EFFICIENT QUANTIZATION OF LSF PARAMETERS BASED ON TEMPORAL DECOMPOSITION

Sung Joo Kim, Sangho Lee, Woo Jin Han, Yung Hwan Oh

Dept. of Computer Science, KAIST, Taejon, Korea
sung@bulsai.kaist.ac.kr

ABSTRACT

In this paper, we present a restricted temporal decomposition method for LSF parameters. The event vectors estimated by this method preserve the ordering property of LSF parameters so that they can be quantized efficiently. Experimental results show that interpolated LSF parameters can be quantized transparently at the rate of 753 bps. We also design a LPC vocoder at 996 bps as an application of the proposed method. According to a listening test, the reconstructed speech of our vocoder has reasonable quality compared with 2400 bps LPC10e.

1. INTRODUCTION

The temporal decomposition is a method of speech coding which decomposes a given vector trajectory into a set of temporally overlapping event functions and corresponding event vectors [1,2,3].

The original temporal decomposition assumes that each event is a superposed component of the given vector trajectory [1], so the distribution of the estimated event vectors is different from that of the given vector trajectory. In case of LAR or cepstrum parameters, this creates no problem, because each order of those parameters is independent and has no boundary value.

However, LSF parameters are dependent to adjacent orders and have the ordering property. Therefore, decomposing LSF into superposed event vectors causes the event vectors not to obtain their respective spectra because they can be unstable, i.e., the event vectors are no longer LSF parameters. To solve this problem, we propose another restriction on event functions so that every event vectors for LSF parameters have their own spectra and the vector trajectory can be interpreted as an interpolation of the estimated events.

2. RESTRICTED TEMPORAL DECOMPOSITION

Let a given vector trajectory be $Y = [\vec{y}(1), \vec{y}(2), \ldots, \vec{y}(N)]$. The temporal decomposition estimates proper event vectors and functions $(\vec{a}_j, \phi_j(n)), j = 1, \ldots, J$, which minimize

$$E = \|Y - A\Phi\|^2 = \sum_{n=1}^{N} \|\vec{y}(n) - \vec{y}'(n)\|^2 , \qquad (1)$$

where $\vec{y}'(n) = \sum_{j=1}^{J} \vec{a}_j \phi_j(n)$.

To estimate proper events for speech, some restrictions on event functions should be enforced as follows.

$$\phi_j(n) = \begin{cases} 0 < x \le 1 & , \text{ if } L(j) \le n \le R(j) \\ 1 & , \text{ if } n = C(j) \\ 0 & , \text{ otherwise} \end{cases} \qquad (2)$$

where $L(j)$, $C(j)$, and $R(j)$ are the left boundary, the center, and the right boundary position of $\phi_j(n)$ respectively, and $L(j) \le C(j) \le R(j)$, for $1 \le j \le J$ and $C(j-1) < C(j)$, for $1 < j \le J$. (2) means that each event function $\phi_j(n)$ has only one lobe and maximum value one at its center $C(j)$ and the centers are ordered [1,2,3].

For the temporal decomposition of LSF parameters, we make event functions have one more restriction:

$$\sum_{j=1}^{J} \phi_j(n) = 1 , \quad \text{for } 1 \le n \le N . \qquad (3)$$

(3) gives a certain constraint on superposing event vectors. Furthermore, the following property is also guaranteed if we consider (2) and (3) simultaneously.

$$\sum_{n=1}^{N} \phi_i(n)\phi_j(n) = 0 , \quad \text{if } |i - j| > 1 \qquad (4)$$

Now, (4) implies that no more than two event functions can have non-zero values for a certain time of the vector trajectory. Consequently, reconstructed vector trajectory becomes a simple piecewise interpolation as follows.

$$\vec{y}'(n) = \vec{a}_j \phi_j(n) + \vec{a}_{j+1}\phi_{j+1}(n) = \vec{a}_{j+1} + (\vec{a}_j - \vec{a}_{j+1})\phi_j(n) , \qquad (5)$$

for $C(j) \le n < C(j+1)$.

Substituting (5) into (1) and setting the partial derivatives of (1) with respect to $\phi_j(n)$ equal to zero, we have

$$\phi_j(n) = \begin{cases} \min(1, \max(0, \hat{\phi}(n))) & , \text{ if } C(j) \le n < C(j+1) \\ 1 - \phi_{j-1}(n) & , \text{ if } C(j-1) < n < C(j) \\ 0 & , \text{ otherwise} \end{cases} \qquad (6)$$

where $\hat{\phi}(n) = \left\langle \vec{y}(n) - \vec{a}_{j+1}, \vec{a}_j - \vec{a}_{j+1} \right\rangle \big/ \|\vec{a}_j - \vec{a}_{j+1}\|^2$.

