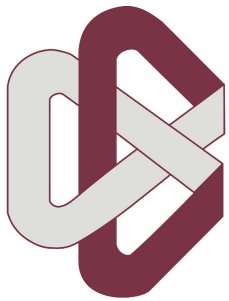
A Markov model for automatic segmentation in audio recodings

Rafael de Jesús Robledo Juárez
rrobledo@cimat.mx

Advisor: Salvador Ruíz-Correa Ph.D

Computer Science Department
Center for Mathematical Research (CIMAT), Guanajuato

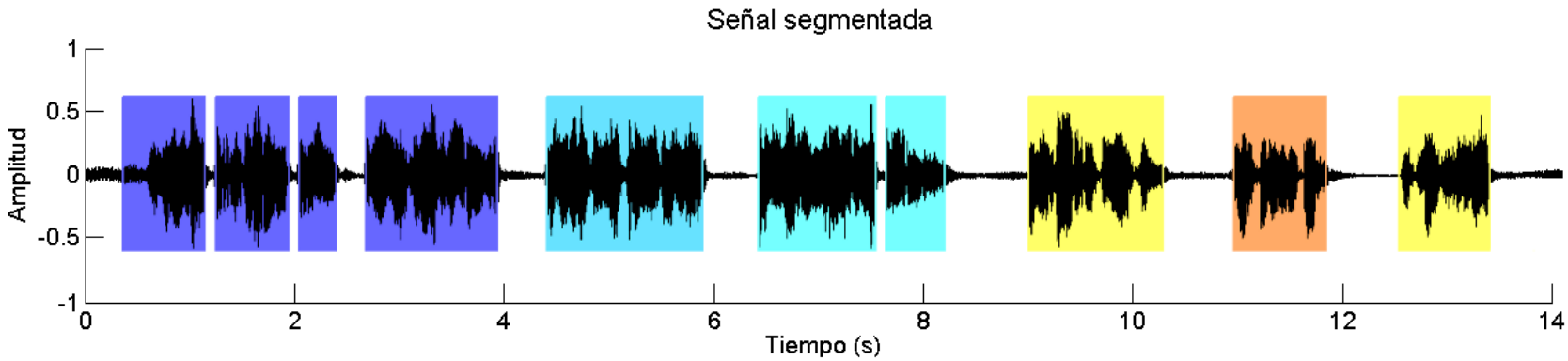


CIMAT

Preliminary thesis presentation, 2013

Main objectives

- Use a Hidden Markov Model to partitioning an audio recording into similar-speaker regions. Each segment should correspond to a different speaker.
- This task is known as Speaker diarization. Has two main stages: segmentation and clustering.



- Each of the segments will be labeled according to different speakers found in the conversation.

