Selective-LPC based Representation of STRAIGHT Spectrum and Its Applications in Spectral Smoothing*

Heng Kang, Wenju Liu

National Laboratory of Pattern Recognition,
Institute of Automation Chinese Academy of Sciences, Beijing
{hkang, lwj}@nlpr.ia.ac.cn

ABSTRACT

In this paper we propose a new method to represent STRAIGHT spectrum. The new method provides STRAIGHT spectral parameters with the capability of interpolation and quantization, which is needed for most speech manipulation, especially for spectral smoothing. The proposed method estimates 2-band selective-LPC whose spectral envelope fits the given STRAIGHT spectrum. With the interpolation properties of LSP, the estimated selective-LPC could be converted to LSP and then simply interpolated. We apply this representation in our spectral smoothing experiments and the results show that this method can get smooth spectral envelope over the segment boundaries. Listening tests prove that this algorithm effectively smooth speech boundaries with little quality degradation.

Index Terms: speech synthesis, STRAIGHT, selective-LPC

1. INTRODUCTION

Today more and more speech systems adopt STRAIGHT[1,2,3] as its basic speech signal representation technique, based on which they can do much manipulation to the original speech signals. Its applications in prosody modification, voice conversion and parameter-based TTS are promising. However there are limitations on the flexibility of the output 3-dimensional time-frequency spectra. As a spectral parameter STRAIGHT spectrum itself has poor capability of interpolation because the spectral formants will rise and fall rather than move smoothly in frequency if directly interpolated. This restricts its application within narrow areas such as duration and pitch modification. In order to provide it with the capability of interpolation, it is a good way to convert it into another intermediate representation before doing modification or processing.

To overcome the limitations some people use mel-cepstral coefficients to which the STRAIGHT spectrum analyzed is converted[5]. It is true that mel-cepstrum is a proper method in their applications but mel-cepstrum's interpolation capability is very limited, especially at high frequency band.

In this paper we propose a new method based on selective-LPC. The proposed method estimates 2-band selective-LPC[6] whose spectral envelope fits the given STRAIGHT spectrum. With the interpolation properties of LSP, the estimated selective-LPC could be converted to LSP and then simply interpolated. This representation not only has little distortion from the original spectrum but also provides for the parameters the capability of interpolation and quantization.

2. INTRODUCTION TO STRAIGHT

The SRAIGHT (Speech Transformation and Representation using Adaptive Interpolation and weighted spectrum) is a channel vocoder technology proposed by Hideki Kawahara. The purpose is to remove spectral interference structure caused by signal periodicity. This method is mainly a pitch-adaptive spectral envelope extractor which can get a smooth time-frequency representation of the spectral envelope of speech signal. It consists of three main components, i.e., F0 extraction, spectral and aperiodic analysis and speech synthesis.

The STRAIGHT first automatically extracts F0 with fixed-point analysis, and using the extracted F0 it applys a pitch-adaptive time-frequency smoothing spectral analysis to remove signal periodicity. An aperiodicity measure on the frequency domain is also extracted. In synthesis stage STRAIGHT uses the weighted sum of a pulse train with phase manipulation and Gaussian noise as the mixed excitation. Then it combines the excitation and smoothed spectrum in frequency domain to synthesize speech waveform. Figure 1 is the schematic diagram of STRAIGHT analysis and synthesis.

