Effects of Reverberation on Pitch, Onset/Offset, and Binaural Cues

DeLiang Wang

Perception & Neurodynamics Lab

The Ohio State University

Outline of presentation

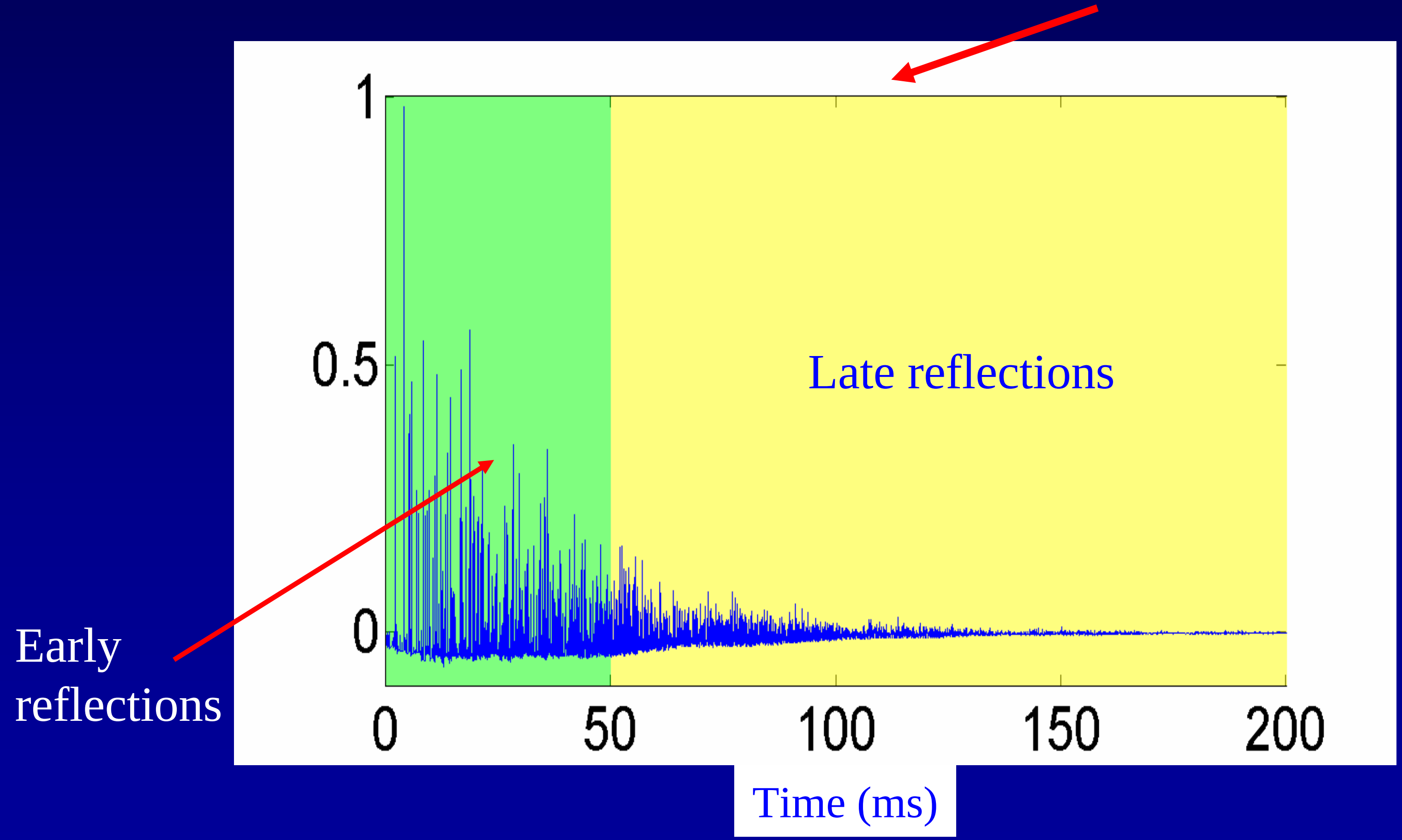
- **Introduction**
- **Human performance**
- **Reverberation effects**
  - On pitch
  - On onset/offset
  - On binaural cues
- **Monaural enhancement of reverberant signal**
- **Binaural segregation of reverberant signal**
- **Discussion and summary**