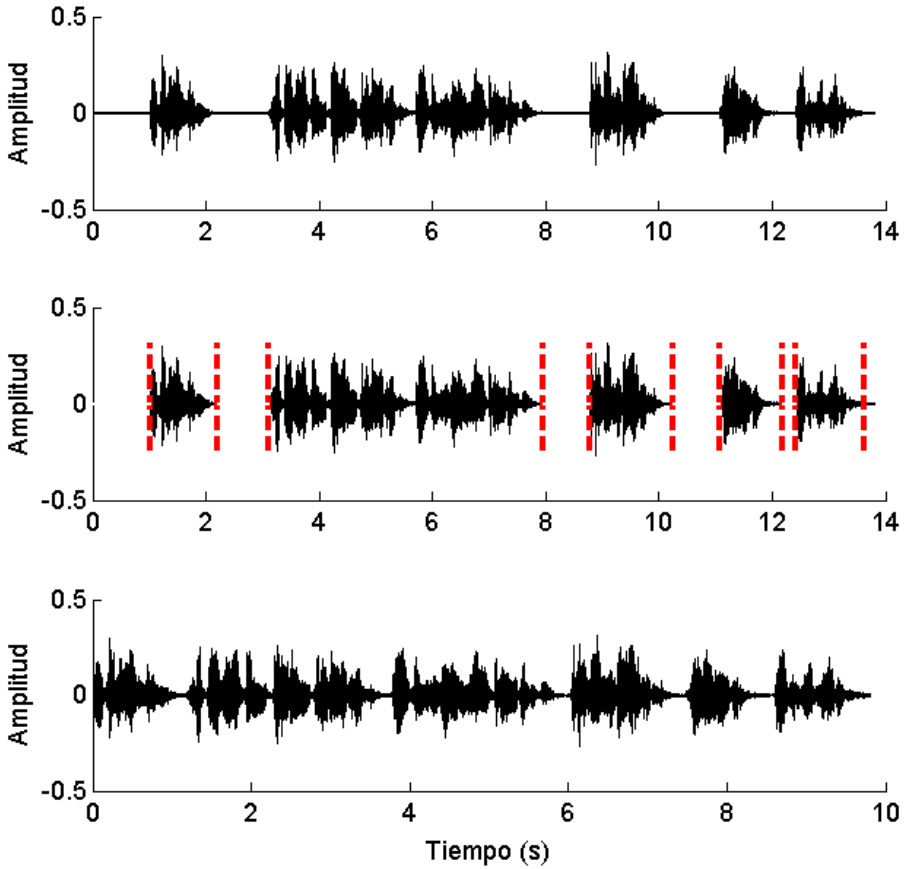
Motivation

- Speaker identification in a audio recording it's an important stage for improving results in different NPL applications such as automatic transcription and speech recognition, as long as it allow us to fit a model for each different found speaker.

- A remarkable point in the automation of this process is to perform segmentation with out needing of priori knowledge on the number or gender of persons involved in recording.

Methodology

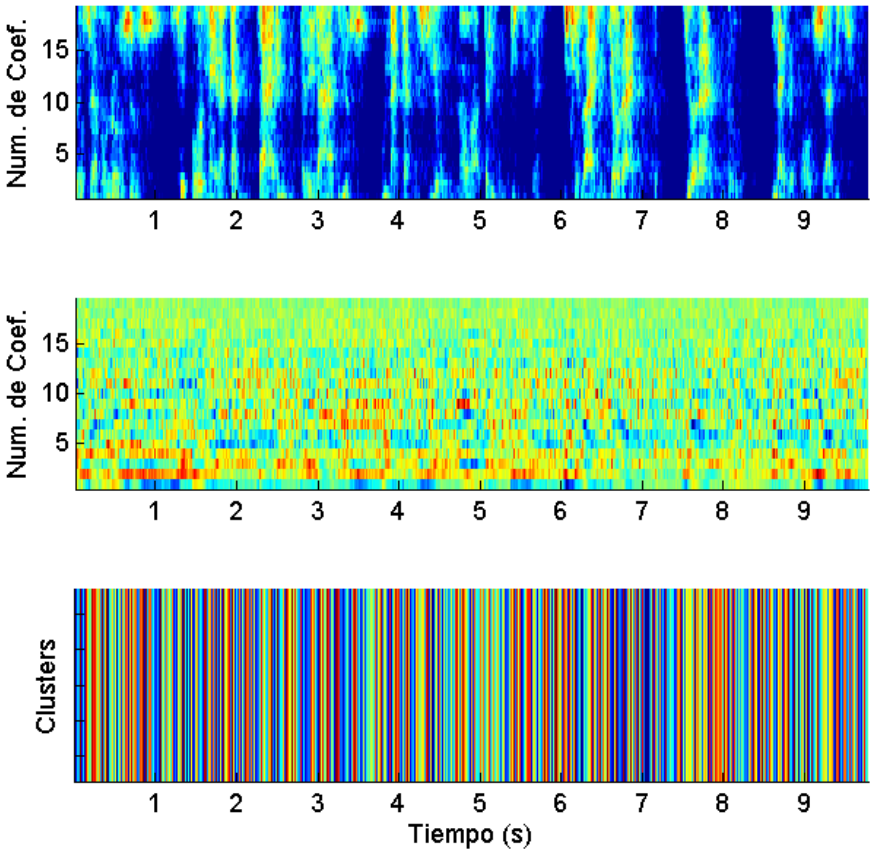
Signal processing (1)



