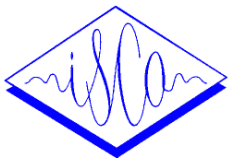

2nd European Conference on
Speech Communication and Technology
EUROSPEECH '91
Genova, Italy, September 24-26, 1991

Energy, Duration and Markov Models

P. Kenny, S. Parthasarathy, V.N. Gupta*, M. Lennig*, P. Mermelstein*, and D. O'Shaughnessy*
INRS–Télécommunications, Montreal, Quebec, Canada

Abstract

We present a new stochastic model for the energy and duration of phone segments which takes account of the speech rate, the loudness of the signal and the effects of stress and pre-pausal lengthening and we show how the block Viterbi decoding algorithm can be used to integrate it with phone-based HMM speech recognizers. The model has been implemented on an isolated-word data-base and a preliminary experiment gives a modest improvement in word recognition accuracy.

I. Introduction

The success of hidden Markov models in automatic speech recognition demonstrates the effectiveness of the Markov property in modelling the time evolution of frame-based features of the speech signal, but a different approach is needed if we are to develop systems capable of extracting the linguistic information encoded in longer-term features such as phone durations and energy and pitch contours. The importance of longer-term features for automatic speech recognition is demonstrated by the results in [5] where the imposition of hard energy and duration constraints on phone segments in Viterbi decoding leads to a substantial increase in recognition rates. It is reasonable to expect that if a stochastic model for these features could be constructed, it would yield better results than hard constraints. This is a difficult problem since a good model has to take account of supra-segmental factors such as speaking rate, loudness and stress and it ought to be constructed in such a way that it can be integrated with frame-based hidden Markov models. (Models which are implemented as post-processors to HMM recognizers are sub-optimal.)

The deficiencies of HMMs in modelling phone durations have long been recognized. It has often been pointed out that the implicit assumption of a geometric distribution on state occupancy times is unrealistic, but this does not preclude the possibility that duration distributions can be faithfully modelled by the transition probabilities of a Markov chain having multiple states. In fact Crystal and House [2] have shown that phone duration histograms can be fitted very closely by this method and the expanded state Markov model of Russel and Cook [9] uses this idea to effectively get around the the assumption of a geometric distribution on the time spent in a HMM state.

However the transition probabilities of Markov chains trained by the Baum-Welch algorithm do not provide a good model for phone durations. For instance, we have observed that while the mean duration for phones predicted by the transition probabilities of phone models are in close agreement with observation, the variances are greatly overestimated. The problem is that if phones are modelled by multi-state HMMs, then the length of time a model spends in a given state is determined more by the match of the frame data with the corresponding output distribution than by the probability of the self-loop transition associated with the state.

Thus it is necessary to divorce the duration model from the spectral matching mechanism. In [8] Levinson accomplishes this in the semi-Markov model framework by the device of using a single state for each phone model. This is unsatisfactory since the spectral matching capability of the phone HMMs is obviously compromised. On the other hand in [6] we suggested a new way of doing this based on the block Viterbi algorithm which allows spectral matching to be done with multi-state phone models. In this paper we will explore the possibility of using this idea to model both energy and duration in a HMM recognizer in a way which takes account of supra-segmental features.

The Γ distribution is commonly used to model phone durations [1, 8] but we believe that the log normal distribution may be a more suitable choice. Like the Γ distribution this is a two-parameter unimodal distribution concentrated on the positive time axis. In so far as the mean and variance of the Γ distribution are independent of each other whereas the variance of the log normal distribution increases with its mean, the log normal distribution is less flexible than the Γ distribution. However, as pointed out by Crystal and House [3], the assumption that the variance increases with the mean is actually quite appropriate for modelling phone durations. Statistical estimation procedures for log normal distributions are much simpler than for Γ distributions (see [8]) and if a multiplicative model for durations such as Klatt's [7] is assumed, it is natural to use log normal distributions since standard linear statistical methods can then be applied. This is an important consideration since it is not very realistic to model phone durations in a context-independent fashion and multiplicative models provide an economical way for dealing with the numerous contextual factors which are known to be involved.

