Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institut für
Technische Informatik und
Kommunikationsnetze

Hidden Markov Models for Speech Segmentation

Patrik Wyss

Semester Thesis SA-2010-10

Speech Processing Group

Computer Engineering and
Networks Laboratory (TIK)

Department of Information Technology
and Electrical Engineering (D-ITET)

Advisors:

S. Hoffmann and Dr. B. Pfister

Supervisor:

Prof. Dr. L. Thiele

December 2010

Abstract

Statistical models are important tools in speech processing. The task of this Semester Thesis was to determine the influence of the underlying statistical model on speech segmentation and to evaluate the precision of phone boundaries for the segmentation.

At first, Hidden-Markov Models were adjusted. In a second step, some modifications of Hidden Semi-Markov Models were implemented. Additionally, the comparison of the results of these implementations with the manually perfectly labelled references was made.

In this thesis, the following was established: None of the examined models led to overall improvements for all transitions. The transitions [n] → [h], [n] → [w], fricatives → [@_r], plosive coronal → fricative, [n] → fricative coronal, [m] → fricative coronal and diphthong to fricative labial showed the best set labels with the original Hidden-Markov Model. With an equally distributed transition probability in the Hidden-Markov Model, the labels were set on a more exact basis for the transition [>] → plosive dorsal. Weighting the observation probability and the transition probability in the Hidden-Markov Model led to more precisely set labels for the transition fricatives coronal → [w]. With Hidden Semi-Markov Models the labels for the transitions to [j] were improved majorly. Furthermore, there were some improvements ascertained for the transitions [>] → plosive labial, [>] → plosive dorsal, plosive dorsal → [sil] and plosive coronal → [sil]. Finally, the labels of the silence model were largely improved when the duration had not been bounded.

Acknowledgement

I would like to thank Prof. Dr. Lothar Thiele and Dr. Beat Pfister for offering me the opportunity to write this Semester Thesis at the Speech Processing Group of the Computer Engineering and Networks Laboratory (TIK) at ETH Zurich.

I would also like to thank my advisor, Sarah Hoffmann, for her guidance throughout this research. Without her support this Semester Thesis would not have been possible.

Contents

1 Introduction

Speech processing divides sentences into smaller parts such as words. A word in turn can be subdivided in so called phones. Phones are the smallest unit for which differences in utterances can be distinguished.

To describe speech, statistical models need to be applied. An important condition has to be fulfilled before the models can be used: The model has to be trained or one has to conduct a training with a part of the existing data before its use. In a usual training session, different training iterations are needed to obtain a satisfying result. To be able to train the parameters of a statistical model, pre-labelled speech material is required, in which the phones need to be located as accurately as possible.

The most qualified models to solve this task are Hidden-Markov Models (HMM). A Hidden-Markov Model is a combination of a hidden sequence state and a visible observation. A tighter explanation of Hidden Markov Models is given in section 1.1.1. The state duration of a HMM is implicitly a geometric distribution [Yu,09]. This might be inappropriate for natural speech or short utterances. Therefore, one requests an explicit implementation of the duration. Models considering this aspect are the so called Hidden Semi-Markov Models (HSMM), which will be outlined more precisely in section 1.1.2.

The aim of this thesis is to determine the influence of the underlying statistical model for the speech segmentation. The existing Hidden-Markov Model segmentation is considered regarding the precision of the segmentation of the transitions of phones, as well as the transitions among the classes. The focus will lie on the determination of which phoneme transitions cause difficulties.
An additional Hidden Semi-Markov Model (HSMM) with some modifications is implemented for speech segmentation to evaluate the progress of the model in comparison to the Hidden-Markov Models (HMM). By comparing the accuracy of the new segmentation with the one of the manually segmented labels, the improvement of the models can be measured.

This thesis is partitioned as listed below.
In chapter 2, an overview of the different Hidden Semi-Markov Models and their applications in speech recognition and speech synthesis will be given.
Thereafter, in chapter 3, the implementation of the system will be described. All results will be discussed in section 4.

1.1 Models and Algorithms

The most important models will be the Hidden-Markov Models (HMM) and the Hidden Semi-Markov Models (HSMM) described in the following.

1.1.1 Hidden-Markov Models (HMM)

Hidden-Markov Models (HMM) are originally made for detecting statistic events. Figure 1 shows a general Hidden-Markov Model. It is a double stochastic process, where the circles denote the states and the boxes illustrate the observations.

The actual state has an effect on the observation, but the states sequence is not observable. For that reason, they are called *Hidden* Markov Models. The probability $a_{(i,j)}$ is the transition probability to change from state i to state j and $b(i)$ indicates the observation probability for the observation $o(i)$.

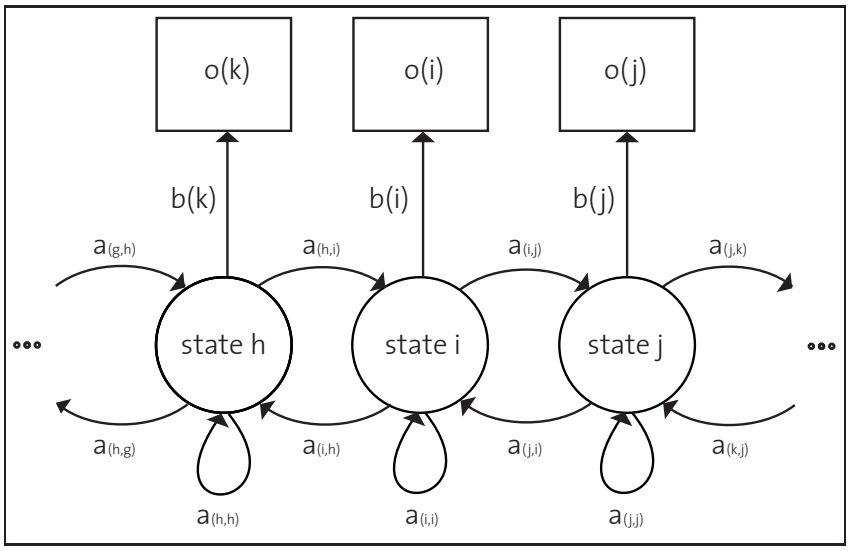

Figure 1: *This is an example of a general Hidden-Markov Model (HMM)*

The advantages of using Hidden-Markov Models are on the one hand that they can represent speech as probability distributions. On the other hand, efficient algorithms (such as the Baum-Welch- or the Viterbi-Algorithm) are provided for estimating the model parameters.

With the Hidden-Markov Models, the state duration probabilities are implicitly modelled by their state transition probabilities. This means, that the duration probability decreases exponentially with time. The found phone boundaries are very imprecise. Thus, one of the major drawbacks is the duration modelling. In practice, a segmentation made with Hidden-Markov Models corresponds to an exponentially distributed phone length.

To avoid this disadvantage, Hidden Semi-Markov Models (HSMM) have been developed and will be described next.

1.1.2 Hidden-Semi-Markov Models (HSMM)

Hidden Semi-Markov Models (HSMM) allow the underlying process to be a Semi-Markov chain. Unlike HMM, with HSMM the duration is explicitly considered and consists of a variable state duration time. Other distributions of phone length such as Gaussian, Gamma or Poisson distributions can be modelled. Other features of phones could as well be included. However, this lies beyond the scope of this thesis.

In figure 2, a general Hidden Semi-Markov Model is shown: The important differences to HMM are that on the one hand the transition probability to change from state i to state j is now depending on both the duration in state i and in state j. On the other hand, the probability of a change between the same states is zero.

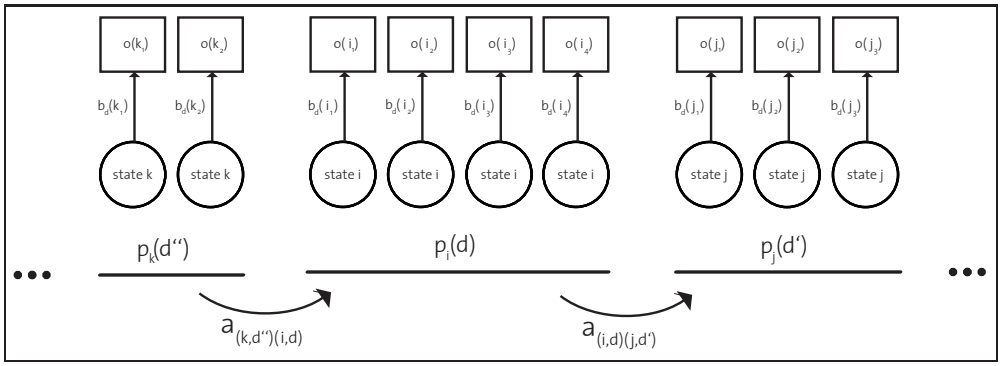

Figure 2: *This is an example of a general Hidden Semi-Markov Model (HSMM)*

The probability of duration being in state i is denoted by $p_i(d)$.

Since the Hidden Semi-Markov Models fit the requirements of consideration of duration and the possibility to change the distributions of observations, they are widely used in speech synthesis. In this work, they will be used for the speech segmentation processing.

1.1.3 Algorithms

Viterbi-Algorithm
The Viterbi-Algorithm deduces the most probable sequence of a given sequence. The formulae are shown in equations (1) - (7) on page 17. A complete overview and the derivation of the Viterbi-Algorithm for Hidden-Markov Models is given in [PK08].

On-line algorithms
Passing all sentences, the mean and variance values are computed with the on-line algorithm proposed by Knuth[1]. For the re-estimation of the parameters for the Gamma distribution, the on-line algorithm proposed by Choi and Wette[2] is used.

1.2 Data description

For this semester thesis, a proprietary corpus of spoken sentences is used. The set of sentences includes 401 sentences in English spoken by one female person.

A list of all phone models used in this thesis is available in the appendix in table 5 in section A.1.

[1]Knuth, D. (1998) , *The Art of Computer Programming*
[2]Choi, S.C. and Wette, R. (1969), *Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias*

2 Related Work

The following articles are the underlying related work for this semester thesis.
Initially, an overview of the work done in the area of Hidden-Markov Models and Hidden Semi-Markov Models is given, which summarises the most important papers concerning the speech segmentation and speech reconstruction. Various approaches have previously been used in this area. The most important ones will be described now.

Levinson described in [Lev86] the modifications made with Hidden-Markov Models using continuous probability density functions for duration. He used the family of Gamma distributions, provided the formulae for forward and backward likelihoods and showed how to estimate the parameters. For the five-state-three dimensional case he achieved convergence with respect to the correct values in seven to ten iterations. He concluded that based on the results obtained, the Gamma distribution was best suiting the durational density. The Gamma function has only two parameters defining both mean and variance.

Nakagawa and Hashimoto described in [NH88] a statistical method of segmentation using a Hidden-Markov Model (HMM) and a Bayes' classifier. Only one HMM represented all phone models. A Gamma distribution approximated the distribution of duration whereby a duration control mechanism was adapted for each state. The maximum duration of a state was among 5 to 10 frames. The Baum-Welch algorithm estimated the parameters. The Viterbi algorithm found the optimal HMM sequence with back tracing. As a result, the rate of segmentation was more than 92% for two male speakers and an improvement up to 97.5% was reached using the duration control mechanism based on a discrete probability distribution. Almost all diphthongs were correctly segmented as two sounds. The phone boundaries were detected implicitly; they assumed a transition from a consonant to a vowel as a phone boundary.

Codogno and Fissore characterised in [CF87] the duration of sub-word sets by suitable probability density functions. Furthermore, they described state-multiplied first-order Hidden Markov Models with continuous probabilistic density function (SMHMMC). With SMHMMC's the duration function for each state became a negative binomial distribution. Additionally, they replaced each loop-state with a number of replications of the same state and the replication factor controlled the shape of the temporal structure. On the other hand, with continuous variable duration HMMC's (CVDHMMC) the distribution was a Gamma or a Dirac-Delta function and an additional restriction was made forbidding usual loops. Hence, Gamma distribution did not match the model for a very short sound. The conclusion was that manual segmentations of samples must be consistent with the length distribution predicted by the model.

In [GTL91], Gu et al. modelled the state durations of Hidden Markov Models with lower and upper bounding parameters for each state (HMM/BSD) in the recognition phase. The likelihoods were the same as in conventional HMM but the lower and upper portions were removed. This prevented every state from occupying too many or too few states. Gu et al. used four parameters to describe the HMM/BSD: the state transition, the observation production as well as the lower and upper bound for the duration of a state. The training phase adjusted the latter two parameters. The probability distribution function (pdf) was geometrically distributed and a relatively small number of training utterances was needed. The number of states was set to be 6 for all experiments. They used the Mandarin syllables to test and train their model. The important point for comparison was the recognition rate of HMM/BSD versus the recognition rate of a conventional HMM, a HMM with Poisson distribution and a HMM with Gamma dis-

tribution. In other words, by comparing the HMM/BSD with these models, they discovered a significant difference in both the discrete and continuous case between HMM/BSD and HMM with Poisson or Gamma distribution. Additionally, all experiments showed a higher recognition rate in the range of 1.9% to 9% in the discrete case. In the continuous case, where partitioned Gaussian mixture modelling was applied, the improvement of the recognition rate was among 1.8% and 6.3%, shown in Table 1.

