MODELLING UNCERTAINTY IN STOCHASTIC VECTOR MAPPING WITH MINIMUM CLASSIFICATION ERROR TRAINING FOR ROBUST SPEECH RECOGNITION

Jian Wu and Qiang Huo

Department of Computer Science and Information Systems
The University of Hong Kong, Pokfulam Road, Hong Kong, China
(Email: jwu@csis.hku.hk, qhuo@csis.hku.hk)

ABSTRACT

Recently, we witness several works of considering the uncertainty of feature compensation module for robust speech recognition. In most of these studies, the modelling and the exploiting of the uncertainty are seldom treated in a unified way. In this paper, we present a new framework, which casts the problem of considering the uncertainty of feature compensation module as the one of designing a new discriminant function, thus the uncertainty parameters of the feature compensation module and other parameters of the discriminant function can be estimated jointly under a consistent criterion of minimum classification error (MCE). It is hoped that such MCE-trained discriminant function can improve the performance of a maximum discriminant function based speech recognition system. The preliminary experimental results on Aurora2 multi-condition tasks have confirmed the above conjecture.

1. INTRODUCTION

Current automatic speech recognition (ASR) systems are always compelled to be used in an unexpected noisy environment which is quite different from the environment where the training speech were collected. In order to alleviate the performance degradation caused by such mismatch, usually either a feature compensation module is embedded as part of the front-end processing which aims at transforming the feature vector of noisy speech to that of pseudo clean speech, or a model compensation procedure is used to derive a more accurate model to model the noisy speech. In both cases, an exact knowledge of the mismatch mechanism is often unavailable, thus the compensation results can not be fully trusted. Based on this reasoning, three research directions were pursued to consider the uncertainty during recognition.

The first direction is to use the so-called robust decision rules, such as the MINIMAX classification rule [12] and the Bayesian Predictive Classification (BPC) rule (e.g., [8, 9]), to take into account the uncertainty of the Hidden Markov Model (HMM) parameters in the model space during recognition. The second direction is the so-called *Bayesian predictive density based model compensation* method, where each Gaussian mixture component in a continuous density HMM (CDHMM) based ASR system is replaced with its predictive density by considering the uncertainty of HMM parameters directly (e.g. [9]) or some transformation parameters indirectly (e.g., [15, 2]). The third direction is to consider the uncertainty in feature space during recognition. Some recent examples include works in [13, 14, 5, 1, 11, 4]. In these approaches, the estimation of the uncertainty model for feature compensation is treated separately from the design of the other components in the speech recognizer. Furthermore, the design criteria used in the above approaches are not directly linked with the objective of the minimum classification error (MCE).

In [16], we have developed an environment compensated MCE training approach for the joint design of the feature compensation module and the recognizer itself. The parameters for both feature compensation and the CDHMMs of recognizer are updated simultaneously or alternatively to minimize the empirical classification error defined on the training set with a specific form of the discriminant function. By combing the ideas in [5, 4] and the one in [16], in this paper, we propose to cast the problem of considering the uncertainty of feature compensation module as the one of designing a new discriminant function. Consequently, the uncertainty parameters of the feature compensation module and other parameters of the discriminant function can be estimated jointly under a consistent MCE criterion. It is hoped that such MCE-trained discriminant function can improve the performance of a maximum discriminant function based speech recognition system.

The rest of paper is organized as follows. In Section 2, we describe the discriminant function in which the parameters of feature compensation are treated as random with a prior distribution and embedded into the discriminant function. In Section 3, we present the update formulae of the parameters which are derived by minimizing the MCE objective function. In Section 4, we report the illustrative results on Aurora2 database to demonstrate the effectiveness of the proposed approach. Finally, we summarize the paper in Section 5.

