The effect of reverberation on speech

Martin Cooke[1], Madhu Shashanka[2]
& Barbara Shinn-Cunningham[3]

(1) Speech and Hearing Research
Department of Computer Science
University of Sheffield

(2) Auditory Neuroscience Lab
Cognitive & Neural Systems
Boston University

(3) Depts. of Cognitive & Neural Systems
and Biomedical Engineering
Boston University

Montreal: November 2004

Outline

- How potential speech cues are affected by reverberation (focus on monaural)

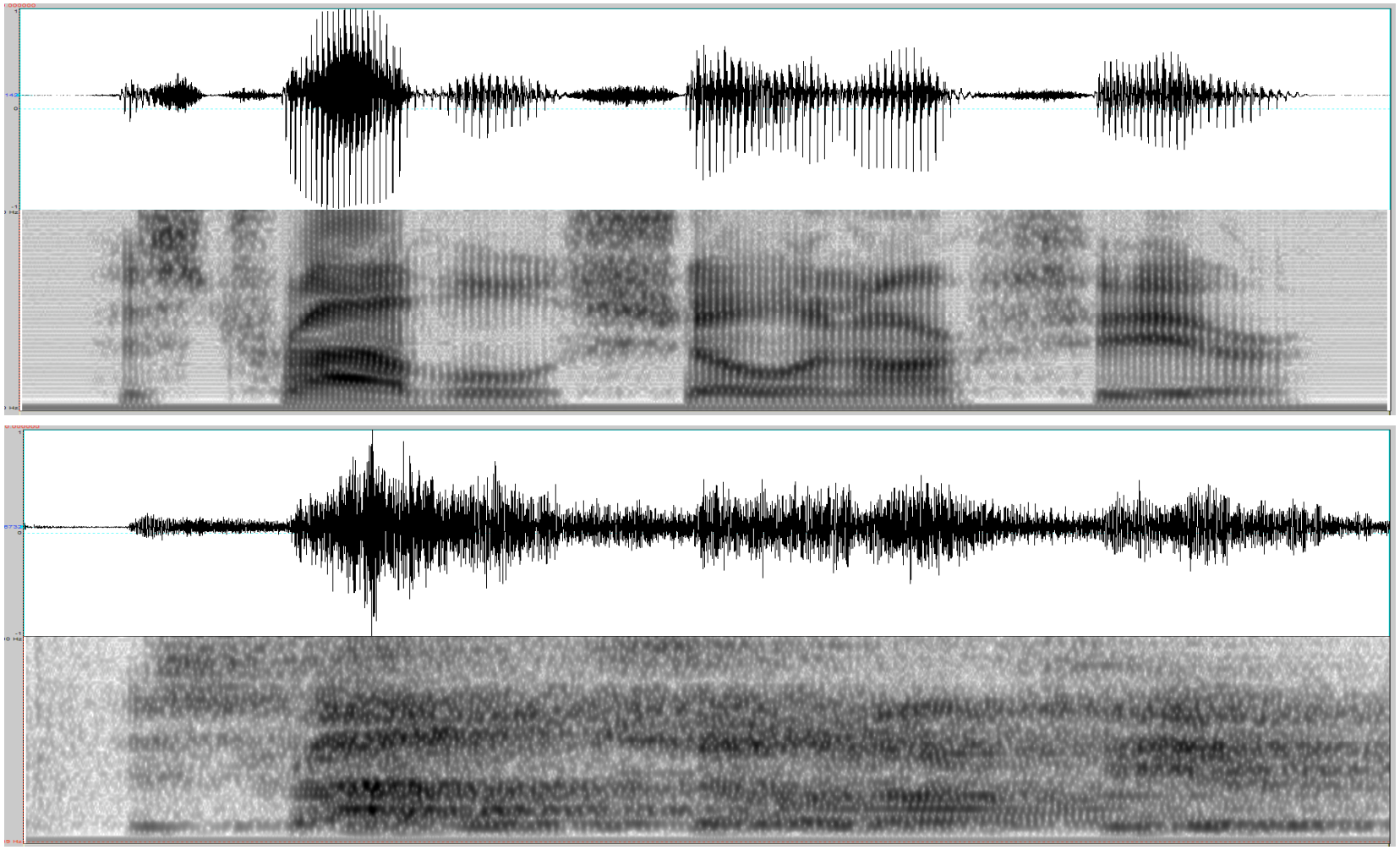
  See: Assmann, P. & Summerfield, Q (2004) The perception of speech under adverse conditions. In: *Speech Processing in the Auditory System* (eds S. Greenberg & W. Ainsworth), Springer Handbook of Auditory Research

- Corruption of (modelled) spectro-temporal excitation patterns
  - Single speech source
  - Two speech sources

Caveats

- Small corpus
- exploratory/descriptive

Clean vs highly reverberant speech (bathroom)



