ISCA Archive
http://www.isca-speech.org/archive

3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

SEGMENTAL HIDDEN MARKOV MODELS

M.J.F. Gales *S.J. Young*

Cambridge University Engineering Department,
Trumpington Street, Cambridge, CB2 1PZ, England.

ABSTRACT

The most popular and successful acoustic model for speech recognition is the Hidden Markov Model (HMM). To use HMMs for speech recognition a series of assumptions are made about the waveform, some of which are known to be poor. In particular, the 'Independence Assumption' implies that all observations are only dependent on the state that generated them, not on neighbouring observations. In this paper, a new form of acoustic model is described called the Segmental Hidden Markov Model (SHMM) in which the effect of the 'Independence Assumption' on the observation likelihood is greatly reduced. In the SHMM all observations are assumed to be independent given the state that generated them but additionally they are conditional on the mean of the segment of speech to which they belong. Re-estimation formulae are presented for the training of both single and multiple Gaussian Inter Mixture models and a recognition algorithm is described. Additionally it is shown that the standard HMM, both in the single Gaussian mixture and multiple Gaussian mixtures cases, is just a subset of the SHMM. The new model is shown to provide better recognition performance on a wider set of synthetic data than the standard HMM.

Keywords: speech recognition, HMM, segment models.

10.21437/Eurospeech.1993-354

1. INTRODUCTION

The Hidden Markov Model (HMM) is the most popular and successful stochastic approach to speech recognition in general use. Its popularity and success are due to the existence of elegant and efficient algorithms for both training and recognition. However the use of HMMs for speech recognition is dependent on certain assumptions. These are

1. Speech may be split into segments, states, in which the speech waveform may be assumed to be stationary. The transition between these states is assumed to be instantaneous.
2. The probability of a certain symbol being generated is only dependent on the current state, not on any previously generated symbols. This is a first order Markov assumption and is usually referred to as the 'Independence Assumption'.

The first assumption implies that the model should have a large number of states. However, this results in estimation problems. So, in most real system a small number of states are used. The second assumption is not valid and is the major drawback to the use of HMMs for speech recognition.

Various methods have been proposed for compensating for the invalidity of this 'Independence Assumption'. Dynamic coefficients, which are less sensitive to the assumption, have been incorporated [6]. The correlation may be modelled explicitly by making the observation probability conditional on the previous observation and state [5] or by altering the structure of the model, as in the Stochastic Segment Model [3]. These methods compensate to varying degrees for the invalid assumption, however, they tend to dramatically increase the number of model parameters.

This paper introduces a new stochastic model of speech based on the HMM. By removing some of the above limitations of the HMM, it is believed that this new model offers considerable promise for improving future systems.

2. THE SEGMENTAL HIDDEN MARKOV MODEL

A model that minimises the effects of the 'Independence Assumption' without significantly increasing the number of parameters is desirable. To this end a new style of acoustic model is introduced, the Segmental Hidden Markov Model (SHMM). In the SHMM all observations are assumed to be independent given the state that generated them, but additionally they are conditional on the mean of the segment of speech to which they belong. The idea behind this assumption is that certain characteristics of the speech, such as speaker or stress condition, are fixed over the whole segment. Hence, when the first frame in the segment is observed, some characteristics are known and fixed. In standard HMMs this information is completely ignored. However, in making the observations conditionally dependent on the mean of the segment the SHMM takes these effects into account, thereby making better use of the acoustic information. With this new assumption the probability of a segment given a particular model must be calculated. For each state of the model $\mathcal{M}$ the output probability distribution will no longer be described by one distribution, but by two. One describing the distribution of the segment mean, the 'inter distribution', the other describing the observation probabilities given that mean, the 'intra distribution'.

$$p(\mathcal{Y}_{t_i,t_j}|s_i,\mathcal{M}) = \int_{\mathcal{R}^n} p(\mu|s_i,\mathcal{M})p(\mathcal{Y}_{t_i,t_j}|\mu,s_i,\mathcal{M})\,d\mu \qquad (1)$$

where $\mathcal{Y}_{t_i,t_j} = [\mathbf{y}_{t_i},...,\mathbf{y}_{t_j}]$, $\mathbf{y}_\tau$ is the observation vector at time τ and s_i indicates state S_i of the model. Using the assumption, that given the mean of the segment, all the observations within that segment are independent

$$p(\mathcal{Y}_{t_i,t_j}|s_i,\mathcal{M}) = \int_{\mathcal{R}^n} p(\mu|s_i,\mathcal{M}) \prod_{\tau=t_i}^{t_j} p(\mathbf{y}_\tau|\mu,s_i,\mathcal{M})\,d\mu \qquad (2)$$

