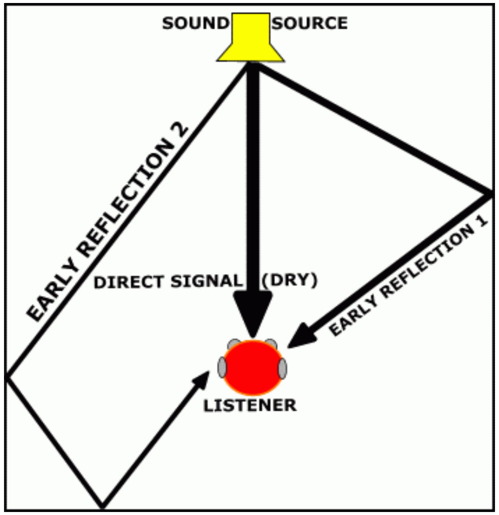
Blind Dereverberation of Audio Signals

Graham Grindlay

E4810 Final Project

Introduction

- **Reverb** is the process of multipath propagation from a sound source to the microphone or listener
- Reverb can be broken into 3 parts:
  - Direct sound (<5ms)
  - Early reflections (5-70ms)
  - Late reflections (>70ms)
- Delay times are frequency-dependent



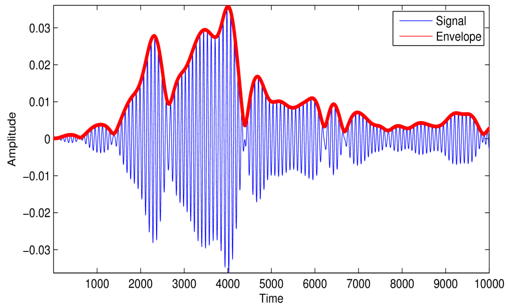
- A room's reverb characteristics can be described with an IR $h[n]$
  → can make a clean sound reverberant: $y_r[n] = h[n] * y_c[n]$

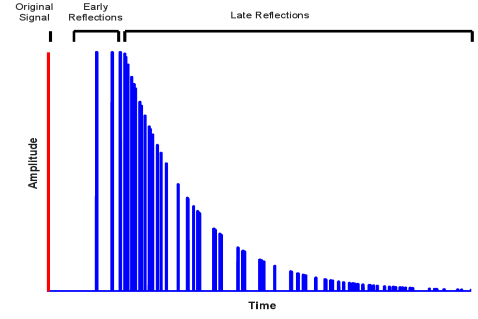
Problem Description

- Given the reverberant signal $y_r[n]$ want to estimate clean signal $\hat{y}_c[n]$
- Applications where we might want to reduce/eliminate reverb:
  - Speech recognition
  - Automatic music transcription
  - Hearing aids
- If we knew $h[n]$, it would be easy to determine $y_c[n]$ (could just use deconvolution)
- But here there is **no** prior knowledge assumed $\rightarrow$ hard problem!

A Simple Blind Dereverberation Approach

An Envelope Follower





- Focus only on late reflections
- Use signal envelopes:

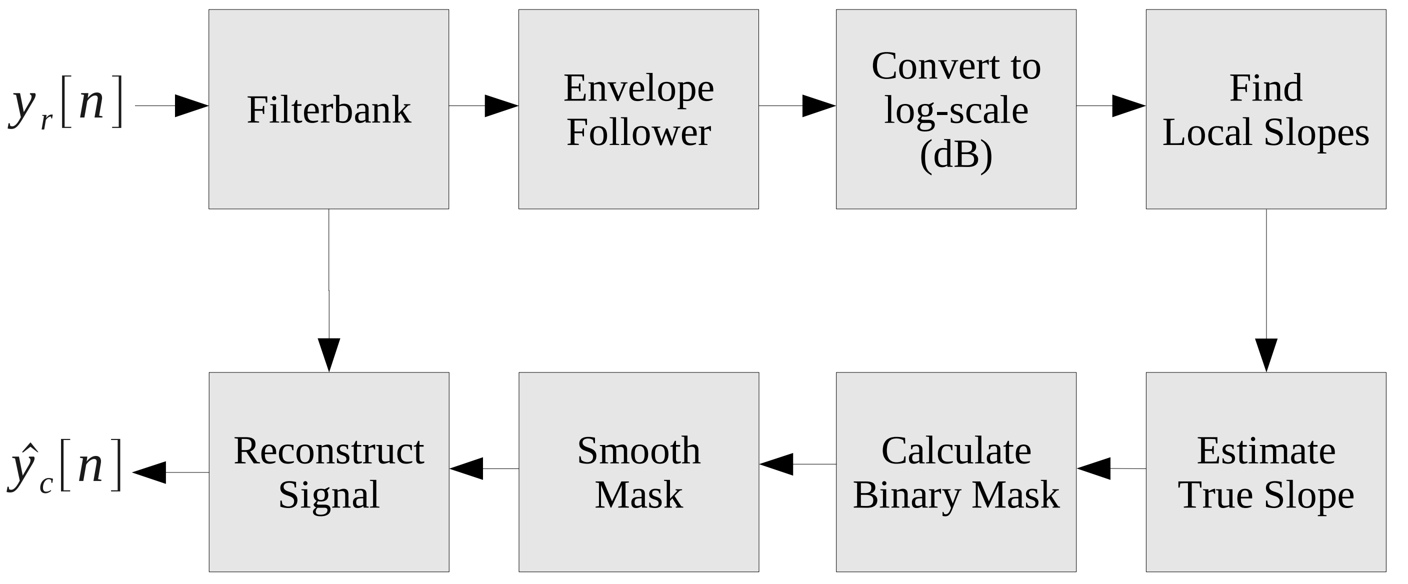
$$e_r[n] = e_h[n] * e_c[n]$$

- Envelope of late reflections has exponential decay form:

$$\exp(-1/\tau)^n = a^n$$

- Basic idea:
  - Estimate *a* from the the envelope of the reverberant signal
  - Use estimate to find and remove regions of $y_r[n]$ that are mostly reverberant decay