Reverberation primer

Reverberated speech = direct speech energy + reflected energy

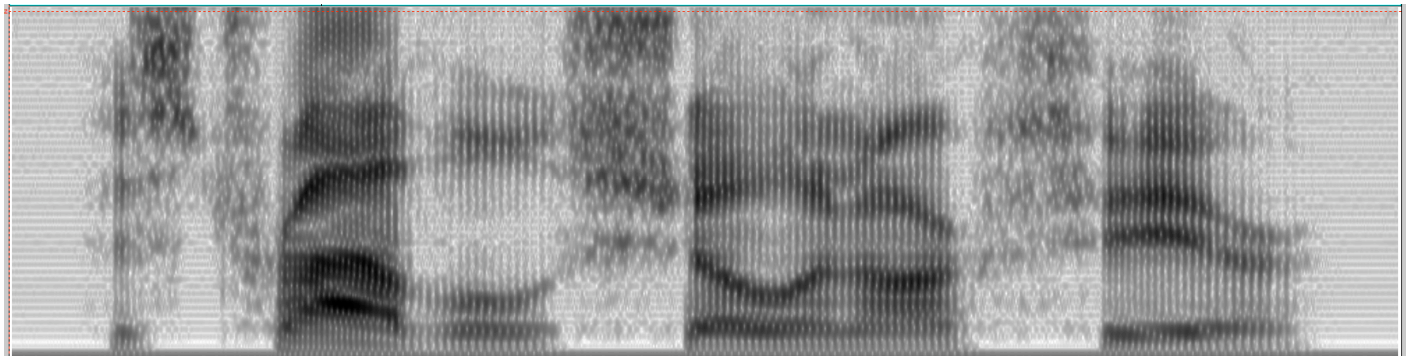
 *Clean, reversed*

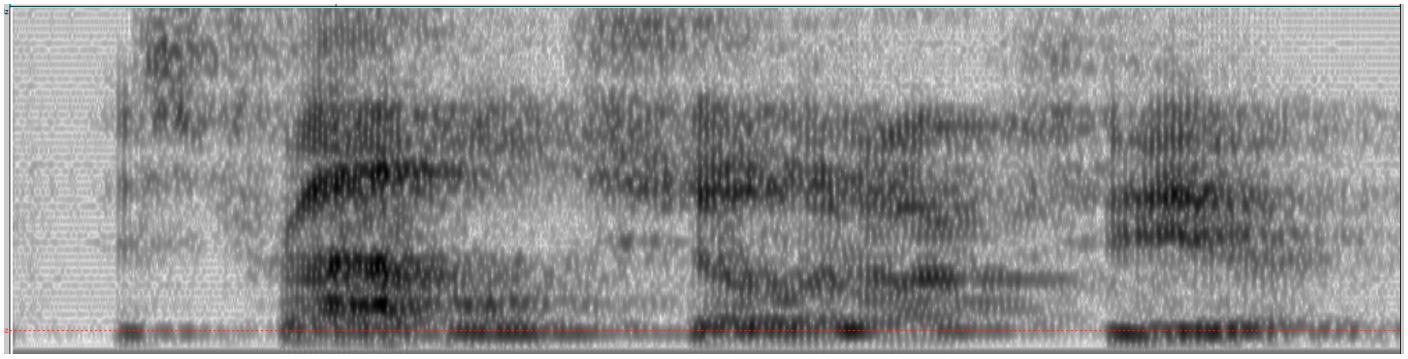
 *Reverberated, reversed*

Factors influencing reverberation

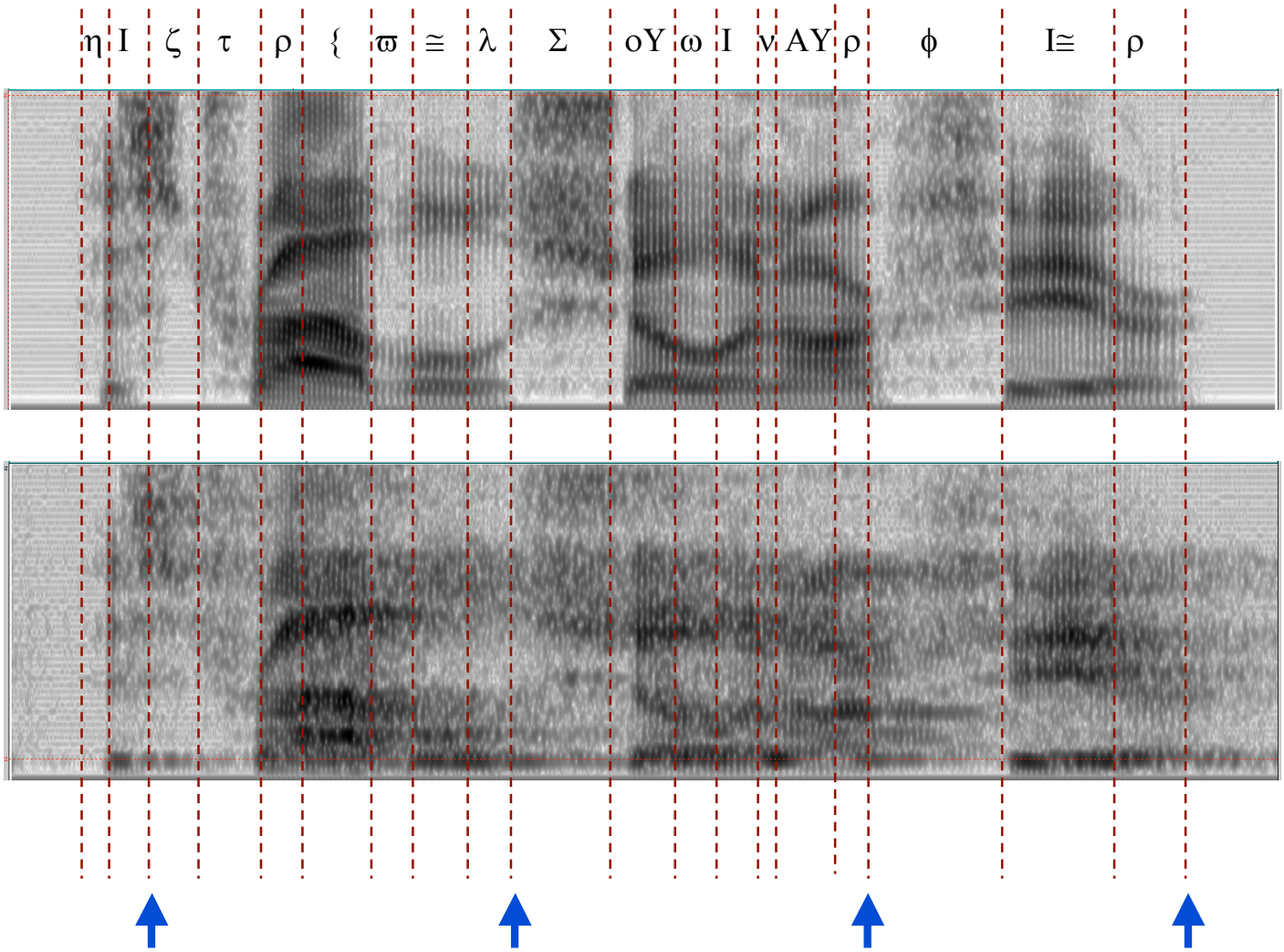
- Volume of the space in which the source and receiver are located
- Material of the reflective surfaces
- Location of source and receiver relative to reflective surfaces
- Distance of receiver from source

Moderate reverberation (ping pong room)

