SPEECH PARAMETER GENERATION CONSIDERING LSP ORDERING PROPERTY FOR HMM-BASED SPEECH SYNTHESIS

Shijun Qian[1,2]*, Huanliang Wang*[2]*, Wenjiang Pei*[1]*, Ping Zou*[2] *and Kai Wang*[1]

[1]School of Information Science and Engineering, Southeast University, Nanjing, China
[2]AI Speech Co., Ltd, Suzhou, China
sjqian@seu.edu.cn, huanliang.wang@aispeech.com

ABSTRACT

LSP has many advantages for speech representation, especially correlates well to spectrum formants as long as the LSP parameters are strictly ordered and bounded. This ordering property cannot be guaranteed during HMM-based speech synthesis when LSP is adopted as the spectrum feature, because diagonal covariance is utilized and correlation between LSP dimensions is ignored, with the result that unstable issue will be caused in synthesized speech. In this paper, we modify the parameter generation criterion to preserve ordering property of generated LSPs, by considering not only the likelihoods for HMM and GV maximized in conventional method but also a mis-orderings penalty. Experimental results show that the proposed method can alleviate the mis-orderings significantly and achieve high quality synthesizing performance when the penalty weight is selected appropriately.

Index Terms— Speech synthesis, hidden Markov model, parameter generation, line spectral pair, ordering property

1. INTRODUCTION

Hidden Markov model (HMM) based speech synthesis has been widely used in recent years [1]. In this method, the frequency spectrum, pitch and duration of speech are modeled simultaneously within a unified framework during the training procedure. At synthesis stage, speech waveforms are reconstructed using acoustic features predicted from trained HMMs by maximum likelihood parameter generation (MLPG) algorithm [2].

Line Spectral Pair (LSP) [3] has been a popular spectrum feature in many speech synthesis systems [4], since it has many advantages for speech representation, especially the distance of adjacent LSPs closely relates with neighboring spectrum formants [5]. According to the definition of LSP [3], every LSP dimension should be exactly ordered and bounded, which means the value of higher dimension should be always larger than that of lower dimension, and value of every dimension should be in the range between zero and π. This ordering property is an important characteristic of LSP feature, and if it is guaranteed, minimum phase property of the reconstructed all-pole filter will be easily preserved [6]; otherwise, unstable issue will emerge and naturalness of synthesized speech will degrade.

However, the correlations among different dimensions of spectrum parameters are usually ignored in the conventional modeling method of HMM-based speech synthesis, due to the usage of diagonal covariance to characterize spectrum model. Thus, the crucial relationship between LSP dimensions, namely ordering property, fails to be considered when LSP is adopted as the spectrum feature. Consequently, disordered LSPs will be unavoidably generated during synthesizing for the effect of static and dynamic feature variance if no further constrain is considered. Besides, the mis-ordering problem becomes much worse when the global variance (GV) [7] likelihood is also considered, for the reason that each dimension of parameter is also modeled independently for GV which usually has much larger distribution intervals. In [8], some methods have been proposed to preserve the the ordering property of generated LSPs for minimum generation error (MGE) training, by introducing mis-ordering related distance measurements into model training criterion. However, since MGE is not a fundamental criterion for model training, the method proposed in [8] cannot deal with the mis-ordering issue in general HMM-based synthesis using maximum likelihood criterion; on the other hand, advancing the training criterion is also not a direct way to control mis-orderings and cannot solve the problem when GV is considered.

In this paper, an improved parameter generation method based on MLPG algorithm is proposed to achieve more direct control of the mis-ordering of generated LSPs. The main strategy is to introduce a penalty function for mis-orderings in conventional parameter generation algorithm. The generated LSP parameter sequence not only maximizes HMM likelihood as well as likelihood of GV if considered, but also minimizes the mis-ordering penalty. Experimental results prove the effectiveness of the proposed method.