2. A DISCRIMINANT FUNCTION ACCOUNTING FOR UNCERTAINTY IN FEATURE COMPENSATION

2.1. General Formulation

Generally speaking, the feature compensation can be regarded as a process of feature transformation $\mathcal{F}_\Theta(\cdot)$ with parameter Θ, which estimates the pseudo clean speech feature vector $\hat{x} = \mathcal{F}_\Theta(y)$ from the noisy speech feature vector y. In our approach, we will consider the uncertainty of the transformation parameters Θ by treating them as if they were random. Their prior uncertainty is modelled by a joint *a priori* probability density function (pdf) $p(\Theta|\varphi_\Theta)$, with $\Theta \in \Omega_\Theta$, where Ω_Θ denotes admissible region of possible Θ, and φ_Θ is the set of unknown parameters of the prior pdf. We also assume that each CDHMM $\Lambda = \{a_{ij}, c_{sm}, \mu_{sm}, \Sigma_{sm}, i, j, s = 1...L, m = 1...M\}$ of the recognizer is fixed but unknown. It con-

This research was supported by grants from the RGC of the Hong Kong SAR (Project Numbers HKU7022/00E and HKU7039/02E).

sists of L states with transition probability a_{ij} from state i to state j. Each state has M Gaussian components with D-dimensional mean vectors $\mu_{sm}(=[\mu_{smd}]_{d=1}^{D})$ and diagonal covariance matrices $\Sigma_{sm}(=diag[\sigma_{smd}^2]_{d=1}^{D})$. c_{sm} denotes the weight of m-th Gaussian component in the s-th state.

Given a noisy speech utterance with a feature vector sequence $Y=(y_1,y_2,\cdots,y_T)$, we define the following discriminant function of Y for word sequence W to take into account the uncertainty of the above transformation parameters Θ:

$$\begin{aligned}\tilde{g}(Y;\Lambda,\varphi_\Theta,W) &\triangleq \log p(\hat{X}|\Lambda,W)=\log\sum_S p(\hat{X},S|\Lambda,W)\\ &= \log\sum_S A_S^*\prod_{t=1}^{T}p(\hat{x}_t|s_t,\Lambda,\varphi_\Theta,W), \quad (1)\end{aligned}$$

where S denotes a possible state sequence,

$$p(\hat{x}_t|s_t,\Lambda,\varphi_\Theta,W)=\int_{\Omega_\Theta}p(\hat{x}_t|s_t,\Lambda,\Theta,W)p(\Theta|\varphi_\Theta)d\Theta \quad (2)$$

is the marginal state observation pdf of $\hat{x}_t$, and $A_S^*=\prod_{t=1}^{T}a_{s_{t-1}s_t}$. The recognizer will choose the word sequence $\hat{W}$ as the recognition result that produces the maximum value of the above discriminant function.

2.2. A Special Case

In our previous work [16], we adopted a specific stochastic vector mapping function for the feature compensation from the SPLICE algorithm developed by Microsoft researchers [3]. Consequently, the feature vector of pseudo clean speech is estimated as follows:

$$\hat{x}\triangleq\mathcal{F}_\Theta(y)=y+\sum_{k=1}^{K}p(k|y)b_k\ , \quad (3)$$

where

$$p(k|y)=\frac{p(k)p(y|k)}{\sum_{j=1}^{K}p(j)p(y|j)}\ , \quad (4)$$

and $\Theta=\{b_k\}_{k=1}^{K}$ is the set of mapping function parameters (also referred to as *correction vectors*) associated with the K Gaussian components $p(y|k)$'s. The noisy speech feature vector y is assumed to have a Gaussian mixture pdf $p(y)=\sum_{k=1}^{K}p(k)p(y|k)$.

In order to make the integration in Eq.(2) tractable, hereinafter we further simplify the above transformation to be

$$\hat{x}\triangleq\mathcal{F}_\Theta(y)=y+b_k\ , \quad (5)$$

where $k=\arg\max_{j=1}^{K}p(j|y)$. Moreover, each correction vector b_k is assumed to follow a normal pdf $\mathcal{N}(b_k;r_k,\Xi_k)$ with mean vector $r_k(=[r_{kd}]_{d=1}^{D})$ and diagonal covariance matrices $\Xi_k(=diag[\tau_{kd}^2]_{d=1}^{D})$. With these assumptions, a specific discriminant function can be derived as

$$\begin{aligned}\tilde{g}(Y;\Lambda,\varphi_\Theta,W) &= \log\sum_S A_S^*\prod_{t=1}^{T}\sum_{m=1}^{M}c_{s_tm}\cdot \quad (6)\\ &\quad \mathcal{N}(y_t;\mu_{s_tm}-r_{k_t},\Sigma_{s_tm}+\Xi_{k_t})\ ,\end{aligned}$$

where k_t denotes the index of correction vector chosen at time t, and $\varphi_\Theta=\{r_k,\Xi_k\}$.

