Characterisation and simulation of telephone channels using the TIMIT and NTIMIT databases

Herman Kamper and Thomas Niesler

Department of Electrical and Electronic Engineering
Stellenbosch University

30 November 2009

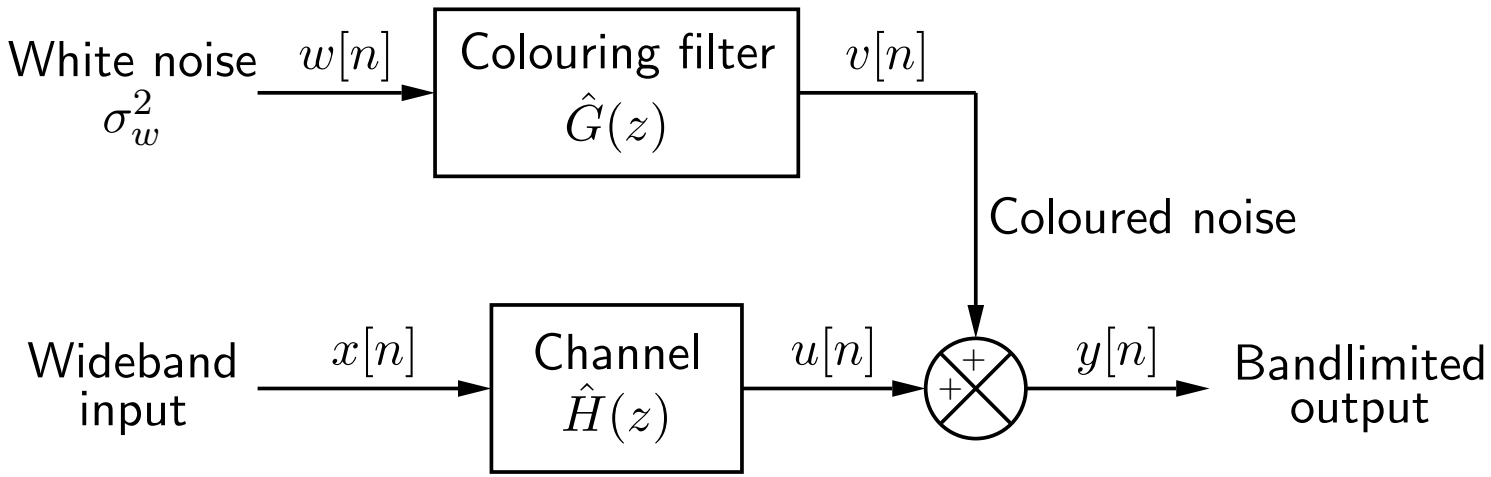
Introduction

- Speech recognition systems are often telephone-based
- Requires speech recorded over a variety of telephone channels
- Compilation of such corpora often expensive or impractical
- Paper describes techniques that allow a variety of telephone channels to be simulated, given wideband recordings

Analysis of telephone channels

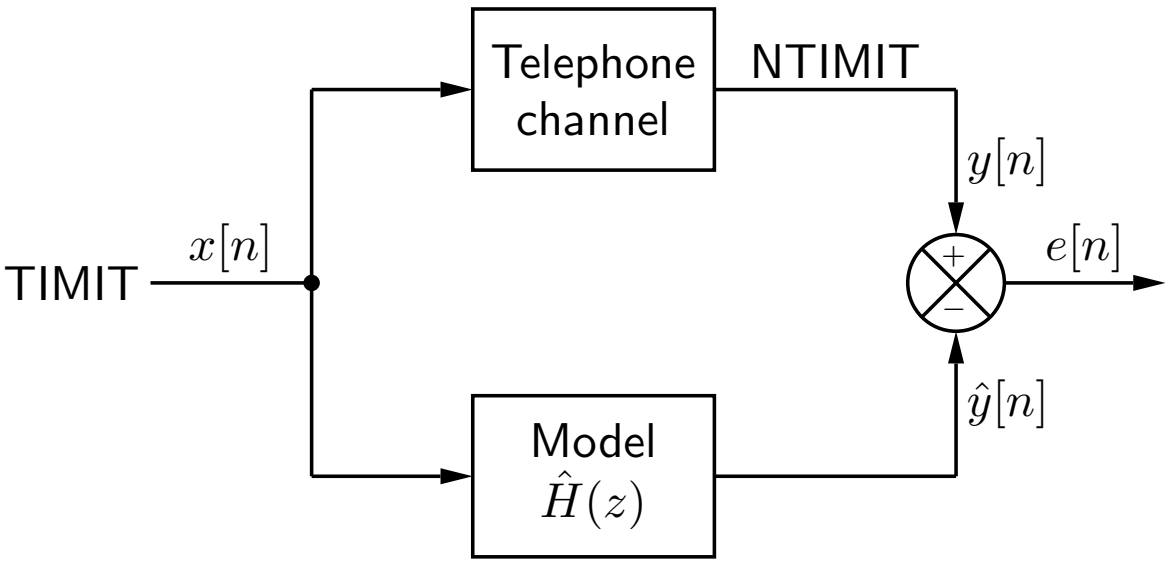
- Used the TIMIT and NTIMIT corpora
- Investigated channel (bandlimiting) characteristics
- Investigated noise which is added by telephone channel

