Design of Bandwidth Scalable LSF Quantization using Interframe and Intraframe Prediction

Hiroyuki Ehara†, Toshiyuki Morii‡, Masahiro Oshikiri†, Koji Yoshida† and Kouichi Honma†

† Next-Generation Mobile Communications Development Center
‡ AV Core Technology Development Center
Matsushita Electric (Panasonic), Japan

{ehara.hiroyuki/morii.to/oshikiri.masahiro/yoshida.koji/honma.kouich}@jp.panasonic.com

Abstract
We developed a bandwidth scalable LSF (Line Spectral Frequency) quantizer utilizing interframe and intraframe prediction. The predictor used in the quantizer was designed as combination of first order AR (auto regressive) prediction and a codebook mapping technique, which maps a narrow band LSF codebook to a wide band LSF codebook. An upsampling process in the autocorrelation domain was utilized for conversion between narrowband LSF and wideband LSF. Codebooks for the mapping, coefficients for the predictor, and codebooks for VQ (vector quantization) of prediction residues were designed by using training database including multi languages and various kinds of background noise conditions. When 16 bits were assigned to an enhancement layer of the scalable LSF quantizer, utilization of the interframe and intraframe prediction improved the quantizer performance in spectral distortion by about 0.3 dB.

1. Introduction
In this article, we present a design of bandwidth scalable LSF (line spectral frequency) quantization in which interframe and intraframe prediction is utilized for predicting WB-LSF (LSF calculated from wide-band speech signals) from NB-LSF (LSF calculated from narrow-band speech signals). In our previous report [1], we explored the exploitation of the previously quantized NB-LSF and WB-LSF, and currently quantized NB-LSF for predicting the current WB-LSF. Our explored scheme is based on the LSF parameters quantized in the past and thus is memory-based. To achieve further improvement without degrading robustness against channel errors, a codebook mapping technique has been newly introduced as an efficient memory-less quantization scheme.

The following sections will present the design of the bandwidth scalable LSF quantizer including description of the enhancement layer of the bandwidth scalable LSF quantizer and training procedures for designing codebooks used in the quantizer. The sections are organized as follows. Section 2 describes the algorithm of the bandwidth scalable LSF quantizer. Section 3 provides the designing procedure of the quantizer. Section 4 shows the objective test results. Finally, Section 5 summarizes the results of our work.

2. Bandwidth scalable LSF quantizer
In this section, the algorithm of the scalable LSF quantization is described. Since the main subject of our study is scalable quantization of WB-LSF using NB-LSF, we focus on the algorithm of a bandwidth (BW) extension quantizer of the BW scalable LSF quantizer. The BW extension quantizer is based on predictive vector quantization (PVQ) and comprises the mapping-based prediction described in Section 2.3.2.

2.1. Overview of BW scalable WB-LSF quantization
Figure 1 shows a simulation setup of the BW scalable LSF quantizer. In an enhancement layer, a WB-LSF vector is quantized in a layered manner using a NB-LSF vector quantized by a NB-LSF quantizer. The quantized NB-LSF vector is upsampled in the autocorrelation domain, and then an error vector between the upsampled NB-LSF and the input WB-LSF vectors is quantized by a switched predictive 3-stage vector quantizer. The quantizer has two prediction modes. One is designed to handle a variety of input LSF vectors, while the other is specialized to quantize LSF vectors in stationary segments. As described later, the former is a memory-less VQ mode, and the latter is a memory-based VQ mode. Selection of the mode is performed in a closed-loop manner, i.e. the quantization is performed in the both modes, and the mode that gives the least quantization error is selected. In our study, the sampling rates of the NB and WB signals are 8 and 16-kHz, and the order of LSF is 12 and 18, respectively. The quantizer is operated on 20-ms frames.