3. MODELLING UNCERTAINTY IN STOCHASTIC VECTOR MAPPING WITH MINIMUM CLASSIFICATION ERROR TRAINING

As mentioned above, in our proposed approach, both the CDHMM parameters and the hyperparameters of the feature compensation parameters are unknown. In order to find their optimal values in terms of achieving the minimum empirical classification error on training set $\mathcal{Y}=\{Y_i\}_{i=1}^{I}$, the objective function should be defined as follows,

$$\ell(\Lambda,\varphi_\Theta)=\frac{1}{I}\sum_{i=1}^{I}l(Y_i;\Lambda,\varphi_\Theta)\ , \quad (7)$$

with $l(Y;\Lambda,\varphi_\Theta)$ being the loss function for the training utterance Y defined as:

$$l(Y;\Lambda,\varphi_\Theta)=\frac{1}{1+exp(-\alpha d(Y;\Lambda,\varphi_\Theta)+\beta)}\ , \quad (8)$$

where α and β are two control parameters. In the above equation, $d(\cdot)$ is a misclassification measure defined as

$$d(Y;\Lambda,\varphi_\Theta)=-\tilde{g}(Y;\Lambda,\varphi_\Theta,W_c)+\tilde{G}(Y;\Lambda,\varphi_\Theta)\ , \quad (9)$$

with

$$\tilde{G}(Y;\Lambda,\varphi_\Theta)=\frac{1}{\eta}log\{\frac{1}{N}\sum_{n=1}^{N}exp[\eta\cdot\tilde{g}(Y;\Lambda,\varphi_\Theta,W_n)]\},$$

where η is a positive control parameter, W_c and $\{W_n\}$ are the correct word sequence and the N-best competing word sequences of the training utterance Y, respectively.

3.1. Updating Parameters Using Sequential Gradient Descent Algorithm

Given the objective function in Eq.(7), the following sequential gradient descent algorithm is usually used to update the parameters iteratively. Let's use Γ to denote generically the parameters to be estimated, $\{\Lambda,\varphi_\Theta\}$. Given $\mathcal{Y}$, we first randomize the ordering of $\{Y_i\}$ and then we present the training samples sequentially. Upon the presentation of the j-th training sample, Γ is updated as follows:

$$\Gamma_{j+1}=\Gamma_j-\epsilon_j\frac{\partial l(Y_j;\Gamma)}{\partial\Gamma}|_{\Gamma=\Gamma_j}\ , \quad (10)$$

where "j" represents the cumulative number of training samples presented so far, ϵ_j is the learning rate, and

$$\begin{aligned}\frac{\partial l(Y;\Gamma)}{\partial\Gamma}&=\alpha l(1-l)\Big\{-\frac{\partial\tilde{g}(Y;\Gamma,W_c)}{\partial\Gamma} \quad (11)\\ &+\sum_{n=1}^{N}\Big[\frac{\exp(\eta\cdot\tilde{g}(Y;\Gamma,W_n))}{\sum_{i=1}^{N}\exp(\eta\cdot\tilde{g}(Y;\Gamma,W_i))}\frac{\partial\tilde{g}(Y;\Gamma,W_n)}{\partial\Gamma}\Big]\Big\}\ .\end{aligned}$$

One pass of the training samples is called an epoch. After the completion of each epoch, we need to randomize the ordering of $\{Y_i\}$ again.

Besides this sequential gradient descent algorithm, a batch-mode approximate second-order optimization algorithm, namely Quickprop [6, 17], may also be used to minimize the above objective function. No matter which approach is used, the partial derivative of the discriminant function with respect to the parameters to be estimated, $\partial\tilde{g}/\partial\Gamma$, need be calculated utterance by utterance. In the following, we present the formulae related to these partial derivatives.