Model of the telephone channel



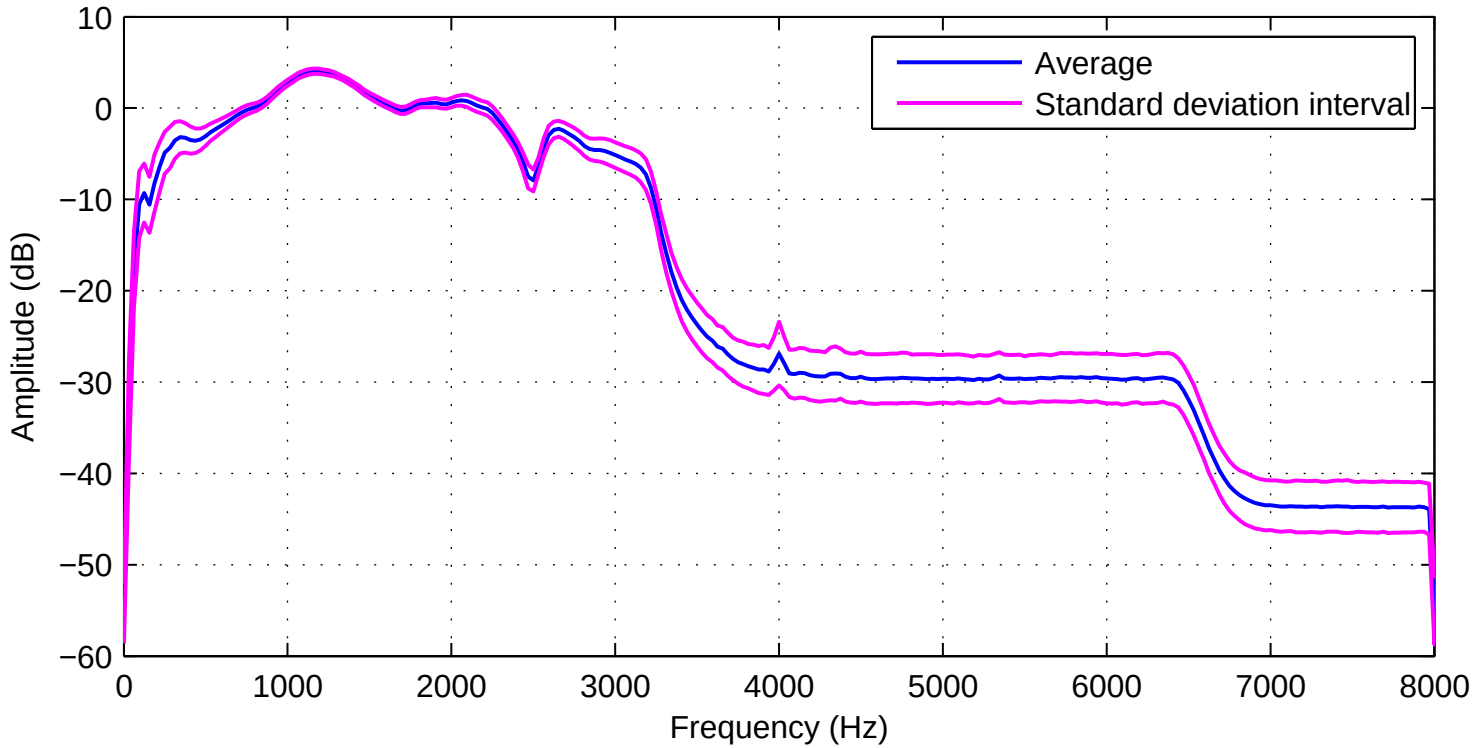
Channel analysis

- Parametric channel modelling was evaluated (below)
- Spectral channel analysis techniques were also evaluated
- Used synthetic filters to evaluate the different techniques