In addition, we can re-estimate event vectors proper to the estimated event functions by using the following formula the same as the original temporal decomposition does [1,2,3].

$$A = Y\Phi^T(\Phi\Phi^T)^{-1} \tag{7}$$

However, a re-estimated event vector may violate the ordering property of LSF parameters, since (1) does not properly measure the error occurred by disordered LSF parameters. So an event vector should be updated within a valid range conserving the ordering property.

In the result, if we know the central positions of the events $C(j), j = 1, \ldots, J$, and initialize the corresponding event vectors with the samples of the vector trajectory $\vec{y}(C(j))$, we can calculate proper event functions and vectors iteratively by using (6) and (7) one after the other. In short, all we need to know for the restricted temporal decomposition (RTD) is the central positions of all events. There are several ways to guess proper locations of events, but in this paper, we suggest the following spectral transition measure based on LSF parameters, where $M = 2$ [7].

$$STM_{LSF}(n) = \left\| \sum_{t=-M}^{M} t \cdot \vec{y}(n+t) \right\|^2$$

A local minimal point of $STM_{LSF}(n)$ denotes the location of minimal spectral transition and can be considered as the central position of corresponding event. However, the number of events found by $STM_{LSF}(n)$ is about 10 per second experimentally and not enough to interpolate the vector trajectory of LSF parameters. It is because $STM_{LSF}(n)$ cannot detect some events with short duration like bursts. Therefore a few events should be inserted to reduce the error E of the initial interpolation result. By this reason, we insert a new event where the error $e(n) = \|\vec{y}(n) - \vec{y}'(n)\|^2$ has a local maximum and larger than a certain threshold θ.

By the way, a weighted error measure is widely used [4] for LSF parameters, and we also use a weighted error E_w rather than E during the RTD of LSF parameters:

$$E_w = \sum_{n=1}^{N}\sum_{k=1}^{p} \frac{\left(y_{k+1}(n) - y_{k-1}(n)\right)\left(y_k(n) - y'_k(n)\right)^2}{\left(y_{k+1}(n) - y_k(n)\right)\left(y_k(n) - y_{k-1}(n)\right)},$$

assuming $y_0(n) = 0$ and $y_p(n) = \pi$. By using this weighted error for LSF parameters, the final spectral distortion error of interpolation was significantly reduced.

3. RTD OF LSF PARAMETERS

We designed two separate experiments to measure the performance of the interpolation and the quantization of LSF parameters based on RTD. We used the prediction gain for the former and the spectral distortion (SD) for the latter [5].

The speech corpus for these experiments was given from the TIMIT database. We chose 1,890 phonetically diverse sentences (SI set) and used 504 sentences for the test and 1,386 sentences for training the threshold θ for event insertion and 10 dLSF quantizers. By using the auto-correlation method, we estimated 10th order LPC parameters with 30ms hamming window and 20ms frame shift. Finally, LPC parameters were converted into 10th order LSF parameters, so the updating rate of LSF parameters was 50 Hz.

First, we measured the average interpolation error E_w of a sample sentence 4.98 seconds long to decide the maximum iteration count for the re-estimation. Because both (6) and (7) are stepwise optimal solutions, the error decreases monotonically and converges rapidly to a local minimum. From this experiment, we found that re-estimating five times is enough to converge the interpolation error. Therefore, we decided to set the maximum iteration count as five.

The prediction gain of given LSF parameters describes how well the LSF parameters model the original spectrum. For the training set, which is 221,586 frames long, the average prediction gain of the original LSF parameters was 9.09 dB and the gain reduction caused by interpolation was only 0.15 dB when we used $\theta = 0.6$. Moreover, the reconstructed speech was almost indistinguishable from the original speech during an informal listening test.

Event functions of the proposed decomposition method are estimated by (6) and satisfy several properties like (2) and (3). Therefore, quantizing only the interval $[C(j), C(j+1)]$ of $\phi_j(n)$ is enough to reconstruct all event functions. Furthermore, $\phi_j(n)$ always starts from one and goes to zero in that interval and the type of decrease can be vector quantized after normalizing the length of $\phi_j(n)$. In this experiment, we took 10 interpolated samples from an event function for length-normalization. Consequently, $\phi_j(n)$ can be quantized by its length $p(j) = C(j+1) - C(j)$ and the type of decrease. During the interpolation of the training set with $\theta = 0.6$, the maximum length of the event function was 11 frames long, i.e., 220 ms. Therefore we set the maximum length of the event function as 15 and quantized $p(j)$ with four bits except $p(j) = 1$ with three bits, '000'.