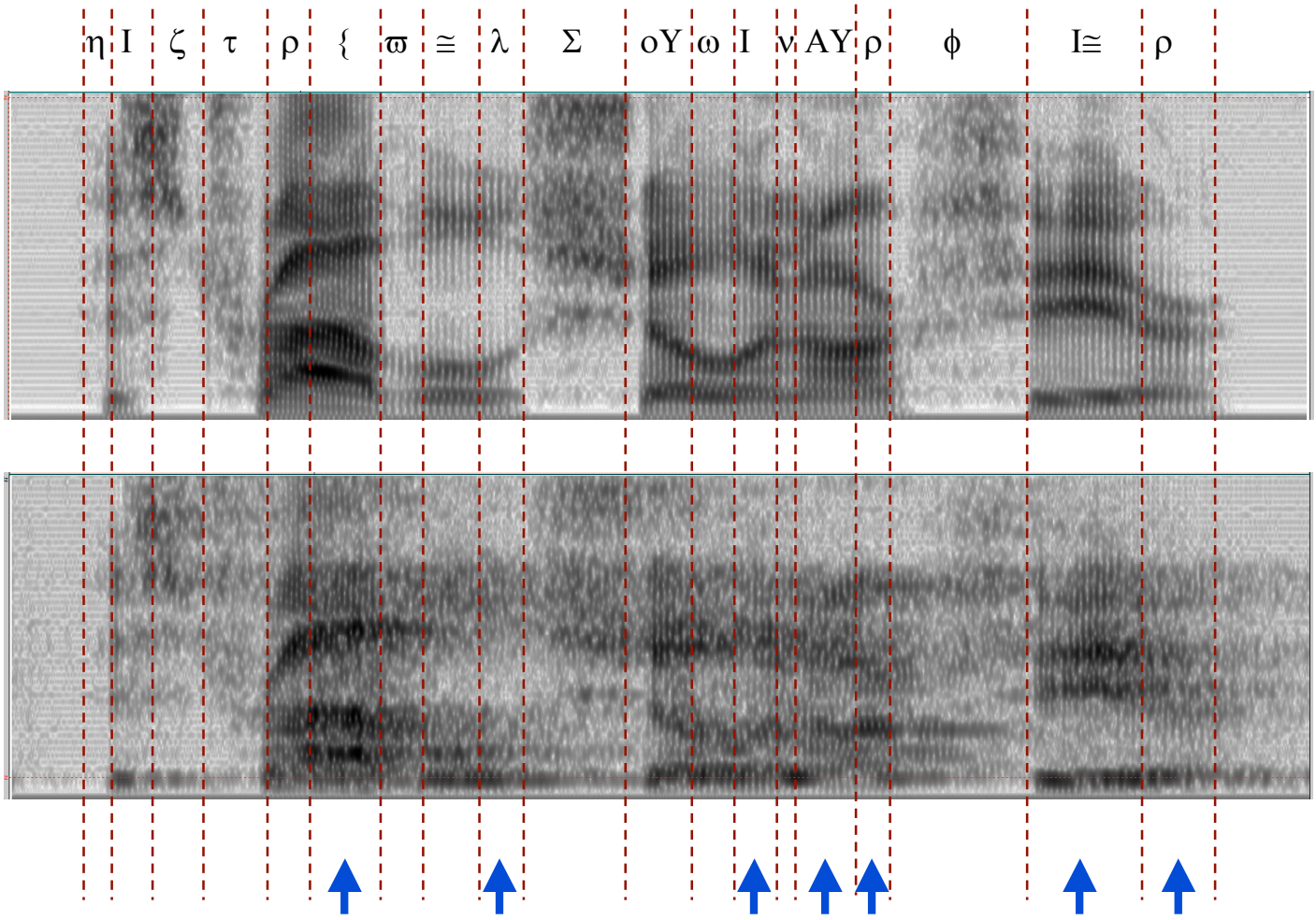




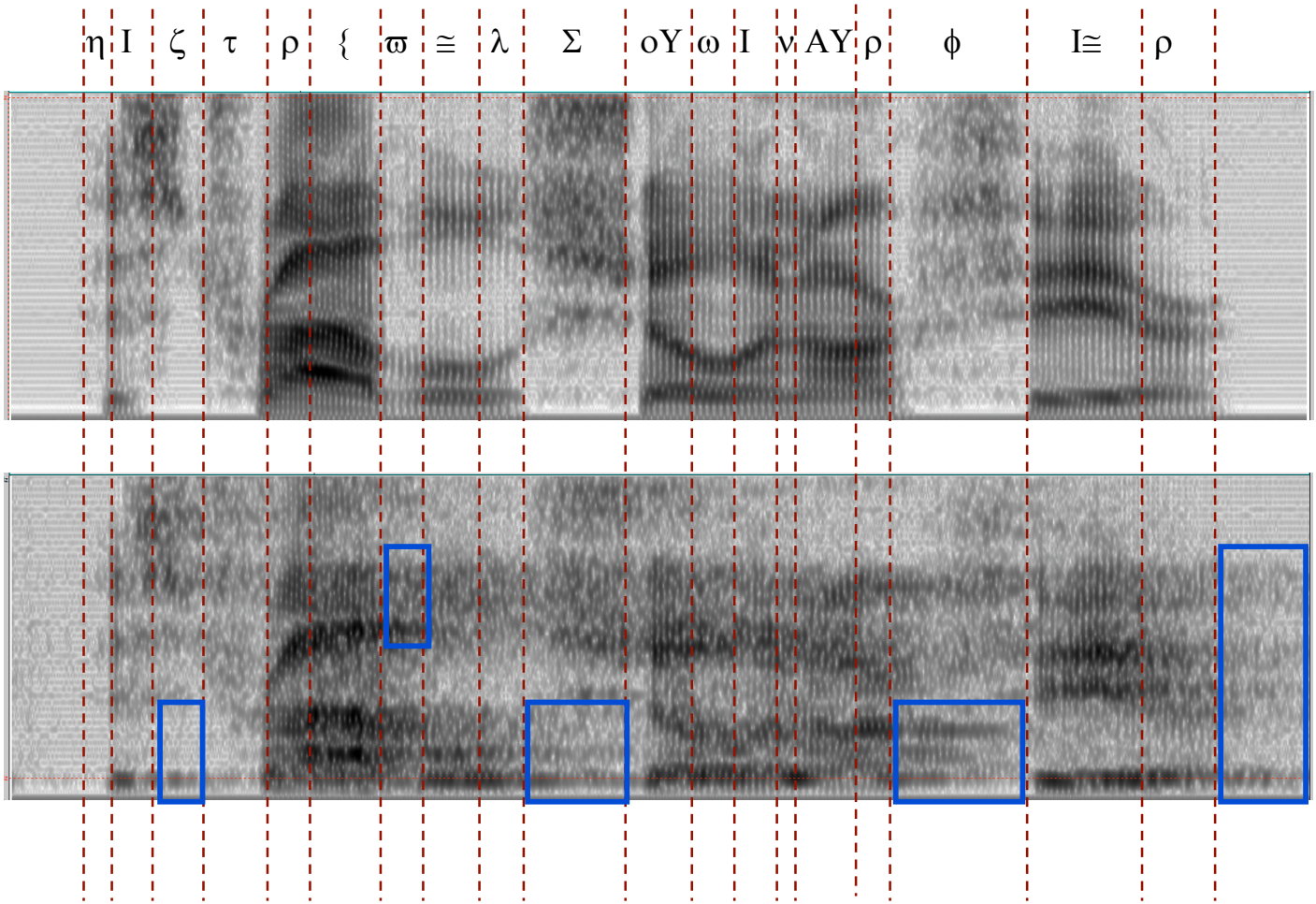
Onsets, and especially offsets, are blurred



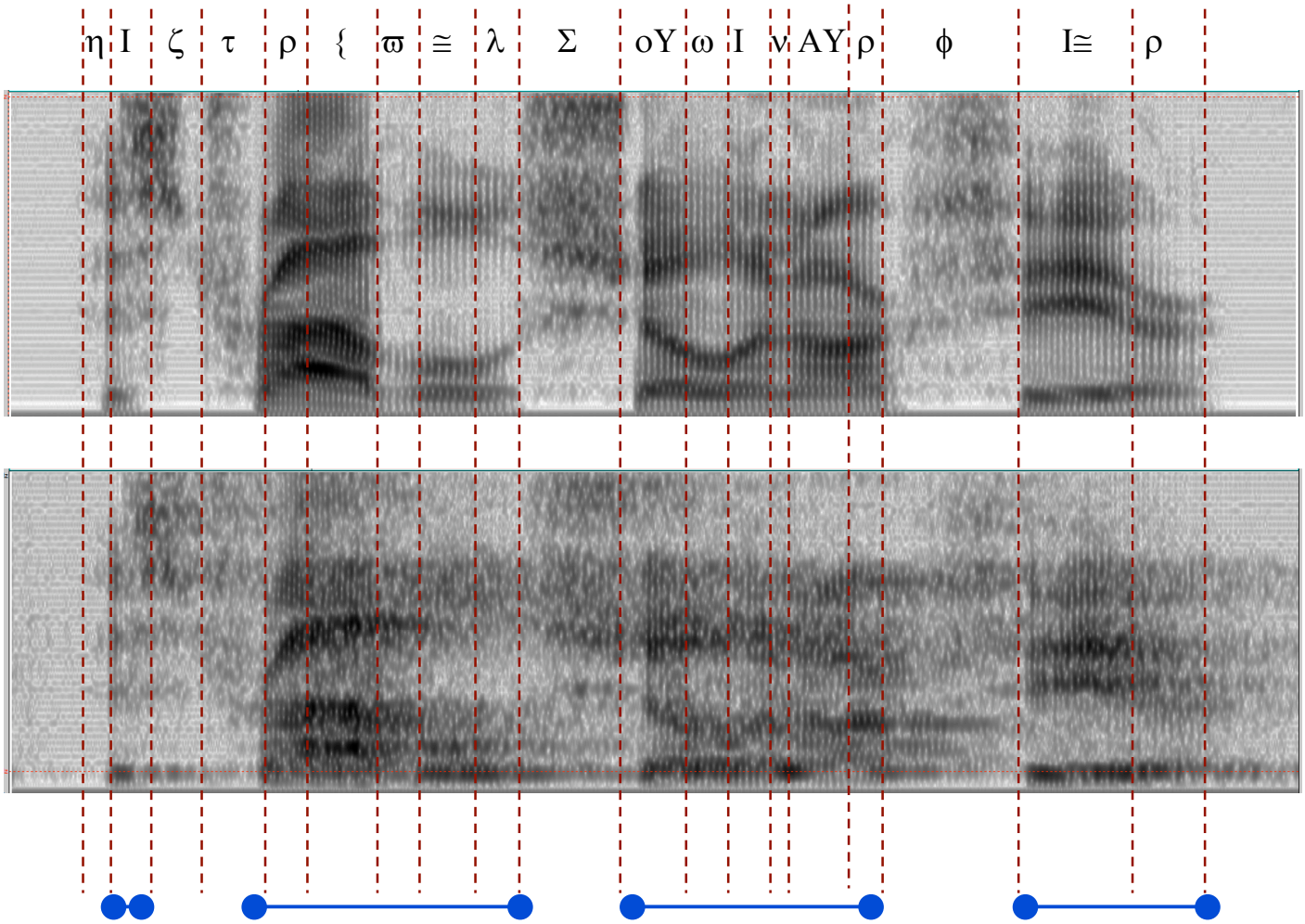
Formant movements are lost (except at onsets), becoming flattened/blurred