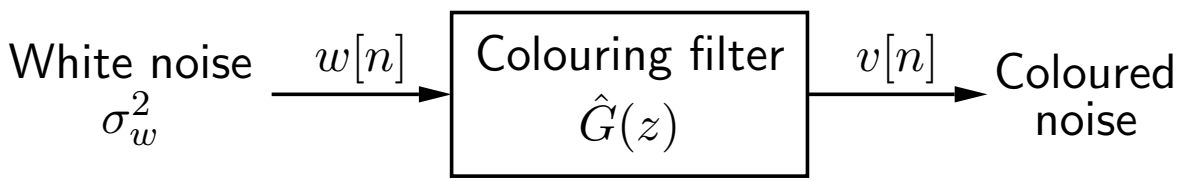
Design of channel model

- Analysed the 253 NTIMIT telephone channels
- Used a spectral analysis technique
- Two possibilities for channel model:
  - Use filter from channel library
  - Generate random filter based on distributions

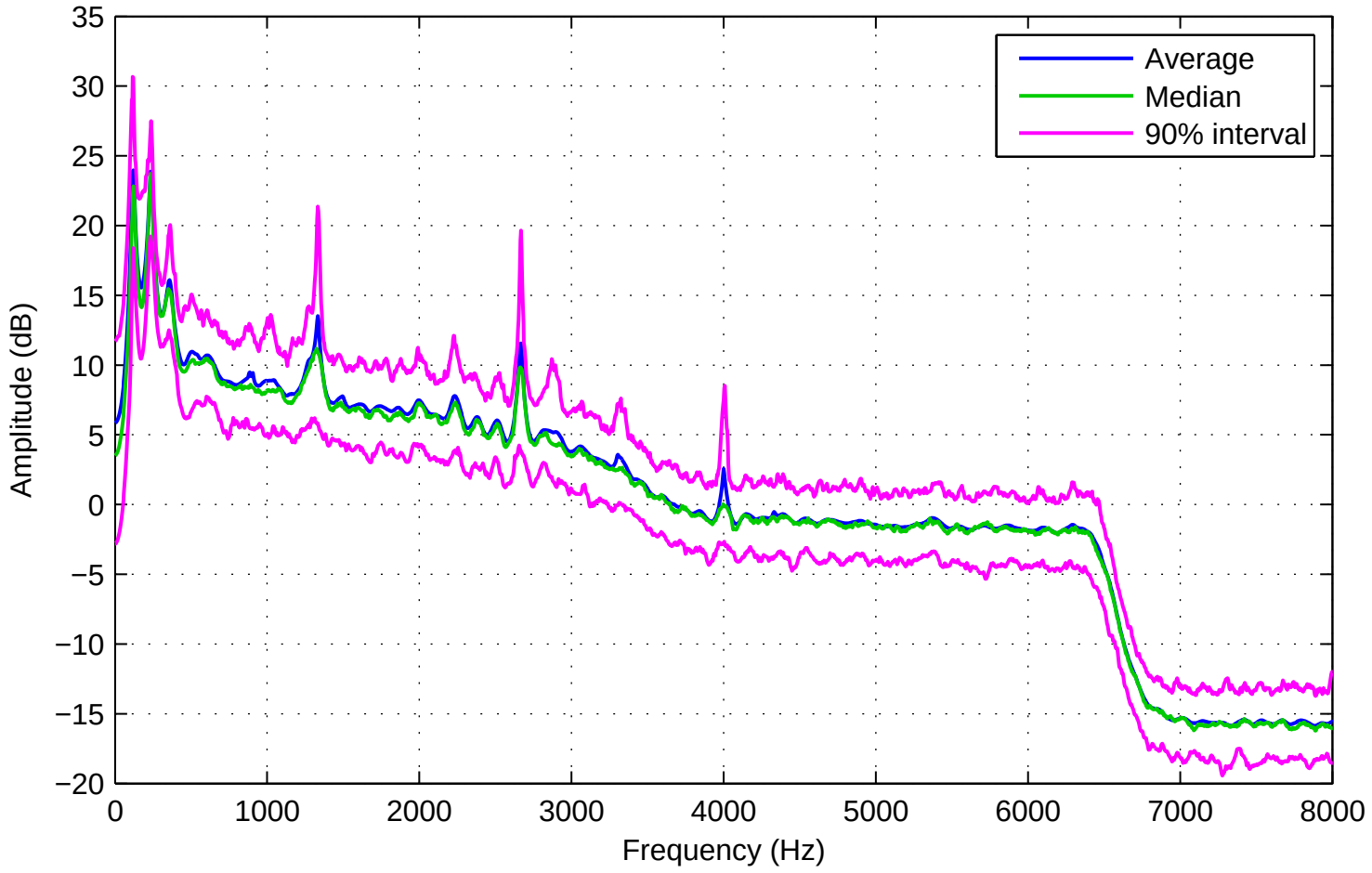


Noise analysis I