2.2. Bit allocation
Bit allocation of the bandwidth scalable LSF quantizer is shown in Table 1. A 29-bit scalable LSF quantizer is used as the NB-LSF quantizer. It consists of a 21-bit and an 8-bit vector quantizer performance in spectral distortion by about 0.3 dB.
Table 1: Bit allocation of the scalable LSF quantizer

<table>
<thead>
<tr>
<th>Layer</th>
<th>Parameter</th>
<th>Bits/frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>NB-core layer</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>NB-enhancement layer</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>29 (1.45 kbit/s)</td>
</tr>
<tr>
<td>WB</td>
<td>Mode</td>
<td>Bits/frame</td>
</tr>
<tr>
<td></td>
<td>1st stage codebook</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2nd stage codebook</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>3rd stage codebook</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>16 (0.8 kbit/s)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>45 (2.25 kbit/s)</td>
</tr>
</tbody>
</table>

2.3. Bandwidth extension quantizer

Figure 2 shows a block diagram of the BW extension quantizer in the WB scalable LSF quantizer. Upsampling process, codebook mapping and 2-mode predictive 3-stage VQ are described in the following subsections.

2.3.1. Upsampling from NB-LSF to WB-LSF

The quantized NB-LSF parameters are converted to linear prediction (LP) coefficients, and the LP coefficients are transformed into autocorrelation coefficients. The autocorrelation coefficients are upsampled in a way that is equivalent to an upsampling process of the input signal in the time domain by using Equations (1) and (2). $r(i)$ is the $i$-th autocorrelation coefficient converted from the NB-LSF, and $R(i)$ is an upsampling version of $r(i)$. This upsampling calculation is equivalent to the upsampling calculation in a time domain on the condition the autocorrelation coefficients have an infinity order. The upsampling autocorrelation coefficients, $R(i)$, are recovered to LSF parameters, and the upsampling version of the NB-LSF parameters are obtained.

$$R(2k) = r(k) + \sum_{m=-\infty}^{+\infty} \sum_{n=0}^{m} r(k-n+m)\sigma(m)\sigma(n)$$

where

$$\sigma(x) = \text{sinc} \left( x + \frac{1}{2} \right) \pi$$

2.3.2. Codebook mapping

One-to-one mapping is adopted as the simplest mapping scheme. The upsampled NB-LSF is vector quantized using a NB-LSF codebook for the mapping. Each codevector in the NB-LSF codebook is associated with a codevector in a WB-LSF codebook. The estimated WB-LSF vector, $\hat{L}_W$, can be expressed by Eq.(4).

$$\hat{L}_W = L_{CBw}(\text{idx})$$

$L_{CBw}(\text{idx})$ is the WB-LSF codevector whose index is idx, and $L_{CBn}$ is an index of the selected codevector from the NB-LSF codebook in the vector quantization for the mapping. Using this mapping scheme, however, the number of possible WB-LSF vectors is limited to the size of the mapping codebook. To increase the number of possible WB-LSF vectors, we considered utilization of the upsampled NB-LSF vector, $\hat{L}_N$, and the selected codevector from the NB-LSF codebook, $L_{CBn}(\text{idx})$. We compared two mapping schemes as follows. Mapping-A simply interpolates the upsampled NB-LSF and the mapped WB-LSF, while information about the distance between the upsampled NB-LSF and the selected NB-LSF codevector is used to estimate WB-LSF in Mapping-B.

[Mapping-A]

$$\hat{L}_W = \beta_1 L_N + \beta_2 L_{CBw}(\text{idx})$$

[Mapping-B]

$$\hat{L}_W = \beta_1 L_N + \beta_2 L_{CBw}(\text{idx}) + \beta_3 L_{CBn}(\text{idx})$$

$\beta_1$, $\beta_2$ and $\beta_3$ are predictive coefficients, and they can be obtained through off-line training so as to minimize the total estimation error, $D_{map}$, for a training database ($n$ is a frame no.).

$$D_{map} = \sum_n \| L^{(n)}_N - \hat{L}^{(n)}_W \|^2$$

Our experimental result showed that Mapping-B scheme gave slightly lower $D_{map}$ than Mapping-A did.