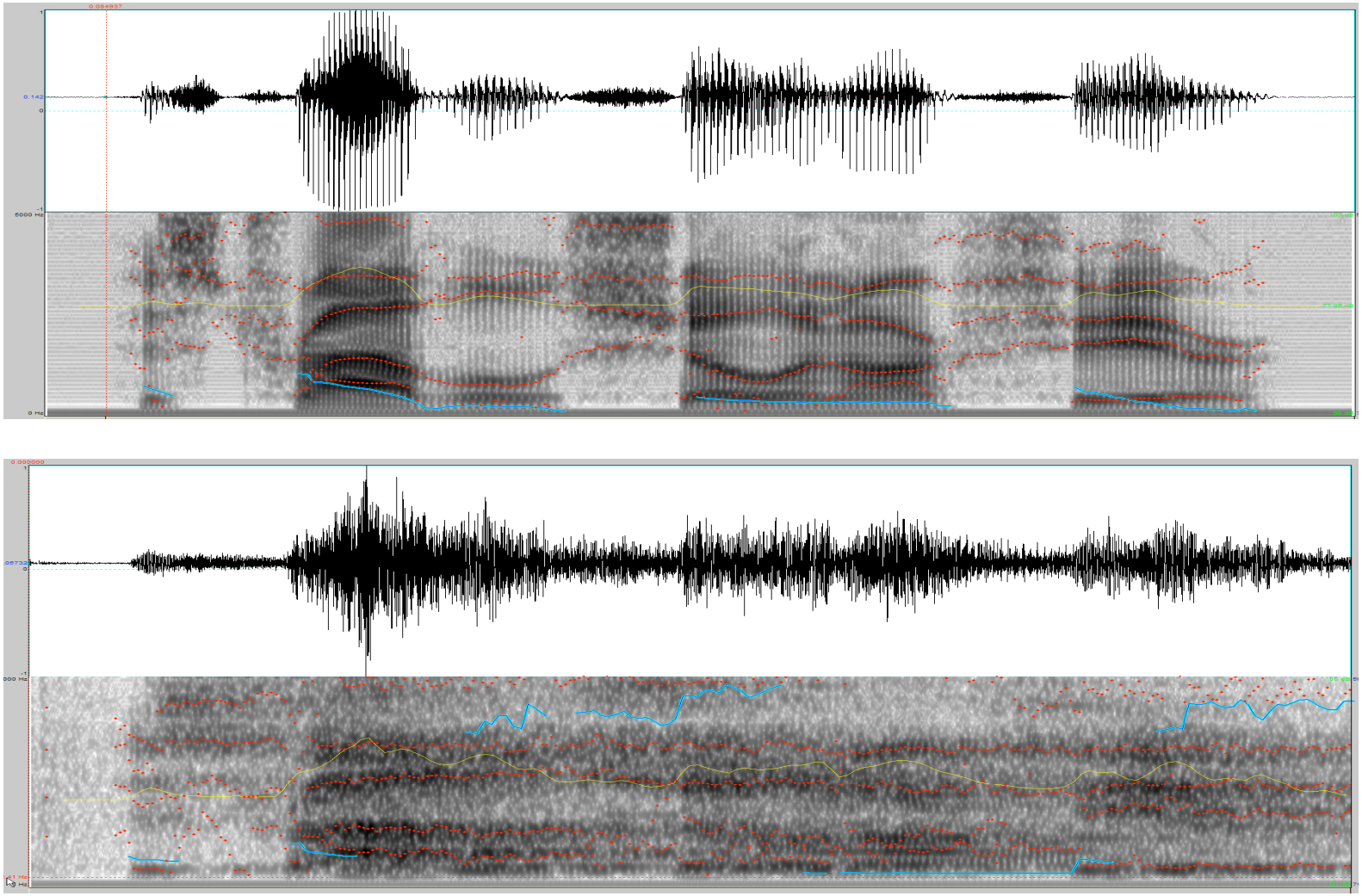
Spectro-temporal gaps are filled



F0-related amplitude modulation is less sharp



Reduction in pitch/formant tracking performance (PRAAT)



Montreal: November 2004

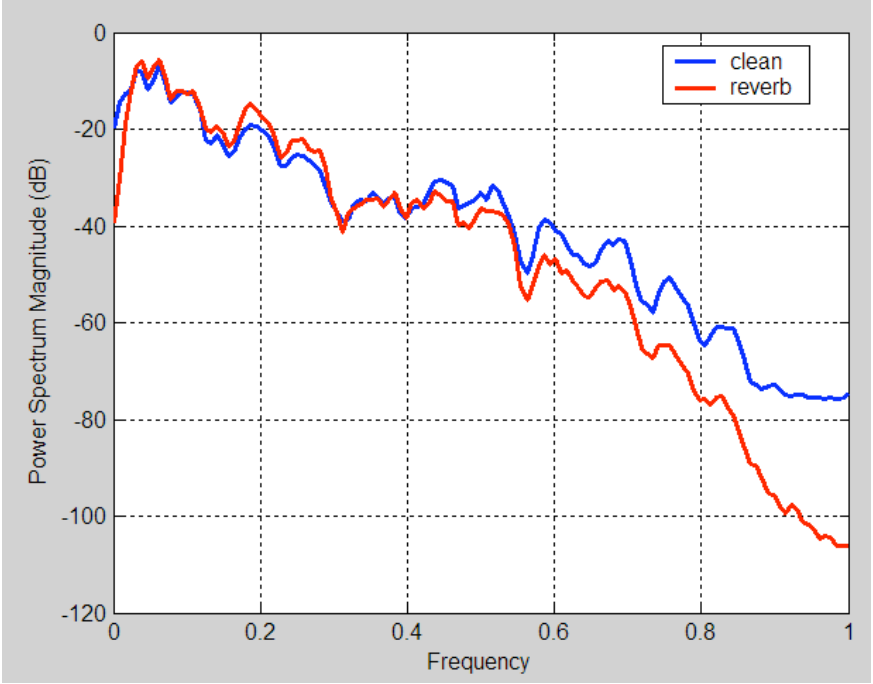
Monaural consequences

The good news …

- Steady-state sounds emerge relatively unscathed

… and the bad

- Traditional **phonetic cues** such as bursts are masked/blurred
- **Relational cues** to phoneme identity eg VOT are less precise
- The salience of **dynamic features** is reduced
- **Rate of change cues** are imprecise, leading to problems in distinguishing stops from liquids from diphthongs
- Blurring of boundaries reduces effectiveness of **durational cues**
- **Change in spectral tilt** since HF energy more likely to be attenuated

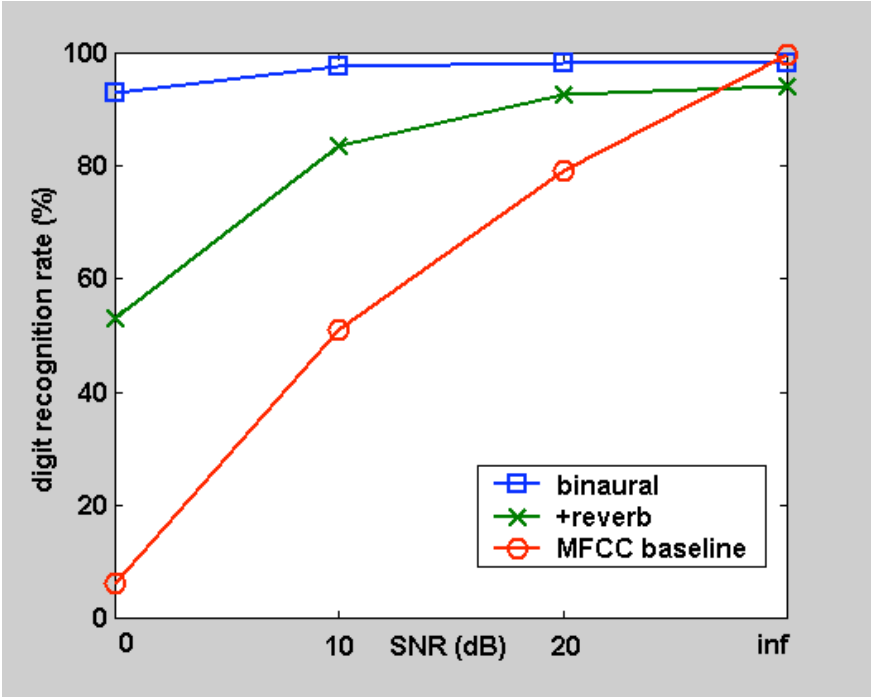


Binaural consequences

- Echoes from non-direct path tend to randomise patterns of interaural phase and level differences …
- … significantly reducing (and in some cases wiping out) any binaural advantage