The above expression and assumptions have also been used by Russell [4]. However he uses an MAP approach to estimate the mean, this will be referred to as MAP SHMM. Here the full form of the above expression is used.

Assuming that $p(\mu|s_i,\mathcal{M})$ and $p(\mathbf{y}_\tau|\mu,s_i,\mathcal{M})$ are both Gaussian probability distributions, the output distribution is described by $\{\Sigma,\mu_c,\Sigma_c\}$, the intra-state variance, the inter-state mean and the inter-state variance respectively. The probability may be written in terms of μ_n and Σ_n^{-1}, which are independent of μ

$$\log(p(\mathcal{Y}_{t_i,t_j}|s_i,\mathcal{M})) = \quad (3)$$
$$t\log(K)+\log(K_c)-\log(K_n)-\frac{1}{2}\mathcal{G}(\mathcal{Y},\mathcal{M})$$

where

$$\mathcal{G}(\mathcal{Y},\mathcal{M}) = \mu_c\Sigma_c^{-1}\mu_c^T + \sum_{\tau=t_i}^{t_j}\mathbf{y}_\tau\Sigma^{-1}\mathbf{y}_\tau^T - \mu_n\Sigma_n^{-1}\mu_n^T \quad (4)$$

and

$$\Sigma_n^{-1} = \Sigma_c^{-1}+t\Sigma^{-1} \quad (5)$$
$$\mu_n^T = \Sigma_n\left(\Sigma_c^{-1}\mu_c^T+\Sigma^{-1}\mu_s^T\right) \quad (6)$$
$$\mu_s^T = \sum_{\tau=t_i}^{t_j}\mathbf{y}_\tau^T \quad (7)$$

K, K_c and K_n are the standard normalising constants associated with Σ, Σ_c and Σ_n respectively and t is the segment length.

The above expressions have assumed Gaussian inter and intra distributions. Any form of of inter and intra distributions may be used. However, if the distributions are restricted to those where the inter distribution is a conjugate prior of the intra distribution, closed forms for the probabilities and sufficient statistics given the complete data set, are obtainable.

2.1. Relationship to MAP SHMM

The SHMM proposed here can be related to the MAP SHMM. For this section the feature vector is assumed to have dimensionality one. For the MAP SHMM

$$\log(p(\mathcal{Y}_{t_i,t_j}|s_i,\mathcal{M})) = \quad (8)$$
$$\log(\mathcal{N}(\hat{c};\mu_c,\Sigma_c)) + \sum_{\tau=t_i}^{t_j}\log(\mathcal{N}(y_\tau;\hat{c},\Sigma))$$

where the target mean, $\hat{c}$ is given by

$$\hat{c} = \frac{\mu_c\Sigma+\mu_s\Sigma_c}{\Sigma+t\Sigma_c} \quad (9)$$

Simplifying the above equation gives

$$\log(p(\mathcal{Y}_{t_i,t_j}|s_i,\mathcal{M})) = t\log(K)+\log(K_c)-\frac{1}{2}\mathcal{G}(\mathcal{Y},\mathcal{M}) \quad (10)$$

This expression is identical to the probability equation for the SHMM proposed here, except for the term $\log(K_n)$. This normalisation constant may be written as

$$\log(K_n) = \log(K)+\frac{1}{2}\left(\log(t)+\log\left(1+\frac{\Sigma}{t\Sigma_c}\right)\right) \quad (11)$$

This term is solely dependent on the length of the segment, not the observations within that segment. Hence the use of the MAP estimate of the mean, instead of the true distribution for the mean, is a model dependent bias on the length of the segmentation.