- Original signal
- Silence identifier
- Truncated signal

Methodology

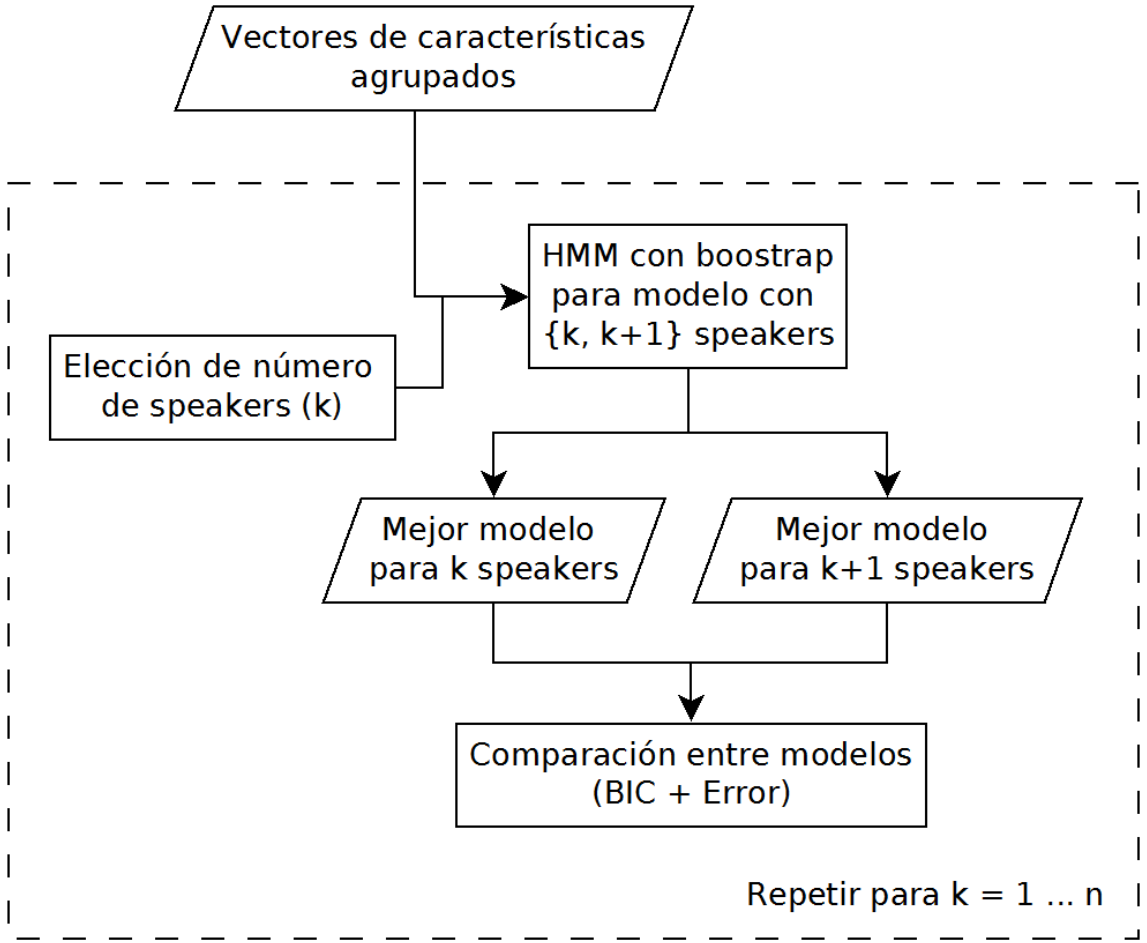
Signal processing (2)