The rest of this paper is organized as follows. In section 2, the conventional parameter generation algorithm as well as LSP properties are reviewed. Section 3 describes the proposed LSP generation method considering ordering property

in detail. The experimental results are shown in Section 4. Finally, section 5 concludes this paper.

2. RELATED TECHNIQUES

2.1. Maximum likelihood parameter generation

In HMM-based speech synthesis, the MLPG algorithm has typically been adopted to predict speech parameters. Let assume a static feature vector $\boldsymbol{c}_t = [c_t(1), c_t(2), ..., c_t(D)]^\top$ at frame t. For given HMM λ and determined state sequence Q, the algorithm is to determine speech parameter vector sequence $\boldsymbol{o} = [\boldsymbol{o}_1^\top, \boldsymbol{o}_2^\top, ..., \boldsymbol{o}_T^\top]^\top$ to maximize $P(\boldsymbol{o}|\lambda, Q)$. In order to keep generated parameters smooth between frames, the dynamic features including velocity and acceleration components $\Delta^{(n)}\boldsymbol{c}_t (n = 1, 2)$ are usually incorporated, that is, $\boldsymbol{o}_t = [\boldsymbol{c}_t^\top, \Delta\boldsymbol{c}_t^\top, \Delta^2\boldsymbol{c}_t^\top]^\top$, and also can be formulated as

$$\boldsymbol{o} = \boldsymbol{W}\boldsymbol{c}, \tag{1}$$

where $\boldsymbol{c} = [\boldsymbol{c}_1^\top, \boldsymbol{c}_2^\top, ..., \boldsymbol{c}_T^\top]^\top$ and $\boldsymbol{W}$ is a $3DT$-by-DT matrix determined by velocity and acceleration components.

By setting $\partial P(\boldsymbol{o}|\lambda, Q)/\partial \boldsymbol{c} = 0$, we can obtain the static feature vector as [2]

$$\boldsymbol{c} = \left(\boldsymbol{W}^\top \hat{\boldsymbol{U}}^{-1} \boldsymbol{W}\right)^{-1} \boldsymbol{W}^\top \hat{\boldsymbol{U}}^{-1} \hat{\boldsymbol{\mu}}, \tag{2}$$

where $\hat{\boldsymbol{U}} = \text{diag}[\boldsymbol{U}_{q_1,i_1}^{-1}, \boldsymbol{U}_{q_2,i_2}^{-1}, ..., \boldsymbol{U}_{q_T,i_T}^{-1}]$, $\hat{\boldsymbol{\mu}} = [\boldsymbol{\mu}_{q_1,i_1}^\top, \boldsymbol{\mu}_{q_2,i_2}^\top, ..., \boldsymbol{\mu}_{q_T,i_T}^\top]^\top$, with $\boldsymbol{U}_{q_t,i_t}$ and $\boldsymbol{\mu}_{q_t,i_t}$ are the covariance matrix and mean vector respectively, associated with i_t-th mixture of state q_t.

2.2. MLPG algorithm considering GV

To deal with the over-smoothing effect of speech parameter sequences generated only by maximizing HMM likelihood, parameter generation method considering global variance are popularly utilized. The GV over T frames static feature $\boldsymbol{v}(\boldsymbol{c}) = [v(1), ..., v(d), ..., v(D)]^\top$ is calculated by

$$v(d) = \frac{1}{T}\sum_{t=1}^{T}\left(c_t(d) - \frac{1}{T}\sum_{\tau=1}^{T} c_\tau(d)\right). \tag{3}$$

At training stage, the GVs are calculated over training sentences and used to train the GV model λ_v with single Gaussian distribution. During parameter generation, the optimal speech parameter sequence $\boldsymbol{c}$ is predicted by maximizing not only the HMM likelihood but also the GV likelihood, that is the following log-scaled likelihood,

$$L = \log[P(\boldsymbol{W}\boldsymbol{c}|Q, \lambda)P(\boldsymbol{v}(\boldsymbol{c})|\lambda_v)^\rho], \tag{4}$$

with the ρ controlling the balance of the two likelihoods. To determine optimal parameters, the gradient methods are used to update $\boldsymbol{c}$ iteratively [7].