In this article we will show how the multiplicative assumption together with log normal distributions can be used to construct a stochastic model for phone durations which takes account of speaking rate and stress. As explained in section III, we will consider the speaking rate as a syllable-level feature, and assume that the expected duration of a phone in a given syllable is inversely proportional to the speech rate in the syllable. The reason for treating the speech rate as a syllable-level feature rather than as a characteristic of the utterance as a whole (as is usually done) is to provide a mechanism for capturing the effects of stress and pre-pausal lengthening on the durations of phones within a syllable. (For instance, phones in stressed syllables are generally longer than phones in unstressed syllables, no matter what the overall speech rate is.) The effects of the overall speech rate on phone durations are accounted for by conditioning the probability distribution of the rate feature in a given syllable on the observed value of the rate feature in the following syllable, as explained in section IV.

The energy contour is another aspect of the speech signal which is inadequately accounted for by HMMs. In our applications, we have not found the per frame log energy to be a useful parameter; this is not surprising since its value depends on recording conditions and there is no clearly satisfactory way of normalizing for these. On the other hand the corresponding dynamic parameter (obtained by differencing the log energies over a 40 ms interval) does not suffer from this drawback and proves to be very useful. Including this parameter in the feature set enables phone models to track the shape of the energy contour within phone segments but not across segment boundaries, so the

[1] This work was supported by the Natural Sciences and Engineering Research Council of Canada.

* Also with Bell-Northern Research, Montreal.

problem of exploiting information contained in the relative energies of neighbouring phone segments remains to be solved.

The problem of energy modelling is therefore similar to that of duration modelling taking account of the speech rate in that in order to score both of these features it is necessary to compare them with the features extracted from neighbouring segments. We will propose a multiplicative model for the mean energy of phone segments which is also based on the log normal distribtion and takes account of the overall loudness or 'volume' (as defined in section III) of the speech signal and the effect of stress. Like the rate, we will treat the volume as a syllable-level feature, rather than as a characteristic of the utterance as a whole. (Stressed syllables are generally louder than unstressed syllables, independently of the overall loudness of the utterance.) In a given syllable, the expected value of mean energy of the frames in a phone segment is assumed to be proportional to the value of the volume feature in the syllable. The effects of the overall loudness of the utterance on the energy of the individual phone segments is accounted for by conditioning the probability distribution of the volume feature in each syllable on the observed value of the volume feature in the following syllable.

II. The block Viterbi algorithm

The basic ingredient in training and recognition algorithms for phone-based statistical speech recognizers is a method for finding the path through a phonetic graph which best accounts for a given sequence of acoustic observations $Y_1, \ldots, Y_T$. Each branch in the graph is a triple (n, f, n') where n and n' are nodes and f is a phone label. We can construct a HMM corresponding to the graph by replacing each of its branches by a copy of the corresponding phone model. The problem is to find the Viterbi path through this HMM.

In [6] we describe a method for doing this called the block Viterbi algorithm which, unlike the standard Viterbi algorithm, enables us to impose hard constraints on phone durations. As we shall see, this algorithm can be extended to accommodate stochastic models for segment-level features as well as hard constraints.

We first describe the block Viterbi algorithm. For each phone f, we construct a 'point score' function h^f which represents the cost of recognizing the phone f starting at a given time. For $1 \le t < t' \le T$ we provisionally define $h^f[t, t']$ to be the Viterbi score of the data $Y_t, \ldots, Y_{t'}$ against the model for f. Note that for each t this function can be evaluated for all t' in the interval $[t+1, T]$ using a single trellis calculation (by the ordinary Viterbi algorithm) and we can impose maximum and minimum constraints on the duration of f by setting $h^f[t, t']$ to be 0 for all values of t' for which the duration (namely $t' - t + 1$) is inadmissible.

Viterbi decoding can be conducted either forwards or backwards in time. We present only the forward version of the algorithm and for simplicity we will assume that the graph we are searching corresponds to a single phonetic transcription $f_1 \ldots f_K$. That is, there are $K+1$ nodes $n_0, \ldots, n_K$ in the graph and the branches are (n_{k-1}, f_k, n_k), $k = 1, \ldots, K$.

For each $k = 1, \ldots, K$ and each time $t = 1, \ldots, T$ define the forward probability $\alpha_t(n_k)$ to be the Viterbi score of the data $Y_1, \ldots, Y_t$ on the optimal path ending at node n_k at time t. Define $\tau_t(f_k)$ by the condition that the time at which the optimal state sequence enters the model for f_k is $\tau_t(f_k) + 1$.