Table 1: *HMM/BSD vs. other models*

	HMM/BSD	HMM	HMM with Poisson	HMM with Gamma
discrete	78.5%	-9.0%	-6.3%	-1.9%
continuous (1)	88.3%	-6.3%	-5.9%	-3.1%
continuous (3)	88.8%	-5.0%	-3.1%	-1.8%
continuous (5)	89.4%	-5.5%		

In [RSS92], Ratnayake et al. used a Hidden Semi-Markov Model which is defined by its state transition probability matrix, its observation probability matrix, the set of probability mass distribution of state occupancy and the set of probabilities to be in a certain state. The state occupancies were described by non-parametric distributions, because the parametric distribution (such as Gamma, Poisson and Binomial) were unable to describe the distribution of phones sufficiently. Allophones were combined to one of the total 46 states, if they had similar distributions or similar observation probability distributions. Using the mentioned HSMM increased the phone recognition accuracy from 48.4% of a conventional HMM up to 53.7%. The increased computational complexity was a drawback, but models for decreasing computational load were proposed. Conclusively, non-parametric distributions needed more training data.

Oura et al. presented in [OZN+06] a fully consistent speech recognition system based on the HSMM framework. They modelled the state duration explicitly into the HMM and obtained the Hidden Semi-Markov Model (HSMM). They estimated both the state output and duration probability distributions based on the HSMM statistics, which were calculated by the generalised forward-backward algorithm. With phonetic decision trees, they clustered individually the state output and duration probability distributions. Finally, weighted finite-state transducers (WFSTs) decoded the speech. A mixture of Gaussians was used to model the state output probability function. The state duration probabilities of each HSMM were modelled by a single multivariate Gaussian.

Since the dimensionality of these multivariate Gaussians was equal to the number of states of the HSMM, the covariance matrix was diagonal. The WFST associated weights (such as probabilities, duration and penalties) to each pair of input and output symbol sequences. This lead to the following results: In a speaker-dependent continuous speech recognition experiment, HSMM-based speech recognition system achieved about 5.9% relative error reduction over the corresponding HMM-based one. The improvement of phone accuracy was confirmed by modelling state duration probability distribution with context dependence.

The results of these papers lead to the implementation of a Hidden Semi-Markov Model with Gamma distribution in this thesis. This model is expected to be the most accurate function to approximate the phone models.

In the next chapter, the system parameters for the implementation will be characterised.

3 Design and Implementation

This chapter explains the design and implementation of the models.
The basis for the implementation is the Hidden-Markov Model word recogniser introduced in the speech processing lecture at ETH Zurich. Firstly, a modification needs to be made to adopt the existing template to recognise sentences. This template contains the following procedure: Initially, all sentences will be examined, whereby all phone models will be initialised, before the actual training will be conducted. During the training the new labels will be calculated according to the Viterbi-Algorithm.
Secondly, the segmentation obtained with the adapted Hidden-Markov Model is compared to manually perfect set labels. The starting point is the segmentation made with the Hidden-Markov Models. This case is called *Baseline*.
In a first approach, the segmentation is achieved by using the Hidden-Markov Models with an equally distributed transition probability for every iteration. An equally distributed transition probability can be reached by not updating the transition matrix A. This line of action is called *Approach 1*.
In a second approach, one has to be mindful of the observation probability and the transition probability. In this thesis, every emitting state is assigned to an observation composed of 26 features and only one transition probability which will have to be weighted accordingly. This approach is named *Approach 2*.
For the Hidden Semi-Markov Models, the duration is explicitly considered. A phone model is said to have three emitting states. Therefore, the maximal duration for each phone model to stay in a single state is one out of three of the entire model duration. Thus, in *Approach 3* the duration of a state is assumed to be a third of the maximal duration of all occurring phones for each phone model.
To intensify the influence of duration, the model is requested to have an unbounded duration or at least the duration should be as large as possible. Due to computational reasons, the duration must be a finite number. In *Approach 4* the duration for every state in every model is set to the global maximal duration.
Summarising the four approaches, the following descriptions are introduced:

- **Reference**: These are the manually perfect labels.
- **Baseline**: These are the labels obtained with the Hidden-Markov Model (HMM) adapted of the existing Hidden-Markov Model word recogniser to be able to recognise sentences.
- **Approach 1**: These are the labels obtained with the Hidden-Markov Model (HMM) as the Baseline, however, with an equally distributed transition matrix in each iteration.
- **Approach 2**: These are the labels obtained with the Hidden-Markov Model (HMM) and weighted observation probability and transition probability.
- **Approach 3**: These are the labels obtained with the implemented Hidden Semi-Markov Model (HSMM) with a Gamma distribution.
- **Approach 4**: These are the labels obtained with the implemented Hidden Semi-Markov Model (HSMM) with a Gamma distribution and unbounded duration.

During the evaluation of the several approaches, the position of the labels in the Baseline or one of the approaches are considered as related to the perfectly set labels. For the evaluation, the transitions from a phone to another are taken into account. As an example, all 65 phones occurring in sentence 401 ('She almost danced with joy as the royal corpse was brought back from Kinghorn') are listed in table 2. The first line denotes the phone number and in the corresponding second line the phone is named.

Table 2: *Example: All phones of sentence 401*

1	36	29	21	19	16	20	3	2	7	15	7	23
[sil]	[S]	[i:]	[?]	[O:]	[l]	[m]	[o_U]	[s]	[>]	[t]	[>]	[d]
14	6	2	7	23	18	9	4	7	37	44	1	40
[A:]	[n]	[s]	[>]	[d]	[w]	[I]	[D]	[>]	[d_Z]	[O_I]	[sil]	[q]
12	4	22	35	44	22	16	7	10	19	7	27	2
[z]	[D]	[@]	[r]	[O_I]	[@]	[l]	[>]	[k]	[O:]	[>]	[p]	[s]
18	22	12	7	8	35	19	7	15	7	8	40	7
[w]	[@]	[z]	[>]	[b]	[r]	[O:]	[>]	[t]	[>]	[b]	[q]	[>]
10	32	35	38	20	7	10	9	46	26	19	6	1
[k]	[f]	[r]	[V]	[m]	[>]	[k]	[I]	[N]	[h]	[O:]	[n]	[sil]

Altogether 64 transitions arise. Several transitions occur more than once such as [>] → [t] or [>] → [d]. For each transition the sum of the deviation between the labels of the particular models and the perfectly set labels is formed. By adding up over all 401 sentences, a mean and a variance value for every occurring transition is computed, as well as the total number of occurrence. These values are listed in the tables 7 to 27. The evaluation and the sketch of difficult transitions will take place in chapter 4.

3.1 Implementation

The aim of this section is to show the differences in the implementation of the Baseline and the four approaches. In this thesis a linear Hidden-Markov Model, respectively, a linear Hidden Semi-Markov Model are assumed. No state can be skipped and no state is re-visited; every state is visited and if it has changed to the preceding state it will not in any case change back.

As already mentioned, the Viterbi-Algorithm determines the optimal path for a given sequence. The same Viterbi-Algorithm is used for both the phone models and the sentences composed of several phone models.

Baseline: Viterbi-Algorithm for Hidden-Markov Models
The algorithm has the following components: The joint probability $\delta_t(i)$, the matrix $\Psi_t(j)$ and the optimal path $\hat{q}$. In the matrix $\Psi_t(j)$ the optimal preceding state is stored. The joint probability of the observation sequence X_1^t and the optimal path $\hat{Q}_q^t$ terminating in state S_j at time t is $\delta_t(j) = \max_{\text{all } Q_1^t \text{ with } q_t = S_j} P(\mathbf{X}_1^t, Q_1^t|\lambda)$.

Initialisation:

$$\delta_1(j) = a_{1j} b_j(\mathbf{x}_1), \qquad 1 < j < N \tag{1}$$

Recursion:

$$\delta_t(j) = \max_{1<i<N} [\delta_{t-1}(i) a_{ij}] b_j(\mathbf{x}_t) \tag{2}$$

$$\Psi_t(j) = \arg\max_{1<i<N} [\delta_{t-1}(i) a_{ij}] b_j(\mathbf{x}_t), \quad 1 < i \leq T \quad 1 < j < N \tag{3}$$

Termination:

$$\delta_T(N) = \max_{1<i<N} [\delta_T(i) a_{iN}] \tag{4}$$

$$\Psi_T(N) = \arg\max_{1<i<N} [\delta_T(i) a_{iN}] \tag{5}$$

Back tracing of optimal path:

$$\hat{q}_T = \Psi_T(N) \tag{6}$$

$$\hat{q}_t = \Psi_{t+1}(\hat{q}_{t+1}), \quad 1 \leq t < T \tag{7}$$

Figure 3: *Viterbi-Algorithm formulae for Baseline*

The algorithm steps are listed in figure 3. To illustrate the central aspects: For every state, at every moment the maximal joint probability $\delta_t(j)$ is computed and the predecessor state is stored. All other values are discarded.

Approach 1:
Viterbi-Algorithm for Hidden Markov Models with equally distributed transition probability
To show the differences, only one component is changed between Baseline and Approach 1. The remaining code of the Baseline HMM is unchanged.
To determine the influence of the transition probability matrix, it is equally distributed; the probability to stay in the actual state as well as the probability to change to the next state of the linear HMM is set to 0.5 by not updating the transition probability matrix A.

Approach 2:
Viterbi-Algorithm for Hidden Markov Models with weighted observation probability and transition probability
In Approach 2, the imbalance between the observation probability and the transition probability is considered. In equation (1) to (3) of figure 3 the weighting is introduced: Every observation consists of a feature vector with 26 components such as the MFCC, the derivation of these coefficients or the energy. The transition probability is a single component and has to be weighted accurately by introducing a weight. The initialisation and recursion steps change as seen in formulae (8) to (10) of figure 4.

Initialisation:

$$\delta_1(j) = a_{1j} \cdot \frac{\mathbf{1}}{\mathbf{26}} \cdot b_j(\mathbf{x}_1), \qquad 1 < j < N \tag{8}$$

Recursion:

$$\delta_t(j) = \max_{1<i<N}[\delta_{t-1}(i)a_{ij}] \cdot \frac{\mathbf{1}}{\mathbf{26}} \cdot b_j(\mathbf{x}_t) \tag{9}$$

$$\Psi_t(j) = \arg\max_{1<i<N}[\delta_{t-1}(i)a_{ij}] \cdot \frac{\mathbf{1}}{\mathbf{26}} \cdot b_j(\mathbf{x}_t), \quad 1 < t \leq T \quad 1 < j < N \tag{10}$$

Figure 4: *Adapted formulae for Approach 2*

Approach 3:
Viterbi-Algorithm for Hidden Semi-Markov Models with Gamma duration distribution
In order to use the duration, in a first step the maximal duration of each phone model needs to be determined. Since the phone model has three emitting states, every state can be assumed to have the same duration, in particular a third of the maximal duration.
Up to now, only one observation has been considered for the computation of joint probability δ. Hence, with the Hidden Semi-Markov Model several observations are taken into account for a single state. As in the original Viterbi-Algorithm, the adapted algorithm has the components joint probability $\delta_t(i)$, matrix $\Psi_t(j)$ and optimal path $\hat{q}$. Additionally, a duration matrix Ω and a matrix $\Lambda_t(j, \Omega_t(j))$ are used. The matrix $\Omega_t(j)$ indicates the duration of being in state j at time t. This value lies between 1 and the maximal duration for the actual phone. In $\Lambda_t(j, \Omega_t(j))$, the probability of being in state j at time t for a duration $\Omega_t(j)$ is stored.
For the re-estimation of the parameters α and β of the Gamma distribution $\Gamma(\alpha, \beta)$, three parameter have to be cumulated for every phone model: the number of duration, the sum of the durations and the logarithmic sum of the durations. With these values, the new parameters of $\Gamma(\alpha, \beta)$ approximating the duration distribution can be computed. Supposing a linear Hidden Semi-Markov Model, one knows for a state j that the prior state was $j - 1$ and the successive one will be $j + 1$. The duration after which the transition takes place has to be noted and is stored in $\Omega_t(j)$.
The algorithm steps are described in figure 5.

Approach 4:
Viterbi-Algorithm for Hidden Semi-Markov Models with unbounded duration
To examine the influence of the duration, in this approach the duration is unbounded. No further restrictions are made. For computational reasons the value is set to the maximal occurring one over all models.

The formulae stay the same as in Approach 3.