2.2. Multiple Gaussian Inter-Mixtures

The above analysis has been performed assuming a single Gaussian inter mixture probability distribution for $p(\mu|s_i,\mathcal{M})$. If in fact this distribution is described by a multiple Gaussian mixture distribution the same style of analysis may be applied. Letting

$$p(\mu|s_i,\mathcal{M}) = \sum_{m=1}^{M}c_m\mathcal{N}(\mu;\mu_{c_m},\Sigma_{c_m}) \quad (12)$$

and substituting this in equation 2

$$p(\mathcal{Y}_{t_i,t_j}|s_i,\mathcal{M}) = \quad (13)$$
$$\sum_{m=1}^{M}c_m\int_{\mathcal{R}^n}\mathcal{N}(\mu;\mu_{c_m},\Sigma_{c_m})\prod_{\tau=t_i}^{t_j}p(\mathbf{y}_\tau|\mu,s_i,\mathcal{M})\,d\mu$$

It can be seen that the analysis for the single mixture case may be directly applied to the multiple mixture case.

2.3. Multiple Gaussian Intra Mixtures

Multiple Gaussian intra mixtures models occur when

$$p(\mathbf{y}_\tau|\mu,s_i,\mathcal{M}) = \sum_{m=1}^{M}c_m\mathcal{N}(\mathbf{y}_\tau;\mu_m,\Sigma_m) \quad (14)$$

where $\mu_m=\mu+\Delta_m$. The delta intra means, Δ_m, are stored as model parameters. By substituting the above expression in equation 2 and assuming a single Gaussian intermixture

$$p(\mathcal{Y}_{t_i,t_j}|s_i,\mathcal{M}) = \quad (15)$$
$$\int_{\mathcal{R}^n}\mathcal{N}(\mu;\mu_c,\Sigma_c)\prod_{\tau=t_i}^{t_j}\left(\sum_{m=1}^{M}c_m\mathcal{N}(\mathbf{y}_\tau;\mu_m,\Sigma_m)\right)d\mu$$

The analysis for the single intra mixture case can be seen to be inappropriate for the multiple Gaussian intra mixture case, due to the product of the weighted sum of Gaussians. In fact there are no sufficient statistics for the multiple intra mixture case [1].

3. RELATIONSHIP TO STANDARD HMMS

As previously stated the observation probability for a segment is given by

$$p(\mathcal{Y}_{t_i,t_j}|s_i,\mathcal{M}) = \int_{\mathcal{R}^n}p(\mu|s_i,\mathcal{M})\prod_{\tau=t_i}^{t_j}p(\mathbf{y}_\tau|\mu,s_i,\mathcal{M})\,d\mu \quad (16)$$

If the inter mixture variance, Σ_c, is set to zero then $p(\mu|s_i,\mathcal{M})=\delta(\mu_c-\mu)$, where $\delta(.)$ is the Dirac delta function, and the above equation may be simplified to

$$p(\mathcal{Y}_{t_i,t_j}|s_i,\mathcal{M}) = \prod_{\tau=t_i}^{t_j}p(\mathbf{y}_\tau|\mu_c,s_i,\mathcal{M}) \quad (17)$$

This is the same equation as a standard single Gaussian mixture HMM with the mean set to the interstate mean, μ_c, and the variance set to the intra mixture variance, Σ. Hence the standard single Gaussian mixture HMM may be viewed as a subset of the SHMM.

An equivalent standard model for the multiple Gaussian inter mixture model described in section 2.2. may be found by setting all the inter mixture variances to zero in equation 14. This gives

$$p(\mathcal{Y}_{t_i,t_j}|s_i,\mathcal{M}) = \sum_{m=1}^{M}c_m\left(\prod_{\tau=t_i}^{t_j}p(\mathbf{y}_\tau|\mu_{c_m},s_i,\mathcal{M})\right) \quad (18)$$

So the multiple Gaussian inter mixture case with zero inter mixture variances is the same as the standard multiple Gaussian mixture case with the added constraint that all observations in a segment associated with a given state are generated by the same Gaussian mixture. An exact form of the standard multiple Gaussian mixture HMM is obtained by setting the inter mixture variance of a multiple Gaussian intra mixture model to zero. Hence

$$p(\mathcal{Y}_{t_i,t_j}|s_i,\mathcal{M}) = \prod_{\tau=t_i}^{t_j}\left(\sum_{m=1}^{M} c_m \mathcal{N}(\mathbf{y}_\tau;\mu_m,\Sigma_m)\right) \quad (19)$$

where $\mu_m = \mu_c + \Delta_m$. This is identical to the standard HMM with the multiple mixture means set to μ_m.