The forward probabilities $\alpha_t(n_k)$ for each of the nodes n_k can be calculated recursively for $t > 0$:

$$\alpha_t(n_k) = \max_{t' < t} \alpha_{t'}(n_{k-1}) h^{f_k}[t'+1, t]$$

and

$$\tau_t(f_k) = \operatorname*{argmax}_{t' < t} \alpha_{t'}(n_{k-1}) h^{f_k}[t'+1, t].$$

Once the forward recursion has been carried out, the Viterbi segmentation of the utterance $Y_1, \ldots, Y_T$ is obtained as follows. Set $t_K = \tau_T(f_K)$ and, for $k < K$, set $t_k = \tau_{t_{k+1}}(f_k)$. Then for each $k = 1, \ldots, K$, the segment corresponding to f_k starts at time $t_k + 1$ The Viterbi score of the utterance is just $\alpha_T(n_K)$.

The block Viterbi algorithm does not require that the phone poin scores be generated by a hidden Markov model (for instance, it can b used with a stochastic segment model [4]). In this paper we take advan tage of this fact to incorporate segmental and supra-segmental feature into the procedure for scoring an utterance against a transcription. I particular, we will treat the sequence of spectral envelopes in a phon segment as being statistically independent of its duration and energ so that the point scores for a phone f are obtained by multiplyin the spectral match (calculated with a HMM) and the energy-duratio score (calculated using log normal distributions).

The spectral features Y_t that we use are vectors of the form $(c_1, \ldots, c_7, \Delta c_0, \ldots, \Delta c_7)$ extracted every 10 ms; here, $c_0, \ldots, c_7$ ar mel-based cepstrum coefficients and $\Delta c_0, \ldots \Delta c_7$ are the correspond ing difference coefficients calculated over an interval of 40 ms. Not that the loudness c_0 is used as a dynamic parameter but not as a stati parameter.

In matching a sequence of spectral features $Y_t, \ldots, Y_{t'}$ with phone model f we have to be careful to factor out the contribution o the duration $t' - t + 1$ to the Viterbi score (otherwise the duration wil be scored twice). We do this by representing each phone by a networ of states having a standard left-to-right topology but no transitio probabilities. In generating a sequence of spectral features we assum that all paths through the model of a given duration d are a prior equally likely, having probability $\frac{1}{N_f(d)}$ where $N_f(d)$ is the number o paths through the model of length d. Thus the score of the spectra features $Y_t, \ldots, Y_{t'}$ given that the duration of the segment is $t' - t + 1$ is

$$N_f(t' - t + 1) V(Y_t, \ldots, Y_{t'} | f)$$

where $V(Y_t, \ldots, Y_{t'} | f)$ is the likelihood of the data on the best path through the model for f of length $t' - t + 1$. This can be calculated using the standard Viterbi algorithm by setting the 'probability' o each transition in the phone model to be 1. (Note that in Levinson's model the need for the correction term does not arise since, in this case, there is only one path through the model of a given duration.)

The segment-level features that we use are the duration d, namely $t' - t + 1$, and the mean energy e, calculated by averaging the quantity 10^{c_0} from t to t', and for each phone f we define the point scores by

$$h^f[t, t'] = P(e, d) N_f(t' - t + 1) V(Y_t, \ldots, Y_{t'} | f)$$

where $P(e, d)$ is the joint pdf of e and d. In the next two sections we explain how to construct a model for $P(e, d)$ using two supra-segmental features called the rate and volume which are modelled at the syllable level, taking stress into account.

III. Segmental Features — Energy and Duration

We wish to model the duration and mean energy of a phone using log normal distributions whose parameters depend on two syllable level features which we call the rate and the volume. To extract these features, we need to know the phonetic transcription of the syllable S, and its endpoints t_1 and t_2. In the present section we will assume that the end-points of the syllable are given.

Let d_S be the duration of the interval $[t_1, t_2]$ and let $\hat{e}_S$ be the mean energy of the frames in the interval. Let f be a phone in the syllable and let d_f be its duration. We will assume that if r_S is the speaking rate in the syllable then the expected value of d_f is inversely proportional to r_S with a constant of proportionality which is characteristic of the phone:

$$E(d_f) = \frac{\delta_f}{r_S}. \qquad (1)$$

(The dimensionless quantity δ_f corresponds to Klatt's 'inherent duration'.)