Initialisation:

$$\delta_1(1) = 1 \tag{11}$$
$$\Omega_1(1) = 1 \tag{12}$$

Recursion:

$$\delta_t(j) = \max_i [\max_{1<d<D} [\delta_{t-d}\Lambda_t(j, \Omega_t(j)) \prod_{t=1}^{d} b_j(\mathbf{x}_t)]] \qquad 2 < i < N-1, \forall i \neq j \tag{13}$$
$$\Psi_t(j) = j - 1 \qquad 2 < j < N-1 \qquad 2 < j < T-1 \tag{14}$$
$$\Omega_t(j) = \arg\max_{1<d<D} \delta_t(j) \tag{15}$$

Termination:

$$\delta_T(N) = \delta_T(N-1) \tag{16}$$
$$\Psi_T(N) = N - 1 \tag{17}$$

Back tracing the optimal path:

$$\hat{q}_T = \Psi_T(N) \tag{18}$$
$$\hat{q}_t = \Psi_{t-d+1}(\hat{q}_{t-d+1}), \quad 1 \leq t < T \tag{19}$$

Figure 5: *Viterbi-Algorithm formulae for Approach 3 and Approach 4*

4 Evaluation

A data sheet of all phone models occurring in the 401 sentences is available in table 5 in appendix A.1. The expression [l] | vowel → [n] | [m] denotes the four transitions [l] → [n] , [l] → [m], vowel → [n] and vowel → [m].
To interpret the results obtained, the phone models are collected in clusters such as vowels, approximants, plosives, fricatives, nasals, trill and silence. Some of them are subdivided into smaller categories. In total, 19 groups exist, all listed in table 6 on page 32.

For a total number of 361 ($19 \cdot 19$) possible transitions among all models, only 211 transitions appear at all. From a statistical point of view, all transitions with a number of occurrence smaller than 10 are omitted, which leads to 160 remaining possible phone model transitions. The table 7 on page 33 in the appendix lists all values.

4.1 Hidden-Markov Models

The tables 8 and 9 in the appendix list all values of the evaluation of Reference versus Baseline model. The following details are important for comparison: There is no category which shows approximately the same difference as for the transition to another category. Therefore, it can be concluded that no further grouping is useful. Even a category containing only few transitions (such as the syllabic r) differs from the main category vowel.

The smallest value occurs at the transition from the fricative labial to the plosives glottal.

Most mean differences between the Reference and Baseline labels are in the range of 4ms to 20ms. 33 transitions have a higher mean than 20ms and five a higher one than 35ms. There are transitions from vowel → [r] | approximants, [r] | [j] | [sil] → vowel, [N] → [w] and fricative labial → [j] where the mean is larger than 20ms, which is actually the sampling length. Particularly, the Reference labels and the Baseline labels differ in transitions from [@_r]|fricative coronal → [w]. They also differ in self-transition, especially for the vowels.

There is a major difference between the plosives labial|plosive dorsal → silence model; this mean value is greater than twice the sampling length. In figure 6 the spectral components are shown on top. Beneath the spectral view the labels are arranged in the following order (from top to bottom): Reference, Baseline, Approach 1(HMM without updating A), Approach 2 (weighted HMM), Approach 3 (HSMM), Approach 4 (HSMM - unbounded duration). The great difference between Baseline and Reference labels for the transition [k] → [sil] is due to inaccuracies in the perfect labels: the end label of [k] were set at a later moment where the energy is too low to be accurately detected.

It is notable that the transition from a [>] to a plosive labial is greater than 25ms. This means the Hidden-Markov Model has great difficulties to train the model even if the spectral view seduces to assume the transition not to be a big problem.

The maximal variance of the considered model transitions is occurring at the transition [l] → [w], which can be explained due to the fact that the transition occurs only 13 times. This means, the occurring phones are very differently labelled.

It is important that the HMM can barely model the self-transitions correctly and even for the transition vowels to approximants the labels are set imprecisely.

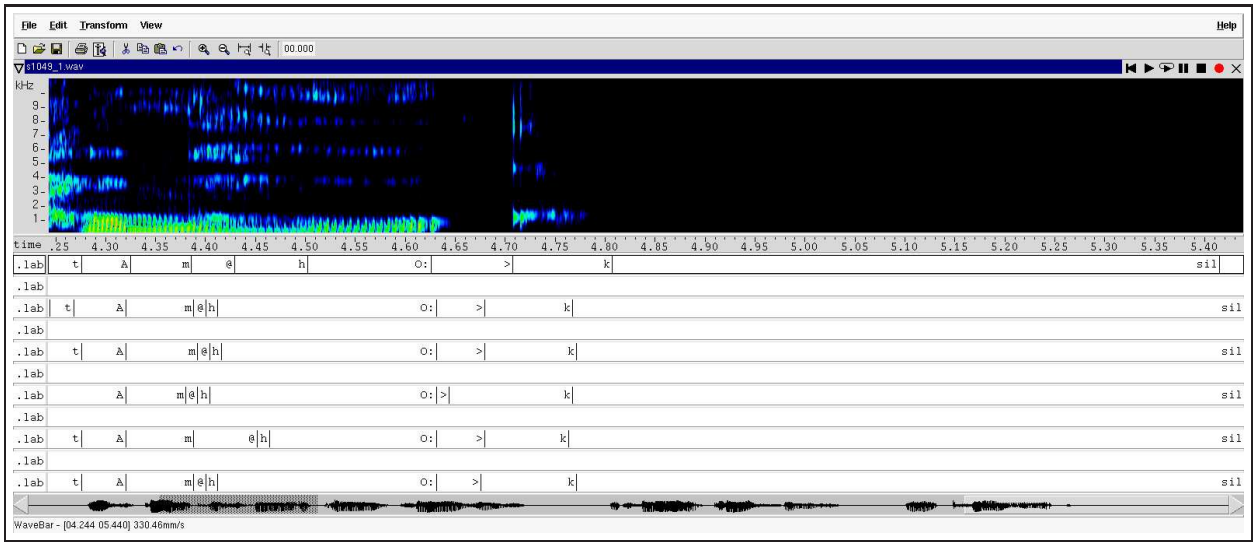

Figure 6: *Example extract: Labels of sentence 10*

4.2 Hidden Markov Models with equally distributed transition probability

The first approach is to ignore the transition probabilities by not updating the transition matrix A. In tables 10 and 11 all values are listed. With this comparison one can see that there are no grouped models which show a similar difference for the transitions to other models either. Therefore, no further grouping is reasonable.

The smallest value occurs again at the transition from the fricative labial to the plosives glottal.

The mean differences between the Reference and Baseline labels lie predominantly in the range of 3ms to 20ms. 32 transitions have a higher mean than 20ms and five transitions have a higher one than 35ms. This time the mean of the transition diphthong → [@_r] is smaller than 35ms whereas the transition [sil] → [>] has both a higher mean and variance value.

47 transitions have both a smaller mean and variance value than with Baseline model. Especially for the transitions to a vowel and the transitions to a plosive labial phone, the model of Approach 1 fits the requirements better than the Baseline. The improvement of the mean is maximal 9.26ms at the transition plosive dorsal → [sil]. Although the mean value (49ms to 40ms) has decreased, there exists still a major difference between the perfect labels and the labels set with Approach 1.

However, there is also the case where Approach 1 is worse than Baseline: at the transition diphthong → [?] the mean differs for 3.61ms.

Except for the transition [sil] → [l], which only occurs 14 times, the same transitions as in the Baseline model are found to have a mean value greater than 20ms. The maximal variance of the

considered model transitions is once again occurring for the transition [l] → [w], which can be explained due to the small number of occurrence.

The differences between the Approach 1 and the Baseline are shown in detail in tables 12 and 13. The main distinctions occur at the transitions to a vowel and at the transitions from plosives → [sil].

Additionally, for 11 transitions (such as the transitions fricative labial → [m]|[h] or [n] → fricative labial|[m]|[r]) both the mean and variance value is exactly the same. In other words, the labels of both Baseline and Approach 1 models are set at the very same position. These transitions occur between 13 and 52 times.

Approach 1 can match the transitions [>] → plosive dorsal better than Baseline. In turn the transition diphthong → [?] is more inexact with Approach 1 than with Baseline.

In Approach 1, there are transitions where the mean gets smaller (the labels are set more precisely) as well as larger.

4.3 Hidden Markov Models with weighted observation and transition probability

The second approach was to consider the different weights of the observation probability and the transition probability. In formula (2) on page 17 the observation $b_j(\mathbf{x}_t)$ has 26 components. The transition probability a_{ij} has to be weighted accordingly.

In tables 14 and 15 all values of the comparison Reference versus Approach 2 are listed. Moreover, it can be shown that there are no grouped models which show similar differences for transitions to other models. So no further grouping is realised.

Except for the smallest value occurring at the transition plosive dorsal → [@_r], all values are larger than 4ms, which is the frame shift size. The biggest value occurs at the transition [j] → diphthong.

77 transitions, or almost half of the values, have a higher mean than 20ms. There are more than 8 values having a mean greater than 50ms. This occurs particularly at the transitions from vowel → approximants|silence, plosives → silence, nasals → nasals and [h] → diphthong, which disagree in both mean and variance. The weighted model shows problems to train the models for the prae-plosive pause.

27 transitions show a smaller mean value than with the Baseline model, while 25 transitions show a lower one than with the Approach 1. The differences lie between 0.05ms and 13.42ms. 18 of these values have both a smaller mean and variance value than with Baseline and 15 values than with Approach 1. A maximal improvement of 13.42ms is made at the transition fricatives coronal → [w] in relation to Baseline and 13.48ms in relation to Approach 1. Mainly, the mean values for the transitions to [r] and to [w] are narrowed. In case of the transition from nasals or plosives to diphthongs, the mean is smaller but for the case approximants → diphthongs the contrary is true. The value increased to a global maximum of 60.47ms. This observation supports the initial remark, that no further grouping is useful.

As seen in tables 16 and 17, the bigger the number of occurrence the more Approach 2 and Baseline differ.

This model does not provide an improvement for all model in general. Only a few transitions indicate a smaller mean and variance values than for Baseline and Approach 1. So the influence of the transition probability a_{ij} is smaller than the observation probability for most transitions. However, the labels for the transition fricatives coronal → [w] are set more precisely than with the usual Hidden-Markov Model.

4.4 Hidden Semi-Markov Models with Gamma duration distribution

The third approach was the implementation of the Hidden Semi-Markov Model considering the duration explicitly by a Gamma distribution.

In tables 18 and 19 all values of the comparison Reference versus HSMM (Approach 3) are listed. Furthermore, no grouped models show similar differences for transitions to other models. So no further grouping is useful.

The smallest value occurs at the transition fricative labial → [?]. The maximal value is reached at the transition plosive labial|dorsal → [sil]. Most mean difference are again in the range of 4ms to 20ms. 30 transitions have a higher mean than 20ms and two transitions have a higher one than 40ms.

The transitions vowels → vowels (including the subgroup) show a mean value between 18ms and 36ms. Also, the transitions vowels → approximants are labelled inaccurately.

Besides, the following transitions are set imprecise with respect to the perfect marked labels: silence → [l]|[>], fricative coronal → [w], fricative labial → [sil].

In total, 78 transitions have a smaller mean value than Baseline, 81 transitions have a greater one and one transition (fricative labial → nasal labial) has exactly the same mean and variance value, though it occurs 13 times.

Only all transitions to [j] have a smaller mean value with the Hidden Semi-Markov Model than with the Hidden Markov Models.

All other transitions have both higher and smaller mean values, for example the transition to vowels which features 9 higher (fricative,nasals,plosives) and equally many lower (vowels,trill,approximants) values.

So explicitly modelling duration improves merely the mean of several transitions.

4.5 Hidden Semi-Markov Models with unbounded duration

The fourth approach was to implement the maximal duration available. The Hidden Semi-Markov Model considers the duration for each state explicitly by a Gamma distribution.

In tables 22 and 23 all values of the comparison Reference versus HSMM (Approach 3) are listed. No simplification can be achieved by further grouping.

The smallest value occurs once again at the transition fricative labial → [?]. The maximal value is reached at the transitions plosive labial|dorsal → [sil].

Most mean difference are in the range of 4ms to 20ms. 34 transitions have a higher mean than 20ms and again two transitions have a higher one than 40ms.

For the transitions vowels → vowels (including the subgroup) the mean value are between 21ms and 32ms. The transitions vowels → approximants are labelled inaccurately.

86 transitions in Approach 4 show a greater mean value than for Baseline while 74 transitions show a smaller one. 95 transitions show a greater mean value than for Approach 3 while 64 transitions show a smaller mean value. For all 13 occurrences of the transition fricative labial → nasal labial, the labels are set at almost the same position as in Baseline and Approach 3. Changing the duration model from exponentially to Gamma distributed does not affect this transition.

Only for the transitions to [j] the mean of all transitions is decreased with Approach 4. All other transitions have both higher and smaller mean values compared to the Baseline model. For example, the transition to fricative coronal features 7 higher ([@_r], fricative coronal, [n], [m], plosive coronal, [l], [sil]) and equally many lower (vowels, diphthongs, fricative labial, [N], plosive labial, plosive dorsal, [>]) mean values. Moreover, within the same group, the subgroup elements act differently and explicitly modelling duration improves just the mean of several transitions.

In an addition attempt, Approach 4 is conducted once again with a ten times longer duration for the silence model. Thereby it was ascertained that a longer duration for the silence model leads to more exact labels.The mean and variance values are listed in tables 26 and 27.