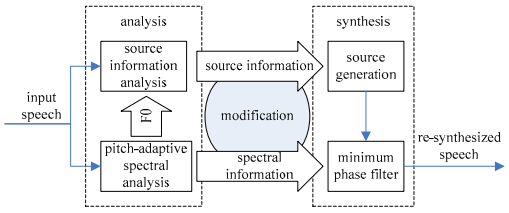

Figure 1: Schematic diagram of STRAIGHT analysis-modification-synthesis system

3. SELECTIVE-LPC BASED REPRESENTATION

Selective-LPC is proposed to use LPC to fit a selected portion of a given spectrum. In our method we divide the full spectrum frequency into 2 bands, and use selective-LPC to fit each band independently. We configure different LPC order

*This work is supported in part by the China National Nature Science Foundation (No. 60172055, No. 60121302), the Beijing Nature Science Foundation (No.4042025) and the National Fundamental Research Program (973) of China (2004CB318105).

on different frequency band in order to control the fitting accuracy and to reduce total LPC order. This method offers some important advantages as the intermediate representation of STRAIGHT spectrum: 1) Since LPC coefficients are equivalent to LSP, it owns all the virtues that LSP has. 2) Because selective-LPC has the ability to emphasize on any sub-band of the spectrum, we can control the LPC order to fit the spectrum of low frequency band more accurately.

In this section we first introduce how to use LPC to fit a given spectrum, then we extend to 2-band selective-LPC method. Finally we will discuss the interpolation property of this representation.

3.1 ESTIMATION LPC FROM STRAIGHT SPECTRUM

Let us assume that we are given a STRAIGHT power spectrum $P(\omega)$ and $0 \le \omega \le 2\pi$. We want to fit $P(\omega)$ by a LPC spectrum $\hat{P}(\omega)$, which could be calculated from the LPC coefficients $\{a_k, 1 \le k \le p\}$ like this

$$\hat{P}(z) = \left| \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} \right|^2 \tag{1}$$

where p is the order of the LPC inverse filter.

Given the power spectrum $P(\omega)$ and the order p, the problem is to determine the parameters $\{a_k\}$ and G.

Define the error between $P(\omega)$ and $\hat{P}(\omega)$:

$$E = \frac{G^2}{2\pi} \int_{-\pi}^{\pi} \frac{P(\omega)}{\hat{P}(\omega)} d\omega \tag{2}$$

$\{a_k\}$ can be determined by minimizing E with respect to each a_k. This is

$$\frac{\partial E}{\partial a_i} = 2(R_i + \sum_{k=1}^{p} a_i R_{|i-k|}) = 0, 1 \le i \le p \tag{3}$$

where

$$R_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega) \cos(k\omega) d\omega \tag{4}$$

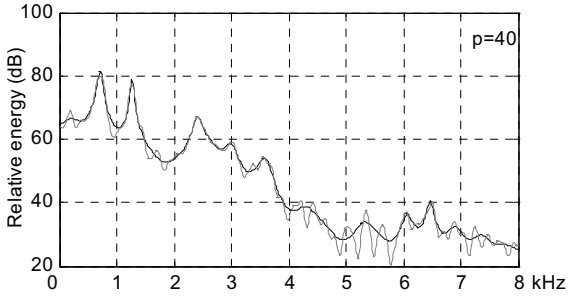

Figure 2: STRAIGHT smoothed spectrum (the thin line) and its LPC spectral envelope (the thick line)

is the autocorrelation corresponding to $P(\omega)$.

From formula 3 we have

$$\sum_{k=1}^{p} a_k R_{|i-k|} = -R_i, \quad 1 \le i \le p \tag{5}$$

Therefore $\{a_k\}$ can be solved from the linear equations by Levinson-Durbin algorithm.

Figure 2 illustrates the STRAIGHT smoothed spectrum and the corresponded LPC spectral envelope.

3.2 TWO-BAND SELECTIVE-LPC

LPC parameters determining method described in last section is equivalent to the autocorrelation method in time domain. As we know that linear prediction tends to fit the spectrum peaks more than spectrum valleys because of the error function, which causes the LPC spectrum envelope not close enough to the original STRAIGHT spectrum. Especially in high frequency, the algorithm wastes much order to fit the detailed curves, but as we all know the spectrum in low frequency carrys most of the information of speech.

Can we estimate the parameters at low/high frequency independently? If that we can use high order for the low frequency band to reserve the detailed spectral information, and in high frequency band we use fewer order.