2.3.3. 2-mode predictive 3-stage VQ

As shown in Figure 2, a quantized LSF parameters, $\hat{L}_W(i), i = 0, \ldots, 17$, are given by Equation (8).

$$\hat{L}^{(n)}_W(i) = \beta_0(i)\hat{C}^{(n)}(i) + \beta_1(i)\hat{L}^{(n)}_N(i) + \beta_2(i)L^{(n)}_{CBw}(i) + \beta_3(i)L^{(n)}_{CBn}(i)$$

$$\beta_4(i)\hat{L}^{(n-1)}_W(i) + \beta_5(i)\frac{L^{(n-1)}_N(i)}{L^{(n-1)}_N(i)}\hat{L}^{(n)}_W(i)$$

$\hat{L}^{(n)}_W(i)$ is the $i$-th quantized WB-LSF parameter at the $n$-th frame. $\hat{L}^{(n)}_N(i)$ is the $i$-th upsampling NB-LSF parameter at the $n$-th frame. $L^{(n)}_{CBw}(i)$ is the $i$-th element of the WB-LSF vector.
selected in the codebook mapping at the n-th frame. \( L_{GBn}(i) \) is the i-th element of the NB-LSF vector selected in the codebook mapping at the n-th frame. \( C(i) \) is the i-th element of an error vector quantized by the 3-stage VQ at the n-th frame. \( \beta_0(i) \) to \( \beta_5(i) \) are predictive coefficients for the i-th element.

This formulation utilizes the "Mapping-B" described in Section 2.3.2. and "Mapping-A" is a special case where \( \beta_3 \) is set to zero.

The last two terms of Eq. (8) include previously quantized LSF. Therefore the VQ given by Eq. (8) is a memory-based VQ unless \( \beta_4(i) = \beta_5(i) = 0 \). To improve robustness against channel errors, we adopt 2-step predictive VQ, and a memory-less VQ is used as one of the two modes. That is, in Figure 2, the coefficient table consists of two sets of coefficients, and one of the sets is for the memory-less VQ mode (Mode0) and having \( \beta_4(i) = \beta_5(i) = 0 \). The 3-stage VQ codebooks also contain two subsets of codebook, one for Mode0 (memory-less) and the other for Mode1 (memory-based, for stationary segments). It should be noted that Mode1 also has memory-less contributions, \( \beta_1, \beta_2 \) and \( \beta_3 \), and thus Model1 is memory-based VQ having "forgetting" capability [1].

3. Codebook training procedure

The algorithms of training 1) a codebook for mapping, 2) a set of predictor coefficients and 3) a three-stage VQ codebook are outlined in the following subsections. Regarding database utilized for those training, total of 794 seconds data, including seven languages, several background noise conditions and music samples, was used.

3.1. Codebook for mapping

The design procedure of the mapping codebook is straightforward and summarized as follows.

(step-1) to prepare training data set of vector pairs of NB-LSF and WB-LSF.

(step-2) to create a codebook of NB-LSF by LBG algorithm using the NB-LSF training data.

(step-3) to perform VQ on the NB-LSF training data using the created NB-LSF codebook and collect paired WB-LSF data for each NB-LSF code space (cluster).

(step-4) to calculate the average of the collected WB-LSF for each cluster.

The NB-LSF and WB-LSF codebooks have 128 (7-bit) codevectors each.

3.2. Predictor coefficients

3.2.1. Initial coefficients

As shown in Eq.(8), the predicted WB-LSF, \( \hat{L}_W(i) \), is expressed by:

\[
\hat{L}_W(i) = \beta_1(i) \hat{L}_W(i) + \beta_2(i) L_{GBn}(i) + \beta_3(i) \hat{L}_W(i) + \beta_4(i) \hat{L}_W(i) - \beta_5(i) \hat{L}_W(i)
\]

Then the total weighted prediction error, \( E \), is given by:

\[
E = \sum_{n} E_n = \sum_{n} \sum_{i} w(n)(i) | L_W(i) - \hat{L}_W(i) |^2
\]

where \( w(n)(i) \) is a weighting coefficient to the i-th LSF at the n-th frame and given by \( w(n)(i) = c_1(i)/L_W(i) + c_2(i)/L_W(i) + c_3(i) \), where \( l(i) = L_W(i+1) - L_W(i-1) \) and \( c_1-c_3 \) are constants.