With the HSMM, an overall improvement is not achieved. Therefore, just considering the duration barely satisfies.

4.6 Global evaluation

Comparing tables 8, 10, 14, 18 and 22 leads to the following result. For the transition from plosive labial → [sil], the unbounded duration in Approach 4 attains the same result as the Hidden-Markov Model with the exponentially distributed length. This is also the case for the transitions from nasals → vowels, fricative → [?], fricative → [j], nasals → [>] and vowels → [>]. For the transitions plosive|fricative → [l] and [sil] → [n]|[m], Approach 4 has a greater mean than with HMM. Thus, considering the duration model is of minor importance for this transitions.

On the contrary, for the transitions [>] → plosive labial|dorsal and plosive dorsal|coronal → [sil] the mean is smaller and the duration model is important.

Self-transitions only occur within the categories vowels and fricative. None of the approaches shows a qualitative improvement of the mean values.

For the transitions fricative coronal|plosive coronal → [w] and [@_r] → [?] the weighted model achieves the best result.

In figure 7, one can see an extract of sentence 45. Only the fourth label (Approach 2) matches the first label (Reference) for the transition [s] → [w]. However, the transition [l] → [I] is completely missed in Approach 2.

Except for the transition fricative coronal → [w], where the means stay constant for Baseline, Approach 1, Approach 3 and Approach 4, all other transitions to [w] are made smaller consid-

ering the duration modelling. Considering the duration, the bounded model achieves improvements for the transition to [w].

The labels of the Hidden-Markov Model are better marked than with all approaches for the following transitions: fricatives → [@_r] , diphthong → fricative labial, [n]|[m] → fricative coronal, plosive coronal → fricative and [n] → [w]|[h]. In these cases no improvement is achieved.

In table 3 on page 27, the difference of the means between Reference labels and Baseline labels and the difference between Reference labels and Approach 1-4 labels is illustrated. For the weighted Hidden-Markov Model (Approach 2) only a half of the values are smaller than 20ms, whereas Baseline and the other approaches achieve approximately a value of 80%. The Approach 2 also has mean values which are larger than 60ms. So, with Approach 2 a general improvement is not achieved. Approach 4 shows a similar behaviour as Baseline. Approach 3 has more values in the range 10ms to 20ms than the others. Approach 1 attains a higher number of values being smaller than 30ms than for Baseline.

Table 4 on page 27 shows the amount of labels, which have absolute label differences between Reference labels and Baseline as well as between Reference and the four approaches.

Consequently, the overall improvement is best for the Hidden Semi-Markov Model. Within the range [0ms. . . 20ms] a 2.4% improvement is achieved. Weighting the components of the Hidden-Markov Model usually leads to a worse labelling, but for a few transitions the labels were set more precisely.

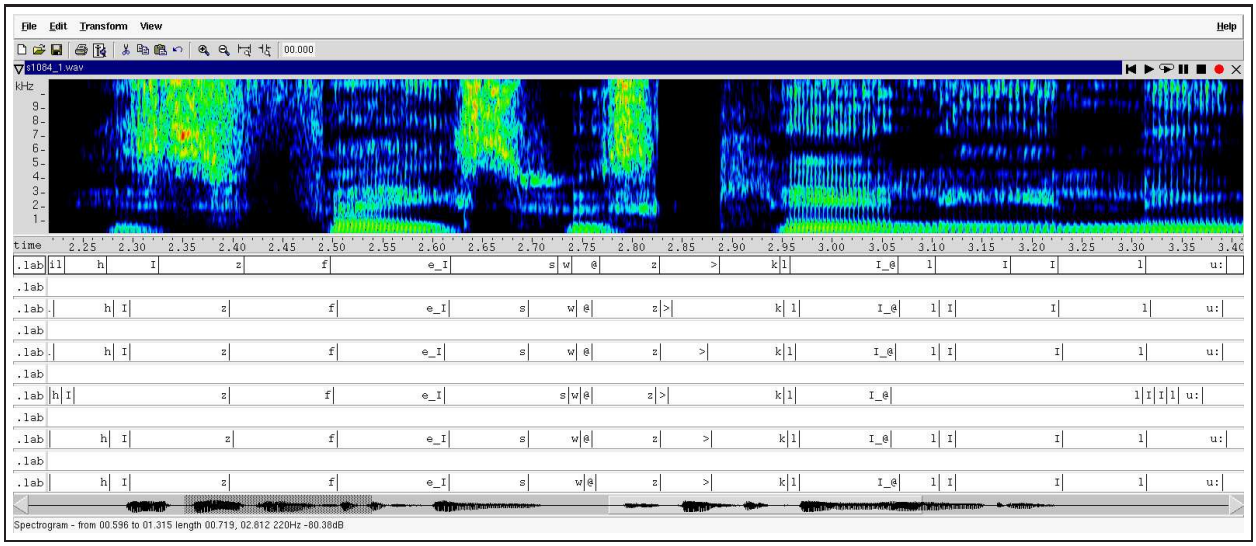

Figure 7: *Example extract: Labels of sentence 45*

Table 3: *Comparison: Mean value difference between the Reference labels and the labels listed*

Δ Mean [ms]	Baseline	Approach 1	Approach 2	Approach 3	Approach 4
[0ms . . . 10ms]	0.4813	0.4938	0.2125	0.4000	0.4375
[0ms . . . 20ms]	0.7938	0.8000	0.5188	0.8125	0.7875
[0ms . . . 30ms]	0.9250	0.9313	0.7625	0.9375	0.9250
[0ms . . . 40ms]	0.9875	0.9938	0.9000	0.9875	0.9875
[0ms . . . 50ms]	0.9938	0.9938	0.9500	0.9938	0.9938
[0ms . . . 60ms]	1	1	0.9813	1	1
[0ms . . . 70ms]	1	1	0.9938	1	1
[0ms . . . 80ms]	1	1	1	1	1

Table 4: *Comparison: Label difference between the Reference labels and the labels listed*

Δ Labels [ms]	Baseline	Approach 1	Approach 2	Approach 3	Approach 4
[0ms . . . 5ms]	0.3230	0.3267	0.2563	0.3326	0.3257
[0ms . . . 10ms]	0.6144	0.6297	0.4621	0.6325	0.6269
[0ms . . . 20ms]	0.8484	0.8617	0.6414	0.8720	0.8655

5 Conclusions and Further Work

5.1 Conclusions

In general, the duration modelling conducted with the Hidden Semi-Markov Model enhances the labelling. The overall improvement reaches its peak for the Hidden Semi-Markov Model in which the duration is bounded and modelled by a Gamma distribution. Using that approach, in part larger improvements are made for the transitions to [j]. The duration model is also important for the transitions [>] → plosive labial | dorsal and plosive dorsal | coronal → [sil] and achieves considerable improvements.

The labels of the silence model are improved if the duration is not bounded. The duration is of minor importance for the transitions nasals → vowel | [>], vowels → [>] and fricative → [?]|[j].

Weighting the observation and the transition probability leads to improvement for the transition fricatives coronal → [w]. An equally distributed transition probability increases the matches for labels of the transition [>] → plosive dorsal.

The transitions [n] | [m] → fricative coronal, [n] → [w] | [h], plosive coronal → fricative, fricatives → [@_r] and diphthong → fricative labial have the best set labels with the Hidden-Markov Model. The exponentially distributed duration tends to match the phone length better than in the case of Gamma-distribution usage.

Finally, none of the approaches were able to improve the labelling of the self-transitions to a larger extent.

5.2 Further work

To examine the influence of the chosen duration model, in a further work other distributions such as Gaussian or Poisson distribution could be implemented.

Another remaining issue will be the possibility to weight the Hidden Semi-Markov Model with adequate coefficients; depending on the duration, the number of observations made and the probability to change to the next state have to be weighted appropriately.

As seen for some transitions, the duration is one of many components to segment speech accordingly. For this reason spectral components or spectral informations have to be considered to accomplish a better segmentation.

References

[CF87] Codogno, M. and Fissore, L. Duration modelling in finite state automata for speech recognition and fast speaker adaptation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'87)*, volume 12, pages 1269–1272. IEEE, April 1987.

[GTL91] Gu, H.-Y., Tseng, C.-Y., and Lee, L.-S. Isolated-utterance speech recognition using hidden markov models with bounded state durations. *IEEE Transactions on Signal Processing*, 39(8):1743–1752, August 1991.

[Lev86] Levinson, S.E. Continuously variable duration hidden markov models for automatic speech recognition. *Computer Speech and Language*, 1(1):29–45, 1986.

[NH88] Nakagawa, S. and Hashimoto, Y. A method for continuous speech segmentation using hmm. In *9th International Conference on Pattern Recognition*, volume 2, pages 960–962, 14th-17th November 1988.

[OZN+06] Oura, K., Zen, Y., Nankaku, Y., Lee, A., and Tokuda, K. Hidden semi-markov model based speech recognition system using weighted finite-state transducer. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-2006)*, volume 1, pages I–33–I–36. IEEE, 14th-19th May 2006.

[PK08] Pfister,B. and Kaufmann, T. *Sprachverarbeitung*. Springer Verlag, 2008.

[RSS92] Ratnayake, N., Savic, M., and Sorensen, J. Use of semi-markov models for speaker-independant phoneme recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-92)*, volume 1, pages 565–568. IEEE, 23rd-26th March 1992.

[Yu,09] Yu, S.-Z. Hidden semi-markov models. *Artificial Intelligence*, 2009.

A Appendix

A.1 Phone Models

The following phone models appear in the sentences used for this thesis:

Table 5: *Phone models*

(1): [sil]	(9): [I]	(17): [u:]	(25): [@_r]	(33): [U]	(41): [t_S]		
(2): [s]	(10): [k]	(18): [w]	(26): [h]	(34): [e_@]	(42): [U_@]		
(3): [o_U]	(11): [A]	(19): [O:]	(27): [p]	(35): [r]	(43): [Z]		
(4): [D]	(12): [z]	(20): [m]	(28): [j]	(36): [S]	(44): [O_I]		
(5): [e]	(13): [a_U]	(21): [?]	(29): [i:]	(37): [d_Z]	(45): [g]		
(6): [n]	(14): [A:]	(22): [@]	(30): [v]	(38): [V]	(46): [N]		
(7): [>]	(15): [t]	(23): [d]	(31): [T]	(39): [e_I]	(47): [3]		
(8): [b]	(16): [l]	(24): [a_I]	(32): [f]	(40): [q]	(48): [I_@]		

A.2 Grouping

Every phone model is grouped according to the International Phonetic Alphabet chart for English dialects.

Table 6: *Grouping phone models for the classification*

categories	No.	subdivision	No. of phone models
vowels			
- vowel	1	(e,I,A,A:,u:,O: @,i:,U,V,q,3)	(5,9,11,14,17,19, 22,29,33,38,40,47)
- diphthong	2	(o_U, a_U, a_I, e_@ e_I, U_@, O_I, I_@)	(3,13,24,34, 39,42,44,48)
- syllabic r	3	(@_r)	(25)
fricatives			
- coronal	4	(s,D,z,T,S,d_Z,t_S,Z)	(2,4,12,31,36,37,41,43)
- labial	5	(v,f)	(30,32)
nasals			
- dorsal	6	(N)	(46)
- coronal	7	(n)	(6)
- labial	8	(m)	(20)
trill			
- coronal	9	(r)	(35)
plosives			
- labial	10	(b,p)	(8,27)
- dorsal	11	(k,g)	(10,45)
- coronal	12	(t,d)	(15,23)
- glottal	13	(?)	(21)
approximants			
- coronal	14	(l)	(16)
- voice labialised	15	(w)	(18)
- laryngeal	16	(h)	(26)
- dorsal	17	(j)	(28)
silence			
- sil	18	(sil)	(1)
- prae-plosive pause	19	(>)	(7)

A.3 Mean and Variance of Baseline and Approach 1-4

Table 7: *Number of transitions in 401 sentences*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	173	21	8	1032	455	216	992	305	157	0	0	0	117	412	123	71	33	62	1731
2	90	10	24	230	64	0	125	67	70	0	0	0	39	80	35	26	15	53	386
3	37	7	0	74	11	0	6	7	4	0	0	0	21	13	10	15	2	19	59
4	1231	220	71	122	47	0	26	47	40	0	0	0	64	41	73	37	12	172	502
5	291	52	41	70	11	0	3	13	62	0	0	0	14	31	9	14	23	19	91
6	25	6	1	40	10	0	1	0	1	0	0	0	5	3	10	10	3	15	86
7	241	99	15	193	36	0	7	25	15	0	0	0	36	30	29	38	25	41	394
8	263	78	5	52	6	0	3	3	3	0	0	0	7	2	7	3	7	18	76
9	474	131	1	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	1
10	368	129	16	39	3	0	0	2	87	0	0	0	1	93	4	3	11	26	26
11	311	86	20	101	9	0	7	5	50	0	0	0	4	50	25	9	15	32	70
12	772	115	68	211	38	0	28	37	106	0	0	0	51	51	62	45	23	122	185
13	281	93	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	365	106	9	85	30	0	8	8	8	0	0	0	10	3	13	6	8	41	123
15	372	67	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	266	52	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
17	175	18	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	173	24	0	153	24	0	18	14	9	0	0	0	0	14	41	44	17	0	89
19	0	0	0	303	0	0	0	0	0	808	794	1914	0	0	0	0	0	0	0