If d_S denotes the duration of the syllable then

$$E(d_S) = \sum_f E(d_f) = \frac{1}{r_S}\sum_f \delta_f$$

where the sum extends over all of the phones in S. Hence we can estimate r_S by setting $E(d_S) = \hat{d}_S$:

$$r_S = \frac{\sum_f \delta_f}{\hat{d}_S}$$

Let $D_f = \log d_f$ and $R_S = \log r_S$. We assume that for each phone f there are numbers Δ_f and $\sigma_{d,f}$ such D_f has a normal distribution with mean $\Delta_f - R_S$ and standard deviation $\sigma_{d,f}$. A straightforward calculation shows that (1) holds with

$$\delta_f = e^{\Delta_f + \frac{1}{2}\sigma^2_{d,f}}.$$

Now let e_f be the mean energy in the phone segment. We assume that

$$E(e_f) = v_S\gamma_f \tag{2}$$

where γ_f is a gain factor associated with the phone and the volume v_S is interpreted as the mean energy in the syllable corrected for its phonetic transcription. (Like the speaking rate, the volume is assumed to be constant within syllables but may vary from one syllable to the next.)

Assume that for each phone f in the syllable, d_f and e_f are conditionally independent given r_S and v_S[1]. If te_S denotes the total energy in the segment corresponding to the syllable,

$$E(te_S) = \sum_f E(d_f e_f) = \sum_f E(d_f)E(e_f) = \frac{v_S}{r_S}\sum_f \gamma_f\delta_f$$

so we can estimate v_S by setting $E(te_S) = \hat{d}_S\hat{e}_S$. This gives

$$v_S = \frac{\sum_f \delta_f}{\sum_f \delta_f\gamma_f}\hat{e}_S$$

Let $V_S = \log v_S$ and $E_f = \log e_f$. We assume that for each phone f there are numbers Γ_f and $\sigma_{e,f}$ such that E_f is normally distributed with mean $V_S + \Gamma_f$ and standard deviation $\sigma_{e,f}$. In order for (2) to hold,

$$\gamma_f = e^{\Gamma_f + \frac{1}{2}\sigma^2_{e,f}}.$$

Thus if we are given that the endpoints of the syllable are t_1 and t_2 we can estimate the rate and volume statistics $R_S[t_1,t_2]$ and $V_S[t_1,t_2]$ using the transcription of the syllable and use them to determine the probability distribution for the duration and mean energy of each of the phone segments in the syllable. Let $X_S[t_1,t_2]$ be the 2×1 vector with components $R_S[t_1,t_2]$ and $V_S[t_1,t_2]$.

For each phone f in the syllable, the point scores are given by

$$h^f[t,t'] = P(D_f = D[t,t']|R_S[t_1,t_2])P(P_f = P[t,t']|V_S[t_1,t_2]) \times N_f(d)V(Y_t,\ldots,Y_{t'}|f)$$

for $t_1 \le t \le t' \le t_2$.

Once the point scores for each of the phones in the syllable have been calculated, the total match of the interval $[t_1,t_2]$ with the syllable S, $H^S[t_1,t_2]$, and the corresponding Viterbi alignment can be found by the block Viterbi algorithm.

IV. Syllabic Features — Rate and Volume

In the preceding section we treated the rate and volume features associated with a syllable as statistics whose values were determined by the phonetic transcription and the end-points of the syllable. In order to complete the description of the model we have to specify how the rate and volume features associated with the different syllables in an utterance are distributed.

We implemented the model on a speaker-dependent data base consisting of isolated words drawn from a 60,000 word vocabulary which have been automatically end-pointed (so there is no silence model in the phone inventory). The transcriptions of the words in the dictionary include syllable markers corresponding to three degrees of stress. Since the rate feature should be differently distributed for syllables in word-final position than for syllables in other positions we distinguished six types of syllable in all.

In order to take account of the speech rate in the utterance as a whole, we assume that if a syllable is not in word final-position then the distribution of the rate statistic depends on the observed value of the rate in the following syllable. (For instance, if the speaking rate in the following syllable is less than expected the same should be true in the current syllable as well.) We use a similar approach in modelling the volume feature, in order to take account of the overall loudness of the utterance. (If the volume in the following syllable is less than expected, the same should be true of the current syllable.) In word-final position we just use an a priori distribution for the rate-volume statistics for each stress level. (In particular, the rate-volume statistics in monosyllabic words are treated this way. This is reasonable since the they cannot be expected to yield much useful information in recognizing monosyllabic words.)