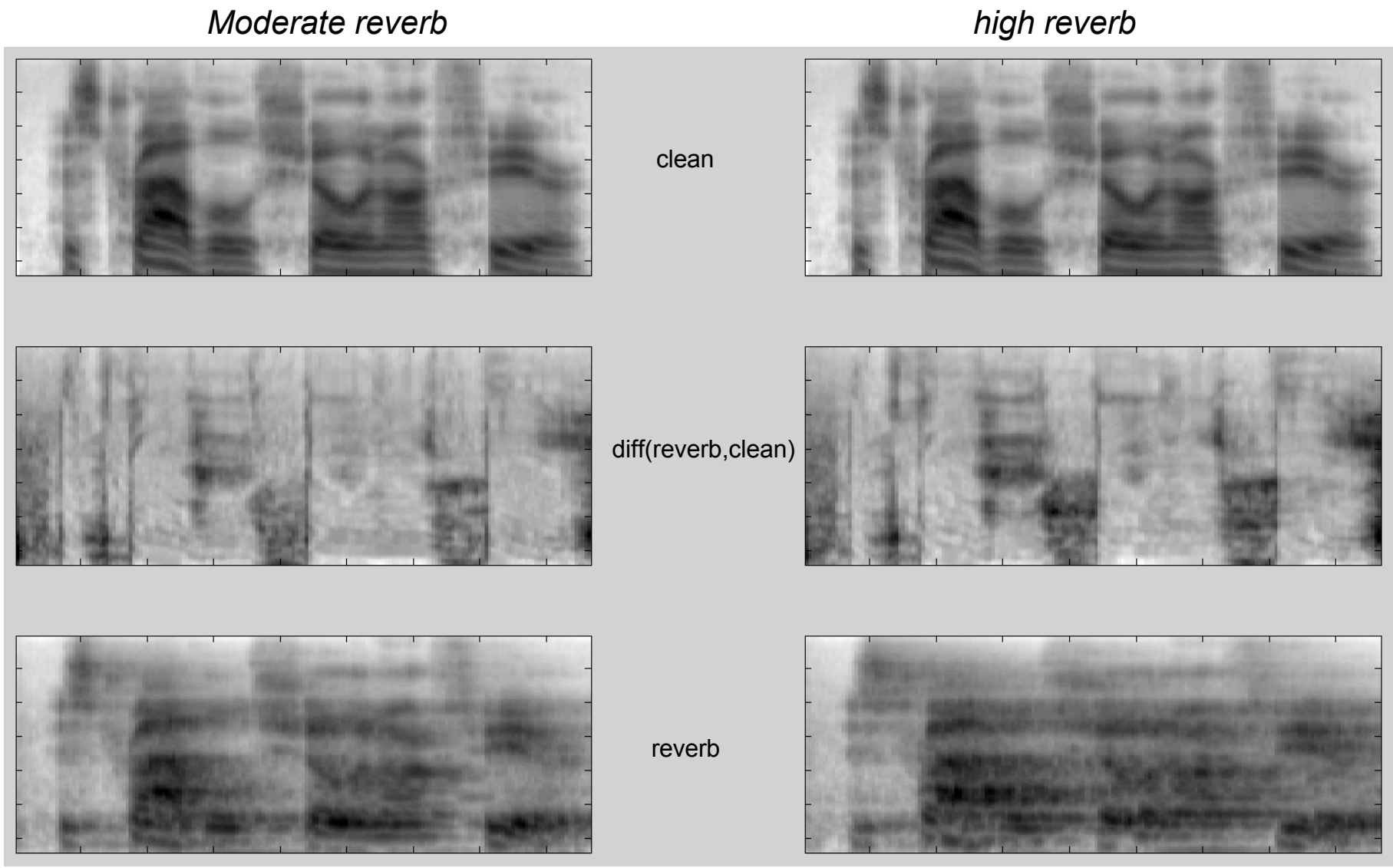
Illustration: effect on the performance of a missing-data based robust ASR system

- Digit sequence identification in the presence of an interfering talker (separation = 40 deg) in anechoic and low-moderate reverb (t_{60} = 300 ms)