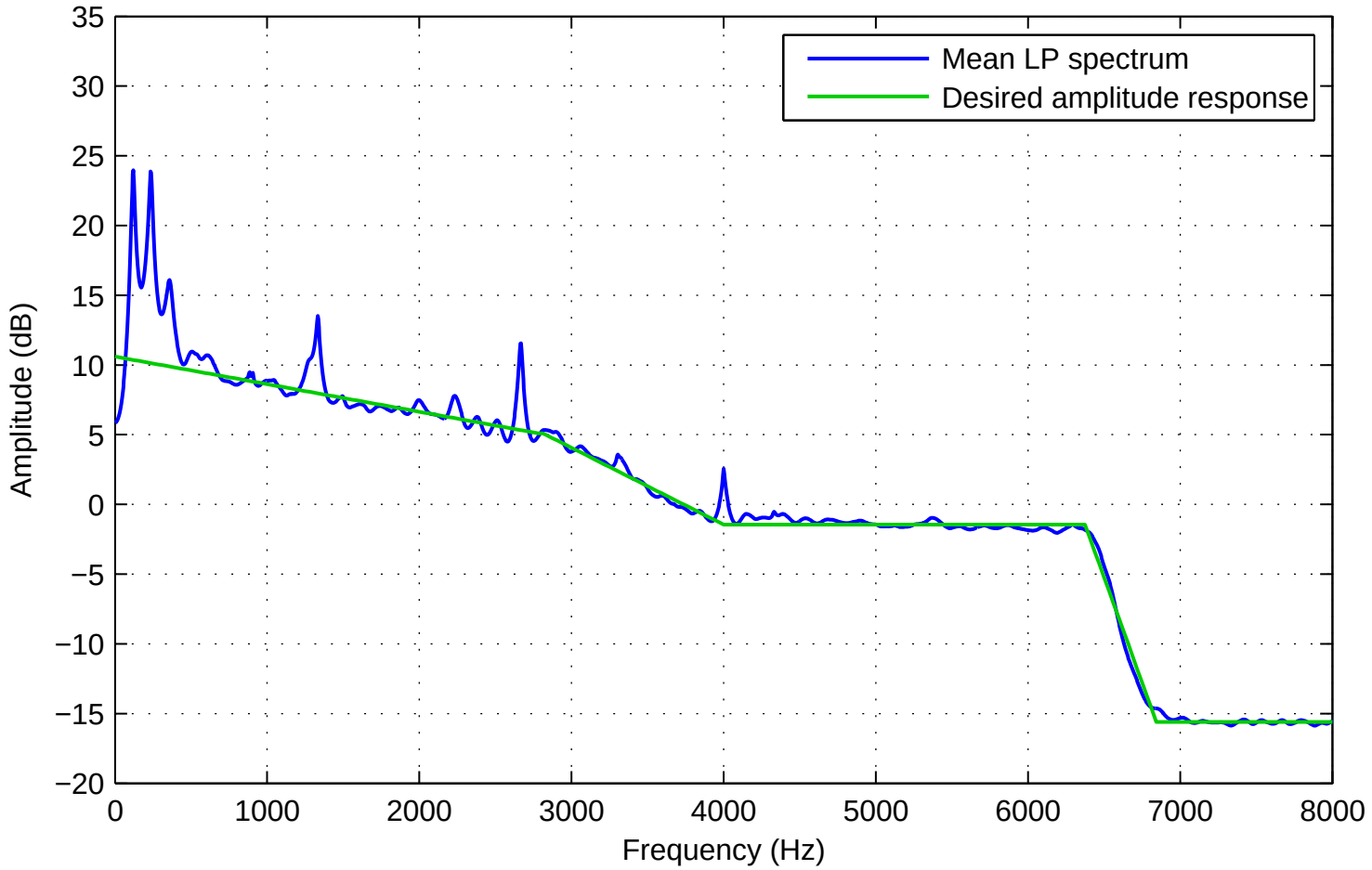
- Used 100 noise segments from arbitrary NTIMIT utterances
- Analysed segments to determine spectral characteristics of additive noise of the NTIMIT telephone channels
- Assumed noise segments to be output from LP filters
- Designed colouring filter based on the mean LP spectrum