Algorithm Outline



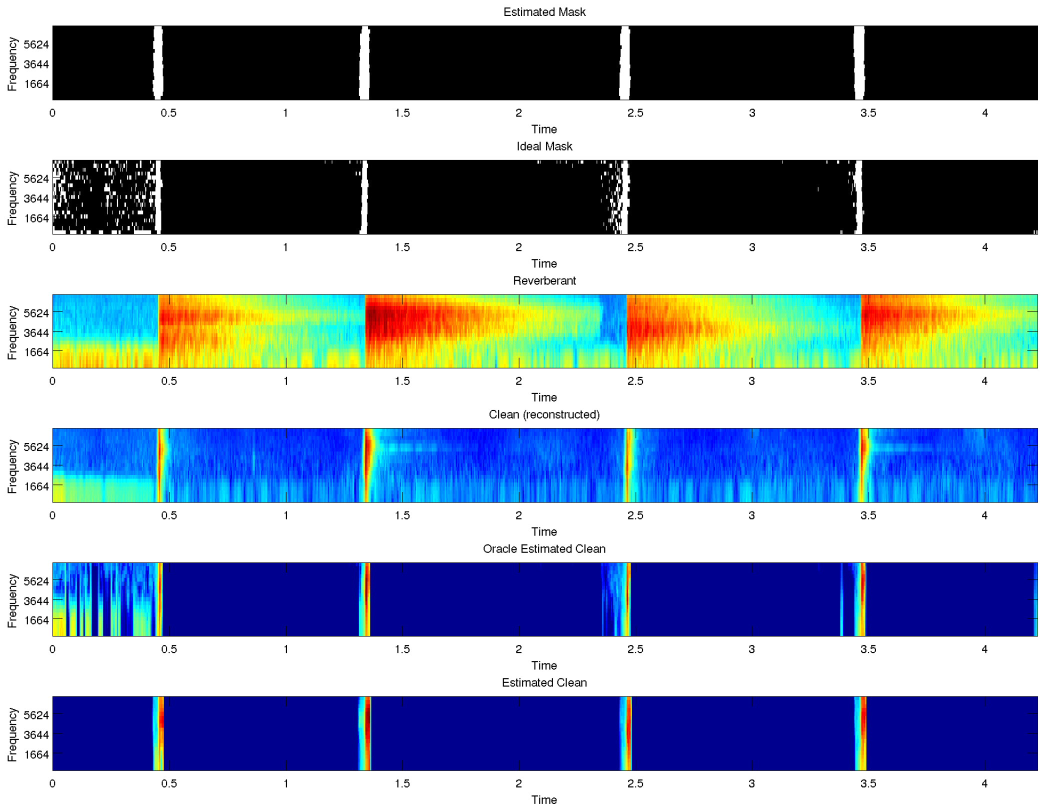
Assessing Performance

- Spectrograms
- Human ears
- Masking accuracy
  - Since we know what the clean signal is, we can create an “ideal” binary mask
  - The ideal binary mask is defined as:

    1 if $\frac{e_r[n]}{e_c[n]} > 2$

    0 otherwise
  - Mask accuracy is percentage of cells that match ideal mask

Experiment 1 (Clapping)



Estimated Mask

Ideal Mask

Reverberant Sound

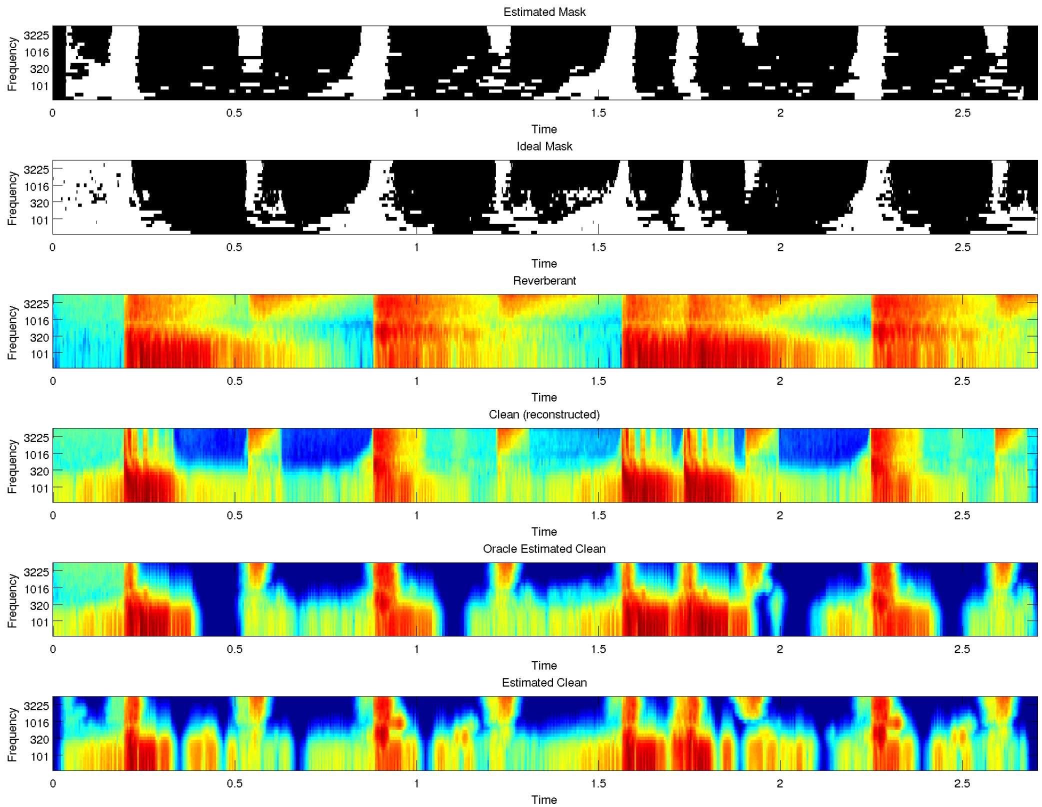
Clean Sound

Ideal Mask Reconstruction

Algorithm Reconstruction

Mask accuracy = 97%

Experiment 2 (Drums)