Baseline (HMM)

Table 8: *Mean: Reference (perfect labels) versus Baseline (HMM)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	22.37	19.59	-	6.63	5.60	16.61	6.49	5.96	26.36	-	-	-	13.23	23.29	21.92	18.60	35.24	12.98	6.62
2	24.82	34.77	37.00	6.72	5.07	-	8.87	5.37	32.59	-	-	-	13.87	31.47	18.77	23.66	19.42	18.04	6.77
3	31.73	-	-	7.03	4.28	-	-	-	-	-	-	-	22.34	16.48	35.93	21.55	-	7.74	7.07
4	8.06	10.49	7.84	18.46	10.93	-	9.80	7.22	8.22	-	-	-	7.82	8.00	34.95	10.33	12.25	14.44	6.20
5	6.39	8.26	5.37	12.87	26.99	-	-	8.11	6.10	-	-	-	4.26	5.89	-	6.84	22.36	21.38	5.96
6	9.61	-	-	6.68	5.12	-	-	-	-	-	-	-	-	-	21.76	5.60	-	8.09	7.04
7	7.95	8.56	9.72	5.50	5.75	-	-	21.68	14.49	-	-	-	6.10	13.22	14.31	8.67	5.73	9.64	5.92
8	9.45	10.20	-	5.98	-	-	-	-	-	-	-	-	-	-	-	-	-	10.85	8.10
9	23.53	28.28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	6.36	8.53	8.23	9.66	-	-	-	-	5.96	-	-	-	-	6.76	-	-	12.87	56.22	10.43
11	6.54	8.90	5.19	11.14	-	-	-	-	4.63	-	-	-	-	7.02	6.97	-	15.19	49.05	14.33
12	8.14	9.66	7.25	14.26	11.71	-	10.44	11.11	8.13	-	-	-	12.13	7.94	29.86	13.88	13.39	30.72	12.84
13	19.26	15.16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	8.81	9.15	-	5.87	7.01	-	-	-	-	-	-	-	6.69	-	28.66	-	-	9.15	5.37
15	17.92	14.79	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	13.10	11.12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	28.07	16.65	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	21.28	17.96	-	8.67	11.52	-	19.44	12.23	-	-	-	-	-	22.78	10.28	9.04	18.35	-	34.93
19	-	-	-	8.54	-	-	-	-	-	25.45	17.06	10.93	-	-	-	-	-	-	-

Table 9: *Variance: Reference (perfect labels) versus Baseline (HMM)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	330.55	287.84	-	27.15	16.88	174.34	42.86	61.82	344.43	-	-	-	202.94	505.70	205.72	179.73	389.41	64.07	35.43
2	414.82	194.96	531.53	21.10	12.95	-	110.72	10.08	480.85	-	-	-	77.89	1001.73	157.04	177.74	133.50	143.55	25.68
3	187.05	-	-	21.93	6.22	-	-	-	-	-	-	-	163.29	252.11	452.23	719.69	-	25.23	18.58
4	22.03	12.30	16.73	171.24	70.84	-	98.78	29.92	43.64	-	-	-	40.88	7.48	169.80	48.01	8.09	102.19	51.85
5	11.95	13.81	17.43	126.59	209.32	-	-	19.89	17.42	-	-	-	11.13	5.67	-	21.78	183.50	275.69	23.16
6	28.90	-	-	17.34	8.89	-	-	-	-	-	-	-	-	-	768.84	15.95	-	172.11	39.28
7	40.51	29.48	50.59	16.74	13.77	-	-	779.05	182.20	-	-	-	30.38	157.06	155.93	59.30	21.72	67.69	59.13
8	26.50	21.17	-	29.46	-	-	-	-	-	-	-	-	-	-	-	-	-	58.34	43.13
9	303.19	554.27	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	15.40	29.62	32.30	26.66	-	-	-	-	17.28	-	-	-	-	22.16	-	-	108.56	840.63	99.65
11	24.25	22.59	7.53	46.30	-	-	-	-	10.83	-	-	-	-	28.11	70.15	-	305.62	346.33	317.92
12	50.29	18.11	10.26	161.45	128.87	-	112.62	103.36	45.30	-	-	-	76.49	46.92	386.11	92.79	64.49	298.53	164.96
13	267.95	83.80	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	270.15	85.17	-	36.60	14.21	-	-	-	-	-	-	-	60.18	-	1700.81	-	-	50.96	21.31
15	370.43	117.41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	215.63	40.29	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	601.48	151.48	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	213.09	71.92	-	36.07	91.10	-	14.09	22.77	-	-	-	-	-	72.05	83.18	48.63	408.69	-	256.67
19	-	-	-	44.58	-	-	-	-	-	388.56	172.44	123.75	-	-	-	-	-	-	-

Approach 1 (HMM without updating A)

Table 10: *Mean: Reference (perfect labels) versus Approach 1 (HMM without updating A)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	21.80	18.02	-	6.67	5.60	16.72	6.40	5.96	26.43	-	-	-	13.32	23.51	22.98	16.65	35.66	12.87	6.60
2	24.34	29.43	32.87	7.18	5.86	-	8.99	5.74	31.92	-	-	-	17.48	31.17	19.46	24.28	19.69	19.27	7.40
3	32.05	-	-	6.86	5.37	-	-	-	-	-	-	-	24.46	14.53	34.73	20.21	-	6.94	7.27
4	7.66	10.18	8.09	19.79	11.27	-	9.87	7.22	8.23	-	-	-	7.18	7.81	35.00	10.44	12.25	14.36	6.23
5	6.27	8.03	6.66	12.98	26.99	-	-	8.11	6.24	-	-	-	3.98	5.76	-	6.84	22.88	20.54	5.58
6	9.31	-	-	6.31	5.52	-	-	-	-	-	-	-	-	-	21.36	5.99	-	8.56	6.99
7	6.91	8.16	8.40	5.89	5.75	-	-	21.68	14.49	-	-	-	6.07	12.76	14.45	8.75	5.89	9.54	5.86
8	9.38	10.11	-	6.26	-	-	-	-	-	-	-	-	-	-	-	-	-	10.41	8.37
9	23.25	29.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	6.25	8.22	8.48	9.14	-	-	-	-	5.96	-	-	-	-	6.63	-	-	12.87	53.61	13.11
11	6.63	9.00	6.02	12.36	-	-	-	-	5.10	-	-	-	-	6.86	7.46	-	14.83	39.79	13.09
12	7.95	9.61	8.48	14.75	12.15	-	10.73	9.49	8.49	-	-	-	10.82	7.94	30.06	13.65	14.08	24.64	13.17
13	18.10	16.09	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	8.48	8.44	-	5.86	7.01	-	-	-	-	-	-	-	6.69	-	28.35	-	-	9.15	5.43
15	18.42	14.74	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	12.12	11.66	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	26.82	17.09	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	21.10	17.96	-	9.57	11.69	-	18.99	11.94	-	-	-	-	-	19.93	10.18	9.13	18.59	-	35.02
19	-	-	-	7.87	-	-	-	-	-	23.77	12.83	9.28	-	-	-	-	-	-	-

Table 11: *Variance: Reference (perfect labels) versus Approach 1 (HMM without updating A)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	336.66	280.62	-	26.29	16.25	170.94	40.42	62.36	333.17	-	-	-	197.85	492.28	335.41	143.70	427.84	59.36	27.94
2	397.20	109.59	478.91	22.23	19.44	-	110.63	14.00	491.22	-	-	-	536.73	993.57	154.14	199.88	134.66	166.73	25.94
3	228.12	-	-	19.61	10.91	-	-	-	-	-	-	-	150.15	169.63	380.62	744.84	-	29.03	17.90
4	21.78	11.34	16.75	168.78	69.24	-	95.42	28.96	43.47	-	-	-	40.73	7.17	170.02	49.17	8.09	109.99	53.58
5	11.73	15.17	18.77	125.12	209.32	-	-	19.89	18.08	-	-	-	10.48	5.71	-	21.78	189.98	218.06	21.86
6	38.73	-	-	17.44	9.47	-	-	-	-	-	-	-	-	-	784.26	16.68	-	141.16	41.57
7	30.12	30.17	53.90	21.24	13.77	-	-	779.05	182.20	-	-	-	29.44	157.84	155.28	57.46	23.83	67.45	43.90
8	25.91	21.20	-	28.63	-	-	-	-	-	-	-	-	-	-	-	-	-	61.43	49.89
9	302.31	574.44	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	13.49	31.49	31.60	29.17	-	-	-	-	17.14	-	-	-	-	21.82	-	-	108.56	891.09	297.83
11	30.78	22.94	7.97	53.43	-	-	-	-	13.72	-	-	-	-	26.24	74.28	-	317.75	282.32	280.13
12	16.32	15.80	9.81	176.60	144.53	-	114.04	57.12	43.40	-	-	-	58.90	46.92	402.38	94.47	68.16	258.64	187.20
13	268.68	256.07	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	263.21	55.85	-	35.34	14.21	-	-	-	-	-	-	-	56.33	-	1702.57	-	-	51.49	19.17
15	401.42	117.46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	224.68	33.04	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	614.00	143.31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	206.80	71.92	-	187.19	92.36	-	14.25	25.25	-	-	-	-	-	31.69	71.45	47.41	415.74	-	271.60
19	-	-	-	36.37	-	-	-	-	-	360.00	119.55	89.07	-	-	-	-	-	-	-

Table 12: *Mean: Baseline versus Approach 1 (HMM without updating A)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	2.15	3.04	-	0.92	0.50	0.67	0.47	0.30	1.86	-	-	-	1.91	0.99	1.78	2.98	2.06	0.51	1.02
2	4.43	13.17	8.81	1.25	1.56	-	0.64	0.77	2.62	-	-	-	5.94	1.15	1.14	1.53	0.27	1.43	1.28
3	4.53	-	-	0.49	1.09	-	-	-	-	-	-	-	3.23	5.22	2.00	1.33	-	1.68	0.27
4	0.69	0.34	1.18	2.16	0.51	-	0.77	0.34	0.10	-	-	-	0.87	0.19	0.05	0.11	0.00	1.32	0.83
5	0.40	0.69	1.46	0.51	0.00	-	-	0.00	0.19	-	-	-	0.29	0.13	-	0.00	1.21	0.84	0.61
6	1.28	-	-	0.60	0.40	-	-	-	-	-	-	-	-	-	0.40	0.40	-	1.06	0.46
7	1.19	0.40	4.52	0.83	0.00	-	-	0.00	0.00	-	-	-	0.67	0.67	0.14	0.53	0.16	0.10	0.85
8	0.38	0.10	-	0.61	-	-	-	-	-	-	-	-	-	-	-	-	-	0.44	1.05
9	1.01	2.77	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	0.66	0.43	0.75	0.72	-	-	-	-	0.18	-	-	-	-	0.39	-	-	0.00	2.92	3.84
11	0.96	0.46	1.40	1.98	-	-	-	-	0.56	-	-	-	-	0.32	1.28	-	2.13	9.73	5.19
12	0.66	0.49	1.29	1.06	2.94	-	0.29	1.94	0.60	-	-	-	1.57	0.00	1.29	1.86	1.04	7.03	4.06
13	2.51	3.26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	0.91	1.88	-	0.52	0.00	-	-	-	-	-	-	-	0.80	-	0.31	-	-	0.19	0.62
15	1.32	0.54	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	2.57	1.15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	2.65	3.55	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	0.65	0.00	-	1.51	0.17	-	0.44	0.29	-	-	-	-	-	2.85	0.29	0.45	0.23	-	1.35
19	-	-	-	1.88	-	-	-	-	-	2.15	5.38	3.08	-	-	-	-	-	-	-