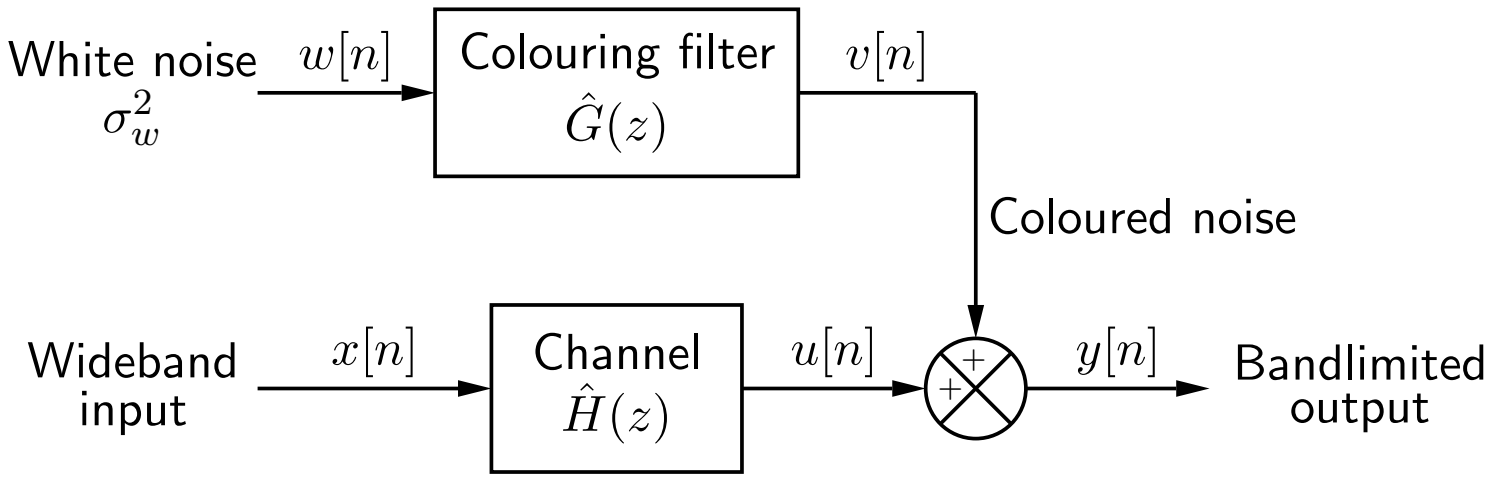
Noise analysis II



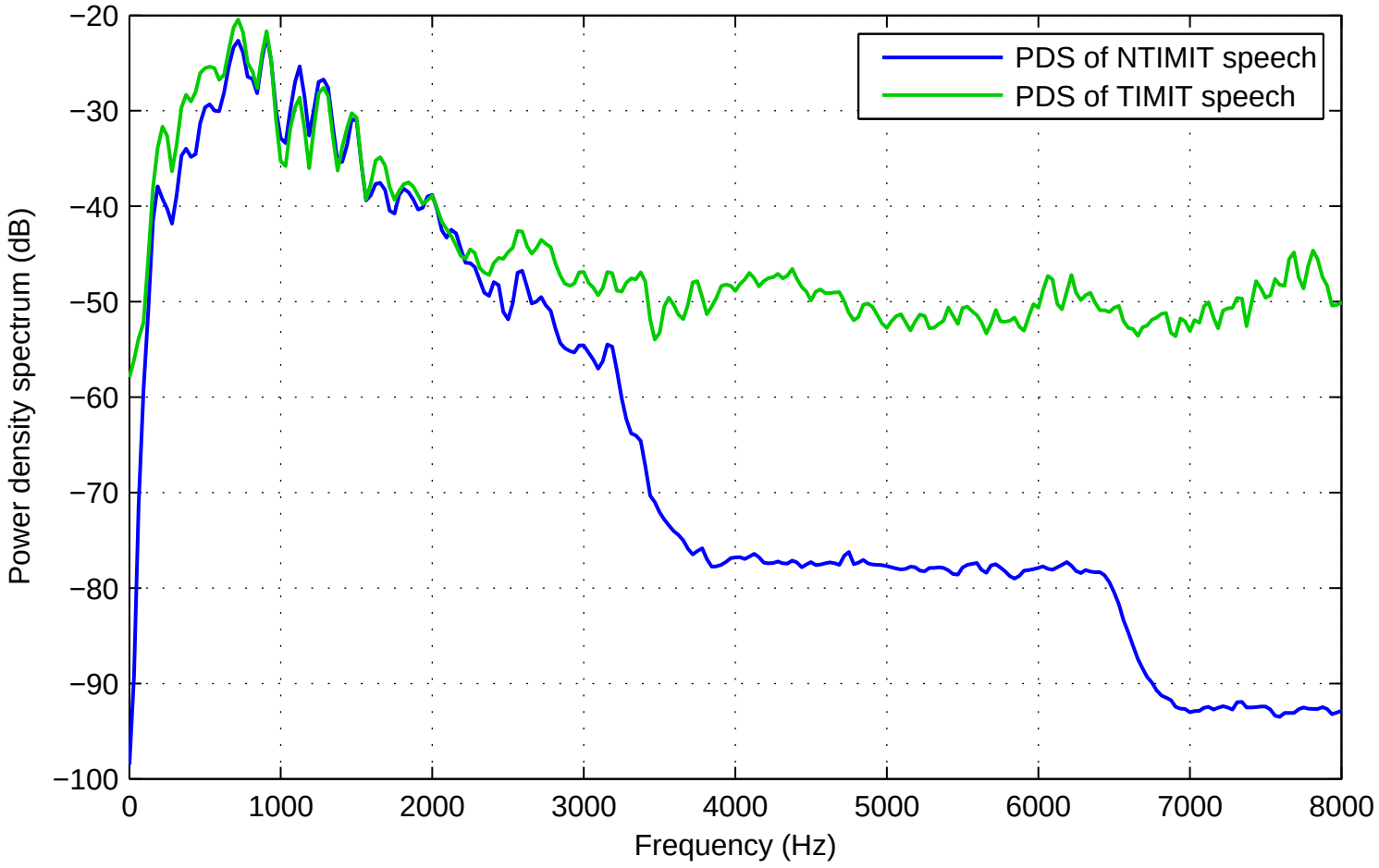
Design of noise model



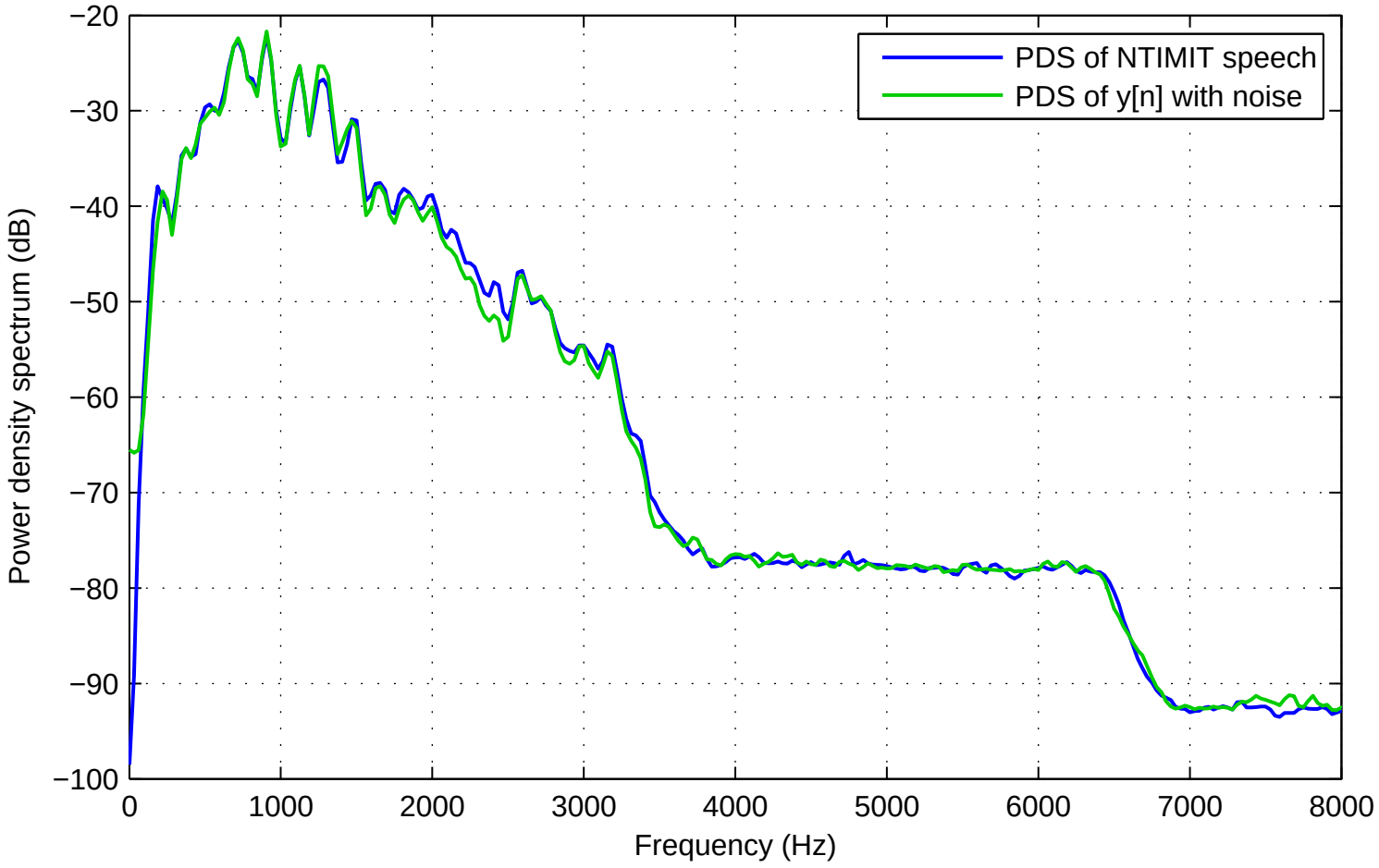
Implementation in software



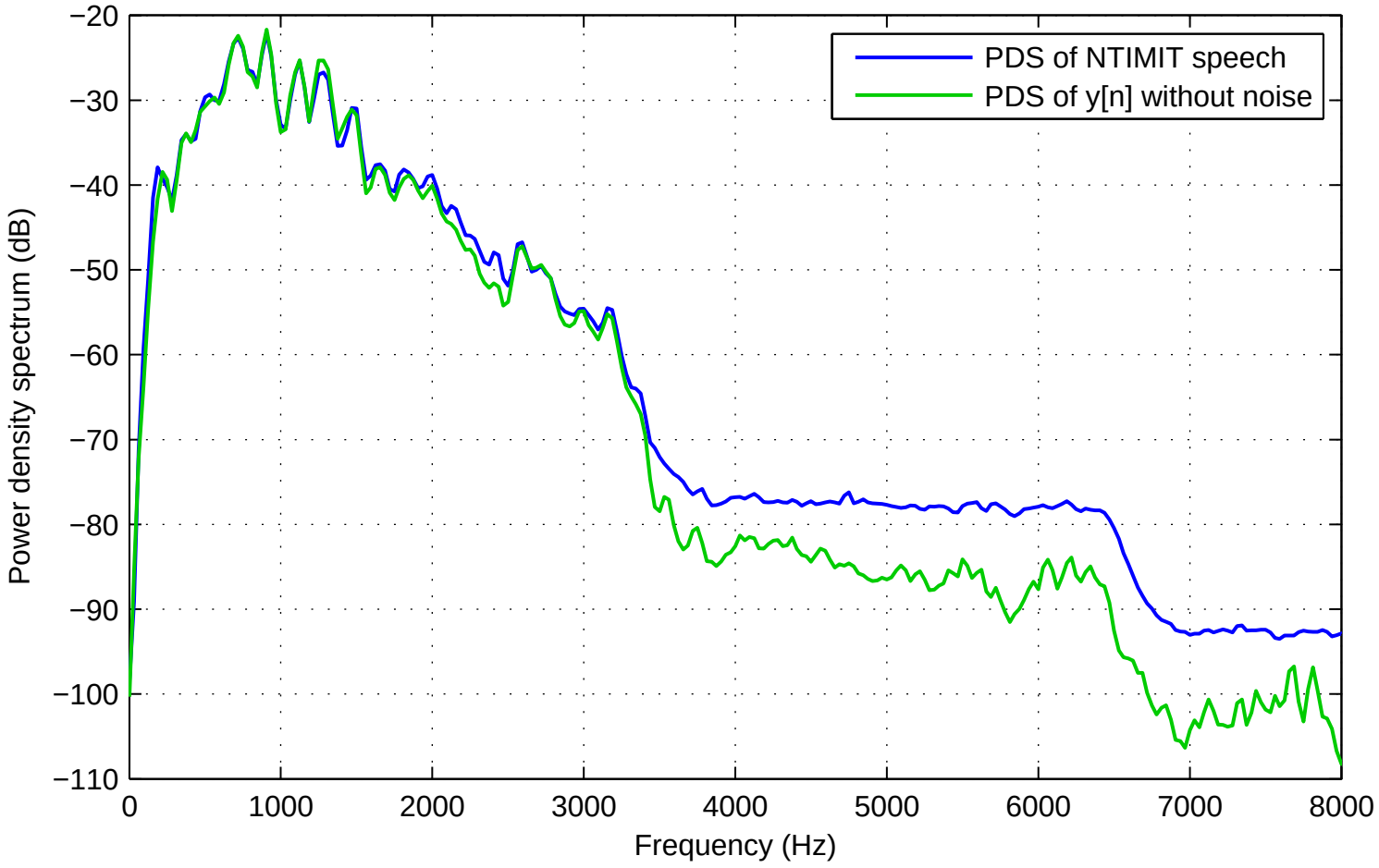
Evaluation: Single NTIMIT channel I