2.3. Ordering property of line spectral pairs

This paper focuses on LSP as the spectrum feature, which is derived from LPC (linear prediction coefficient) as an alternative LPC spectral representations. For a given M-th order LPC analysis polynomial $A(z)$, a pair of $(M+1)$-th order LSP symmetric and anti-symmetric polynomials can be derived, which are

$$\begin{aligned} P(z) &= A(z) + z^{-(M+1)}A(z^{-1}), \\ Q(z) &= A(z) - z^{-(M+1)}A(z^{-1}). \end{aligned} \tag{5}$$

The LSP coefficients are defined as those values of frequency ω such that $\{\omega \in (0, \pi) | P(e^{j\omega}) = 0 \text{ or } Q(e^{j\omega}) = 0\}$.

There is an important property of LSP: all the corresponding zeros of the symmetric and anti-symmetric LSP polynomials are interlaced on the unit circle, in the sense that

$$0 < \omega_1 < \omega_2 < \cdots < \omega_M < \pi, \tag{6}$$

where ω_{2k-1} are the roots of $P(e^{j\omega})$ while ω_{2k} are the zeros of $Q(e^{j\omega})$ with $k = 1, ..., \lfloor M/2 \rfloor$, and the reconstructed LPC all-pole filter preserves its minimum phase property if this ordering property is kept intact.

Due to limited training data and consideration of computation complexity, conventional HMM-based speech synthesis methods adopt models with diagonal covariance to characterize spectrum features, which ignore the cross-dimensional correlation of LSP. Since each dimension of LSP has overlapped distribution with neighboring dimensions [6], the ordering property of generated LSPs cannot be guaranteed. Besides, incorporating dynamic features also introduce dynamic variation into generated LSPs, and the GV likelihood adjusts each dimension independently with larger intervals if considered, both of which inevitably damage the ordering property and result in unstable synthesis filters. Thus it is important to effectively control the LSP mis-ordering issue during HMM-based speech synthesis.

3. LSP GENERATION METHOD CONSIDERING ORDERING PROPERTY

Some methods has been proposed to deal with the mis-ordering issue by introducing mis-ordering related distance measurements into MGE model training criterion [8]. Considering that mis-ordering issue is only observed in generated LSPs and usually the GV likelihood is also considered, it is more reasonable to make modifications in parameter generation stage rather than the model training procedure. Here, we propose an improved parameter generation method to alleviate the mis-ordering more directly based on maximum likelihood criterion. In order to preserve the ordering property cooperating with the MLPG algorithm, a cost function for mis-orderings could be well designed, working as a penalty term like the GV likelihood.

3.1. Mis-ordering penalty of LSP

In order to control the mis-ordering issue effectively, appropriate mis-ordering penalty (MOP) must (1) be a monotonically increasing function respect to the number of mis-orderings, (2) increase faster than the mis-orderings, (3) response 1 to keep HMM and GV likelihood intact when no mis-ordering generated, and (4) work efficiently with the log-scaled likelihood maximization. Based on these conditions, the cost function as a mis-ordering penalty is defined to be the exponential function of mis-ordering numbers, that is,

$$F_P(\boldsymbol{c}) = \exp\left\{N_{moc}(\boldsymbol{c})\right\}, \tag{7}$$

where $N_{moc}(\boldsymbol{c})$ is the mis-ordering counting (MOC) function defined later.

We exploit the summation of logistic function as a common sigmoid curve to be our counting function like [8],

$$N_{moc}(\boldsymbol{c}) = \sum_{t=1}^{T}\sum_{d=1}^{D+1} 1/\left(1+\exp[\kappa(\omega_{d,t}-\omega_{d,t-1}-\epsilon_d)]\right), \tag{8}$$

where $\omega_{d,t} = c_t(d)$ stands for d-th dimension of LSP at frame t, with $\omega_{0,t} = 0$ and $\omega_{D+1,t} = \pi$ fixed, κ is coefficient to control the logistic curve shape and ϵ_d works as the differential threshold for d-th dimension.