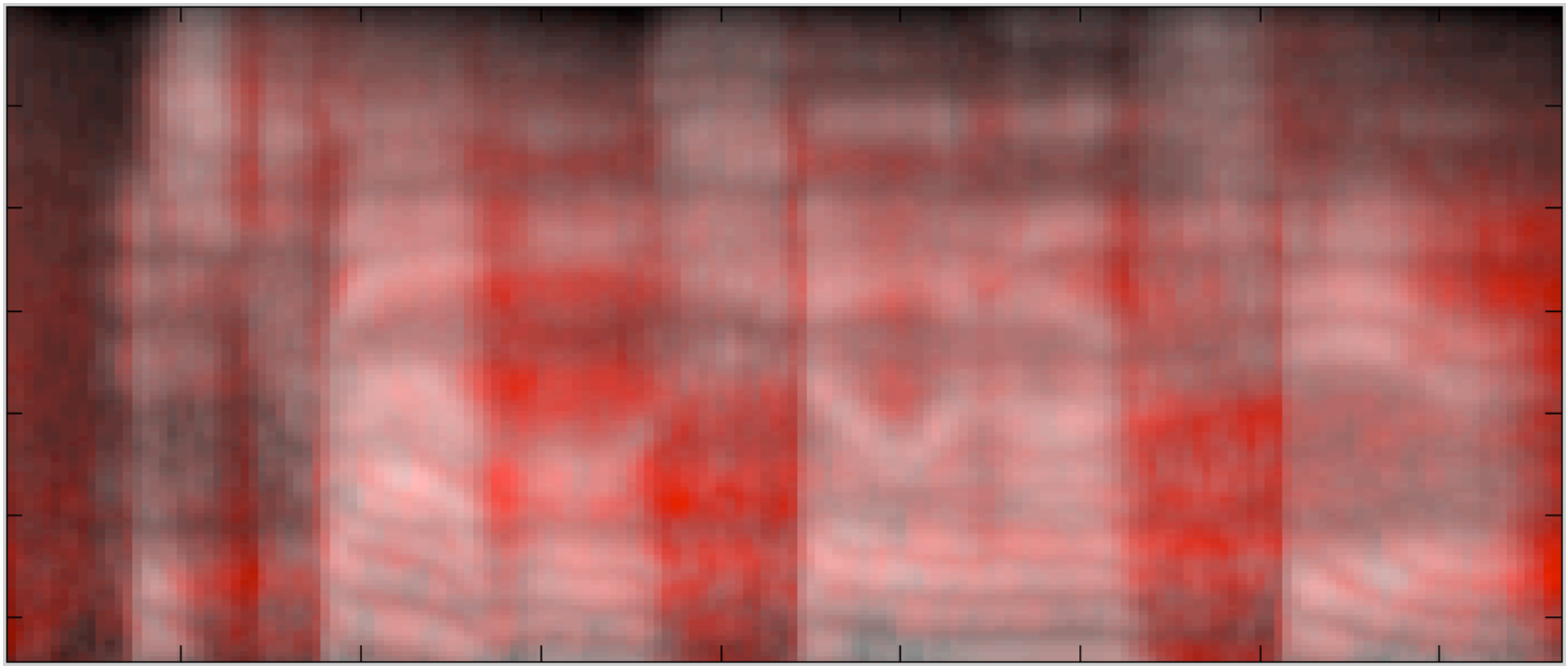


Source: Palomaki, Brown & Wang (2004) Speech Communication

Where is the reverberant energy?
I. Spectro-temporal excitation pattern model



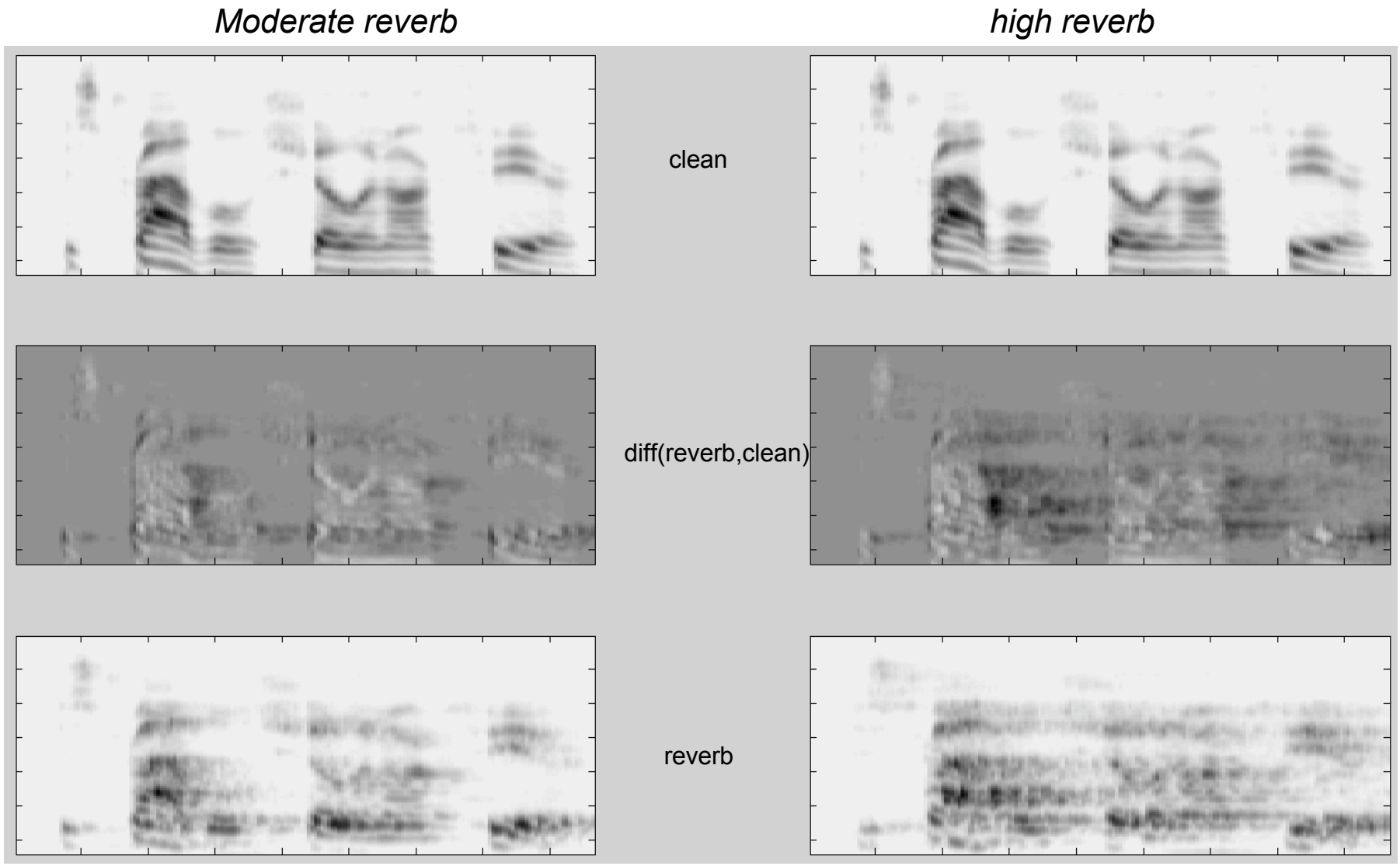
Visualising reverberant energy



HSV representation

- ‘Hue’ = red
- ‘Saturation’ proportional to difference in log energy between reverb and clean
- ‘Value’ is log energy of reverberant signal

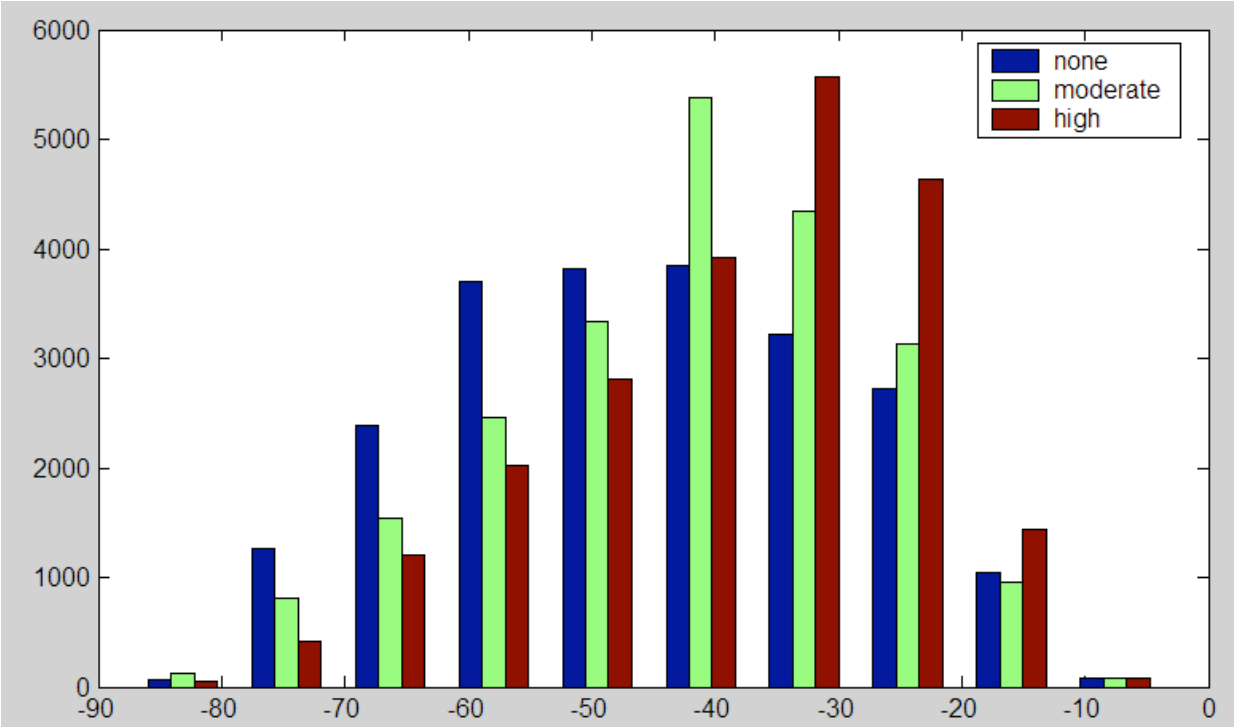
Where is the reverberant energy?
II. STEP + forward masking model



Distribution of energy

Two effects

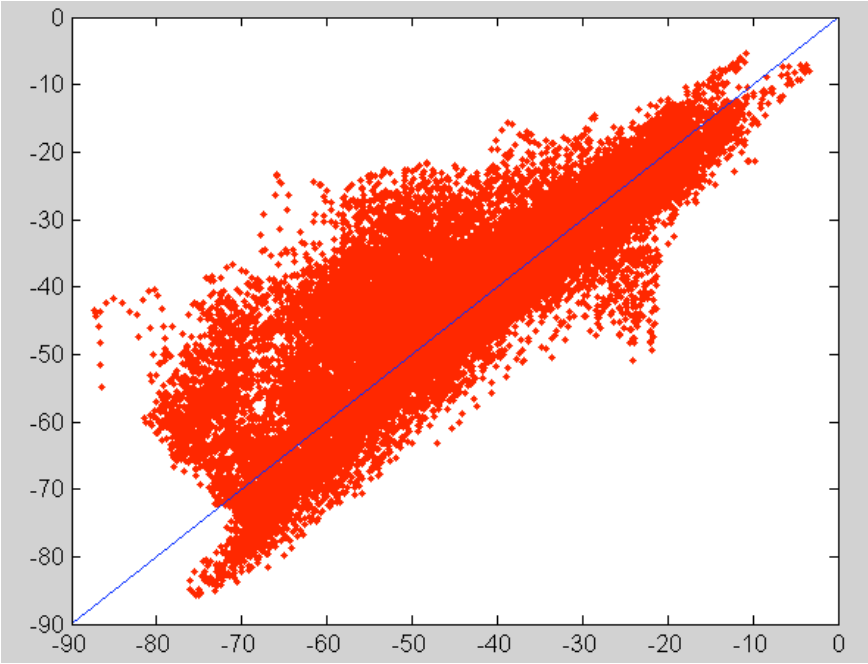
- Shift in mean of distribution by 3.4 dB (moderate reverberation) and 7.2 dB (high reverberation) due to additional reflected energy
- Distribution becomes increasingly skewed due to filling of low-energy regions



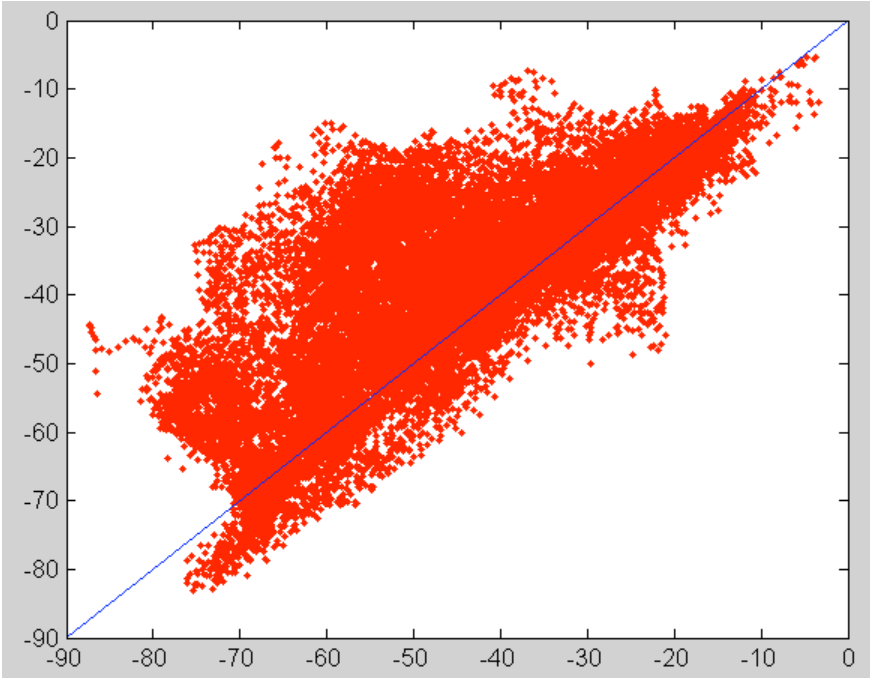
log energy (dB)

