n)}(\gamma))^T \\ \hat{\omega}^{(n)}(\gamma) &= (\hat{\omega}_1^{(n)}(\gamma), \dots, \hat{\omega}_p^{(n)}(\gamma))^T. \end{aligned} \quad (10)$$

The superscript n indicates the frame number of the data. The conversion model eq. (7) can also be written as

$$\hat{\omega}^{(n)}(\gamma_2) = L^{(n)} \mathbf{x} + \omega^{(n)}(\gamma_1) \quad (11)$$

where

$$L^{(n)} = \begin{bmatrix} d_1^{(n)} & d_2^{(n)} & & & & & 0 \\ & d_1^{(n)} & d_2^{(n)} & d_3^{(n)} & & & \\ & & d_2^{(n)} & d_3^{(n)} & d_4^{(n)} & & \\ 0 & & & d_2^{(n)} & d_3^{(n)} & d_4^{(n)} & \\ & & & & & & \ddots \end{bmatrix}$$

$$d_i^{(n)} = \left(\omega_i^{(n)}(\gamma_1) - \frac{\pi i}{p+1} \right) (\gamma_2 - \gamma_1)$$

$$\mathbf{x} = (x_1, y_1, z_2, x_2, y_2, \dots, z_p, x_p)^T. \quad (12)$$

Therefore, the objective function of eq. (9) can be transformed as

$$\sum_n \left\{ \mathbf{x}^T L^{(n)T} L^{(n)} \mathbf{x} - 2 \Delta^{(n)T} L^{(n)} \mathbf{x} + \Delta^{(n)T} \Delta^{(n)} \right\} \equiv G(\mathbf{x})$$

$$\Delta^{(n)} \equiv (\omega^{(n)}(\gamma_2) - \omega^{(n)}(\gamma_1)). \quad (13)$$

Since this function $G(\mathbf{x})$ is convex, it can be minimized at the stationary point of \mathbf{x} found as:

$$\begin{aligned} \frac{d}{d\mathbf{x}} G(\mathbf{x}) &= \sum_n \left\{ 2L^{(n)T} L^{(n)} \mathbf{x} - 2L^{(n)T} \Delta^{(n)} \right\} = 0 \\ \iff \mathbf{x} &= \left(\sum_n L^{(n)T} L^{(n)} \right)^{-1} \left(\sum_n L^{(n)T} \Delta^{(n)} \right) \end{aligned} \quad (14)$$

By changing the training data, this model can approximate not only the perceptual weighting but also the inverse weighting or a more complex weighting mentioned in the following section.

5. APPLICATIONS OF THE DIRECT LSP CONVERSION

5.1. Inverse weighting in low-bit-rate TCX coder

The low-bit-rate TCX coder used in the hybrid TCX and ACELP coding is one possible application of the conversion scheme proposed in this paper.

The TCX coder, as mentioned above, quantizes and codes LSP parameters and spectra divided by their weighted envelopes. Therefore, the TCX coder itself needs only the weighted parameters. However, in the hybrid coder, the unweighted LSP parameters are also used to interpolate between the next frame when the mode is switched to ACELP.

Especially in low-bit-rate situations where sufficient bits cannot be allocated for the parameters, the accuracy of the weighted parameters is more important for the TCX coder. As explained above, LSP parameters become farther away from each other when weighted, which decreases the dynamic range of the vectors of the parameters. For these reasons, it can be more efficient to quantize the weighted LSP parameters instead of unweighted ones. When the weighted parameters are quantized, the unweighted ones must also be calculated by inverse weighting. Here, the conversion scheme makes a difference in the computational complexity. The conversion of LPC to LSP costs much more than the inverse conversion: LPC to LSP costs about 0.4 weighted Million Operations Per Second (wMOPS [6]), while LSP to LPC and

the devised conversion in the LSP domain cost about 0.06 wMOPS and 2.4×10^{-3} wMOPS respectively, in the case of 16th order at 20 ms per frame at a 16-kHz sampling rate. By inverse weighting the parameters in the LSP domain directly, we can reduce the additional conversion of LPC to LSP and thereby lower the computational costs. Moreover, it is much easier in the LSP domain to detect or modify the parameters that are destabilized by the inverse weighting.