The counting function responses the number of mis-orderings when negative or too small differential LSPs are generated and responses 0 on the contrary. Note that the MOC function and consequently the MOP are functions of $\boldsymbol{c}$.

3.2. LSP generation method considering MOP

The LSPs preserving ordering property are generated by not only maximizing the conventional likelihood but also minimizing the mis-ordering penalty defined above. That is,

$$L = \log[P(\boldsymbol{W}\boldsymbol{c}|Q,\lambda)F_P^{-\gamma}(\boldsymbol{c})], \tag{9}$$

and if GV likelihood is also considered,

$$L = \log[P(\boldsymbol{W}\boldsymbol{c}|Q,\lambda)P(\boldsymbol{v}(\boldsymbol{c})|\lambda_v)^{\rho}F_P^{-\gamma}(\boldsymbol{c})], \tag{10}$$

where ρ and γ are the GV weight and MOP weight respectively, and the LSP parameters are determined by maximizing the proposed L,

$$\boldsymbol{c} = \underset{\boldsymbol{c}}{\operatorname{argmax}}\, L. \tag{11}$$

In this paper, we focus on Equation (10) and keep ρ to be $3T$ as in [7]. Let L_{hmm} and L_{gv} denote the HMM and GV log-likelihood respectively, Equation (10) can be further expanded as

$$L = L_{hmm} + \rho L_{gv} - \gamma N_{moc}. \tag{12}$$

To determine the optimal parameter vector sequence, $\boldsymbol{c}$ is updated iteratively with the gradient method,

$$\boldsymbol{c}^{(i+1)}\text{-th} = \boldsymbol{c}^{(i)}\text{-th} + \alpha\cdot\delta\boldsymbol{c}^{(i)}\text{-th}, \tag{13}$$

where α is a step size parameter.

We investigate the Newton-Raphson method for our new objective Equation (10), since it usually converges quickly when the initial trajectory is close to the optimal one, with the step $\delta\boldsymbol{c}^{(i)}$-th written as

$$\delta\boldsymbol{c} = -[\boldsymbol{H}(L)]^{-1}\frac{\partial L}{\partial \boldsymbol{c}}, \tag{14}$$

where the first order derivative is formulized by

$$\frac{\partial L}{\partial \boldsymbol{c}} = \frac{\partial(L_{hmm}+\rho L_{gv})}{\partial \boldsymbol{c}} - \gamma\frac{\partial N_{moc}}{\partial \boldsymbol{c}}, \tag{15}$$

and the Hessian matrix is

$$\boldsymbol{H}(L) = \frac{\partial^2(L_{hmm}+\rho L_{gv})}{\partial \boldsymbol{c}\partial \boldsymbol{c}^{\top}} - \gamma\frac{\partial^2 N_{moc}}{\partial \boldsymbol{c}\partial \boldsymbol{c}^{\top}}. \tag{16}$$

Details of derivation for $\partial(L_{hmm}+\rho L_{gv})/\partial\boldsymbol{c}$ in Equation (15) and $\partial^2(L_{hmm}+\rho L_{gv})/(\partial\boldsymbol{c}\partial\boldsymbol{c}^{\top})$ in Equation (16) can be found in [7]. We need to calculate the new term $\partial N_{moc}/\partial\boldsymbol{c}$ additionally by

$$\partial N_{moc}(\boldsymbol{c})/\partial\boldsymbol{c} = [\boldsymbol{\theta}_1^{\top}, \boldsymbol{\theta}_2^{\top}, ..., \boldsymbol{\theta}_T^{\top}]^{\top}, \tag{17}$$

$$\boldsymbol{\theta}_t = [\theta_t(1), \theta_t(2), ..., \theta_t(D)]^{\top}, \tag{18}$$