We can divide frequency axis into 2 sub-bands: low frequency band Q_{low} ($0 \le \omega < \omega_s$) and high frequency band Q_{high} ($\omega_s \le \omega < \pi$). Then we use selective-LPC method to fit each band independently.

In low frequency band Q_{low}, we first define a new spectrum

$$P'_{low}(\omega) = \begin{cases} P(\omega), 0 \le \omega < \omega_s \\ undefined, otherwise \end{cases} \tag{6}$$

Transform ω into ω' such that

$$\omega' = \frac{\omega}{\omega_s} \pi \tag{7}$$

This makes P'_{low} spans the whole frequency axis. Estimate LPC coefficients $\{a^k_{low}\}$ by the algorithm described in last section. The spectral envelope of the estimated $\{a^k_{low}\}$ will fit the low sub-band of the spectrum $P(\omega)$.

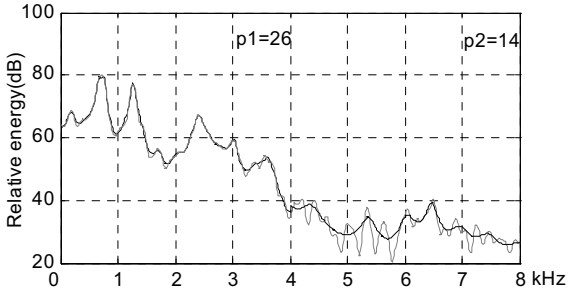

Figure 3: STRAIGHT smoothed spectrum (the thin line) and its 2-band selective-LPC spectral envelope (the thick line)

Similarly define a new spectrum from high sub-band of the original spectrum and use the transforming equation

$$\omega' = \frac{\omega - \omega_s}{\pi - \omega_s}\pi \tag{8}$$

to make P'_{high} spans the whole frequency axis, we can estimate LPC coefficients $\{a^k_{high}\}$.

Now we have two sets of LPC coefficients $\{a^k_{low}\}$ and $\{a^k_{high}\}$ which fit low/high frequency bands respectively.

Comparing figure 3 to figure 2, we can see that the spectral envelope of the 2-band selective-LPC is closer to the STRAIGHT at the low sub-band.

3.3 INTERPOLATION PROPERTY

The interpolation property is a very import feature needed in speech processing systems such as speech coding and spectral smoothing[7]. By "interpolation properties" we mean that the interpolated parameters yield smooth evolution of spectra, and small variation of the parameters yields small variation of the spectral envelope.

LPC itself is not good at interpolation because it is not guaranteed to yield stable filters after interpolation. But LSP (line spectral pair)[8,9] as an alternative LPC spectral representation has some good properties such that it always yields stable filters after interpolation. And studies show that LSP interpolation gives better results than most other representations when used for interpolation in coding[10,11].

Therefore to interpolate the estimated two-band LPC, we can convert them into LSP and simply interpolate the LSP in each corresponding frequency band, respectively.

4. APPLICATION IN SPECTRAL SMOOTHING

4.1 SPECTRAL SMOOTHING

In concatenative speech synthesis system, spectral smoothing often need to be applied when the speech segments have different spectral or formant structures. If the segments are concatenated directly, noise or discontinuity can be perceived. In order to eliminate this discontinuity, effective spectral smoothing should be applied to the segment boundaries.

Some speech synthesis systems use LP smoothing, which first decomposes speech signal of segments into LPC parameters (represent for spectral envelope) and residual (represent for excitation), and then interpolate LPC and residual independently. But as we know that LPC spectral envelope and residual are not separated completely, in other words, they are correlated in some ways. Therefore interpolating them independently causes evident degradation in speech quality [12].

4.2 NEW PROPOSED METHOD

STRAIGHT is proposed to eliminate spectral interference structure caused by signal periodicity. Comparing to linear prediction, it completely decomposes speech signal into two incorrelate components: spectral information and source information. Therefore we could interpolate the spectrum and source independently and they will not influence each other in re-synthesis.