3.2. Partial Derivatives of The Discriminant Function

In order to maintain the constraints for both the CDHMM parameters and the hyperparameters of the *a priori* density of the correction vectors, the following parameter transformation, which is similar to that described in [10], are applied to the relevant parameters during updating (For simplicity, we only list the formulae related to μ_{sm}, Σ_{sm}, r_k and Ξ_k):

$$\tilde{\mu}_{smd} = \frac{\mu_{smd}}{\sigma_{smd}} , \quad (12)$$

$$\tilde{\sigma}_{smd} = \log \sigma_{smd} , \quad (13)$$

$$\tilde{r}_{kd} = \frac{r_{kd}}{\tau_{kd}} , \quad (14)$$

$$\tilde{\tau}_{kd} = \log \tau_{kd} . \quad (15)$$

Therefore, given a training utterance $Y = \{y_1, y_2, ..., y_T\}$, the current value of the partial derivatives of discriminant function with respect to the above parameters are as follows:

$$\frac{\partial \tilde{g}}{\partial \tilde{\mu}_{smd}} = \sum_{t=1}^{T} \tilde{\zeta}_t(s,m) \left(\frac{y_{td} + r_{k_t d} - \mu_{smd}}{\sqrt{\sigma_{smd}^2 + \tau_{k_t d}^2}} \right) , \quad (16)$$

$$\frac{\partial \tilde{g}}{\partial \tilde{\sigma}_{smd}} = \sum_{t=1}^{T} \tilde{\zeta}_t(s,m) \left[\frac{(y_{td} + r_{k_t d} - \mu_{smd})^2}{\sigma_{smd}^2 + \tau_{k_t d}^2} - 1 \right] \cdot \frac{\sigma_{smd}^2}{\sigma_{smd}^2 + \tau_{k_t d}^2} , \quad (17)$$

$$\frac{\partial \tilde{g}}{\partial \tilde{r}_{kd}} = -\sum_{t=1}^{T} \sum_{s,m} \tilde{\zeta}_t(s,m) \left(\frac{y_{td} + r_{k_t d} - \mu_{smd}}{\sqrt{\sigma_{smd}^2 + \tau_{k_t d}^2}} \right) \cdot \delta(k_t - k) , \quad (18)$$

$$\frac{\partial \tilde{g}}{\partial \tilde{\tau}_{kd}} = \sum_{t=1}^{T} \sum_{s,m} \tilde{\zeta}_t(s,m) \left[\frac{(y_{td} + r_{k_t d} - \mu_{smd})^2}{\sigma_{smd}^2 + \tau_{k_t d}^2} - 1 \right] \cdot \frac{\tau_{k_t d}^2}{\sigma_{smd}^2 + \tau_{k_t d}^2} \cdot \delta(k_t - k) , \quad (19)$$

where $\tilde{\zeta}_t(s,m)$ denotes the occupation probability of Gaussian component m in state s at time t calculated with the newly defined marginal density in Eq.(2), and $\delta(\cdot)$ denotes the Kronecker delta function.

Using the above equations, the transformed parameters can be updated iteratively. After each update of the transformed parameters, the original parameters will be obtained by applying the inverse transform of Eq.(12-15). For example, if the $(j+1)$-th update of $\tilde{\mu}_{smd}(j+1)$, $\tilde{\sigma}_{smd}(j+1)$, $\tilde{r}_{kd}(j+1)$ and $\tilde{\tau}_{kd}(j+1)$ are calculated, then the $(j+1)$-th update of the original parameters will be

$$\mu_{smd}(j+1) = \tilde{\mu}_{smd}(j+1)\sigma_{smd}(j) , \quad (20)$$

$$\sigma_{smd}(j+1) = \exp\{\tilde{\sigma}_{smd}(j+1)\} , \quad (21)$$

$$r_{kd}(j+1) = \tilde{r}_{kd}(j+1)\tau_{kd}(j) , \quad (22)$$

$$\tau_{kd}(j+1) = \exp\{\tilde{\tau}_{kd}(j+1)\} , \quad (23)$$

and are used to estimate $\tilde{\zeta}_t(s,m)$ for the next update of the transformed parameters.