Estimated Mask

Ideal Mask

Reverberant Sound

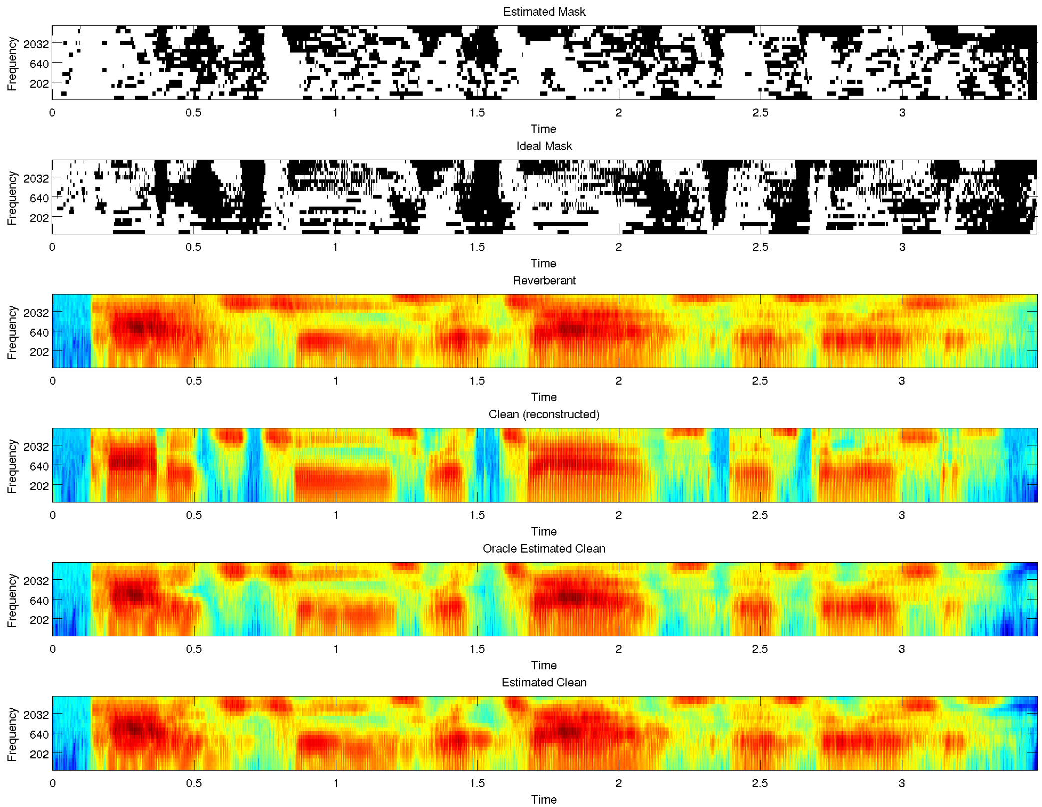
Clean Sound

Ideal Mask Reconstruction

Algorithm Reconstruction

Mask accuracy = 82%

Experiment 3 (Speech)