$$\theta_t(d) = -\frac{\kappa E_{d,t}}{(1+E_{d,t})^2} + \frac{\kappa E_{d+1,t}}{(1+E_{d+1,t})^2}, \tag{19}$$

with $E_{d,t} = \exp[\kappa(\omega_{d,t}-\omega_{d,t-1}-\epsilon_d)]$, and the additional second order derivative can be expanded as

$$\frac{\partial^2 N_{moc}(\boldsymbol{c})}{\partial\boldsymbol{c}\partial\boldsymbol{c}^{\top}} = \text{diag}[\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, ..., \boldsymbol{\Phi}_T], \tag{20}$$

in which each $\boldsymbol{\Phi}_t = \{\theta'(i,j)\}_{D\times D}$ is a symmetric tridiagonal matrix with

$$\theta'(d,d) = \kappa^2[\Gamma_{d,t} + \Gamma_{d+1,t}], \tag{21}$$

$$\theta'(d,d-1) = \theta'(d-1,d) = -\kappa^2\Gamma_{d,t}, \tag{22}$$

$$\Gamma_{d,t} = -\frac{E_{d,t}}{(1+E_{d,t})^2} + \frac{2E_{d,t}^2}{(1+E_{d,t})^3}. \tag{23}$$

Note that $\partial^2(L_{hmm}+\rho L_{gv})/(\partial\boldsymbol{c}\partial\boldsymbol{c}^{\top})$ is approximated by a diagonal matrix in [7] to keep positive definite. Due to the subtraction between the second-order derivatives here, the positive definite property of $\boldsymbol{H}(L)$ cannot be guaranteed, which suggests the Newton-Raphson method hard to be directly employed for Equation (10). We find that a simplification of the gradient method can achieve good convergence performance, when starting from the following trajectory,

$$c_t'(d) = \sqrt{\mu_v(d)/v(d)}(c_t(d)-\bar{c}(d)) + \bar{c}(d), \tag{24}$$

where $\mu_v(d)$ is the GV mean obtained from GV model, while $\bar{c}(d)$ and $v(d)$ is the mean and variance calculated over results of Equation (2) respectively for d-th dimension.

We implement the simplified method by keeping Hessian matrix as that in [7], and using the new first order derivative with MOP components. This is reasonable since the positive defined Hessian matrix just work as a fine tuning factor to make iteration smarter, and we notice that the mis-ordering penalty in Equation (10) usually decrease significantly during the first few iterations if mis-ordering happens, and then the problem develops into the GV iteration.

4. EXPERIMENTS

4.1. Experimental conditions

The training data consists of 1132 phonetically balanced sentences of an US English female speaker (slt) from CMU ARCTIC databases, with the speech waveforms sampled at the rate of 16kHz. The acoustic features, including F0, aperiodicity measure and spectral parameters which were 40-order LSP and an extra gain dimension, were extracted by STRAIGHT [9] analysis with a 5ms frame shift. Feature vectors consisted of log-scaled F0 vector, LSP vector and aperiodicity measures vector, each of which also included velocity and acceleration coefficients. A 5-state left-to-right with no skip multi-space probability distribution HMM (MSD-HMM) structure was adopted and GV model were trained for F0 and spectrum vectors.

In synthesis part, the log-scaled F0, log-scaled predictive gain and aperiodicity measures were generated by the conventional MLPG algorithm. Method proposed in this paper was used to generate 40-order LSP parameters which then were converted into LPCs. ϵ_d was simply fixed to be 0.0 to focus on the mis-ordering issue, and κ was set to be 500.0 to ensure ideal logistic curve shape. The speech was synthesized using all-pole filter with the generated acoustic features. 593 sentences from BL2009 was used as our test set.