By solving the simultaneous equation, \( \frac{\partial D}{\partial \beta_0(i)} = \frac{\partial D}{\partial \beta_1(i)} = \frac{\partial D}{\partial \beta_2(i)} = \frac{\partial D}{\partial \beta_3(i)} = \frac{\partial D}{\partial \beta_4(i)} = \frac{\partial D}{\partial \beta_5(i)} = 0 \), the updated set of predictor coefficients, \( \{\beta_0(i), \beta_1(i), \beta_2(i), \beta_3(i), \beta_4(i), \beta_5(i)\} \), are obtained.

3.2.2. Update of predictor coefficients

The total quantizing distortion, \( D \), is expressed by:

\[
D = \sum_{n} D_n = \sum_{n} \sum_{i} w(n)(i) | L_W(i) - \hat{L}_W(i) |^2
\]

By solving the simultaneous equation, \( \frac{\partial D}{\partial \beta_0(i)} = \frac{\partial D}{\partial \beta_1(i)} = \frac{\partial D}{\partial \beta_2(i)} = \frac{\partial D}{\partial \beta_3(i)} = \frac{\partial D}{\partial \beta_4(i)} = \frac{\partial D}{\partial \beta_5(i)} = 0 \), the updated set of predictor coefficients, \( \{\beta_0(i), \beta_1(i), \beta_2(i), \beta_3(i), \beta_4(i), \beta_5(i)\} \), are obtained.

3.3. Three-stage VQ codebook

3.3.1. Initial codebook

Prediction residual vectors, \( L_R(i) \), are calculated using unquantized WB-LSF vectors as follows (note that unquantized WB-LSF, \( L_R(i) \) is used instead of quantized WB-LSF, \( \hat{L}_R(i) \)).

\[
L_R(i) = L_W(i) - \beta_1(i) L_N(i) - \beta_2(i) L_{GBn}(i) - \beta_3(i) L_W(i) - \beta_4(i) L_N(i) - \beta_5(i) L_W(i)
\]

Then the total weighted prediction error, \( E \), is given by:

\[
E = \sum_{n} E_n = \sum_{n} \sum_{i} w(n)(i) | L_W(i) - \hat{L}_W(i) |^2
\]

Finally, the prediction residual vectors, \( L_R(i) \), are calculated using the unquantized WB-LSF vectors as follows (note that unquantized WB-LSF, \( L_R(i) \) is used instead of quantized WB-LSF, \( \hat{L}_R(i) \)).
Then the total quantizing distortion $D$ is given by:

$$D = \sum_n D_n = \sum_n w^{(n)} \| \hat{L}_W^{(n)} - L_W^{(n)} \|^2$$  \hspace{1cm} (14)$$

$w^{(n)}$ is a diagonal matrix whose elements are $w^{(n)}(i), i = 0, \ldots, 17$, and $L_W^{(n)}$ and $\hat{L}_W^{(n)}$ are the unquantized and the quantized WB-LSF vectors at the $n$-th frame respectively. The updated codebook is obtained by minimizing Eq. (14) using a projection method [2].