Reverberation as linear transmission system

$$x(t) = \int_{-\infty}^{\infty} h(\tau) s(t-\tau) d\tau$$

$x(t)$: reverberant signal; $s(t)$: source signal

$h(\tau)$: room impulse response function

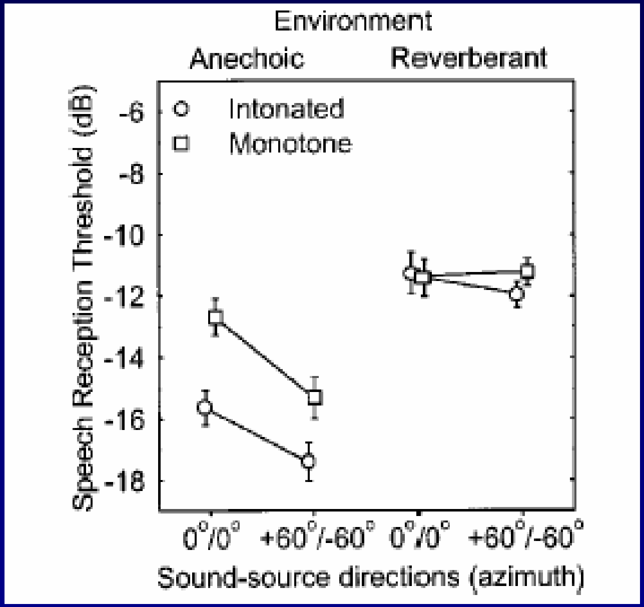


Reverberation and speech quality

- **Room reverberation causes two distinct perceptual effects on speech quality**
  - Early reflections lead to coloration or spectral deviation, determined by signal-to-reverberant energy ratio; it also boasts loudness
  - Late reflections (long-term reverberation) smear the time-frequency components of speech, and are characterized by the reverberation time (T60)

Human performance

- **Though speech perception in quiet seems robust to reverberation, speech intelligibility in noise suffers in the presence of reverberation (Plomp'76; Culling et al.'03)**
  - Culling et al. showed that reverberation (T60 = 0.4 s) produces 5 dB increase in speech reception threshold when naturally intonated speech is presented together with a competing talker
  - Hearing impaired listeners are particularly susceptible to reverberation
- **The binaural advantage for speech perception in noise is diminished by reverberation**
  - The Culling et al. study found no advantage at all



Culling et al. (2003)