4. EXPERIMENTS AND RESULTS

4.1. Aurora2 Database and Experimental Setup

The task used to verify our idea is the speaker independent recognition of connected digit strings. The recognition results presented in this section are produced on the Aurora2 database using the reference of Aurora front-end version 2.0 [7]. In this front-end, for each frame, a 39-dimensional feature vector is generated, which consists of 12 MFCCs (MFCC of order 0 is not included) and logarithmic frame energy, plus their first and second order derivatives. Whole word left-to-right CDHMMs are created for all digits. The CDHMM consists of 18 states, each having 3 Gaussian mixture components with diagonal covariance matrices. Besides, two pause models, "sil" and "sp", are created to model the silence before/after the digit string and the short pause between any two digits. In the BASELINE system shown in Table 1, all of the CDHMMs are trained from the collection of 8440 utterances that come from 20 subsets representing 4 different noise scenarios (i.e., *suburban train, babble, car* and *exhibition hall*) at 5 different SNRs (i.e., 20dB, 15dB, 10dB, 5dB and the *clean condition*). The test set consists of three different parts. For the Test Set A, the same four types of noises as those in training set are added to its subsets, but with 7 different SNRs. For the Test Set B, another 4 types of noises (i.e., *restaurant, street, airport* and *train station*) are added to its subset with also 7 SNRs. For the Test C, *suburban train* and *street* noises are used as the additive noise sources but the speech and noise are filtered with a MIRS characteristic while the G.712 characteristic is used in training set as well as the first two test sets. During the recognition, an utterance can be modelled by any sequence of digits with the possibility of a "sil" model at the beginning and at the end and a "sp" model between two digits. All of the recognition experiments are performed with the search engine of HTK3.0 toolkit [18] and follow exactly the default scripts provided in Aurora2 CD-ROM.

4.2. Experimental Results

In order to illustrate the potential of considering the uncertainty in the procedure of stochastic vector mapping, a reference system using the deterministic parameters for stochastic vector mapping is built first which is labeled as "VM-DET" in Table 1. The value of the deterministic parameters are estimated using the SPLICE algorithm. The CDHMM parameters are estimated by using an ML criterion. The details about the construction of this reference system can be found in [16].

In the experiment considering the uncertainty of the stochastic vector mapping, which is labeled as "VM-UNCERTAIN" in Table 1, we use the discriminant function defined in Eq.(6) to characterize the decision function of the recognizer. In order to estimate the hyperparameters of the prior distribution of the correction vectors used in stochastic vector mapping, for each noisy training feature vector y_t, we estimate its pseudo-clean version $\hat{x}_t$ by using Eq. (3). The corresponding bias vector $b_t = \hat{x}_t - y_t$ is treated as a sample of $p(b_k) = \mathcal{N}(b_k; \{r_{kd}\}, \{\tau_{kd}^2\})$, where $k = \arg\max_{j=1}^{K} p(j|y_t)$. In this way, we can collect a set of *bias vector samples* for each $p(b_k)$. The hyperparameter r_k is set to be the same as the corresponding estimate in conventional SPLICE algorithm and τ_{kd}^2 is estimated as a sample variance with r_{kd} as mean. The CDHMM parameters are the same as in reference system. From the experimental results of VM-UNCERTAIN and VM-DET, it is observed that although VM-UNCERTAIN can perform

Table 1. Aurora2 Word Error Rate (Multicondition Training)

	Set A	Set B	Set C	Overall
BASELINE	11.93%	12.78%	15.44%	12.97%
VM-DET	9.99%	12.92%	13.34%	11.83%
VM-UNCERTAIN	9.37%	13.24%	13.03%	11.65%
UNC-MCE(Ξ)	8.95%	12.21%	12.46%	10.96%

slightly better than that of VM-DET in terms of overall accuracy, the accuracies in Test Set B are actually degraded. This result is not surprising because the success of the VM-UNCERTAIN is highly dependent on the goodness of the assumed prior distribution for the specific testing scenario. We guess that the proper distribution of correction vectors for Test Set B is quite different from the one derived as above from the training samples. That is why we propose the MCE training approach with the newly designed discriminant function in this paper, hoping that the direct link between the setting of the relevant hyperparameters and the MCE objective of the recognizer may offer a better estimation.

To examine the correctness of the above statement, the algorithm described in Section 3 is partly implemented and the corresponding experiment is labeled as "UNC-MCE(Ξ)" in Table 1. In this experiment, the CDHMM parameters and hyperparameters r_k are the same as in the experiment "VM-UNCERTAIN", while the variance hyperparameters, Ξ_k, are estimated using MCE criterion. It is observed from Table 1 that the MCE training of Ξ_k helps reduce the word error rate from that of 11.65% by VM-UNCERTAINTY to 10.96%. Compared with that of baseline system, it represents a relative word error rate reduction of 15.5%.