### 4. Objective evaluation

This section presents results of an objective evaluation test. The performance of the BW scalable LSF quantization is evaluated in terms of spectral distortion (SD), which is given by Eq. (15). $SD$ is calculated as follows:

$$SD = \left( \frac{1}{N} \sum_{n=1}^{N} \left( 10 \log_{10} \frac{|A(e^{j2\pi n/N})|^2}{|\hat{A}(e^{j2\pi n/N})|^2} \right)^2 + \frac{1}{N} \sum_{n=1}^{N} \left( 10 \log_{10} \frac{|A(e^{j2\pi n/N})|^2}{|\hat{A}(e^{j2\pi n/N})|^2} \right)^2 \right)^{1/2}$$  \hspace{1cm} (15)$$

Table 2 shows configurations of the LSF quantizer tested. VQ1 and VQ3 are studied in our previous report [1]. Eight Japanese sentence pairs (four females and four males, 64 seconds in total) were used as test material. Another set of Eight Japanese sentence pairs corrupted with car-noise at SNR=15dB was also used as test data. The test results are shown in Table 3. Observation remarks on the results are summarized as follows.

1. More than 0.1 dB improvement in SD is brought about by introducing the mapping-based prediction. (Comparison between VQ1 and VQ2)
2. Our memory-based prediction [1] achieves lower SD than the mapping-based prediction does, in the case of the car-noise condition in particular. (Comparison between VQ2 and VQ3)
3. Combination of the memory-based and mapping-based predictions gives the lowest SD performance. (VQ4)
4. Mapping-B provides better performance than Mapping-A does. (Comparison between VQ4 and VQ5)
5. As for clean speech, SD 1.3 dB is achievable by using 16 bits. (VQ5)

Regarding the test results of the car-noise condition, the SD values are lower than the clean speech condition. According to our observation, this is because most segments have car-noise-like spectrum characteristics, i.e. much energy in low frequency bands. In the clean speech condition, large SD values are typically found on unvoiced segments, in which high frequency bands have large energy. However, such unvoiced characteristics can be buried by the car-noise. As a result, those segments where large SD values are found almost disappear, and thus the average SD value decreases. Due to the nature of stationary characteristics of car-noise, the memory-based prediction is more effective to the car-noise condition, and the effectiveness of the mapping-based prediction becomes relatively low.

### Table 2: Tested configurations

<table>
<thead>
<tr>
<th>Mapping-based pred.</th>
<th>Memory-based pred.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VQ1</td>
<td>not used ($\beta_2=\beta_3=0$)</td>
</tr>
<tr>
<td>VQ2</td>
<td>Mapping-A ($\beta_3=0$)</td>
</tr>
<tr>
<td>VQ3</td>
<td>not used ($\beta_4=\beta_5=0$)</td>
</tr>
<tr>
<td>VQ4</td>
<td>Mapping-A ($\beta_3=0$)</td>
</tr>
<tr>
<td>VQ5</td>
<td>Mapping-B</td>
</tr>
</tbody>
</table>

### Table 3: Test results

<table>
<thead>
<tr>
<th>Clean speech, 64 sec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VQ1</td>
</tr>
<tr>
<td>VQ2</td>
</tr>
<tr>
<td>VQ3</td>
</tr>
<tr>
<td>VQ4</td>
</tr>
<tr>
<td>VQ5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Car-noise@SNR15dB, 69 sec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VQ1</td>
</tr>
<tr>
<td>VQ2</td>
</tr>
<tr>
<td>VQ3</td>
</tr>
<tr>
<td>VQ4</td>
</tr>
<tr>
<td>VQ5</td>
</tr>
</tbody>
</table>

Our results demonstrated that a one-to-one mapping model was useful for bandwidth scalable LSF quantization. A one-to-many mapping model suggested in [3] may bring about further improvement at the cost of increasing complexity.

### 5. Conclusion

This article presented a design of a bandwidth scalable LSF quantization using intermediate frame and interframe prediction. Our study focused on a bandwidth extension quantizer, which quantizes wideband LSF using quantized narrowband LSF. A combined predictive VQ scheme, memory-based VQ with memory-less VQ, was adopted to the bandwidth extension quantizer. A codebook mapping technique was utilized to improve the performance of the memory-less VQ. Objective test results showed that the designed quantizer achieved 1.31 dB in spectral distortion using 16 bits for the bandwidth extension quantizer when 29 bits were used for quantizing narrowband LSF.

### 6. References