As figure 4 illustrates, our method adopts the selective-LPC based STRAIGHT as the intermediate of the smoothing procedure. First the speech frames are analyzed into sources and spectra. The sources could be interpolate directly, and the spectra are converted into 2-band selective-LPC parameters and then are interpolated by LSP. After that new sources and spectra are combined to re-synthesize the smoothing frames. With the new representation's capability of interpolation and the perfect speech quality, the spectral smoothing is expected to get good results.

4.3 EXPERIMENTAL RESULTS

In our experiments we test our new method on a speech database uttered by a male speaker. We partitioned all utterances into speech segments according to the boundaries labels. Then 60,000 segments are selected to form 30,000 segments pairs. Statistical ANBM scores[13] of all 30,000 pairs can be seen in figure 5.

We perform our smoothing algorithm to boundaries of all the segments pairs on duration of 4 pitch periods, and compute the ANBM scores between each pair, before and after

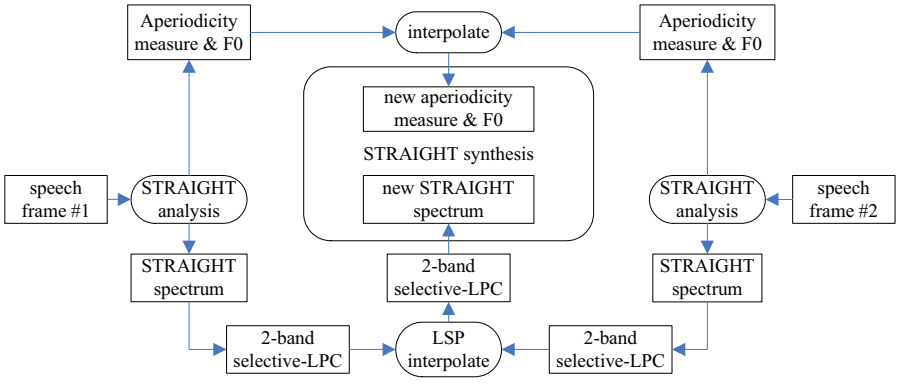

Figure 4: Diagram of spectral smoothing method based on selective-LPC represention of STRAIGHT spectrum

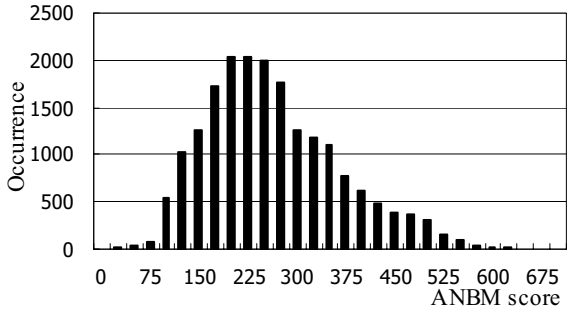

Figure 5: ANBM distribution of all segments pairs

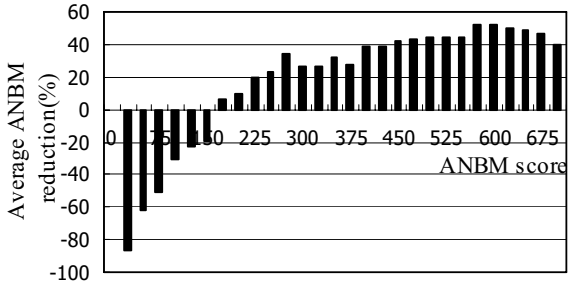

Figure 6: Percentage of ANBM scores reduction of all segments pairs after smoothing

the smoothing. Statistical result of the change in scores could be acquired. We calculate the overall average ANBM scores and find it drops 27.78% after smoothing. Figure 6 shows the percentage of ANBM scores reduction of all pairs after smoothing. Note when the ANBM score is too small (below 200) this method behaves not well. The reason is that the distortion introduced by the modification counteracts the smoothing effect.