Table 13: *Variance: Baseline versus Approach 1 (HMM without updating A)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	14.89	15.86	-	2.81	2.00	6.29	3.58	1.14	38.50	-	-	-	9.25	13.92	151.58	33.94	26.97	1.73	16.22
2	86.72	354.83	141.89	2.64	10.63	-	2.83	3.46	28.32	-	-	-	782.98	15.19	16.21	5.40	0.76	3.61	8.66
3	74.65	-	-	1.48	6.17	-	-	-	-	-	-	-	39.21	178.88	15.04	17.44	-	4.10	1.02
4	3.94	0.78	3.07	40.96	3.69	-	6.16	1.93	0.39	-	-	-	2.83	0.70	0.22	0.38	0.00	23.02	5.90
5	1.79	1.93	3.78	3.02	0.00	-	-	0.00	0.74	-	-	-	1.09	0.51	-	0.00	7.47	6.83	2.76
6	2.34	-	-	1.89	1.57	-	-	-	-	-	-	-	-	-	1.59	1.59	-	3.34	2.78
7	12.25	1.27	40.65	7.68	0.00	-	-	0.00	0.00	-	-	-	2.28	5.58	0.55	4.45	0.64	0.39	19.05
8	1.45	0.29	-	1.78	-	-	-	-	-	-	-	-	-	-	-	-	-	1.67	7.38
9	12.30	36.35	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	5.72	1.32	2.58	2.98	-	-	-	-	0.70	-	-	-	-	1.41	-	-	0.00	36.37	294.34
11	7.97	0.72	3.56	10.32	-	-	-	-	1.89	-	-	-	-	1.19	11.34	-	7.58	117.51	232.22
12	33.26	1.69	3.43	9.64	65.27	-	1.09	91.72	2.24	-	-	-	5.66	0.00	27.61	25.00	3.19	60.93	138.72
13	20.72	345.26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	13.07	47.70	-	1.58	0.00	-	-	-	-	-	-	-	2.83	-	1.23	-	-	0.76	2.88
15	48.82	1.76	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	53.06	3.31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	27.90	22.17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	1.86	0.00	-	177.41	0.66	-	1.67	1.14	-	-	-	-	-	42.71	1.11	1.64	0.94	-	4.69
19	-	-	-	8.66	-	-	-	-	-	24.03	66.26	57.17	-	-	-	-	-	-	-

Approach 2 (weighted HMM)

Table 14: *Mean: Reference (perfect labels) versus Approach 2 (weighted HMM)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	39.69	23.77	-	26.49	25.63	36.87	23.68	15.48	35.72	-	-	-	19.10	33.25	37.62	49.73	50.43	50.05	17.77
2	31.85	39.16	52.11	11.23	10.85	-	15.08	7.00	35.39	-	-	-	18.92	31.31	24.28	40.16	18.96	28.73	10.61
3	36.14	-	-	22.22	9.46	-	-	-	-	-	-	-	13.28	25.19	48.30	31.29	-	19.33	17.01
4	12.04	11.91	12.46	34.45	20.48	-	8.40	9.76	6.70	-	-	-	11.86	8.79	21.52	26.95	10.59	19.89	14.79
5	9.99	5.67	10.52	30.02	31.03	-	-	11.73	5.56	-	-	-	16.54	5.96	-	18.91	23.05	23.83	13.30
6	36.71	-	-	11.20	7.13	-	-	-	-	-	-	-	-	-	24.95	20.91	-	23.06	4.70
7	30.64	12.40	25.65	10.22	8.96	-	-	47.88	24.03	-	-	-	8.08	29.60	39.22	14.65	7.33	25.46	8.23
8	15.81	9.48	-	8.64	-	-	-	-	-	-	-	-	-	-	-	-	-	35.39	8.66
9	29.27	25.88	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	7.75	5.83	21.87	9.96	-	-	-	-	5.20	-	-	-	-	7.13	-	-	18.77	76.79	21.51
11	7.44	6.24	3.45	10.93	-	-	-	-	4.16	-	-	-	-	8.84	5.75	-	28.68	59.65	17.40
12	14.63	13.08	10.77	29.72	23.66	-	29.21	19.84	17.61	-	-	-	23.43	12.96	25.05	29.79	45.95	69.92	26.76
13	19.09	15.11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	40.43	14.52	-	7.33	4.55	-	-	-	-	-	-	-	9.78	-	28.49	-	-	28.34	6.46
15	27.64	15.14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	49.93	60.47	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	38.09	26.65	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	18.67	14.80	-	11.16	11.37	-	22.10	11.09	-	-	-	-	-	30.76	14.53	10.74	21.17	-	38.01
19	-	-	-	32.34	-	-	-	-	-	50.82	41.97	28.46	-	-	-	-	-	-	-

Table 15: *Variance: Reference (perfect labels) versus Approach 2 (weighted HMM)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	634.09	325.96	-	774.11	677.93	548.51	635.20	499.45	1046.04	-	-	-	601.26	1109.30	827.82	1147.50	1684.11	2204.35	684.74
2	404.04	192.50	2850.40	82.39	42.16	-	508.60	43.45	482.06	-	-	-	1077.78	871.83	133.34	563.84	109.57	324.19	422.74
3	232.26	-	-	341.37	67.95	-	-	-	-	-	-	-	158.90	3492.50	524.88	902.51	-	85.91	941.20
4	487.70	274.42	206.32	367.81	1056.37	-	18.65	31.90	42.63	-	-	-	150.71	6.73	144.01	299.24	19.00	163.57	661.85
5	968.96	8.49	670.49	244.31	377.35	-	-	127.50	25.52	-	-	-	1024.23	7.11	-	211.84	186.20	215.35	329.10
6	1090.41	-	-	60.03	27.61	-	-	-	-	-	-	-	-	-	673.07	345.75	-	478.68	18.07
7	1141.09	363.82	317.67	108.51	32.01	-	-	2698.87	424.63	-	-	-	102.19	978.60	976.69	508.85	131.50	258.42	196.44
8	471.45	27.11	-	33.77	-	-	-	-	-	-	-	-	-	-	-	-	-	530.30	371.28
9	756.28	544.24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	292.04	33.03	4491.15	114.99	-	-	-	-	11.80	-	-	-	-	16.53	-	-	105.31	2219.47	972.58
11	328.10	24.23	2.43	918.65	-	-	-	-	9.82	-	-	-	-	34.97	27.78	-	621.16	1233.29	485.60
12	711.62	302.56	635.63	1378.76	283.36	-	1921.52	743.34	396.17	-	-	-	797.76	705.62	753.31	533.53	694.87	1323.80	544.20
13	393.73	642.08	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	1795.58	681.04	-	43.68	19.66	-	-	-	-	-	-	-	31.99	-	1303.14	-	-	308.27	99.30
15	906.01	139.38	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	1483.89	2131.39	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	1727.64	294.01	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	170.33	28.61	-	269.38	52.96	-	132.83	16.87	-	-	-	-	-	62.02	126.66	167.49	480.00	-	256.58
19	-	-	-	558.41	-	-	-	-	-	884.22	859.48	842.08	-	-	-	-	-	-	-

Table 16: *Mean: Baseline versus Approach 2 (weighted HMM)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	31.97	20.52	-	23.43	24.42	24.72	22.54	12.89	24.10	-	-	-	14.60	21.00	19.70	39.80	25.15	40.04	15.62
2	13.39	9.18	32.43	7.48	8.17	-	11.37	3.93	21.38	-	-	-	22.51	10.78	10.26	31.01	6.39	13.86	7.97
3	7.66	-	-	18.01	9.43	-	-	-	-	-	-	-	19.38	31.31	14.77	21.28	-	19.32	12.72
4	11.24	10.38	15.29	28.00	14.94	-	16.12	12.65	6.59	-	-	-	6.11	2.34	19.85	23.30	2.33	9.23	11.19
5	9.08	3.38	8.66	30.84	18.87	-	-	16.27	4.31	-	-	-	14.82	1.42	-	16.82	2.43	9.24	13.24
6	32.25	-	-	6.39	9.58	-	-	-	-	-	-	-	-	-	3.99	18.36	-	20.75	6.68
7	25.70	7.82	20.22	6.58	6.98	-	-	30.33	11.17	-	-	-	4.77	26.87	28.62	12.60	6.86	18.69	7.09
8	9.38	2.00	-	5.76	-	-	-	-	-	-	-	-	-	-	-	-	-	30.60	9.08
9	14.15	12.73	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	6.77	4.05	20.95	13.92	-	-	-	-	2.66	-	-	-	-	1.54	-	-	7.62	20.88	19.95
11	5.43	3.20	2.39	13.20	-	-	-	-	2.16	-	-	-	-	2.23	6.70	-	16.76	15.34	14.08
12	15.57	13.50	11.56	29.66	22.16	-	27.08	17.80	17.77	-	-	-	21.99	8.92	37.46	24.57	40.78	39.78	24.79
13	12.21	15.62	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	36.44	11.33	-	7.32	7.45	-	-	-	-	-	-	-	7.98	-	15.04	-	-	25.31	9.28
15	18.02	5.24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	49.47	63.55	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	13.18	11.97	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	9.80	4.82	-	5.66	6.65	-	4.88	1.71	-	-	-	-	-	11.97	7.69	7.89	3.29	-	11.52
19	-	-	-	26.03	-	-	-	-	-	28.48	27.84	24.56	-	-	-	-	-	-	-

Table 17: *Variance: Baseline versus Approach 2 (weighted HMM)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	854.68	26.35	-	824.43	665.23	625.64	637.15	473.01	807.38	-	-	-	491.46	902.01	653.80	1329.64	1668.82	1855.79	717.02
2	224.49	7.96	1868.22	67.12	20.96	-	459.71	39.35	740.02	-	-	-	1000.50	428.92	172.79	887.46	33.22	69.61	465.22
3	84.60	-	-	357.70	27.94	-	-	-	-	-	-	-	164.51	2022.39	134.68	531.68	-	30.55	1201.96
4	502.58	336.68	211.26	552.73	1132.69	-	139.19	40.90	23.74	-	-	-	148.73	2.69	226.54	281.46	10.93	43.31	640.65
5	948.45	8.79	617.89	304.56	280.48	-	-	98.55	15.62	-	-	-	1114.55	4.69	-	97.27	8.28	140.15	335.69
6	931.78	-	-	50.29	35.39	-	-	-	-	-	-	-	-	-	7.08	386.51	-	89.19	40.27
7	1237.68	402.03	469.79	93.02	12.54	-	-	1299.42	191.58	-	-	-	74.47	923.72	1091.11	395.63	134.75	67.25	128.24
8	487.96	5.53	-	38.59	-	-	-	-	-	-	-	-	-	-	-	-	-	243.60	391.86
9	540.81	177.10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	295.25	13.37	4016.85	117.19	-	-	-	-	6.10	-	-	-	-	5.63	-	-	30.06	1369.86	913.60
11	263.34	3.68	6.01	997.84	-	-	-	-	6.01	-	-	-	-	10.15	99.46	-	468.25	1021.03	734.28
12	681.20	310.30	578.13	1629.67	195.61	-	1590.50	579.88	437.80	-	-	-	534.03	653.97	778.96	505.14	523.15	922.18	676.50
13	298.24	680.97	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	1810.12	713.33	-	32.41	19.46	-	-	-	-	-	-	-	70.79	-	143.76	-	-	135.07	115.54
15	583.39	21.87	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	1495.94	2306.31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	1118.17	166.58	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	160.98	50.52	-	313.22	91.28	-	92.86	4.20	-	-	-	-	-	31.86	99.86	97.41	32.32	-	449.39
19	-	-	-	394.14	-	-	-	-	-	588.42	781.31	831.31	-	-	-	-	-	-	-

Approach 3 (HSMM)

Table 18: *Mean: Reference (perfect labels) versus Approach 3 (HSMM)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	21.20	18.80	-	6.76	5.16	11.98	6.23	5.87	24.28	-	-	-	12.04	25.72	14.56	13.71	30.26	25.86	6.90
2	20.84	28.78	36.12	7.09	9.15	-	8.52	10.45	26.26	-	-	-	15.27	30.87	17.86	20.88	15.87	24.08	6.89
3	30.68	-	-	6.92	4.28	-	-	-	-	-	-	-	20.63	15.92	21.16	15.32	-	10.91	7.13
4	8.48	11.18	8.27	16.43	9.50	-	9.29	8.64	8.51	-	-	-	7.23	11.78	32.38	10.44	10.26	13.94	7.63
5	6.14	9.09	10.38	12.48	27.23	-	-	8.11	6.76	-	-	-	4.12	10.07	-	7.12	14.17	34.33	4.50
6	11.06	-	-	6.86	4.72	-	-	-	-	-	-	-	-	-	16.33	9.52	-	9.85	6.43
7	8.43	9.55	8.67	6.33	5.26	-	-	9.42	14.76	-	-	-	6.21	12.83	17.95	10.68	5.41	13.96	5.29
8	10.52	10.66	-	6.94	-	-	-	-	-	-	-	-	-	-	-	-	-	15.68	7.64
9	21.03	27.26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	6.82	9.54	10.23	8.75	-	-	-	-	11.41	-	-	-	-	12.06	-	-	6.70	52.84	9.81
11	11.35	13.42	8.62	10.25	-	-	-	-	5.85	-	-	-	-	12.99	8.42	-	10.65	47.02	15.79
12	9.71	10.68	8.89	14.54	14.51	-	10.59	9.75	10.31	-	-	-	11.07	12.36	26.34	15.27	12.70	21.40	13.54
13	14.93	14.57	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	8.82	9.26	-	4.98	6.27	-	-	-	-	-	-	-	7.61	-	17.85	-	-	17.88	4.71
15	17.49	15.75	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	10.07	10.17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	21.74	15.56	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	24.05	26.44	-	9.39	14.17	-	26.75	19.36	-	-	-	-	-	36.18	13.98	9.75	13.93	-	35.10
19	-	-	-	6.22	-	-	-	-	-	15.18	10.77	8.24	-	-	-	-	-	-	-