We conducted an experiment to simulate the situation mentioned above. LSP parameters are often quantized by vector quantization [7–9]. Therefore, we used a two-stage vector quantization with first order moving average inter-frame prediction as shown in Fig. 4. This quantizer has several kinds of codebooks: one in the first stage stores vectors for all the orders of LSP parameters and the others in the second stage store vectors for partial orders of them. The outputs in frame t of the quantizer $\xi_1^{(t)}, \xi_2^{(t)}, \dots, \xi_p^{(t)}$ are calculated by adding together for the respective orders the vectors from the codebooks in each stage. Then, the outputs are added with biases and the outputs of the previous frame as:

$$\hat{\omega}_i^{(t)} = \xi_i^{(t)} + a\xi_i^{(t-1)} + b_i, \quad (i = 1, \dots, p) \quad (15)$$

where b_i is a bias of the i th order. The target parameters are quantized to the closest $\hat{\omega}$ in Euclidean distance. We used 0.3 for a .

Table 1(a) shows distortion in the weighted spectral envelopes by either quantizing the unweighted LSP parameters or the weighted ones. The spectral distortion (SD) was calculated from the mean square errors of the log spectral envelopes [10]. It can be seen that especially at lower bit rates, quantizing weighted parameters has more advantages compared with quantizing unweighted ones in representing the weighted envelopes.

Table 1(b) compares the inverse weighting after quantizing the weighted LSP parameters. The weighted LSP parameters were inverse weighted in two ways after quantized. In the first case, the parameters were transformed into LPC parameters and then the LPC parameter of respective order a_i was divided by the i th power of the weighting parameter γ . In the second case, the parameters were inverse weighted directly in the LSP domain by the proposed conversion with optimal coefficients. The direct inverse weighting in the LSP domain performed the same as the inverse weighting in the LPC domain. Fig. 5 shows an example of the results for the inverse weighting. The red solid line is the envelope given by inverse weighting the blue dotted line in the LSP domain. The spectrum and its envelopes are presented in a linear amplitude domain since the spectrum are quantized linearly in the base coder. The inverse weighted envelope became similar to the black dashed line of the original unweighted envelope.

5.2. Weighting resolution-warped envelopes

Another possible application of the conversion scheme is for envelopes with warped frequency resolution [11]. The envelopes are calculated from a model that uses frequency warping based on sparse non-negative matrices, which make it possible to warp the frequency resolution arbitrarily. The model

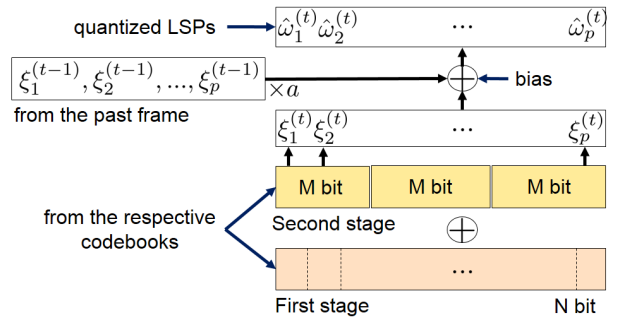


Fig. 4. Outline of an example of two-step vector quantizer for LSP parameters with first order moving average interframe prediction. $\hat{\omega}$ indicates the quantized parameters of the respective order.

Table 1. Spectral distortion (SD) by the LSP quantization calculated in each criteria. 15044 vectors of 16th-order LSP were tested at 16-kHz sampling rate.

(a) SD in weighted envelope given by eq. (2).