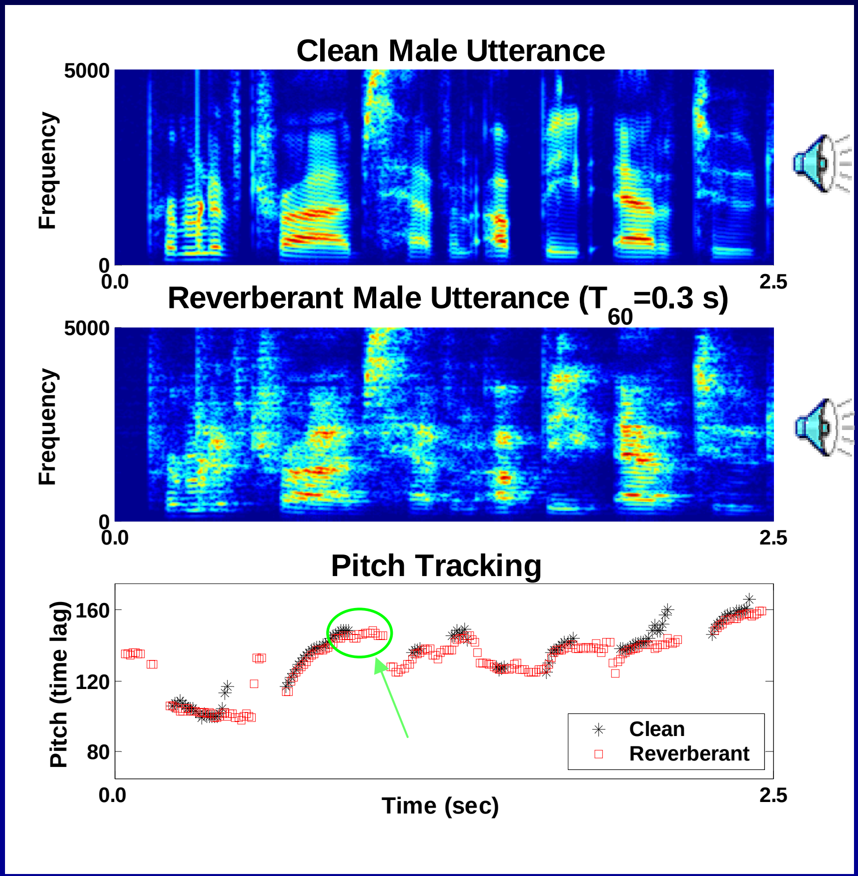
Human performance

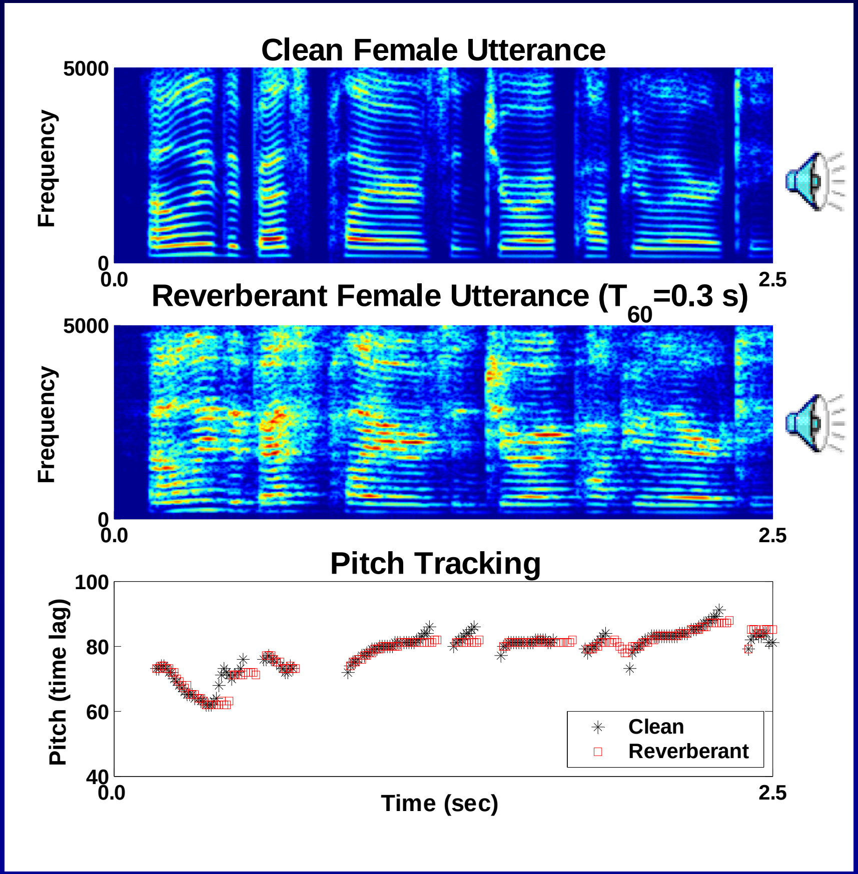
- **Darwin and Hukin (2000) compared reverberation effects on spatial, pitch, and vocal-tract size cues for sequential organization and found that**
  - ITD cues are seriously impaired by reverberation
  - Pitch cues (F0 trajectory) are more resistant
  - A combination of pitch and vocal-tract size cues is very resistant to reverberation

Outline of presentation

- **Introduction**
- **Human performance**
- **Reverberation effects**
  - On pitch
  - On onset/offset
  - On binaural cues
- **Monaural enhancement of reverberant signal**
- **Binaural segregation of reverberant signal**
- **Discussion and summary**