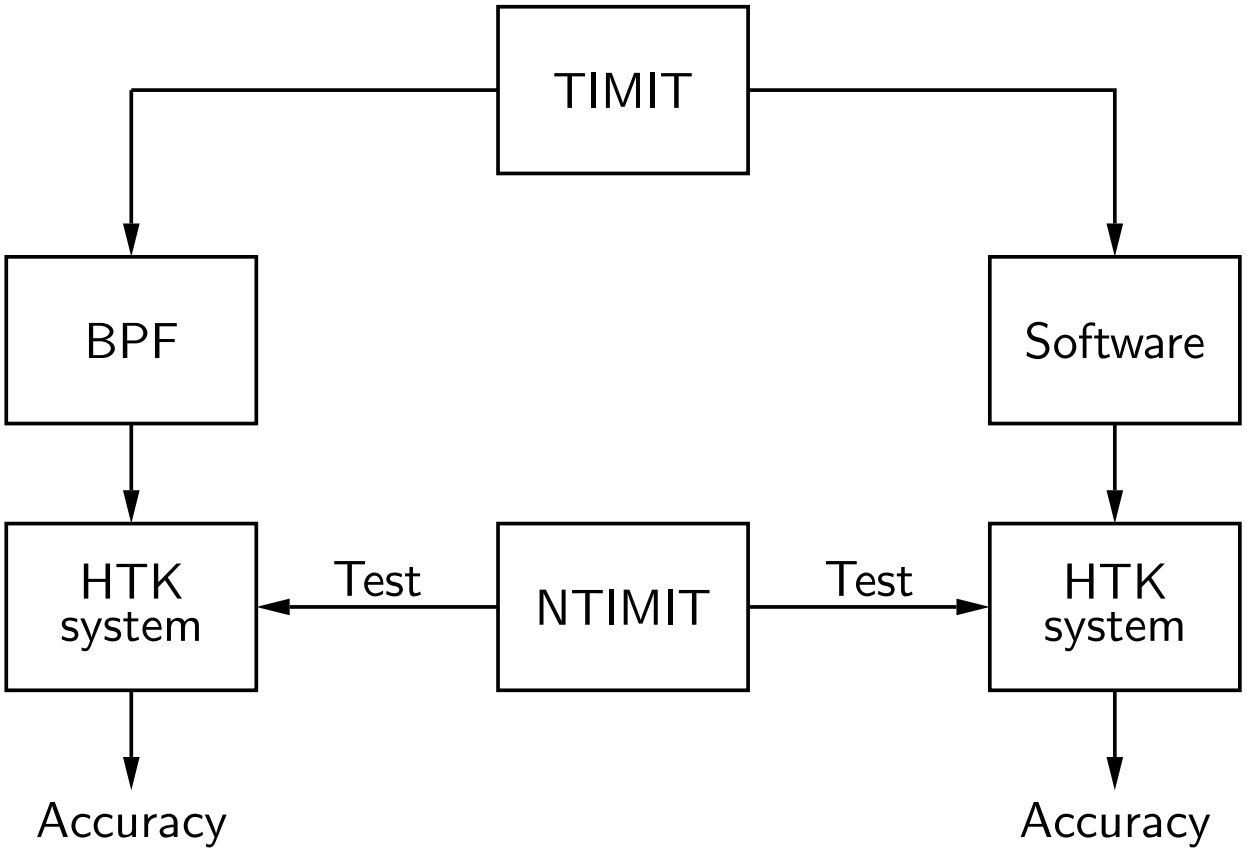
Evaluation: Single NTIMIT channel II



Evaluation: Single NTIMIT channel III



Evaluation: ASR systems I



Evaluation: ASR systems II

Training set	Test Set	% Accuracy
NTIMIT	NTIMIT	40.65%
TIMIT narrowband	NTIMIT	32.56%
Filtered TIMIT, 30 dB noise	NTIMIT	36.34%
Filtered TIMIT, no noise	NTIMIT	32.19%

Conclusion I

- Accuracy obtained using the third system 10.6% lower than accuracy using the NTIMIT training set
- 11.6% increase in accuracy from basic bandpass approach
- When no noise is added, performance is not much different from the TIMIT approach

Conclusion II

- Leads to the conclusion that the noise model is the most important aspect of the complete model
- Possible reasons for this:
  - Cepstral mean normalization
  - Stationarity of channel models
- Experiments to confirm and investigate the above are the subject of ongoing work