Bits used	Unweighted LSP quantized			Weighted LSP quantized		
	Av.SD (in dB)	Outliers (in %)		Av.SD (in dB)	Outliers (in %)	
		2-4 dB	> 4 dB		2-4 dB	> 4 dB
10	1.76	22.7	2.30	1.63	18.3	1.81
15	1.58	15.8	1.33	1.51	14.4	0.96
20	1.41	9.87	0.72	1.40	10.1	0.53
25	1.22	5.58	0.17	1.18	5.93	0.09
30	1.04	3.12	0.05	1.01	3.01	0.01
35	0.87	1.67	0.01	0.83	1.12	0.00

(b) SD in unweighted envelope given by eq. (1)

Bits used	Unweighted LSP quantized		
	Av.SD (in dB)	Outliers (in %)	
		2-4 dB	> 4 dB
10	3.26	67.7	20.8
15	3.02	66.6	16.5
20	2.77	64.7	12.3
25	2.42	49.4	8.20
30	2.08	31.9	5.00
35	1.76	20.3	2.75

Bits used	Weighted LSP quantized					
	Av.SD (in dB)	LSP→LPC→inv. weight→LSP (conventional inverse weighting)		Av.SD (in dB)	Inv. weighted directly in LSP (proposed inverse weighting)	
		Outliers (in %)			Outliers (in %)	
		2-4 dB	> 4 dB		2-4 dB	> 4 dB
10	3.30	66.0	22.1	3.33	66.4	22.6
15	3.21	66.1	20.3	3.21	66.8	20.3
20	3.17	69.3	18.7	3.13	70.0	18.1
25	2.97	62.0	15.1	2.85	62.6	14.3
30	2.66	55.1	10.8	2.56	55.0	9.82
35	2.38	40.2	8.08	2.22	40.3	6.39

of the envelopes is written as

$$H_k = \sum_l U_{kl} \left| 1 + \sum_{n=1}^p a_n e^{-j \frac{2\pi l}{N} n} \right|^{-1}, \quad (k = 0, \dots, N-1)$$

where N and U indicate the length of the envelopes and the inverse warping matrix, respectively. To approximate the perceptual weighting in the warped domain, the weighting parameter γ has to be modified as

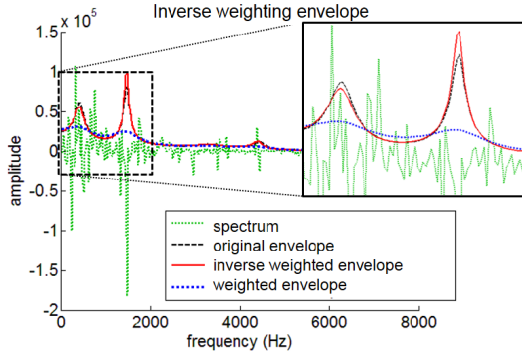


Fig. 5. 16th-order envelope inverse weighted in the LSP domain.

$$\tilde{H}_k = \sum_l U_{kl} \left| 1 + \sum_{n=1}^p \gamma \frac{f^{(l)}(k)}{l} a_n e^{-j \frac{2\pi l}{N} n} \right|^{-1} \quad (16)$$

where

$$f(k) = \sum_{n=1}^N U_{kn} (n-1). \quad (17)$$

This means the extent of the weighting changes over the frequencies so that fast Fourier transform cannot be used for calculating the envelopes, which makes this method complex in computation. However, since the shapes of envelopes are influenced by the arrangement of the neighboring LSP parameters, this weighting can be performed in the LSP domain.

For example, if the envelopes are warped logarithmically, the weighting must be greater for the lower frequencies, and this can be achieved by using higher values for the upper left coefficients of K in the model eq. (7).