The vector quantizer for the shape of length-normalized event function and the scalar quantizers for each order of dLSF of event vectors were trained by the interpolation results of training set. Finally, we measured the quantization errors of test set, which is 83,910 frames long, varying the bit allocations. As shown in Table 1, the LSF parameters can be quantized satisfying the conditions for transparent coding [5] when 33 bits are used for an event vector and 6 bits for the shape of an event function. In Table 2, the average bit rate for LSF parameters is calculated. Note that there is no need for the type of decrease when $p(j) = 1$.

SD[dB] (2-4 dB [%]) (> 4 dB [%])		**LSF SQ Bit Allocation**		
		31 bits (3,3,3,3,4, 3,3,3,3,3,)	32 bits (3,3,3,3,4, 3,4,3,3,3,)	33 bits (3,3,3,3,4, 3,4,3,4,3,)
Event Function Shape VQ Bit Allocation	4 bits	1.126 (4.89) (0.16)	1.082 (3.60) (0.08)	1.054 (2.92) (0.04)
	5 bits	1.070 (3.97) (0.16)	1.023 (2.86) (0.07)	0.994 (2.20) (0.04)
	6 bits	1.011 (3.25) (0.15)	0.963 (2.17) (0.07)	0.933 (1.57) (0.03)

Table 1: Results of the quantization based on RTD.

	LSF SQ	**Event function**		**Freq. (Hz)**	**Total (bps)**
		Position	Shape		
p(j) > 1	33	4	6	14.04	604
p(j) = 1	33	3	0	4.12	149
Total	753 bps				

Table 2: The bit rate of the proposed LSF parameter quantizer.

We present results of RTD for a word, /hanguk/ in Figure 1. You can see the spectra of event vectors in (a), the event functions in (b), the original signal in (c), the original LSF parameters in (d), and the interpolated LSF parameters in (e).

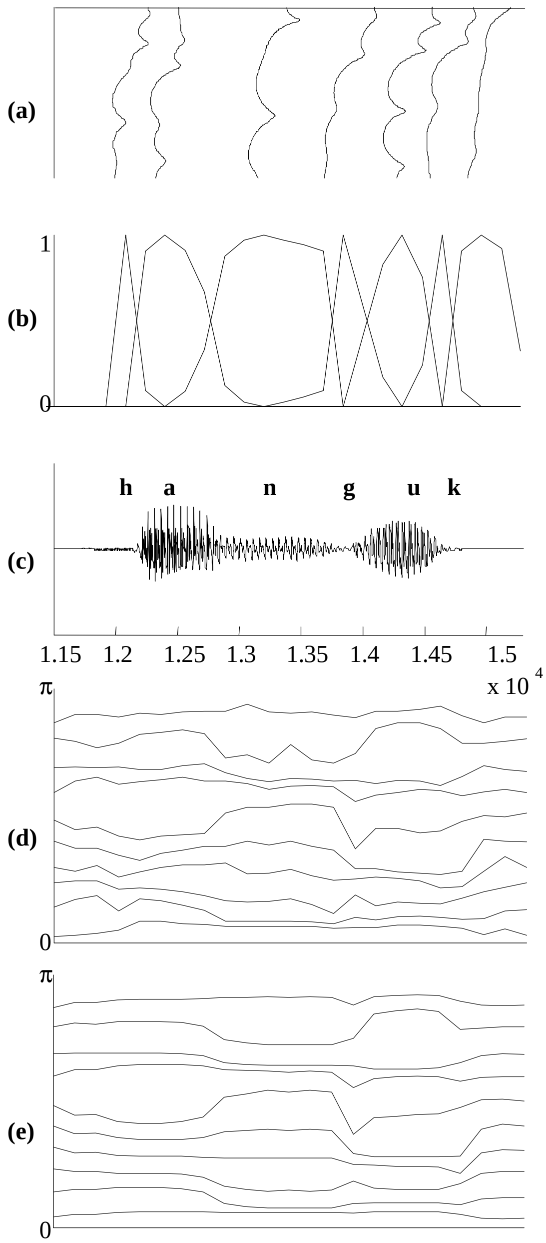

Figure 1: Example of RTD results.

4. A LPC VOCODER USING RTD

In this section, we design a LPC vocoder that uses RTD of LSF parameters as an application of the proposed method. Instead of developing a whole system, we use FS1015 LPC10e as a base system of our vocoder and just modify the quantization method of LPC parameters. Table 3 summarizes the main features of the base system and the modified one [6].

Event functions are quantized as described in Section 3, but for event vectors, we use a split vector quantization method to reduce the bit rate more. We split an event vector into 3, 3, and 4 dimensional ones and then quantize each vector with 8 bits. We use weighted error E_w as the distance measure of quantizers. In this case, the average SD of the test set is 1.28 dB and the percentages of type1 and type2 errors are 6.84 % and 0.12 %.