Table 19: *Variance: Reference (perfect labels) versus Approach 3 (HSMM)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	192.66	131.62	-	163.24	14.55	83.55	39.88	64.32	1030.91	-	-	-	102.38	824.96	100.69	109.63	303.00	783.54	103.21
2	206.68	233.34	490.80	33.79	919.16	-	101.09	1578.99	407.69	-	-	-	103.46	786.73	107.08	179.03	122.56	192.31	28.02
3	314.89	-	-	19.64	5.20	-	-	-	-	-	-	-	163.75	347.58	179.57	292.38	-	69.34	17.93
4	35.94	12.63	16.96	103.62	40.22	-	90.43	33.64	27.07	-	-	-	18.04	4.41	145.65	43.57	8.67	91.63	431.59
5	13.18	13.18	722.22	119.79	174.12	-	-	19.89	18.48	-	-	-	5.43	5.92	-	30.67	128.08	348.98	16.29
6	49.00	-	-	23.93	6.08	-	-	-	-	-	-	-	-	-	175.26	134.21	-	211.67	32.23
7	37.58	32.53	48.73	20.94	18.16	-	-	93.53	254.68	-	-	-	34.34	139.02	212.41	85.13	17.17	134.11	40.01
8	337.96	18.80	-	30.32	-	-	-	-	-	-	-	-	-	-	-	-	-	111.65	43.60
9	518.16	1567.17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	16.11	31.93	41.94	26.09	-	-	-	-	2025.63	-	-	-	-	19.35	-	-	10.77	345.88	138.24
11	773.37	37.89	38.54	55.21	-	-	-	-	14.88	-	-	-	-	30.85	58.27	-	53.14	301.28	321.78
12	253.17	16.54	12.51	299.77	136.29	-	102.04	34.80	29.02	-	-	-	38.26	47.42	284.18	84.09	46.98	247.68	187.21
13	167.24	86.65	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	195.90	177.94	-	22.36	12.93	-	-	-	-	-	-	-	34.88	-	254.87	-	-	1643.65	13.36
15	174.06	110.88	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	138.56	68.24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	281.55	149.50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	232.60	322.68	-	62.64	147.05	-	205.02	145.68	-	-	-	-	-	63.12	125.86	137.04	82.64	-	332.08
19	-	-	-	22.58	-	-	-	-	-	347.34	372.68	138.33	-	-	-	-	-	-	-

Table 20: *Mean: Baseline versus Approach 3 (HSMM)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	9.48	4.94	-	2.73	2.19	6.19	2.12	1.50	17.34	-	-	-	6.48	14.39	14.63	9.84	13.79	17.44	2.98
2	9.40	7.58	14.13	1.51	4.99	-	1.85	6.31	19.04	-	-	-	7.57	12.32	7.53	8.44	5.06	9.94	1.86
3	12.73	-	-	0.70	0.73	-	-	-	-	-	-	-	5.13	11.36	14.77	17.29	-	9.87	0.14
4	1.47	1.51	2.64	7.29	4.16	-	3.53	1.95	2.20	-	-	-	1.25	4.38	2.95	3.88	3.33	3.39	3.85
5	1.44	2.23	5.35	4.73	11.25	-	-	0.00	0.97	-	-	-	1.71	4.25	-	2.00	10.24	16.80	2.68
6	6.07	-	-	1.90	0.40	-	-	-	-	-	-	-	-	-	11.97	7.58	-	5.06	1.72
7	1.97	1.65	3.19	1.72	1.44	-	-	15.33	6.12	-	-	-	1.44	3.19	8.95	2.73	0.96	6.13	2.57
8	2.05	0.56	-	2.23	-	-	-	-	-	-	-	-	-	-	-	-	-	7.98	3.05
9	7.12	9.69	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	1.71	1.83	2.00	3.68	-	-	-	-	5.73	-	-	-	-	5.62	-	-	6.89	12.89	3.84
11	6.67	5.94	3.99	4.07	-	-	-	-	1.44	-	-	-	-	6.31	3.67	-	16.23	7.23	8.10
12	3.54	1.60	1.70	4.88	4.83	-	1.71	4.53	2.86	-	-	-	1.96	5.79	6.63	6.39	9.37	11.74	8.28
13	8.46	3.78	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	7.59	5.76	-	2.35	1.73	-	-	-	-	-	-	-	6.78	-	23.02	-	-	15.09	2.56
15	7.58	3.93	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	5.97	6.29	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	11.95	3.10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	6.09	9.48	-	3.76	8.81	-	7.32	7.13	-	-	-	-	-	13.40	6.33	5.62	7.04	-	12.51
19	-	-	-	9.11	-	-	-	-	-	13.06	12.10	8.86	-	-	-	-	-	-	-

Table 21: *Variance: Baseline versus Approach 3 (HSMM)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	380.39	27.94	-	152.67	7.48	58.63	20.85	10.45	1330.31	-	-	-	145.43	927.31	297.14	127.92	205.83	717.09	98.98
2	291.73	108.84	135.73	10.87	930.97	-	14.66	1628.20	460.81	-	-	-	115.56	390.59	239.00	99.84	47.18	14.43	17.28
3	466.09	-	-	2.46	2.46	-	-	-	-	-	-	-	73.49	318.14	276.25	304.14	-	13.04	0.53
4	17.11	3.59	10.75	94.64	31.19	-	16.71	12.52	24.30	-	-	-	13.70	2.14	22.15	21.48	6.88	15.21	423.12
5	4.91	3.52	807.42	76.18	302.17	-	-	0.00	4.16	-	-	-	13.72	4.27	-	3.12	31.86	421.53	9.97
6	14.05	-	-	7.82	1.57	-	-	-	-	-	-	-	-	-	212.37	167.95	-	12.44	8.82
7	18.64	4.13	11.83	9.65	6.02	-	-	550.40	106.64	-	-	-	3.78	10.33	269.24	51.75	3.03	15.21	32.43
8	377.43	1.92	-	7.42	-	-	-	-	-	-	-	-	-	-	-	-	-	43.10	10.14
9	451.13	956.38	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	9.39	11.91	12.74	11.22	-	-	-	-	2001.60	-	-	-	-	11.13	-	-	74.66	513.35	33.13
11	764.32	24.41	19.00	17.31	-	-	-	-	3.74	-	-	-	-	7.19	14.48	-	294.28	43.89	246.71
12	275.15	3.92	4.20	221.02	63.99	-	6.40	156.56	12.10	-	-	-	20.06	22.52	97.33	147.25	76.70	77.75	177.69
13	160.76	56.20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	422.21	133.18	-	7.78	4.04	-	-	-	-	-	-	-	102.82	-	1834.73	-	-	1993.74	7.87
15	276.48	26.88	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	167.98	109.78	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	509.11	5.60	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	134.01	314.91	-	76.83	227.07	-	148.50	115.61	-	-	-	-	-	21.09	98.73	79.50	480.87	-	603.87
19	-	-	-	32.16	-	-	-	-	-	448.15	425.22	171.50	-	-	-	-	-	-	-

Approach 4 (HSMM with unbounded duration)

Table 22: *Mean: Reference (perfect labels) versus Approach 4 (HSMM with unbounded duration)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	21.40	22.37	-	6.43	5.44	12.99	6.00	5.96	26.68	-	-	-	13.87	24.54	20.93	16.40	37.04	14.48	6.92
2	25.24	30.78	32.18	6.32	9.49	-	9.86	5.23	32.06	-	-	-	25.26	32.59	20.32	21.36	16.40	16.17	6.66
3	30.89	-	-	8.08	5.37	-	-	-	-	-	-	-	22.75	16.95	39.52	18.80	-	6.92	6.87
4	8.59	11.45	9.24	19.02	10.40	-	8.58	8.00	9.01	-	-	-	7.80	12.58	33.50	10.05	11.25	12.91	6.97
5	5.95	9.47	12.29	12.36	25.78	-	-	7.80	6.95	-	-	-	4.20	11.74	-	7.12	19.03	21.89	4.60
6	9.67	-	-	6.57	4.72	-	-	-	-	-	-	-	-	-	12.09	8.25	-	8.46	6.73
7	7.94	9.19	9.45	5.64	5.20	-	-	18.65	16.09	-	-	-	6.65	16.36	17.72	9.53	6.03	11.59	5.86
8	9.06	10.56	-	6.12	-	-	-	-	-	-	-	-	-	-	-	-	-	12.41	7.70
9	21.55	28.25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	7.01	9.63	9.48	8.83	-	-	-	-	6.59	-	-	-	-	11.76	-	-	8.55	56.53	12.34
11	8.59	11.88	8.96	9.60	-	-	-	-	6.56	-	-	-	-	13.95	8.47	-	12.96	44.78	16.51
12	9.03	10.79	9.77	14.50	14.11	-	10.03	9.75	10.09	-	-	-	11.30	12.63	28.24	14.48	11.15	23.03	13.91
13	16.91	13.30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	9.77	8.68	-	5.96	6.87	-	-	-	-	-	-	-	7.03	-	25.12	-	-	9.39	5.35
15	19.63	16.22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	11.06	11.43	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	26.21	15.34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	24.12	27.44	-	11.94	13.50	-	24.31	18.79	-	-	-	-	-	33.90	14.12	8.85	14.40	-	35.68
19	-	-	-	6.27	-	-	-	-	-	21.12	10.15	8.10	-	-	-	-	-	-	-

Table 23: *Variance: Reference (perfect labels) versus Approach 4 (HSMM with unbounded duration)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	315.67	146.73	-	25.15	14.65	87.04	34.11	73.91	325.56	-	-	-	210.61	653.90	194.91	155.09	1341.65	67.81	36.24
2	509.69	67.49	398.72	20.89	915.65	-	234.89	14.97	595.35	-	-	-	5306.63	1010.10	145.40	251.06	132.95	124.44	27.60
3	183.50	-	-	106.66	12.00	-	-	-	-	-	-	-	161.64	167.37	305.09	360.54	-	29.26	19.25
4	31.41	12.08	15.67	173.34	39.56	-	93.68	30.14	43.76	-	-	-	26.62	6.95	183.59	42.38	5.33	92.54	71.01
5	12.79	13.15	707.16	111.76	225.51	-	-	19.48	18.76	-	-	-	7.71	5.48	-	26.95	156.20	300.20	22.33
6	34.27	-	-	13.39	6.08	-	-	-	-	-	-	-	-	-	84.53	58.90	-	196.22	32.90
7	34.20	37.17	40.83	15.26	14.06	-	-	726.95	283.96	-	-	-	37.98	410.11	242.01	64.86	23.08	88.54	45.54
8	14.18	18.05	-	27.71	-	-	-	-	-	-	-	-	-	-	-	-	-	81.95	40.17
9	268.29	460.52	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	19.16	28.99	30.29	36.73	-	-	-	-	18.21	-	-	-	-	22.02	-	-	43.28	411.13	316.50
11	27.81	23.39	14.57	51.87	-	-	-	-	22.94	-	-	-	-	51.03	59.18	-	58.56	308.94	304.76
12	17.62	22.38	8.08	196.37	174.61	-	105.77	54.82	44.24	-	-	-	51.18	46.19	385.87	97.75	50.94	262.47	219.43
13	980.03	75.11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	464.40	125.78	-	34.04	14.16	-	-	-	-	-	-	-	35.17	-	1431.80	-	-	72.06	18.53
15	436.27	116.28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	156.41	61.66	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	520.98	147.75	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	240.48	361.60	-	484.65	148.07	-	107.63	171.50	-	-	-	-	-	48.65	119.23	128.02	90.07	-	320.02
19	-	-	-	32.71	-	-	-	-	-	297.07	88.64	88.82	-	-	-	-	-	-	-

Table 24: *Mean: Baseline versus Approach 4 (HSMM with unbounded duration)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	5.38	6.84	-	1.59	1.42	4.36	1.69	1.36	4.42	-	-	-	6.21	8.76	5.58	7.64	13.54	4.06	1.99
2	7.76	10.38	5.82	1.30	5.49	-	2.55	0.83	7.07	-	-	-	17.91	8.83	6.96	5.68	4.52	3.77	2.16
3	3.67	-	-	1.94	1.09	-	-	-	-	-	-	-	3.80	7.06	6.78	14.10	-	1.89	0.20
4	1.25	1.96	2.19	5.01	3.65	-	1.84	1.02	0.90	-	-	-	1.75	5.35	2.19	3.13	1.66	4.76	2.27
5	1.33	1.69	7.59	3.65	7.62	-	-	0.31	1.16	-	-	-	2.85	5.92	-	0.86	5.03	7.56	2.24
6	4.31	-	-	1.60	0.40	-	-	-	-	-	-	-	-	-	14.77	4.79	-	2.66	0.84
7	1.67	2.18	1.86	0.70	0.55	-	-	5.27	4.79	-	-	-	2.00	6.25	4.68	1.37	0.96	3.02	1.61
8	0.93	0.97	-	0.92	-	-	-	-	-	-	-	-	-	-	-	-	-	4.43	1.84
9	4.19	6.49	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	1.62	2.26	1.25	3.68	-	-	-	-	0.92	-	-	-	-	5.24	-	-	4.35	11.36	4.60
11	4.02	4.87	4.39	4.11	-	-	-	-	2.23	-	-	-	-	7.26	3.35	-	9.58	5.24	8.67
12	2.82	2.22	2.58	2.78	3.99	-	1.57	3.24	2.26	-	-	-	1.72	6.10	4.89	5.41	3.99	9.22	8.09
13	9.27	5.11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	4.95	4.18	-	0.61	0.13	-	-	-	-	-	-	-	4.39	-	6.45	-	-	2.53	0.88
15	6.00	4.05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	5.01	3.30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	3.97	3.55	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	5.65	10.14	-	6.31	9.15	-	4.88	6.56	-	-	-	-	-	11.12	5.84	5.35	7.51	-	12.20
19	-	-	-	9.22	-	-	-	-	-	6.25	11.36	8.31	-	-	-	-	-	-	-