In this context, the training data for weighted LSP parameters $\omega(\gamma_2)$ to optimize the coefficients of the model cannot be strictly given. Therefore, we approximate the training data using the LSP parameters calculated from the autocorrelation function given by Fourier transforming the power of the weighted envelopes before they are inverse warped:

$$\left| 1 + \sum_{n=1}^p \gamma \frac{f^{(k)}(k)}{k} a_n e^{-j \frac{2\pi k}{N} n} \right|^{-2}. \quad (18)$$

Fig. 6 shows an example of applying the optimized conversion in the LSP domain. The red solid line is the envelope weighted in the LSP domain, while the black dashed line is the envelope strictly weighted by eq. (16). The weighting model succeeded in approximating the weighting in the resolution-warped envelope. This method of weighting enables us to calculate the weighted envelopes by fast Fourier transform as in the case of normal envelopes and costs only additional $3p$ operations of multiplication. When $p = 16$, this method reduces the complexity if the length of envelopes is less than 256 points.

6. SUMMARY

We devised a method of direct conversion in the LSP domain for perceptual weighting with low computational complexity. By considering the useful properties of LSP, the weighting or the inverse weighting can be simply approximated in a

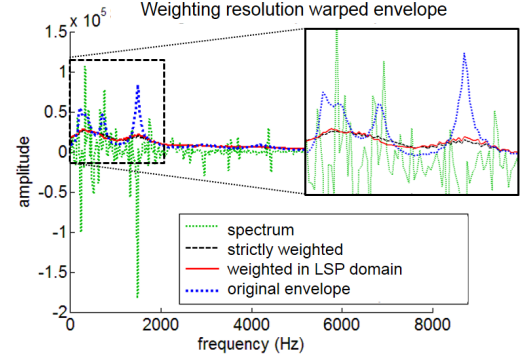


Fig. 6. 16th-order frequency-resolution-warped envelope weighted in the LSP domain.

linear operation form. This method can reduce the computational costs either in the case of a low-bit-rate TCX coder or frequency-resolution-warped envelopes. Furthermore, weighting or inverse weighting in the LSP domain also allows us to easily check the stability of the parameters.

LSP domain perceptual weighting enables us to weight the envelopes adaptively among the frequencies with a small amount of computation. A future issue is whether there are any weights that make the perceptual control of the quantization noise more efficient.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 26730100.

REFERENCES

- [1] 3GPP TS 26.290 version 11.0.0 Release 11, 3GPP, 2012.
- [2] M. Neundorff, et al., "MPEG Unified Speech and Audio Coding - The ISO/MPEG standard for high-efficiency audio coding of all content types," AES, 2012.
- [3] G. Fuchs, et al., "MDCT-based coder for highly adaptive speech and audio coding," EUSIPCO, pp.1264-1268, 2009.
- [4] F. Itakura, "Line spectrum representation of linear predictive coefficients," J. Acoust. Soc. Amer., vol.57, Suppl. no.1, S35, 1975.
- [5] S. Sagayama, "Stability conditions of LSP speech synthesis digital filters," Proc. of ASJ, pp. 153-154, 1982.
- [6] ITU-T Software Tool Library 2009 User's Manual, Geneva, 2009.
- [7] R. Laroia, et. al., "Robust and efficient quantization of speech LSP parameters using structured vector quantizers," ICASSP, Toronto, vol.1, pp.641-644, 1991.
- [8] W. Gardner and B. Rao, "Theoretical analysis of the high-rate vector quantization of LPC parameters," Trans. Speech Audio Process., IEEE, vol.3, no. 5, 1995.
- [9] H. Ohmuro, et. al., "Vector quantization of LSP parameters using moving average interframe prediction", Electronics and Communications in Japan, vol.77, pp.12-26 1994.
- [10] W. Kleijn and K. Paliwal, "Speech coding and synthesis," Elsevier, pp.433-466, 1995.
- [11] R. Sugiura, et al., "Frequency warped spectral envelope representation for audio coding," Proc. of ASJ, no. 1-6-8, pp.281-284, 2014.