**Title**

A comparison of different approaches to automatic speech segmentation. In Text, Speech and Dialogue (pp. 277-284)

**Aims**

Comparison of manual methods of segmentation to the following automatic methods

1. Viberti algorithm
2. Forward-backward algorithm
3. Manual-auto adjustments

**Key Concepts**

SEGMENTATION ALGORITHMS, ML

**Finding**

* In a large body of speech recordings or speech corpora, timestamps can be used to provide segment annotation
* Segments can be selected based on acoustic features of the recording

**Segmentation Types**

1. Acoustic model based segmentation
2. Transcript based …
3. Orthography based

**Orthography based segmentation**

Includes the following Phases

1. Phonetic transcriptions based on speech recognition
2. Phonetic representation matching

**Phase 1: Phonetic Transcription**

This involves

***Grapheme2Phoneme Transcription:***

Involved a pronunciation dictionary of about 200000 entries and training was based on Induction Decision Tree (ID3) mechanism.

***Lexicon Matching:***

The lexicon was done based on word-level segmentation. However this research is phoneme-based.

**Phase 2: Segment Matching**

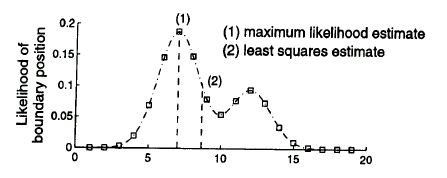
This was done using the following algorithms

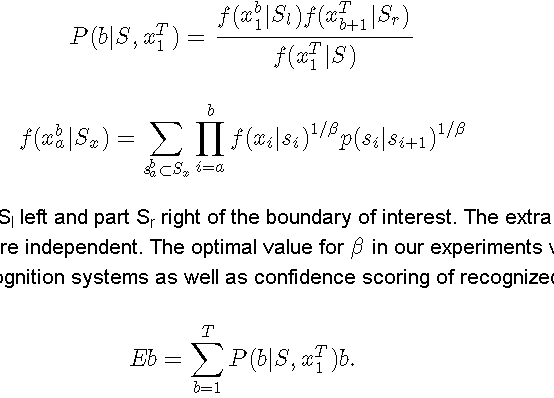
***Viterbi Segmentation***

This is a hidden markov model (HMM) finite state machine (FSM) such that the best state-path is returned when given an input speech signal xT1 or a corresponding sequence of feature vectors. Thus

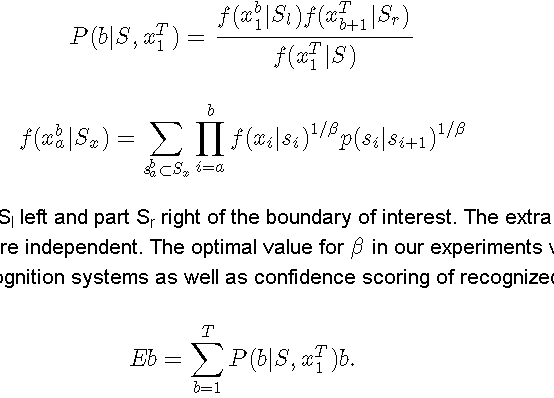


***Forward-backward Segmentation***

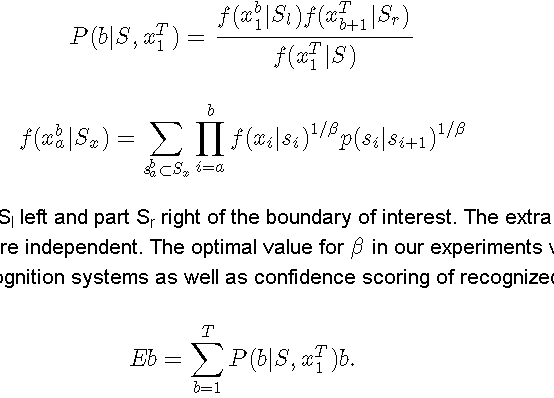
*The Viterbi algorithm provides an approximation of the quantity being searched for in figure 1 below**The Viterbi algorithm generates the boundary corresponding to (1), whereas the optimal boundary in a least squares sense matches up with (2). To find the best possible estimate of the boundary in least squares sense the probaility function of each boundary must be calculates as follows*



with



\*In the above equations, sentence S is divided in part Sl left and part Sr right of the boundary of interest. The extra parameter β compensates for the ill-matched assumption made by HMMs that the observations xi are independent. The optimal value for β in our experiments was 10, but its exact value was not at all critical. The same compensation factor is found in recognition systems as well as confidence scoring of recognized words. Given the probability desnity function, the least squares estimate now equals:

.

**Post Processing**

Demuynck & Laureys (2002) also observed that errors can be introduced to the algorithm from the following sources

1. Broken-off Words *also with broken off orthography*
2. Mispronounced words
3. Dialectical pronunciations
4. cross-word phonetical phenomena *such as assimilation, degemenation and inserted linking phonemes*

These can be handled by a context sensitive knowledge base read-write rules. In addition post processing could be used to remove biases in boundary detection between human and hmm cues.

**Limitation**

Uses phonetic transcriptions

**Relation**

Segmentation algorithms

**Comment**

Demuynck, K., & Laureys (2002) use 2 algorithms to compare manual speech segmentation to automatic speech segmentation. Both algorithms requires the use of Hidden Markov Models. The strength of these methods is also a weakness. It uses dependent probability to more accurately determine the predicted value based on trained data. The more dependencies the more accurate the prediction will be within the limit of accuracy.

**REFERENCE**

Demuynck, K., & Laureys, T. (2002, January). A comparison of different approaches to automatic speech segmentation. In Text, Speech and Dialogue (pp. 277-284). Springer Berlin Heidelberg.