To measure the quality of the reconstructed speech, we did pair comparison tests with LPC10e. We gathered 10 different spoken sentences from five males and five females, and 10 listeners were in testing. Figure 2 shows the result of pair comparison tests and the total number of preferred times are 26 and 48 for modified and original LPC vocoder respectively and there are 26 times of no preferences.

Note that θ for event insertion of RTD is 1.0 and it is much larger than that of previous section. We use larger θ to reduce the number of events and the bit rate. Of course, the interpolation and quantization errors increase as a result of using larger θ, but the errors do not affect the quality of synthesized speech seriously. So the modified LPC vocoder can give a reasonable quality of reconstructed speech although it uses only 462 bps for quantizing LPC parameters.

5. CONCLUSION

This paper showed that the restricted temporal decomposition can interpolate the original vector trajectory of LSF parameters very well and we can quantize the resulting events efficiently. As a result, we can quantize the interpolated LSF parameters at 753 bps while satisfying the conditions for transparent coding.

	LPC10e	Modified LPC vocoder
Sampling rate	8 kHz	
Frame rate (Event rate)	44.4 Hz	Frame: 44.4 Hz Event: 12 Hz (p(j)>1) 2 Hz (p(j)=1)
Assigned bits/frame (bits/event)	54	12 bits/frame (v/uv, pitch ,energy) 34 bits/event (p(j)>1) 27 bits/event (p(j)=1)
Pitch	AMDF method	
Gain	RMS value	
LPC analysis	Semi-pitch-synchronous Covariance method	
LPC order	10	
LPC parameter coding	Generalized reflection coefficients: k_i LAR : k_1, k_2 Linear : $k_3 - k_{10}$	line spectral frequencies: ω_i RTD: $\theta = 1.0, M = 2$ SVQ: event vectors VQ: event function shape Linear: event function position
Bit rate	2400 bps	average 996 bps

Table 3: Main features of the base and the modified vocoder.

We also developed a 996 bps LPC vocoder by using the proposed LSF quantization method.

Currently, we are studying on the properties of the RTD and trying to apply this method to a very low bit rate speech coding and also to a speech recognition system.

6. REFERENCES

1. Atal, B.S., "Efficient Coding of LPC Parameters by Temporal Decomposition," *Int. Conf. on Acoustics, Speech and Signal Processing*: 81-84, 1983.

2. Van Dijk-Kappers, A.M.L., and Marcus, S.M., "Temporal Decomposition of Speech," *Speech Communication, Vol.8, No.2*: 125-135, 1989.

3. Cheng, Y.-M., and O'Shaughnessy, D., "On 450-600 b/s Natural Sounding Speech Coding," *IEEE Trans. on Acoustics, Speech and Signal Processing, Vol.1, No.2*: 207-219, 1993.

4. Ramachandran, R.P., Sondhi, M.M., Seshadri, N., and Atal, B.S., "A Two Codebook Format for Robust Quantization of Line Spectral Freqkuencies," *IEEE Trans. on Speech and Audio Processing, Vol.3, No.3*: 157-167, 1995.

5. Lemma, A.N., Kleijn, W.B., and Deprettere, E.F., "LPC Quantization Using Wavelet Based Temporal Decomposition of the LSF," *EUROSPEECH'97*: 1259-1262, 1997.

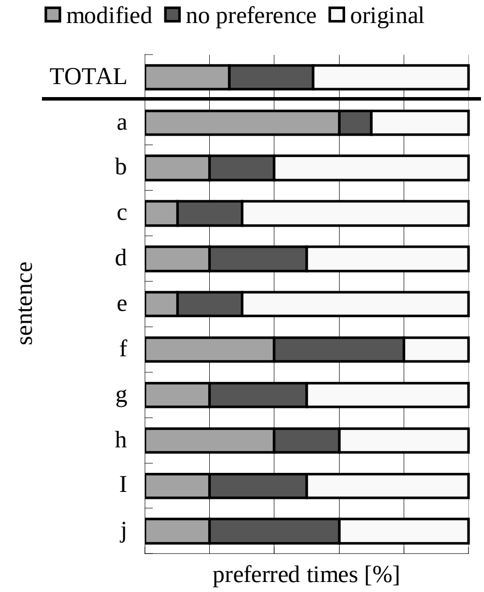

Figure 2: Result of pair comparison tests.

6. Papamichalis, P.E., *Practical Approaches To Speech Coding*, Prentice-Hall, Inc., New Jersey, 1987.

7. Juang, B.-H., *Fundamentals Of Speech Recognition*, Prentice-Hall International, Inc., New Jersey, 1993.