Thus with each stress level s we associate two 2×1 mean vectors. $\mu_1(s)$ and $\mu_2(s)$ and two 2×2 covariance matrices $\Sigma_1(s)$ and $\Sigma_2(s)$. We assume that

1. if S is a syllable in word-final position with stress s then X_S is normally distributed with mean $\mu_1(s)$ and covariance matrix $\Sigma_1(s)$.
2. if S is a syllable of stress s which is not in word-final position and S' is the following syllable, the conditional distribution of X_S given $X_{S'}$ is normal with mean $\mu_2(s) + X^0_{S'}$ and covariance matrix $\Sigma_2(s)$ where $X^0_{S'}$ is the deviation of $X_{S'}$ from its expected value.

Now suppose that the transcription of an utterance $Y_1,\ldots,Y_T$ is given in the form of a string of syllables $S_1 \ldots S_N$ with corresponding stresses $s_1,\ldots,s_N$. We can use the block Viterbi algorithm to incorporate the rate-volume statistics into the decoding at the syllable level in exactly the same way as we incorporate the energy-duration statistics into the decoding at the phone level.

For $t = 0,\ldots,T-1$, define $\beta_t(S)$ to be the joint likelihood of $Y_{t+1},\ldots,Y_T$ on the best path which starts at the begining of S at time $t+1$. The β's can be calculated recursively as follows; the Viterbi score of the utterance is just $\beta_0(S_1)$.

If S is the word-final syllable then

$$\beta_t(S) = P(X_S[t+1,T])H^S[t+1,T].$$

If S is not the word-final syllable,

$$\beta_t(S) = \max_{t'>t} P(X_S[t+1,t']|X_{S'}([t'+1,t''])H^S[t+1,t']\beta_{t'}(S')$$

where S' is the syllable following S and for each t', t'' denotes the end point of S' on the optimal path starting at the begining of S' at time $t'+1$.

V. Implementation on an isolated word data-base

As mentioned in section IV we implemented the model on an isolated-word data-base using transcriptions containing primary, secondary and tertiary stress markers. Spectral matching was done with left-to-right hidden Markov models whose output distributions are Gaussian mixtures having up to 25 components. We experimented with data from a single speaker using 2,705 words to train the model. The parameters of the mixture HMMs together with the parameters of the duration, energy, rate and volume distributions were estimated by Viterbi alignment of the words in the training set with their transcriptions; we are prevented by lack of space from giving the re-estimation formulas here.

Some idea of how well the model fits the data can be obtained by inspecting the parameters γ_f, δ_f, $\mu_1(s)$ and $\mu_2(s)$.

When phones are ranked in order of decreasing inherent energy (γ_f) we obtain the following series: ʊ, ɑr, ɛ, ʌ, æ, ɑ-ɔ[1], aj, ɔr, aw, ɪ, o, u, e, ə, i, r, w, j, l, ɔj, m, ʒ, ŋ, initial breath noise, n, b, θ, dʒ, g, ð, d, t, v, f, k, ʃ, h, z, p, s, tʃ, final breath noise. Obviously ʊ should not appear on the top of the list and we would expect weak fricatives to appear on the bottom, but otherwise the ranking is quite reasonable since we have open vowels > close vowels > liquids and glides > consonants. (The anomalous position of ɔj is accounted for by the fact that there are only four tokens in the training set.)

Ranking phones in order of decreasing inherent duration (δ_f) gives: aw, s, ɔr, ʃ, u, aj, o, tʃ, e, z, ɑr, ʒ, i, ɑ-ɔ, æ, θ, k, f, dʒ, ŋ, m, p, ɔj, t, ɛ, n, ʌ, v, h, l, ɪ, ʊ, g, w, r, d, j, ð, b, ə, initial breath noise, final breath noise. This ranking seems reasonable in that generally speaking we have diphthongs > long vowels > short vowels > liquids and glides, and voiceless obstruents are always predicted to be longer than the corresponding voiced obstruents.