Scatter plots of clean vs reverb energy

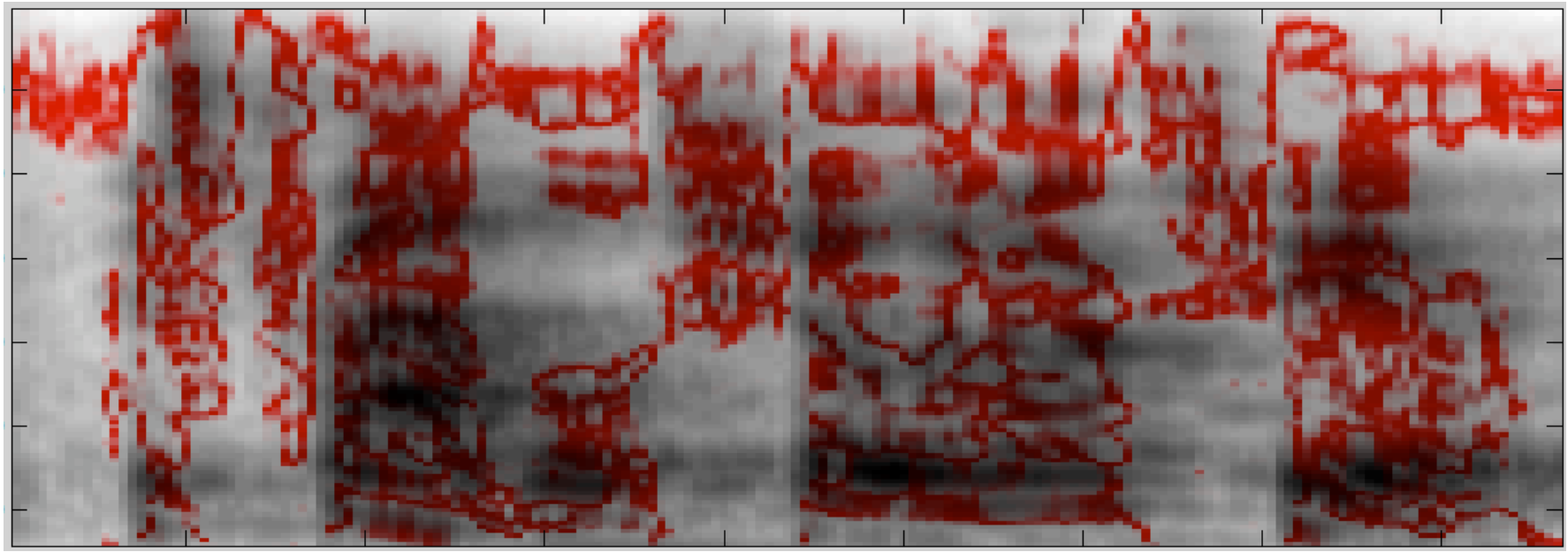
Moderate reverberation



high reverberation



Resynthesis from least corrupted parts



Saturation = N(reverb - clean; μ = 0dB, σ = 3dB)

reverb

clean+ssn (SNR=6dB)

least corrupted +ssn (SNR=6dB)

Reverberation in multisource environments

Two main consequences

1. Sources are now masked by i. each other, ii. their own reverberant energy, and iii. the reverberant energy of the other sources
   - Reduction in number and size of glimpses due to reverberant energy filling spectro-temporal dips

2. Reduced effectiveness of potential grouping cues:
   - binaural cues: due to randomisation of ILD and ITD pattern
   - dynamic F0 differences: due to blurring of harmonic locations (Culling et al, 1994)
   - onset/offset synchrony: blurring of onsets/offsets (though less so for onsets)

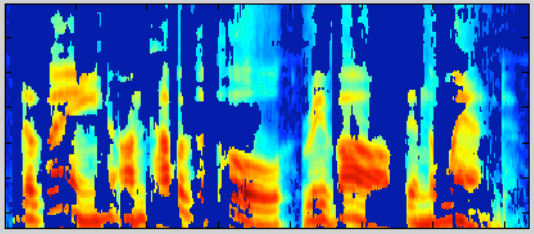
high reverb 

moderate reverb 

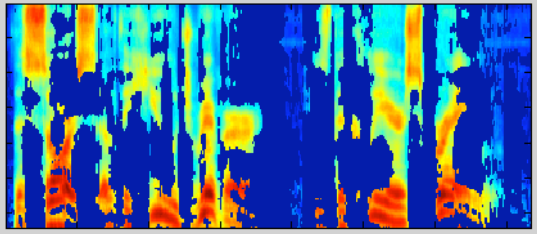
no reverb 

Corruption in moderate reverberation

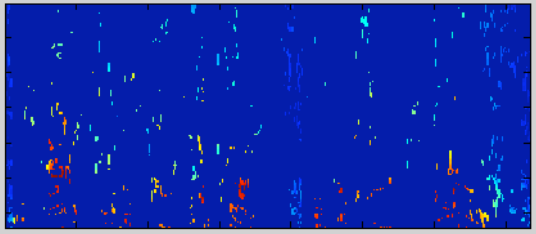
talker 1 close to mix value (50)



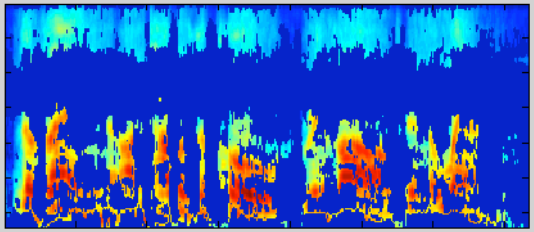
talker 2 close to mix value (46)



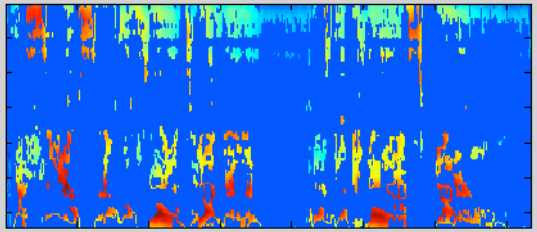
both talkers close to mix value (4)



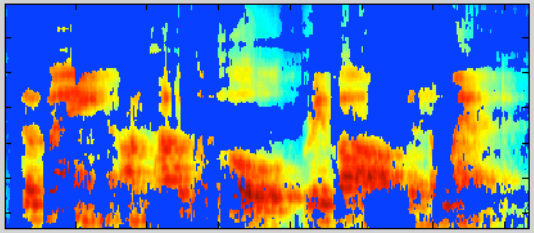
talker 1 close to reverb value (35)



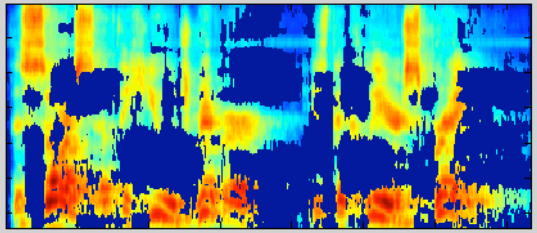
talker 2 close to reverb value (19)



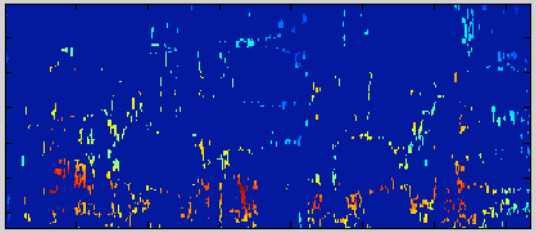
talker 1 reverb close to mix reverb value (42)



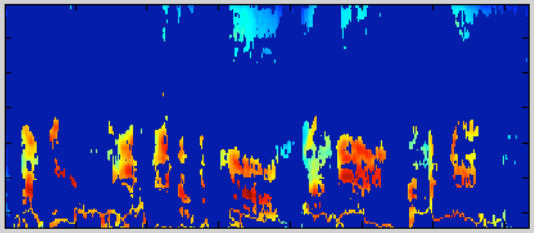
talker 2 reverb close to mix reverb value (57)



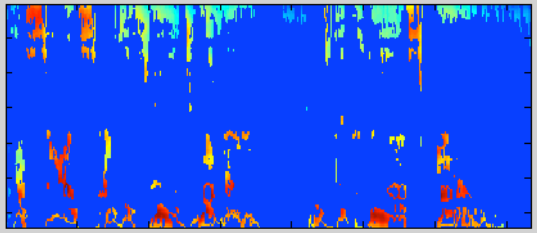
both talkers reverb close to mix reverb value (5)



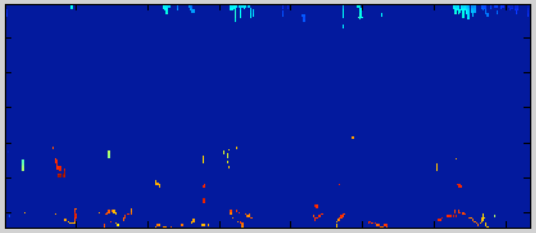
talker 1 close to reverb mix value (12)



talker 2 close to reverb mix value (11)