We investigated the number of both mis-ordering frames and sentences of generated LSPs, considering number of sentences directly correlated with subjective evaluation. Averaging log spectrum distortion (LSD) was calculated as objective measurement over not only normal sentences but also the overall training set. Note that when the MOP weight was set to be 0, the proposed method degenerated into the conventional algorithm, which we employed as baseline. Subjective comparison was also conducted, which was by 8 Chinese postgraduate students over two groups of sentences from test set, i.e. sentences with mis-orderings and with no mis-orderings under conventional algorithm respectively, in each of which 20 sentences were selected randomly.

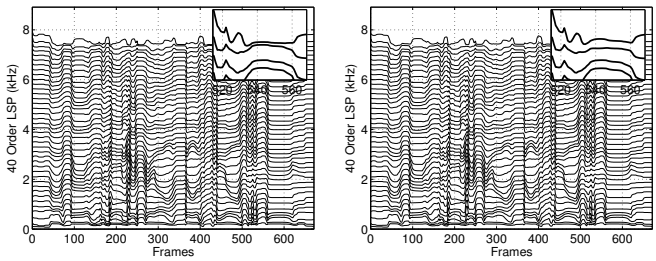
4.2. Experimental results

Different γ were tested to control the contribution of MOP. Objective results including number of mis-orderings over training set and over test set, LSD over normal training sentences and over all training set, are shown in Table 1.

Table 1. Different γ for method with MOP

γ	Number of Mis-orderings				LSD (dB)	
	Training Set		Test Set		Over	Over
	Sens	Frms	Sens	Frms	Normal	All
Baseline	61	963	33	384	7.2243	7.2333
1	43	573	29	293	7.2243	7.2330
10	18	114	17	69	7.2238	7.2320
50	8	21	12	46	7.2230	7.2313
100	3	9	12	43	7.2226	7.2308
200	1	4	12	40	7.2220	7.2302
1000	1	4	10	36	7.2200	7.2285
10000	1	2	9	29	7.2180	7.2267

We can see the number of mis-orderings is reduced on both training set and test set, when the MOP weight becomes larger, that is, the contribution of MOP in MLPG becomes more important compared to HMM and GV likelihood. Notice that the proposed method is more sensitive to γ when it is relatively small. This is because most of the mis-orderings are not serious, that is to say, the negative differential LSPs usually have small absolute values, which need not much MOP contribution to solve. It is also suggested in Table 1 that $\gamma = 200$ be enough to eliminate almost all the mis-orderings on training set except for one particular sentence which also can be removed when γ is large enough to achieve 10^5.



(a) $\gamma = 0, \text{LSD} = 8.2255dB$; (b) $\gamma = 10^4, \text{LSD} = 8.2212dB$.

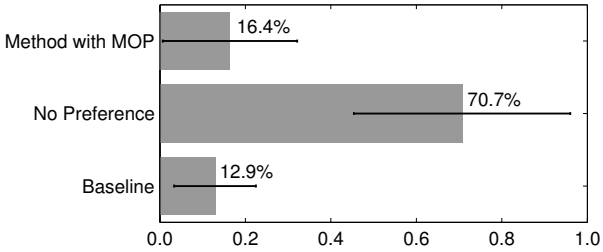
Fig. 1. Example of LSP digram for different γ.

On the other hand, larger MOP weight also brings better LSD over frames of both normal sentences and the overall training set according to Table 1. Improvement of LSD on normal sentences seems harder to understand than that on abnormal sentences, since the MOP term weakens the HMM and GV likelihood. This can be explained as that the MOP term also prevents adjacent LSPs to be very close, since our sigmoid curve is not that ideal to be a step function, and it responses non-zero values for too small differential LSPs. According to [5], adjacent LSP dimensions can also not too close, for that will cause very large response in spectral envelope, i.e. extremely sharper formant. Figure 1 gives a typical example, where too close adjacent LSPs were separated by increasing γ from 0 to 10^4, and LSD was improved.