Looking at the parameters of the distributions for the volume feature we find that the model correctly predicts that syllables carrying primary stress will be louder than those carrying secondary stress and that these will be louder than syllables carrying tertiary stress. This is true both of syllables in word-final and non-word-final position and it confirms that the effects of stress on energy are being modelled appropriately.

Similarly, the effects of pre-pausal lengthening on syllable durations are captured by the model: the parameters of the distribution for the rate feature show that for each of the three stress levels the model predicts that syllables in word-final position will be longer than those in non-word-final position.

As for the effects of stress on duration, the parameters of the distribution for the rate feature show that the model correctly predicts that syllables carrying primary stress will be longer than syllables carrying secondary stress, but it does not predict that these will be longer than syllables carrying tertiary stress. This pattern is observed both in word-final and non-word-final positions, and it does not agree with the data in Crystal and House [3]. (However there is a substantial overlap between the two distributions both in our model and the Crystal and House data.)

At the time of writing, we are not in a position to evaluate the usefulness of the new model for speech recognition. However we have carried out a preliminary experiment on the isolated word data base. We took a set of 458 words uttered by the same speaker and first ran a benchmark experiment using a set of phonetic mixture HMMs and the A* algorithm described in [6] together with the 60,000 word dictionary to generate 25 recognition hypotheses for each of the words in the test set. The correct word was included in the top 25 choices 95.9% of the time and the recognition rate using the phonetic HMMs without the energy and duration model was 73.8%. When the top 25 choices for each word were rescored using the new model the recognition rate increased to 74.7%. Of course, this result is inconclusive and we plan to carry out much more extensive tests.

[1] ɑ and ɔ are merged.

VI. Conclusion

We have presented a new stochastic model for the energy and duration of phone segments which takes account of the speech rate, the loudness of the signal and the effects of stress and pre-pausal lengthening. We have shown how the model can be completely integrated with hidden Markov models in training and recognition by means of the block Viterbi algorithm. The model has been implemented on an isolated-word data-base and a preliminary experiment has led to a modest improvement in word recognition accuracy. We chose to work with the isolated word data-base since it made it easy to debug the implementation to obtain recognition results without a complicated search strategy. However it is clearly not an appropriate test bed since most of the words in it are monosyllabic. We believe that the model will yield larger improvements in recognition accuracy under conditions which more closely resemble natural speech.

References

[1] T.H. Crystal and A.S. House, "Segmental durations in connected-speech signals: Preliminary results", *Journal of the Acoustical Society of America*, **72**, pp. 705—716, 1982.

[2] T.H. Crystal and A.S. House, "Segmental durations in connected-speech signals: current results", *Journal of the Acoustical Society of America*, **83**, pp. 1553—1573, 1988.

[3] T.H. Crystal and A.S. House, "Segmental durations in connected-speech signals: syllabic stress", *Journal of the Acoustical Society of America*, **83**, pp. 1574—1585, 1988.

[4] V. Digalakis, M. Ostendorf and J.R. Rohlicek, "Fast Search Algorithms for Connected Phone Recognition Using the Stochastic Segment Model", *Proceedings of the DARPA Speech and Natural Language Workshop*, pp. 173—178, June 1990.

[5] V.N. Gupta, M. Lennig, P. Mermelstein, P. Kenny, F. Seitz and D. O'Shaughnessy, "Using phoneme duration and energy contour information to improve Large Vocabulary Isolated Word Recognition", in press, *Proc. ICASSP* , 1991.

[6] P. Kenny, R. Hollan, V. Gupta, M. Lennig, P. Mermelstein and D. O'Shaughnessy, "A*-admissible heuristics for rapid lexical access", submitted to *IEEE Transactions on Acoustics, Speech and Signal Processing.*

[7] D.H. Klatt, "Interaction between two factors that influence vowel duration", *Journal of the Acoustical Society of America*, **54**, pp. 1102—1104, 1973.

[8] S.E. Levinson, "Continuously Variable Duration Hidden Markov Models for Speech Analysis", *Computer Speech and Language*, vol. 1, no. 1, pp. 29—46, March 1986.

[9] M.J. Russell and A.E. Cook, "Experimental evaluation of duration modelling techniques for automatic speech recognition", *Proc. ICASSP*, pp. 2377—2379, 1987.

[10] M.J. Russell and R.K. Moore, "Explicit modelling of state occupancy in hidden Markov models for automatic speech recognition", *Proc. ICASSP*, pp. 5—8, 1985.