4. RE-ESTIMATION FORMULAE

In order to re-estimate the parameters of the SHMM an auxiliary function, $\mathcal{Q}(\mathcal{M},\hat{\mathcal{M}})$, is introduced.

$$\mathcal{Q}(\mathcal{M},\hat{\mathcal{M}}) = \sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})\log(p(\mathbf{Y}_T,\mathbf{s}_T|\hat{\mathcal{M}})) \quad (20)$$

where the summation on s is over every possible segmentation of $\mathbf{Y}_T$. It is shown by Baum [2] that if $\mathcal{Q}(\mathcal{M},\hat{\mathcal{M}}) \geq \mathcal{Q}(\mathcal{M},\mathcal{M})$ then $p(\mathbf{Y}_T|\hat{\mathcal{M}}) \geq p(\mathbf{Y}_T|\mathcal{M})$. For all the re-estimation formulae derived, diagonal covariance matrices are assumed and, for simplicity of notation, n is assumed to be 1 so that the vector notation can be dropped. In addition, the notation $\mathcal{Y}$ to represent $\mathcal{Y}_{t_i,t_j}$ is used and the means and variances will be assumed to relate to state S_i. For the following analysis left-to-right models are assumed. Hence only one segment from each utterance is associated with a particular state. This assumption is not necessary, but it further simplifies the notation.

4.1. Re-estimation for $\hat{\mu}_c$

Taking the partial derivative with respect to $\hat{\mu}_c$

$$\frac{\partial}{\partial\hat{\mu}_c}\mathcal{Q}(\mathcal{M},\hat{\mathcal{M}}) = \sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})\left[\left(\frac{\mu_s}{t}-\hat{\mu}_c\right)\mathcal{K}(s_i)\right] \quad (21)$$

where t is the length of the segment s_i. Equating the above equation to zero and solving for μ_c

$$\hat{\mu}_c = \frac{\sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})\left[\frac{\mu_s}{t}\mathcal{K}(s_i)\right]}{\sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})\left[\mathcal{K}(s_i)\right]} \quad (22)$$

where

$$\mathcal{K}(s_i) = \frac{t}{t\hat{\Sigma}_c + \hat{\Sigma}} \quad (23)$$

4.2. Re-estimation for $\hat{\Sigma}_c$

Taking the partial derivative with respect to $\hat{\Sigma}_c$

$$\frac{\partial}{\partial\hat{\Sigma}_c}\mathcal{Q}(\mathcal{M},\hat{\mathcal{M}}) = \quad (24)$$
$$\frac{1}{2}\sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})\left[\mathcal{K}(s_i)^2(\hat{\mu}_c - \frac{\mu_s}{t})^2 - \mathcal{K}(s_i)\right]$$

Equating the above equation to zero yields no closed form for $\hat{\Sigma}_c$.

4.3. Re-estimation for $\hat{\Sigma}$

Taking the partial derivative with respect to $\hat{\Sigma}$

$$\frac{\partial}{\partial\hat{\Sigma}}\mathcal{Q}(\mathcal{M},\hat{\mathcal{M}}) = \sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})\frac{1}{2\hat{\Sigma}^2} \quad (25)$$
$$\left[\sum_{\tau=t_i}^{t_j} y_\tau^2 - t\hat{\Sigma} + \hat{\Sigma}\hat{\Sigma}_c\mathcal{K}(s_i) - \frac{\mu_s^2}{t} + \frac{\hat{\Sigma}^2\mathcal{K}(s_i)^2(\hat{\mu}_c t - \mu_s)^2}{t}\right]$$

Again equating the above expression to zero yields no closed expression for $\hat{\Sigma}$.