Pitch tracking of a single utterance

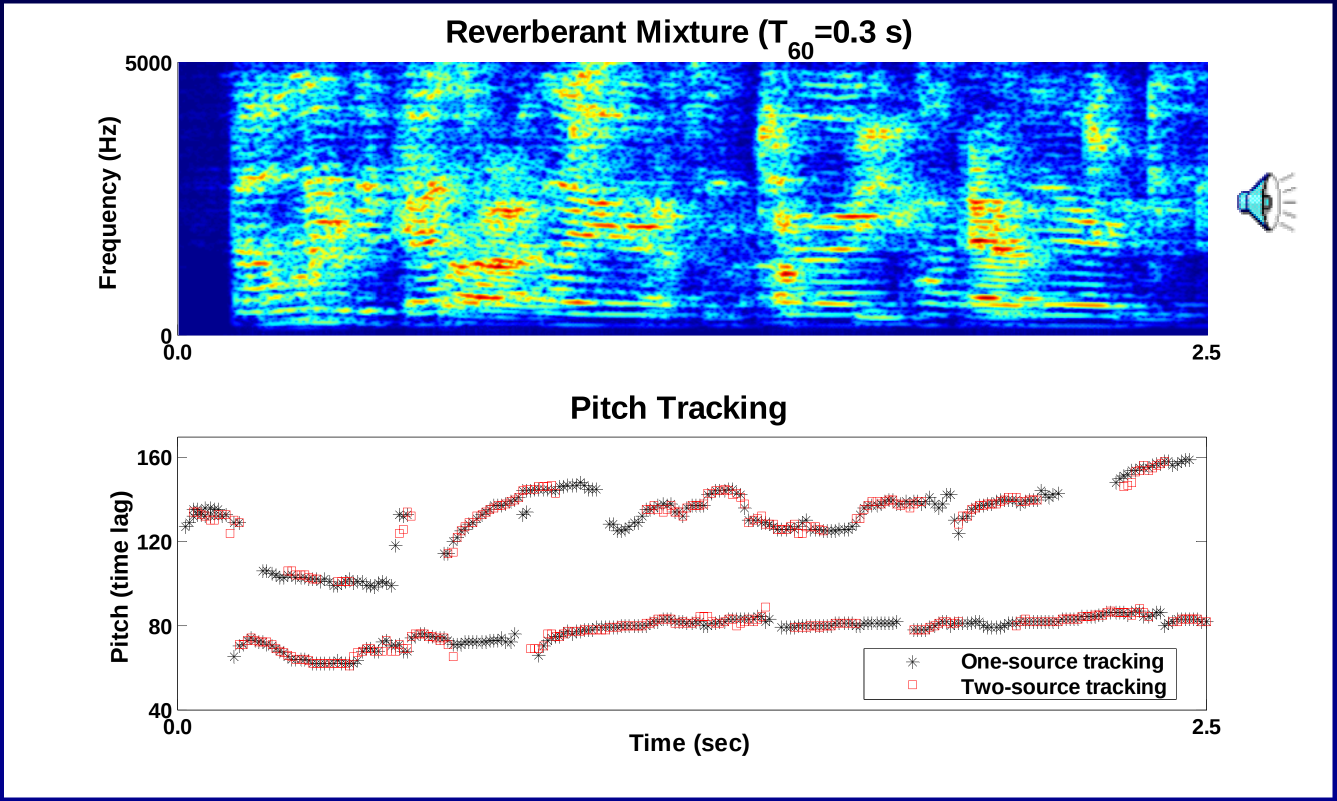




- Pitch is pretty robust to reverberation, especially for slowly changing pitch tracks and long voiced speech segments