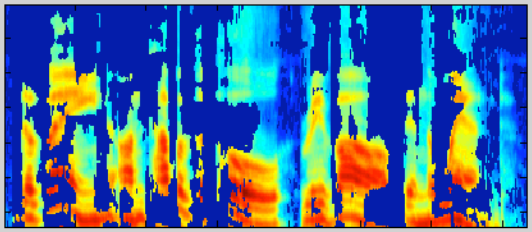


both talkers close to reverb mix value (1)

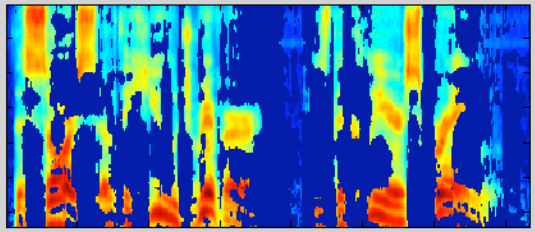


Corruption in high reverberation

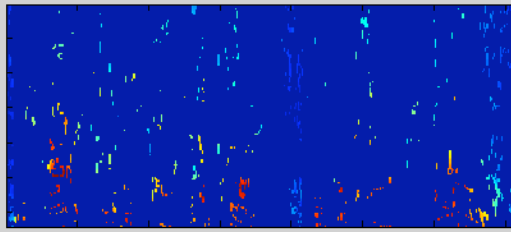
talker 1 close to mix value (50)



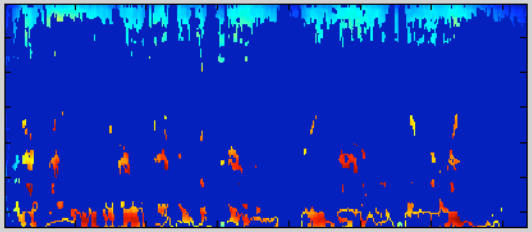
talker 2 close to mix value (46)



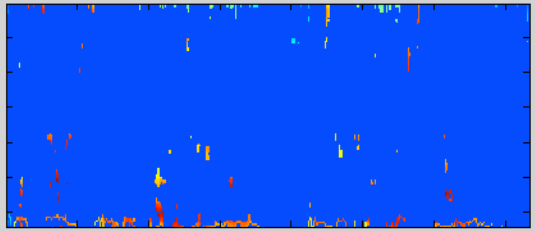
both talkers close to mix value (4)



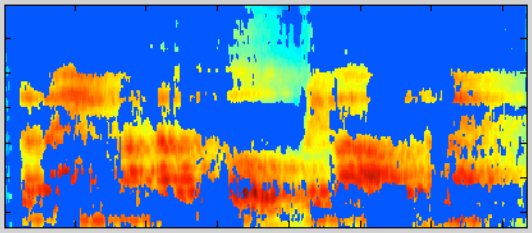
talker 1 close to reverb value (12)



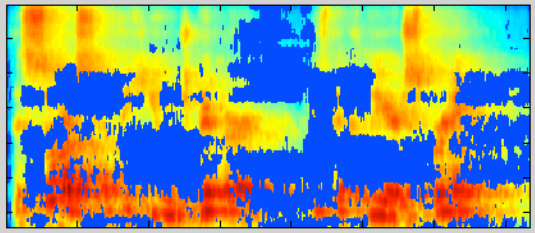
talker 2 close to reverb value (2)



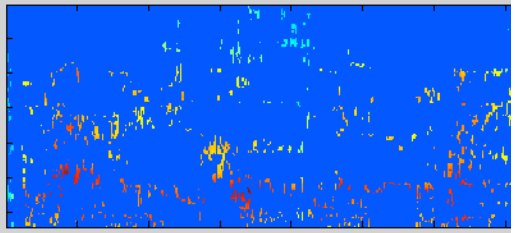
talker 1 reverb close to mix reverb value (39)



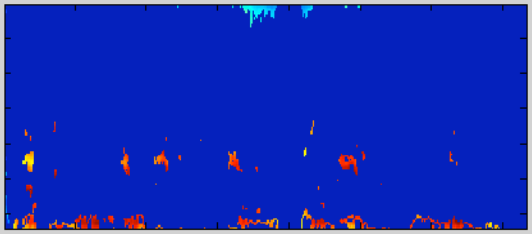
talker 2 reverb close to mix reverb value (60)



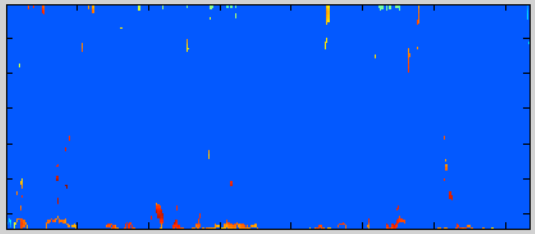
both talkers reverb close to mix reverb value (5)



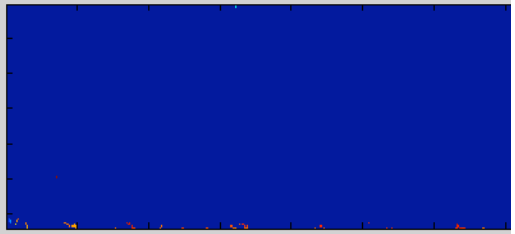
talker 1 close to reverb mix value (3)



talker 2 close to reverb mix value (1)



both talkers close to reverb mix value (0)



Montreal: November 2004

Cumulative distribution of energy corruption in highly-reverberant single and multisource environments

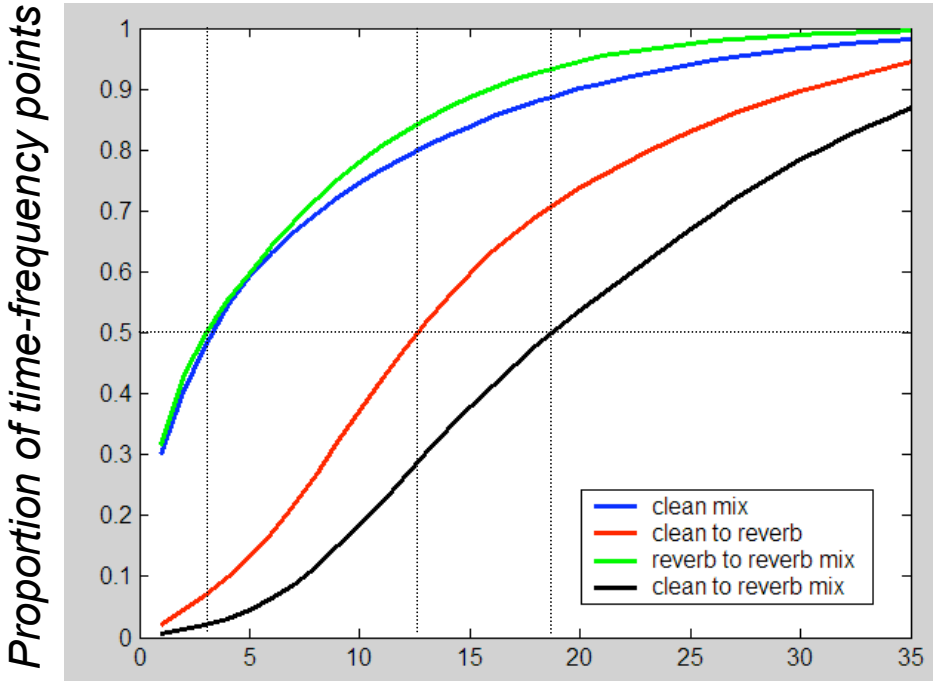
50% of points:

… in clean are within ~ 3 dB of their values in the nonreverberant mixture

… in reverberant speech are within ~ 3 dB of their values in the reverberant mixture

… in clean are within ~ 12 dB of their values in the reverberant signal

… in clean are within ~ 18 dB of their values in the reverberant mixture



Summary

- **In single source environments**, reverberation has a quite well understood effect on the speech signal and can be understood in terms of increased energetic masking
  - Intelligibility well predicted by STI
  - Can employ noise-robust features such as RASTA (Hermansky & Morgan, 1994) or modulation-filtered reps (Kingsbury et al, 1998)

- **In multisource environments**, reverberation additionally reduces the effectiveness of grouping cues
  - Not yet clear which speech features to use and how best to compensate for reverb when more than one source is present