4.4. Approximate Solution

In the previous sections an attempt was made to find closed form expressions for re-estimating the parameters of the SHMM. It was only possible to find such an expression for $\hat{\mu}_c$. Closed loop forms for estimating Σ and Σ_c would be desirable. Rewriting equation 23

$$\mathcal{K}(s_i) = \frac{1}{\Sigma_c}\left(\frac{1}{1+\frac{\Sigma}{t\Sigma_c}}\right) = \frac{1}{\Sigma_c}\left(1 - \frac{\Sigma}{t\Sigma_c} + \left(\frac{\Sigma}{t\Sigma_c}\right)^2 - \ldots\right) \quad (26)$$

By assuming that $t\hat{\Sigma}_c \gg \hat{\Sigma}$, ie the between segment variability is far greater than the within segment variability, and ignoring terms in $\left(\frac{\Sigma}{t\Sigma_c}\right)$ and higher, $\mathcal{K}()$ is independent of the segmentation. Rewriting equation 22

$$\hat{\mu}_c = \frac{\sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})\left[\frac{\mu_s}{t}\right]}{\sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})} \quad (27)$$

Looking at equation 24 and setting the derivative to zero

$$\hat{\Sigma}_c = \frac{\sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})\frac{1}{t^2}(\hat{\mu}_c t - \mu_s)^2}{\sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})} \quad (28)$$

Finally rewriting equation 25 and additionally ignoring terms in $\left(\frac{\Sigma}{\Sigma_c}\right)^2$

$$\hat{\Sigma} = \frac{\sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})\left[\sum_{\tau=t_i}^{t_j} y_\tau^2 - \frac{\mu_s^2}{t}\right]}{\sum_s p(\mathbf{Y}_T,\mathbf{s}_T|\mathcal{M})(t-1)} \quad (29)$$

These may all be shown to be maxima.

5. RECOGNITION ALGORITHM

A modified version of the Viterbi decoder is required for the SHMM. A new variable is defined as

$$\phi_t(j,\tau) = \max_s[p(\mathbf{Y}_t, s_j(t',t), \overline{q}_j(t+1)|\mathcal{M})] \quad (30)$$

where $t' = t - \tau + 1$ and $\overline{q}_j(t+1)$ indicates that the model has left state S_j at time $t+1$. Values of $\phi_t(j,\tau)$ may be calculated using the following recursive equation

$$\phi_t(j,\tau) = \max_{1\leq i\leq N, i\neq j}\left[\max_{1<\gamma<t}[\phi_{t-\tau}(i,\gamma)a_{ij}]\right] p(\mathcal{Y}_{t',t}|s_j,\mathcal{M}) \quad (31)$$

This iterative expression is similar to that of a hidden semi-Markov model [7]. The standard Viterbi recogniser for the HMM has a computational overhead of $\mathcal{O}(T)$. It can be seen that the above expression will result in a cost of $\mathcal{O}(T^2)$. This may be reduced by assuming that there is a maximum duration in any one state of t_{max}. With this assumption the cost is $\mathcal{O}(Tt_{max})$.

6. SYNTHETIC DATA

In order to test the training algorithms and the Viterbi decoder, artificial data was generated using a set of three SHMMs. A total of 1000 isolated utterances were generated, approximately evenly distributed over the three models, using the underlying assumptions of the SHMM. These were then split into a training set of 500 utterances and a test set of 500.

6.1. Training Procedure

The following training procedure was used to generate the models. The number of emitting states in all cases was set to 3, the same as all the source models.

1. Generate a standard HMM, having the same number of states as the SHMM to be trained.
2. Using initial segmentation generated by the standard HMM, the models were updated using a Viterbi style re-estimation scheme. The initial start point for the maximisation was obtained from the approximate estimate and optimisation was performed using conjugate gradient descent. For these experiments the transition matrix, **A**, was not updated and was set to the correct value.

All training and testing used version 1.4 of the portable HTK HMM toolkit [8] with suitable extensions to support SHMMs.

6.2. Results on Synthetic Data

Model	*No. Mixtures*	*% Correct*
Source	1 (1)	77.2
HMM	1	51.6
HMM	2	57.4
SHMM	1 (1)	78.2

Table 1. Synthetic Data Performance from SHMM data

In table 1 *Source* indicates that the models were used to generate the data. Where there is a bracketed value in the *No. Mixtures* column this indicates the number of intra mixtures of an SHMM.