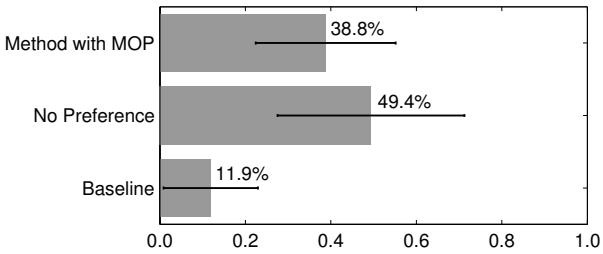
Table 2. LSD (dB) for very large γ

lgγ	LSD over normal	LSD over all
Baseline	7.2243	7.2333
4	7.2180	7.2267
5	7.2182	7.2274
6	7.2221	7.2216
7	7.2310	7.2403

Actually, very large γ will of course damage the synthesized spectrum, shown in Table 2. When γ is increased to 10^7, the LSDs become worse than that of baseline. This is also understandable based on our observation that the value of log-scaled HMM likelihood usually 10^6 times that of MOP.



(a) Results over normal sentences.



(b) Results over abnormal sentences.

Fig. 2. Subjective preference scores between new method and baseline with 95% condence interval.

Figure 2 illustrates subjective preference scores between the method with MOP (γ was 10^4) and the conventional algorithm. There was no obvious difference between the performance of baseline and that of proposed method over the normal sentences that no mis-orderings was observed. However, to those sentences where mis-orderings happened using conventional method, our new method performed better than the conventional algorithm.

5. CONCLUSION

In this paper, a parameter generation method for LSP is proposed to preserving ordering property in HMM-based speech synthesis. The proposed method generating LSP features sequence not only maximizes the conventional HMM and GV likelihood, but also minimizes a mis-ordering penalty. The experimental results show the proposed method can alleviate the generated mis-orderings significantly with also better synthesizing performance. Further, the proposed method actually inspires an unified parameter generation framework, in which we can replace the MOP term with any other likelihood for specific end in future work.

6. ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of China under Grant Nos. 60972165, 61105048, 60901012, the Doctoral Fund of Ministry of Education of China under Grant No. 20100092120012, 20090092120012, 20110092110008, the Foundation of High-Technology Project in Jiangsu Province, the Natural Science Foundation of Jiangsu Province under Grant No. BK2011060, BK2010240, SBK201140040, Open Fund of Jiangsu Province Key Laboratory of Remote Measuring and Control (YCCK201005).

7. REFERENCES

[1] A.W. Black, H. Zen, and K. Tokuda, "Statistical parametric speech synthesis," in *Proc. of ICASSP*. IEEE, 2007, pp. 1229–1232.

[2] K. Tokuda, T. Kobayashi, and S. Imai, "Speech parameter generation from hmm using dynamic features," in *Proc. of ICASSP*. IEEE, 1995, vol. 1, pp. 660–663.

[3] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Am.*, vol. 57, pp. S35, 1975.

[4] H. Zen, T. Toda, and K. Tokuda, "The nitech-naist hmm-based speech synthesis system for the blizzard challenge 2006," *IEICE transactions on information and systems*, vol. 91, no. 6, pp. 1764–1773, 2008.

[5] I.V. McLoughlin, "Line spectral pairs," *Signal processing*, vol. 88, no. 3, pp. 448–467, 2008.

[6] F. Soong and B. Juang, "Line spectrum pair (lsp) and speech data compression," in *Proc. of ICASSP*. IEEE, 1984, vol. 9, pp. 37–40.

[7] T. Tomoki and K. Tokuda, "A speech parameter generation algorithm considering global variance for hmm-based speech synthesis," *IEICE transactions on information and systems*, vol. 90, no. 5, pp. 816, 2007.

[8] M. Lei, Z.H. Ling, and L.R. Dai, "Preserve ordering property of generated lsps for minimum generation error training in hmm-based speech synthesis," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 4712–4715.

[9] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigné, "Restructuring speech representations using a pitch adaptive time-frequency-based f0 extraction: Possible role of a repetitive structure in sounds," *Speech Communication*, vol. 27, no. 3-4, pp. 187–207, 1999.