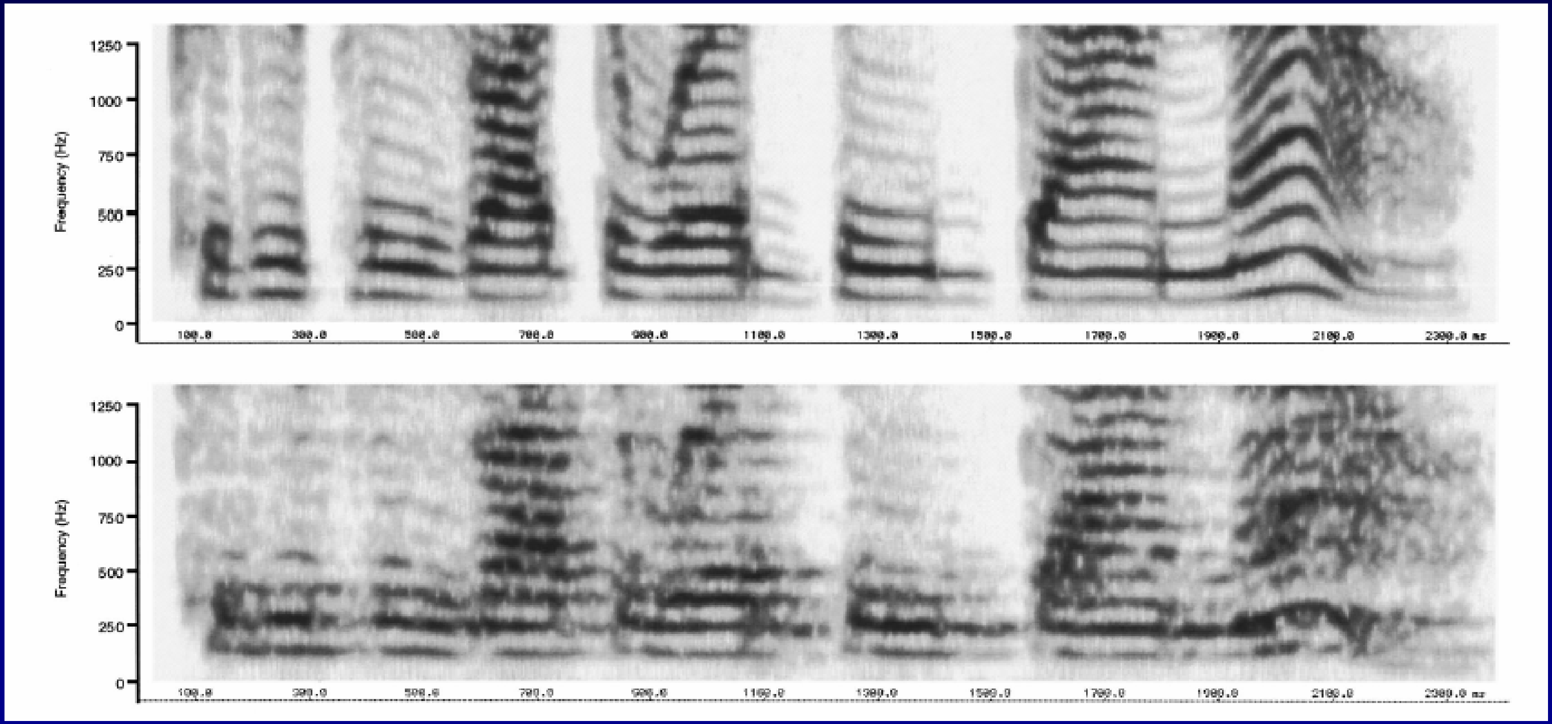
- Noticeable artifacts: elongated pitch tracks

Pitch tracking of two utterances



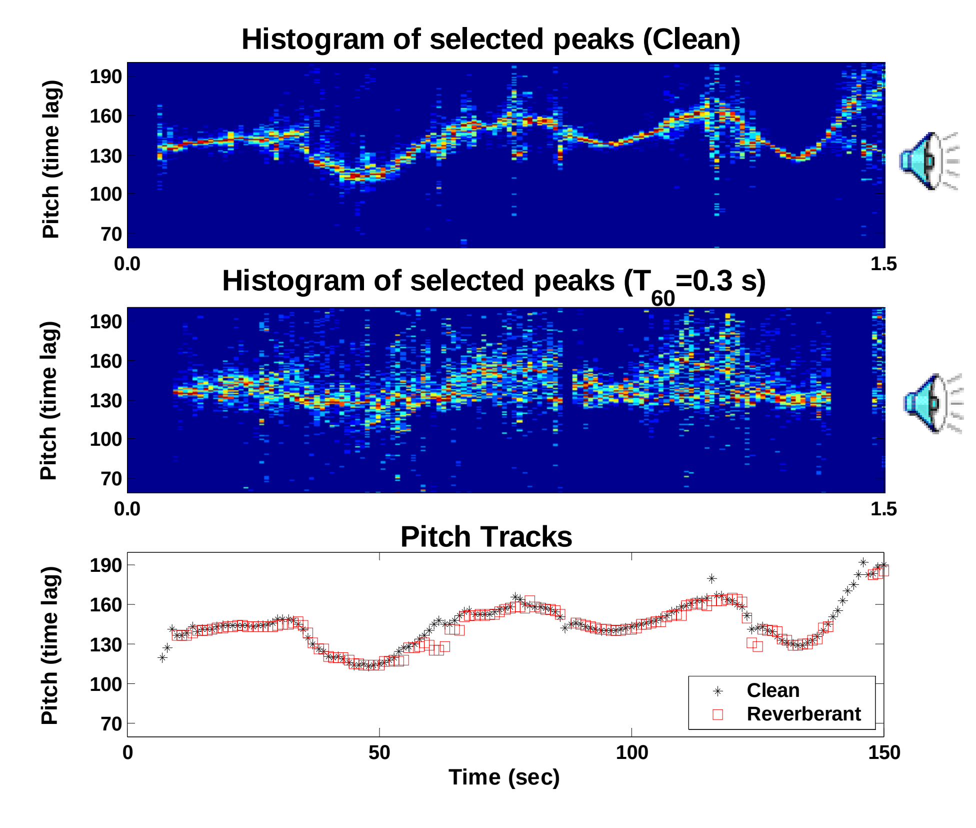
Multipitch tracking using the Wu et al. algorithm (2003). Even with multiple reverberant sources, pitch tracking works reasonably well