From table 1 it can be seen that if the source model is in the form of a SHMM it is not possible to dramatically increase performance by adding additional mixtures to a standard HMM. Standard HMMs were generated using all the data, both training and test, with Viterbi and Baum-Welch re-estimation. The best performance was 58.6% with three mixture models. For acoustic waveforms where the assumptions of the SHMM are valid the SHMM performs significantly better than the standard HMM, even if the standard HMM has over twice as many parameters.

As a comparison, a set of standard HMMs were then used as the sources. The above procedure was repeated and a set of HMMs and SHMMs were generated.

Model	*No. Mixtures*	*% Correct*
Source	1	95.9
HMM	1	96.0
SHMM	1 (1)	95.8

Table 2. Synthetic Data Performance from HMM data

The performances of all three models are approximately the same. When trained, the SHMM set all the inter variances to low values, the largest inter to intra ratio being 0.07. This agrees with the discussion in section 3. where the HMM is shown to be equivalent to the SHMM when the inter variance is zero.

7. PRELIMINARY TIMIT RESULTS

The segment based models were tested on a subset of TIMIT. The 48 KFL phone set was folded down to the standard 39 phone set for scoring. In order to reduce the computation time, a maximum possible duration, t_{max}, in any one state was set. For these experiments a maximum duration of 40 frames was used. This was assumed to be sufficient for all models other than the silence model. Hence it was necessary to map multiple observations of silence onto a single observation. All the results given include the SA sentences in both training and testing.

Dialect region 1 was used for the experiments. The *sub* subscript indicates that the models were generated using the suboptimal approximate training routine. From the results in table 3, it can be seen that the SHMM does not improve the performance over a standard single mixture HMM. In addition the optimal training was superior to the suboptimal training. Though the recognition results are worse for the SHMM, the log likelihoods for the test sentences were higher than for the standard HMM.

Model	*No. Mixtures*	*%Correct*	%Accuracy
HMM	1	56.87	50.94
$SHMM_{sub}$	1 (1)	53.98	47.61
SHMM	1 (1)	55.99	49.87

Table 3. Recognition performance on Dialect Region 1

8. CONCLUSIONS

A new acoustic model for speech has been proposed, the SHMM. Both re-estimation formulae and recognition algorithms have been derived for this new model. We have shown that the standard HMM, is a subset of the SHMM. For synthetic data the new model has been shown to perform better on a wider set of acoustic waveforms than standard HMMs, even when the standard HMM has over twice the number of parameters. However preliminary results on real data are disappointing. The most likely reason for the poor performance on real data is the assumption that the segment means are independent. For the synthetic data this is true. However, for real speech data this is unlikely to be true, but cannot be modelled by the present SHMM. Modifications to the SHMM to allow for this problem are presently being worked on.

ACKNOWLEDGEMENT

The authors are grateful to Dr. M.J. Russell for many helpful discussions about the work presented here. M. Gales is funded by a SERC studentship and a CASE award with DRA Malvern.

REFERENCES

[1] Gauvain J and Lee C. Improved acoustic modelling with bayesian learning. In *Proceedings ICASSP*, pages 481–484, 1992.

[2] Baum L.E., Petrie T., Soules G., and Weiss N. A maximisation technique occuring in the statistical analysis of probabilistic functions of Markov chains. In *Ann. of Math. Stat.*, volume 41, pages 164–171, 1970.

[3] Ostendorf M. and Roukos S. A stochastic segment model for phoneme-based continuous speech recognition. In *IEEE Transactions ASSP*, volume 37, pages 1857–1869, 1989.

[4] Russell MJ. A segmental statistical model for speech pattern recognition. In *Proceedings IOA*, volume 14, pages 503–510, 1992.

[5] Brown P.F. The acoustic-modelling problem in automatic speech recognition. Technical report, IBM Thomas J. Watson Research Centre, August 1987.

[6] Furui S. Speaker independent isolated word recognition using dynamic features of speech spectrum. In *IEEE Transactions ASSP*, pages 52–59, 1986.

[7] Levinson SE. Continuous speech recognition by means of acoustic/phonetic classification obtained from a hidden Markov model. In *Proceedings ICASSP*, pages 93–96, 1987.

[8] Young SJ. *HTK Version 1.4: Reference Manual and User Manual.* Cambridge University Engineering Department Speech Group, 1992.