- FFT + Filter Bank
- log + DCT
- k-means ++

Methodology

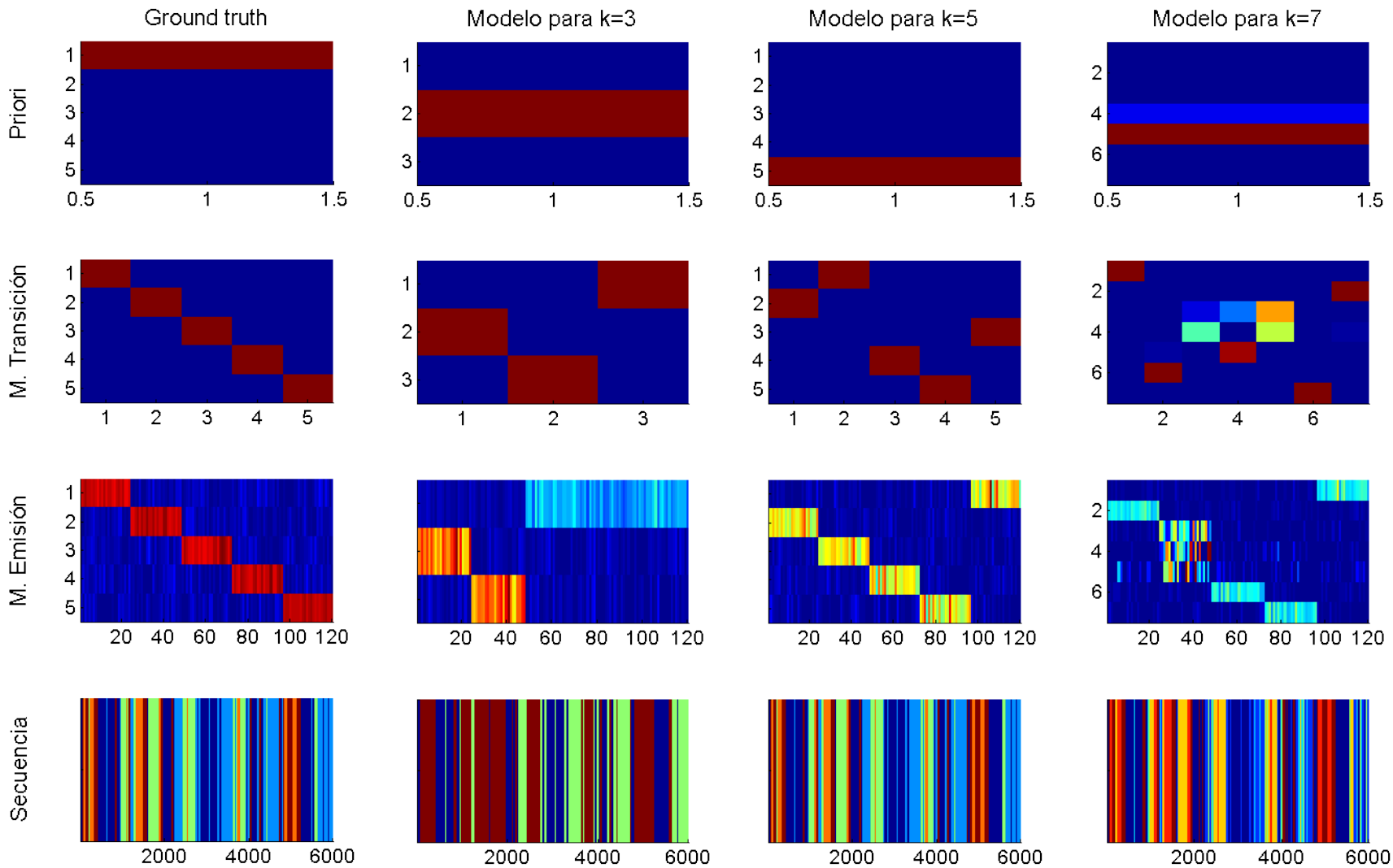
Model



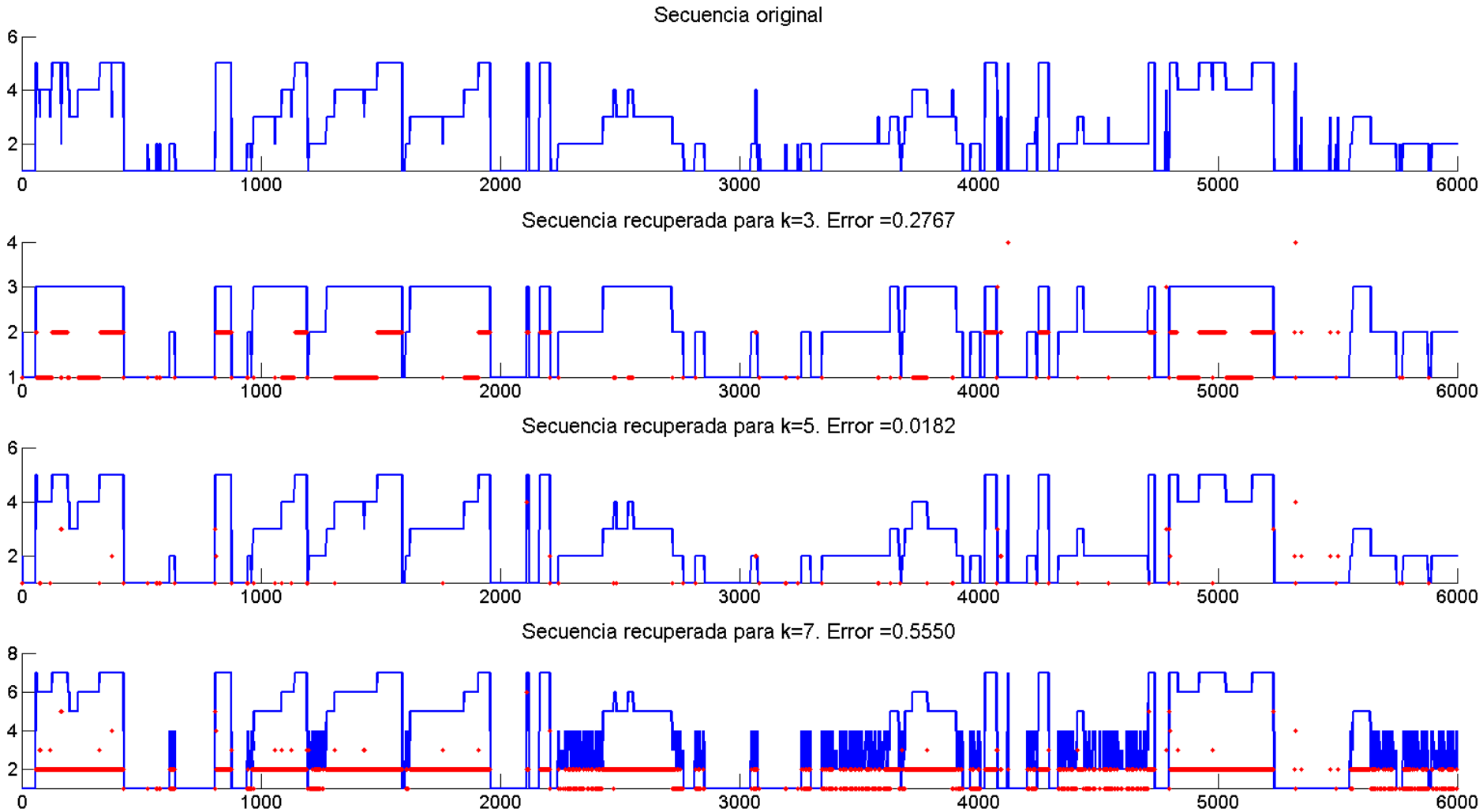
Progress

- System implementation completed.
- Results for synthetic data (randomly generated).
- Starting tests for synthetic voice recordings.