Estimated Mask

Ideal Mask

Reverberant Sound

Clean Sound

Ideal Mask Reconstruction

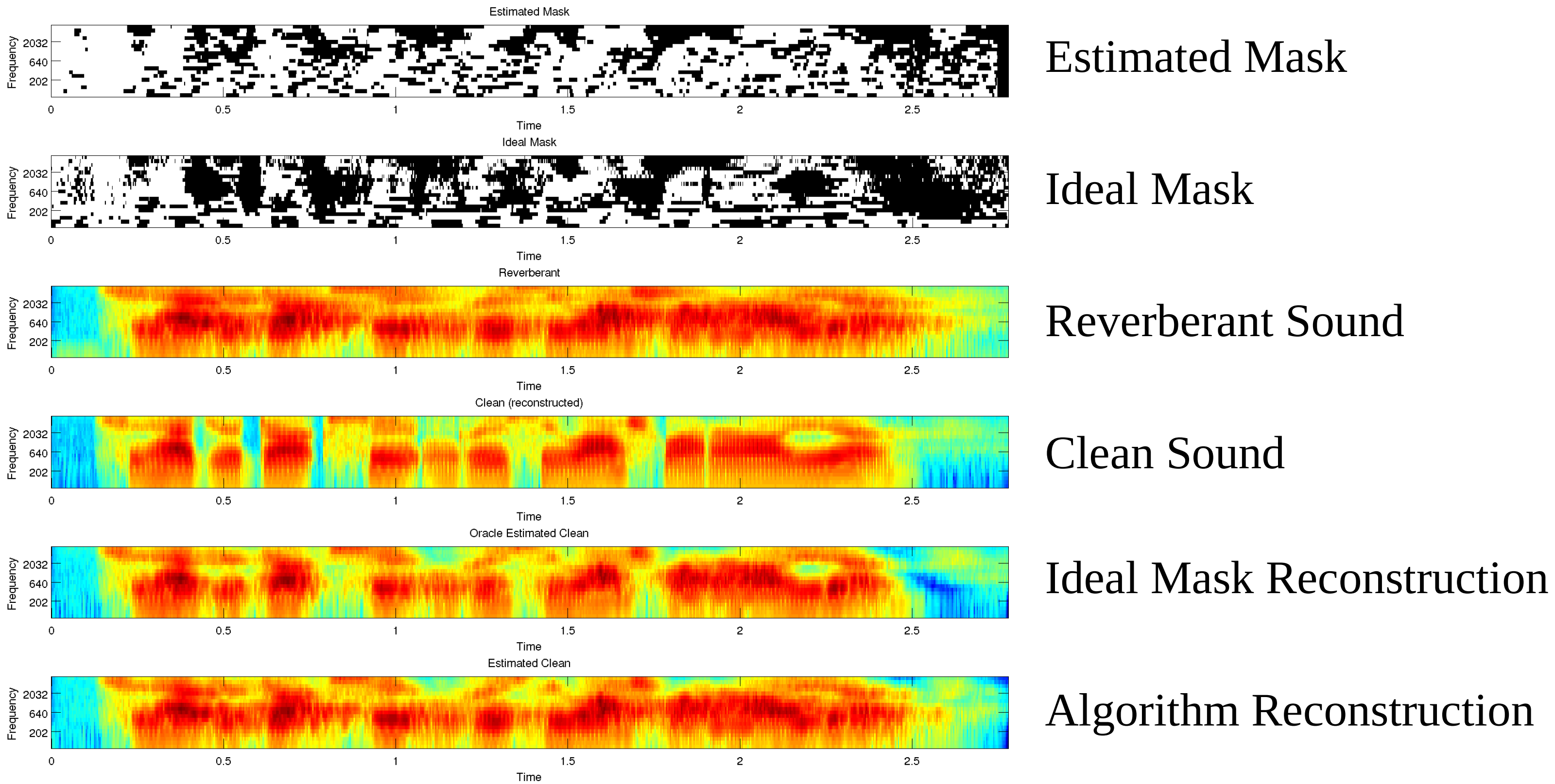
Algorithm Reconstruction

Mask accuracy = 70%

Conclusions

- Method seems to work reasonably well, particularly when audio is sparse
- Better ways to calculate the binary mask?
  - Trained classifiers
  - Source models
- Several parameters to tweak; how to set automatically?

Experiment 4 (Speech [f2])



Mask accuracy = 69%