In addition we make a subjective listening test to investigate the speech quality of the smoothed segments. An AB test is performed on 50 sentences with 8 listeners. The listeners are requested to decide whether the smoothed segments in sentence A or B is better in terms of speech quality. The results showing in figure 7 indicate that our proposed method gives better speech quality than both direct LSP smoothing method and full-band LPC fitting method.

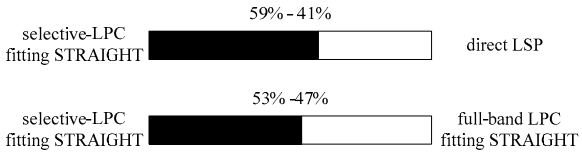

Figure 7: Results for listening test comparison

5. CONCLUSION

A new method to represent STRAIGHT spectrum is proposed based on selective-LPC, for the purpose of providing STRAIGHT spectral parameters with the capability of interpolation. We apply this representation in our spectral smoothing experiments and the results show that this method can get smooth spectral envelope over the segment boundaries. Listening tests prove that this algorithm effectively smooth speech boundaries with little quality degradation.

6. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Hideki Kawahara of Wakayama University for permission to use STRAIGHT vocoding method [14].

7. REFERENCES

[1] Kawahara, H. Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited, *Proceedings of ICASSP'97*, Vol. 2, pp.1303-1306.

[2] Kawahara, H., Masuda-Katsuse, I., Cheveigné, A.D., Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in soulds, *Speech Communication* 27(3-4), pp.187-207.

[3] Kawahara, H., Katayose, H., Cheveigné, A.D., Patterson, R. Fixed points analysis of frequency to instantaneous frequency mapping for accurate estimation of f0 and periodicity. *Proceedings of EUROSPEECH'99*, Vol. 6, pp. 2781–2784.

[4] Tokuda, K.; Matsumura, H.; Kobayashi, T.; Imai, S., Speech coding based on adaptive mel-cepstral analysis, *Proceedings of ICASSP'94*, Vol. 1, pp. 197-200.

[5] Heiga Zen, Tomoki Toda, An overview of Nitech HMM-based speech synthesis system for Blizzard Challenge 2005, *Proceedings of EUROSPEECH 2005*, pp.93-96.

[6] John Makhoul, Spectral linear prediction: properties and applications, *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 23, No. 3, June 1975.

[7] Shadle V.H., Atal B.S., Speech synthesis by linear interpolation of spectral parameters between dyad boundaries, *J. Acoust. Soc. Am.*, Vol. 66, pp.1325-1332, 1979.

[8] Itakura, F., Line spectrum representation of linear predictive coefficients of speech signals, *J. Acoust. Soc. Am.*, Vol. 57, pp.535, 1975.

[9] Soong, F., Juang, B., Line spectrum pair (LSP) and speech data compression, *Proceedings of ICASSP'84*, Vol. 9, pp.37-40.

[10] Paliwal, K. K., Interpolation properties of linear prediction parametric representations, *Proceedings of EuroSpeech'95*, Vol. 2, pp.1029-1032.

[11] Paliwal, K.K., Kleijn, W.B.,Quantization of LPC parameters, *Speech Coding and Synthesis*, Elsevier, Amsterdam, 1995, pp. 433-466.

[12] Chappell D.T., Hansen J., A comparison of spectral smoothing methods for segment concatenation based speech synthesis. *Speech Communication*, 2002, 36:343–374.

[13] Chappell D.T., Hansen J., An auditory-based measure for improved phone segment segment concatenation, ICASSP'97, Vol. 3, pp.1639-1642

[14] *http:// www.wakayama–u.ac.jp / kawahara / STRAIGHT demo/* (Demonstration of STRAIGHT applications).