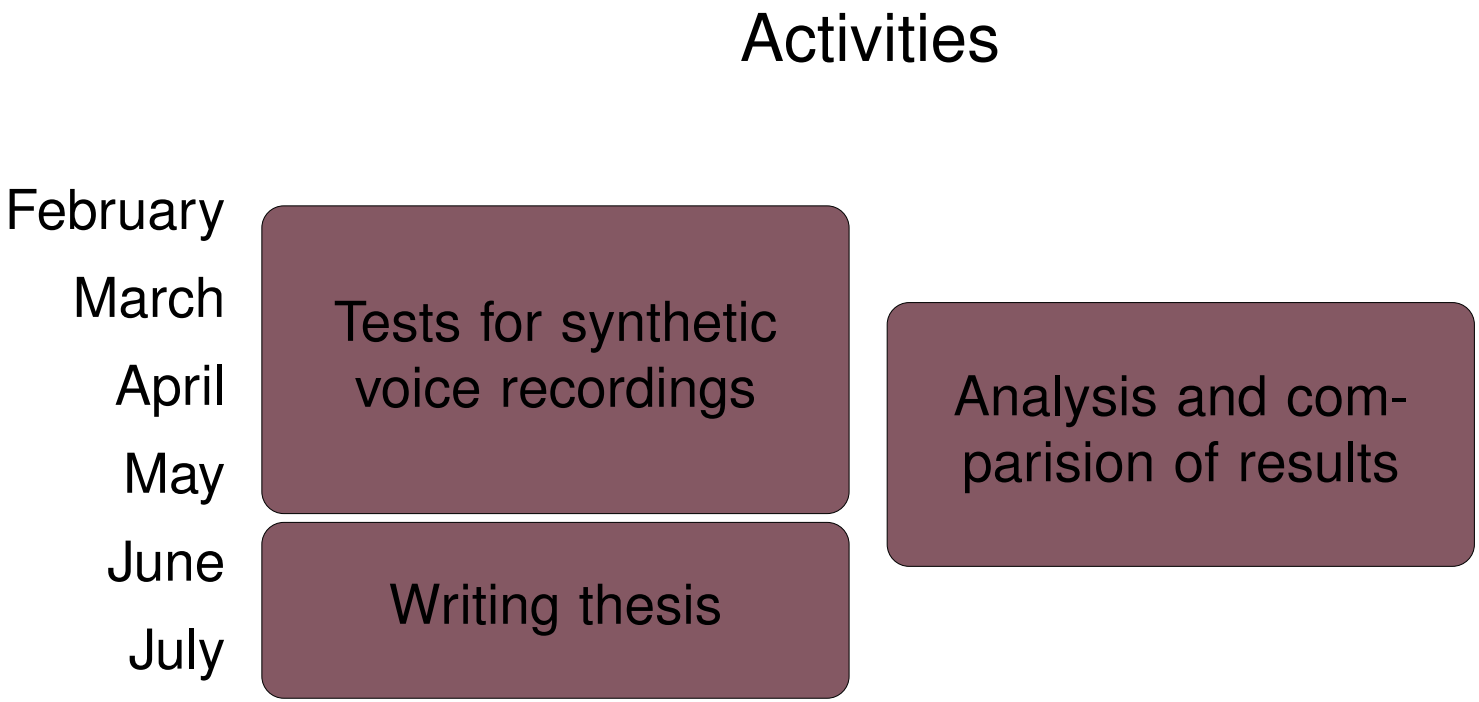
Synthetic data tests



Synthetic data tests



Schedule


Activities
February
March
April
May
June
July
Tests for synthetic voice recordings
Writing thesis
Analysis and comparision of results

Reference material

L.R. Rabiner, B.H. Juang

Fundamentals of speech recognition

Pearson Education India, 2008.

C. M. Bishop.

Pattern Recognition and Machine Learning.

Springer, 2006.

L.R. Rabiner

A tutorial on hidden Markov models and selected applications in speech recognition

Proceedings of the IEEE, 1989.

T. Rydén.

EM versus Markov chain Monte Carlo for Estimation of Hidden Markov Models: A Computational Perspective

Bayesian Analysis (2008) 3, Number 4, p. 659-688

E.B. Fox, E.B. Sudderth, M.I. Jordan, A.S. Willsky.

A sticky HDP-HMM with application to Speaker diarization.

Annals of Applied Statistics, 2011.