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

\[
\text{Wideband input } x[n] \rightarrow \text{Channel } \hat{H}(z) \rightarrow u[n] \rightarrow y[n] \rightarrow \text{Bandlimited output}
\]

\[
\text{White noise } w[n] \rightarrow \text{Coloured noise } v[n]
\]

\[
\sigma^2_w
\]

\[
\text{Coloured noise filter } \hat{G}(z)
\]

\[
\text{Bandlimited output } y[n]
\]
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

\[
\begin{align*}
\text{White noise} & \quad w[n] \quad \xrightarrow{\hat{G}(z)} \quad v[n] \\
\sigma_w^2 & \xrightarrow{\text{Colouring filter}} \xrightarrow{\text{Coloured noise}}
\end{align*}
\]
Noise analysis II

![Noise analysis graph](image)

- Frequency (Hz)
- Amplitude (dB)

Legend:
- Blue: Average
- Green: Median
- Pink: 90% interval

Average, Median, and 90% interval plots over frequency range from 0 Hz to 8000 Hz.
Design of noise model

![Graph showing frequency response and desired amplitude response]

- **X-axis**: Frequency (Hz)
- **Y-axis**: Amplitude (dB)

Legend:
- Mean LP spectrum
- Desired amplitude response
Implementation in software

\[
\begin{align*}
\text{Wideband input} & \quad x[n] \\
\text{White noise} & \quad \sigma_w^2 \quad w[n] \\
\text{Coloured noise} & \quad \hat{G}(z) \quad v[n] \\
\text{Channel} & \quad \hat{H}(z) \quad u[n] \\
\text{Bandlimited output} & \quad y[n]
\end{align*}
\]
Evaluation: Single NTIMIT channel I

Power density spectrum (dB) vs. Frequency (Hz)

- Blue line: PDS of NTIMIT speech
- Green line: PDS of TIMIT speech
Evaluation: Single NTIMIT channel II

- Power density spectrum (dB)
- Frequency (Hz)

- PDS of NTIMIT speech
- PDS of $y[n]$ with noise
Evaluation: Single NTIMIT channel III

![Power density spectrum (dB) vs Frequency (Hz)](image)

- **PDS of NTIMIT speech**
- **PDS of y[n] without noise**
Evaluation: ASR systems I

TIMIT

BPF

HTK system

NTIMIT

Software

HTK system

Accuracy

Test

Test

Accuracy
## Evaluation: ASR systems II

<table>
<thead>
<tr>
<th>Training set</th>
<th>Test Set</th>
<th>% Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTIMIT</td>
<td>NTIMIT</td>
<td>40.65%</td>
</tr>
<tr>
<td>TIMIT narrowband</td>
<td>NTIMIT</td>
<td>32.56%</td>
</tr>
<tr>
<td>Filtered TIMIT, 30 dB noise</td>
<td>NTIMIT</td>
<td>36.34%</td>
</tr>
<tr>
<td>Filtered TIMIT, no noise</td>
<td>NTIMIT</td>
<td>32.19%</td>
</tr>
</tbody>
</table>
Conclusion 1

- 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
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