5. DISCUSSION AND CONCLUSION

In this paper, we present a novel framework to introduce the uncertainty into the classifier design and the parameter estimation for the uncertain feature compensation under a consistent MCE principle. The preliminary experimental results on Aurora2 have shown the potential of this new approach. Considering the fact of that our system in [16] has provided a relative error reduction of 38.42% compared with the baseline system, which jointly update the correction vectors and CDHMM parameters through MCE training without accounting for uncertainty, it would be interesting to verify in our future work whether an even better performance can be achieved by a joint MCE estimation of all the hyperparameters as well as the CDHMM parameters using the new discriminant function as defined in this paper. Furthermore, if we treat the CDHMM parameters also as if they were random, then the uncertainties in both the feature compensation and the model compensation can be embedded together by designing a new discriminant function accounting for both uncertainties, which might provide more space for improving our ASR system in dealing with mismatch conditions where the accurate information about the mismatch are not available.

6. REFERENCES

[1] J. A. Arrowood and M. A. Clements, "Using observation uncertainty in HMM decoding," *Proc. ICSLP-2002*, 2002, pp.1561-1564.

[2] J.-T. Chien, "Combined linear regression adaptation and Bayesian predictive classification for robust speech recognition," *Proc. Eurospeech-2001*, Denmark, Sept. 2001.

[3] L. Deng, A. Acero, M. Plumpe, and X.-D. Huang, "Large-vocabulary speech recognition under adverse acoustic environments," *Proc. ICSLP-2000*, Oct. 2000.

[4] L. Deng, J. Droppo, and A. Acero, "Exploiting variances in robust feature extraction based on a parametric model of speech distortion," *Proc. ICSLP-2002*, 2002, pp.2449-2452.

[5] J. Droppo, A. Acero, and L. Deng, "Uncertainty decoding with SPLICE for noise robust speech recognition," *Proc. ICASSP-2002*, May 2002, pp.I-57-60.

[6] S. E. Fahlman, "An empirical study of learning speed in back-propagation networks," *Technical Report CMU-CS-88-162*, Carnegie Mellon University, 1988.

[7] H. G. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions", *ISCA ITRW ASR2000*, Paris, France, September 2000.

[8] Q. Huo and C.-H. Lee, "A Bayesian predictive classification approach to robust speech recognition," *IEEE Trans. on Speech and Audio Processing*, Vol.8, pp.200-204, 2000.

[9] H. Jiang, K. Hirose and Q. Huo, "Robust speech recognition based on a Bayesian prediction approach," *IEEE Trans. on Speech and Audio Processing*, Vol.7, pp.426-440, 1999.

[10] B.-H. Juang, W. Chou and C.-H. Lee, "Minimum classification error rate methods for speech recognition", *IEEE Trans. on Speech and Audio Processing*, Vol. 5, pp.257-265, 1997.

[11] T.T. Kristjansson and B.J. Frey, "Accounting for uncertainty in observations: a new paradigm for robust automatic speech recognition," *Proc. ICASSP-2002*, May 2002, pp.I-61-64.

[12] N. Merhav and C.-H. Lee, "A minimax classification approach with application to robust speech recognition," *IEEE Trans. Speech and Audio Processing*, Vol.1, pp.90-100, 1993.

[13] S. Moon and J.-N. Hwang, "Robust Speech Recognition Based on Joint Model and Feature Space Optimization of Hidden Markov Models," *IEEE Trans. on Neural Networks*, Vol. 8, No. 2, pp. 194-204, 1997.

[14] M. Roch and R. Hurtig, "The integral decode: a smoothing technique for robust HMM-based speaker recognition," *IEEE Trans. on SAP*, Vol.10, No.5, pp.315-324, 2002.

[15] A. C. Surendran and C.-H. Lee, "Transformation-based Bayesian prediction for adaptation of HMMs," *Speech Communication*, Vol. 34, pp.159-174, 2001.

[16] J. Wu and Q. Huo, "An environment compensated minimum classification error training approach and its evaluation on Aurora2 database," *Proc. ICSLP-2002*, 2002, pp.I-453-456.

[17] J. Wu and Q. Huo, "A comparative study of quickprop and GPD optimization algorithms for MCELR adaptation of CDHMM parameters," *Proc. International Symposium on Chinese Spoken Language Processing*, 2002, pp.355-358.

[18] S. Young et al., *The HTK Book (for HTK V3.0)*, July 2000.