Table 25: *Variance: Baseline versus Approach 4 (HSMM with unbounded duration)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	112.75	18.38	-	5.47	5.55	41.61	11.25	13.35	49.36	-	-	-	69.24	304.04	114.67	74.10	889.24	9.20	29.87
2	310.55	122.11	137.15	2.85	918.96	-	146.20	2.38	96.86	-	-	-	5524.58	219.35	213.36	82.44	34.74	4.83	16.61
3	20.89	-	-	107.01	6.64	-	-	-	-	-	-	-	18.28	42.88	32.03	486.62	-	4.19	0.78
4	16.35	3.99	4.02	81.97	25.59	-	1.43	5.33	2.73	-	-	-	23.77	2.76	24.60	23.19	1.74	23.27	24.12
5	3.94	2.38	789.75	45.36	262.58	-	-	1.22	4.45	-	-	-	24.78	5.62	-	2.67	13.68	135.85	8.14
6	8.22	-	-	5.39	1.57	-	-	-	-	-	-	-	-	-	863.80	73.62	-	6.07	4.92
7	16.76	10.35	6.52	3.10	2.13	-	-	177.49	48.24	-	-	-	7.74	380.81	125.22	8.85	3.03	3.81	19.00
8	11.14	2.29	-	2.64	-	-	-	-	-	-	-	-	-	-	-	-	-	9.16	4.86
9	40.01	151.89	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	9.58	4.25	5.42	14.34	-	-	-	-	2.68	-	-	-	-	11.24	-	-	73.07	420.00	291.79
11	12.22	7.51	4.86	16.95	-	-	-	-	4.63	-	-	-	-	23.93	10.94	-	115.57	20.48	280.65
12	36.52	11.44	3.53	16.71	58.99	-	5.08	110.14	5.13	-	-	-	12.78	21.89	102.90	267.60	36.20	63.50	213.20
13	883.38	30.75	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	248.86	67.88	-	1.90	0.50	-	-	-	-	-	-	-	19.29	-	136.81	-	-	5.38	3.53
15	160.14	19.31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	143.65	30.27	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	182.96	9.97	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	153.30	355.34	-	541.23	263.02	-	83.49	150.96	-	-	-	-	-	20.04	77.33	35.52	475.60	-	637.96
19	-	-	-	45.49	-	-	-	-	-	109.67	130.64	110.18	-	-	-	-	-	-	-

Table 26: *Mean: Reference versus Approach 4 (HSMM with unbounded duration - special case: [sil])*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	21.46	18.26	-	6.42	5.12	11.76	6.29	5.75	21.26	-	-	-	12.18	24.44	15.01	15.41	28.83	26.64	6.74
2	20.58	25.19	34.18	7.21	7.85	-	8.80	5.83	26.59	-	-	-	14.45	30.50	17.63	20.59	16.14	25.19	6.92
3	31.19	-	-	7.31	4.64	-	-	-	-	-	-	-	20.63	15.00	20.76	14.13	-	12.26	7.00
4	8.29	11.22	8.38	16.36	9.59	-	7.50	7.62	8.11	-	-	-	7.23	11.70	32.60	9.87	9.26	14.41	6.42
5	5.95	8.78	10.38	12.01	27.23	-	-	7.50	6.82	-	-	-	3.84	9.94	-	7.98	13.65	34.75	4.59
6	10.95	-	-	6.76	4.76	-	-	-	-	-	-	-	-	-	16.33	13.34	-	11.25	5.94
7	8.54	9.51	8.67	6.20	5.26	-	-	9.42	14.23	-	-	-	6.43	12.69	17.68	11.21	5.41	14.95	5.17
8	9.49	10.76	-	6.79	-	-	-	-	-	-	-	-	-	-	-	-	-	16.12	7.56
9	20.28	25.01	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	6.78	9.57	9.98	8.87	-	-	-	-	6.59	-	-	-	-	11.80	-	-	6.70	51.77	9.66
11	9.83	13.52	8.62	10.28	-	-	-	-	5.77	-	-	-	-	12.83	8.42	-	10.65	49.02	15.49
12	9.10	10.61	8.95	14.44	14.79	-	9.60	9.37	10.27	-	-	-	11.20	12.31	24.70	14.83	12.87	25.76	14.98
13	14.70	14.22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	9.35	9.20	-	4.94	6.34	-	-	-	-	-	-	-	7.61	-	18.46	-	-	11.56	5.12
15	17.29	15.57	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	10.19	10.67	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	21.87	15.12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	15.86	14.97	-	9.06	9.94	-	17.22	10.80	-	-	-	-	-	31.62	9.42	8.71	11.82	-	34.72
19	-	-	-	6.28	-	-	-	-	-	14.42	10.19	8.23	-	-	-	-	-	-	-

Table 27: *Variance: Reference versus Approach 4 (HSMM with unbounded duration - special case: [sil])*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	198.30	141.52	-	23.78	14.61	82.04	39.71	62.75	230.18	-	-	-	95.74	498.62	102.96	128.45	279.56	210.94	32.89
2	207.37	233.34	445.31	35.68	388.63	-	109.88	18.87	383.45	-	-	-	67.12	797.44	104.16	181.97	122.61	202.28	27.58
3	204.88	-	-	22.08	10.04	-	-	-	-	-	-	-	163.60	363.38	192.52	266.57	-	70.80	18.99
4	21.71	13.12	19.59	103.67	40.68	-	77.03	29.98	27.47	-	-	-	18.04	4.41	145.05	47.34	11.17	98.68	49.93
5	12.81	11.31	721.97	106.05	174.12	-	-	22.16	18.28	-	-	-	6.20	6.31	-	23.15	86.06	346.11	15.93
6	50.76	-	-	23.96	5.88	-	-	-	-	-	-	-	-	-	175.26	202.00	-	239.29	32.11
7	37.23	32.80	48.73	17.82	18.16	-	-	93.46	211.56	-	-	-	36.57	136.69	209.00	79.35	17.17	134.59	34.22
8	35.41	18.21	-	29.41	-	-	-	-	-	-	-	-	-	-	-	-	-	137.18	43.07
9	202.80	396.47	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	15.45	35.01	36.63	27.69	-	-	-	-	18.76	-	-	-	-	18.80	-	-	9.93	333.88	140.02
11	40.62	38.81	42.65	57.34	-	-	-	-	13.16	-	-	-	-	32.16	58.27	-	57.69	268.66	309.86
12	17.72	16.66	13.54	146.29	119.65	-	107.28	35.66	27.24	-	-	-	43.20	47.60	279.54	74.81	48.04	284.76	238.34
13	163.98	82.62	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	221.61	120.67	-	22.81	14.69	-	-	-	-	-	-	-	34.88	-	262.82	-	-	128.71	20.60
15	179.39	111.90	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	102.81	80.62	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	288.84	142.61	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	94.18	34.41	-	33.16	26.20	-	11.22	24.70	-	-	-	-	-	45.02	66.61	41.74	86.18	-	248.31
19	-	-	-	24.20	-	-	-	-	-	137.92	92.86	80.69	-	-	-	-	-	-	-

Institut für
Technische Informatik und
Kommunikationsnetze

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich
Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo

(SA-2010-10)

SEMESTERARBEIT

für

Herrn Patrick Wyss

Betreuer: S. Hoffmann, ETZ D97.5
Dr. B. Pfister, ETZ D97.6

Ausgabe: 04. Oktober 2010
Abgabe: 24. Dezember 2010

HMM-Spracherkenner mit diskreten oder kontinuierlichen Merkmalen

Einleitung

Statistische Modelle in der Sprachverarbeitung benötigen annotiertes Sprachmaterial für das Training. Speziell die Prosodiesteuerung in der Sprachsynthese benötigt unter anderem Sprachsignale, in denen die Laute möglichst präzise lokalisiert und annotiert sind. Da die manuelle Segmentierung des Sprachsignals sehr aufwändig ist, werden heutzutage meist semi-automatische Methoden eingesetzt, in denen das Signal mit Hilfe automatischer Methoden vorsegmentiert wird, und dann manuell nachbearbeitet. Für die automatische Segmentierung bringt die Verwendung von Hidden-Markov-Modellen (HMM) die besten Ergebnisse. Weil sie jedoch dafür gemacht wurden, statische Ereignisse zu detektieren, werden Lautgrenzen gefunden, die sehr unpräzise sind.

Ein Nachteil der HMMs ist der, dass der zugrundeliegende statistische Prozess nur eine exponentiell verteilte Aufenthaltswahrscheinlichkeit für jeden Zustand modellieren kann. Im konkreten Fall der Segmentierung entspricht das einer exponentiellen Verteilung der Lautlänge.

Um diese Einschränkung zu umgehen, wurden Hidden-Semi-Markov-Modelle (HSMM) entwickelt, die erlauben, dass der zugrundeliegende statistische Prozess als Semi-Markov-Kette modelliert wird. Dadurch können andere Verteilungen der Lautlänge modelliert werden oder es können auch ganz andere Eigenschaften des Lautes einbezogen werden. [1]

gibt einen Überblick über die verschiedenen Varianten der HSMM und verweist auch auf Anwendungen in Spracherkennung und -synthese.

Problemstellung

Im Rahmen dieser Arbeit soll untersucht werden, welchen Einfluss das unterliegende statistische Modell auf die Genauigkeit der Lautgrenzen bei der Segmentierung hat. Dabei soll nicht nur die globale Verbesserung der Segmentierung im Allgemeinen betrachtet werden, sondern auch speziell untersucht werden, welche Lautübergänge besonders problematisch sind, und welche weniger. Dazu sollen verschiedene Varianten des HSMM implementiert werden und die Ergebnisse der Segmentierung auf einem Testkorpus verglichen werden.

Vorgehen

Für diese Semesterarbeit wird das folgende Vorgehen empfohlen:

1. Zuerst sollte sich in der Literatur ein kurzer Überblick über die verschiedenen HSMM-Varianten und ihre bisherige Anwendung in Spracherkennung und -synthese verschafft werden. Ausgangspunkt sollte dabei [1] sein.

2. Dann ist die bestehende HMM-Segmentierung bezüglich Präzision der Segmentierung für verschiedene Lautübergänge zu untersuchen. Dabei empfiehlt es sich, die Laute ähnlich wie in [2] in grobe Klassen einzuteilen (Vokale, Approximanten, Nasale, etc.) und Übergänge zwischen den Klassen zu betrachten. Ein Sprachkorpus mit einer Referenzsegmentierung wird zur Verfügung gestellt.

3. Mit Hilfe der Betreuer sind relevante HSMM-Varianten auszuwählen. Diese sollen dann implementiert und ihre Eigenschaften bezüglich Segmentierungsqualität untersucht werden.

Die ausgeführten Arbeiten und die erhaltenen Resultate sind in einem Bericht zu dokumentieren (siehe dazu [3]), der in gedruckter und in elektronischer Form (als PDF-Datei) abzugeben ist. Zusätzlich sind im Rahmen eines Kolloquiums zwei Präsentationen vorgesehen: etwa zwei Wochen nach Beginn soll der Arbeitsplan und am Ende der Arbeit die Resultate vorgestellt werden. Die Termine werden später bekannt gegeben.

Literaturverzeichnis

[1] Shun-Zheng Yu, "Hidden semi-markov models," *Artificial Intelligence*, 2009, (http://people.cs.ubc.ca/~murphyk/Teaching/CS540-Spring10/projects/Yu-hsmm09.pdf).

[2] Ladan Baghai-Ravary, Greg Kochanski, and John Coleman, "Precision of phoneme boundaries derived using hidden markov models," in *INTERSPEECH-2009*, 2009, (http://kochanski.org/gpk/papers/2009/IS090244.pdf).

[3] B. Pfister, "*Richtlinien für das Verfassen des Berichtes zu einer Semester- oder Diplomarbeit*," Institut TIK, ETH Zürich, Februar 2009,
(http://www.tik.ee.ethz.ch/~spr/SADA/richtlinien_bericht.pdf).

[4] B. Pfister, "*Hinweise für die Präsentation der Semester- oder Diplomarbeit*," Institut TIK, ETH Zürich, März 2004,
(http://www.tik.ee.ethz.ch/~spr/SADA/hinweise_praesentation.pdf).

Zürich, den 17. September 2010