Reverberation effects on harmonic structure



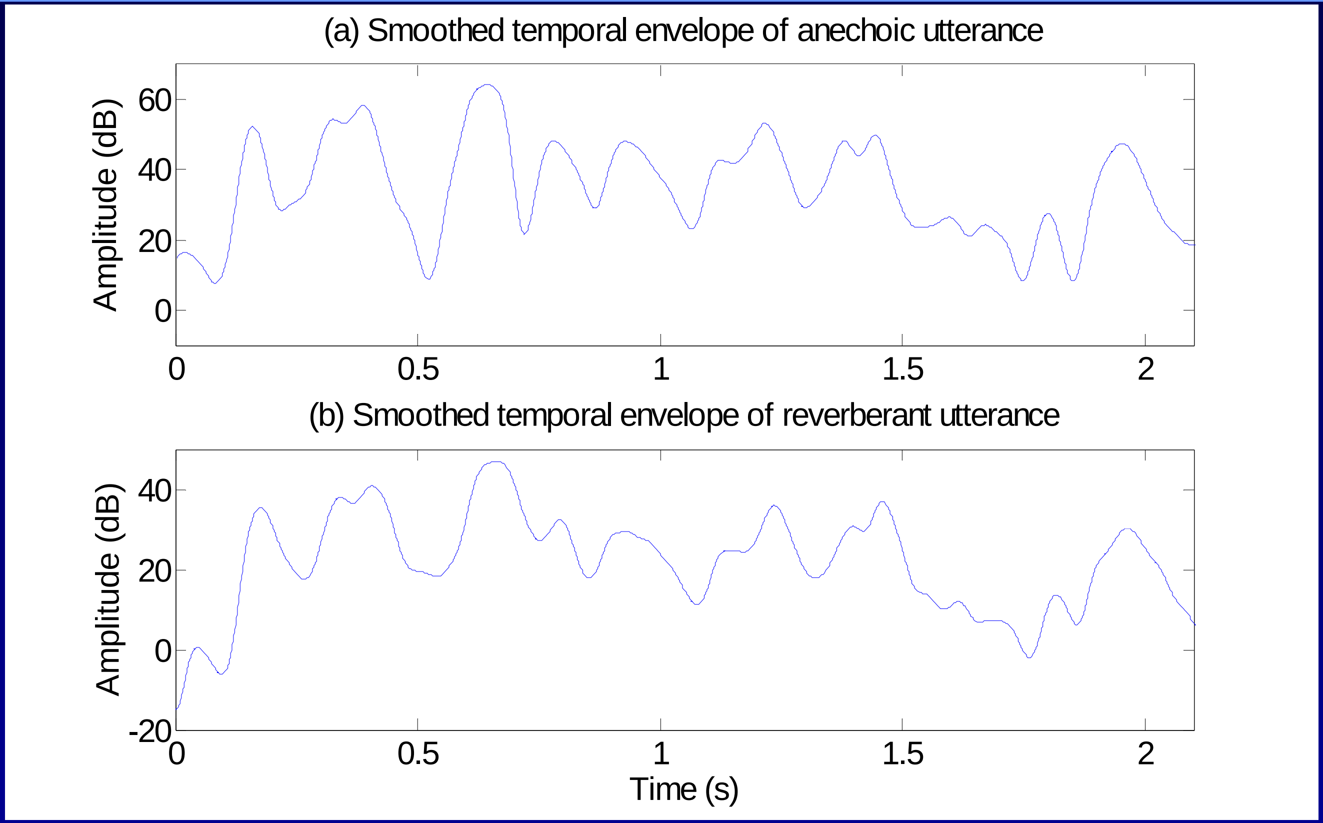
- From Darwin and Hukin (2000). The utterance is “Could you please write the word bead down now.” T60 = 0.4 s
- Primarily in the low-frequency range

Implications on pitch-based grouping



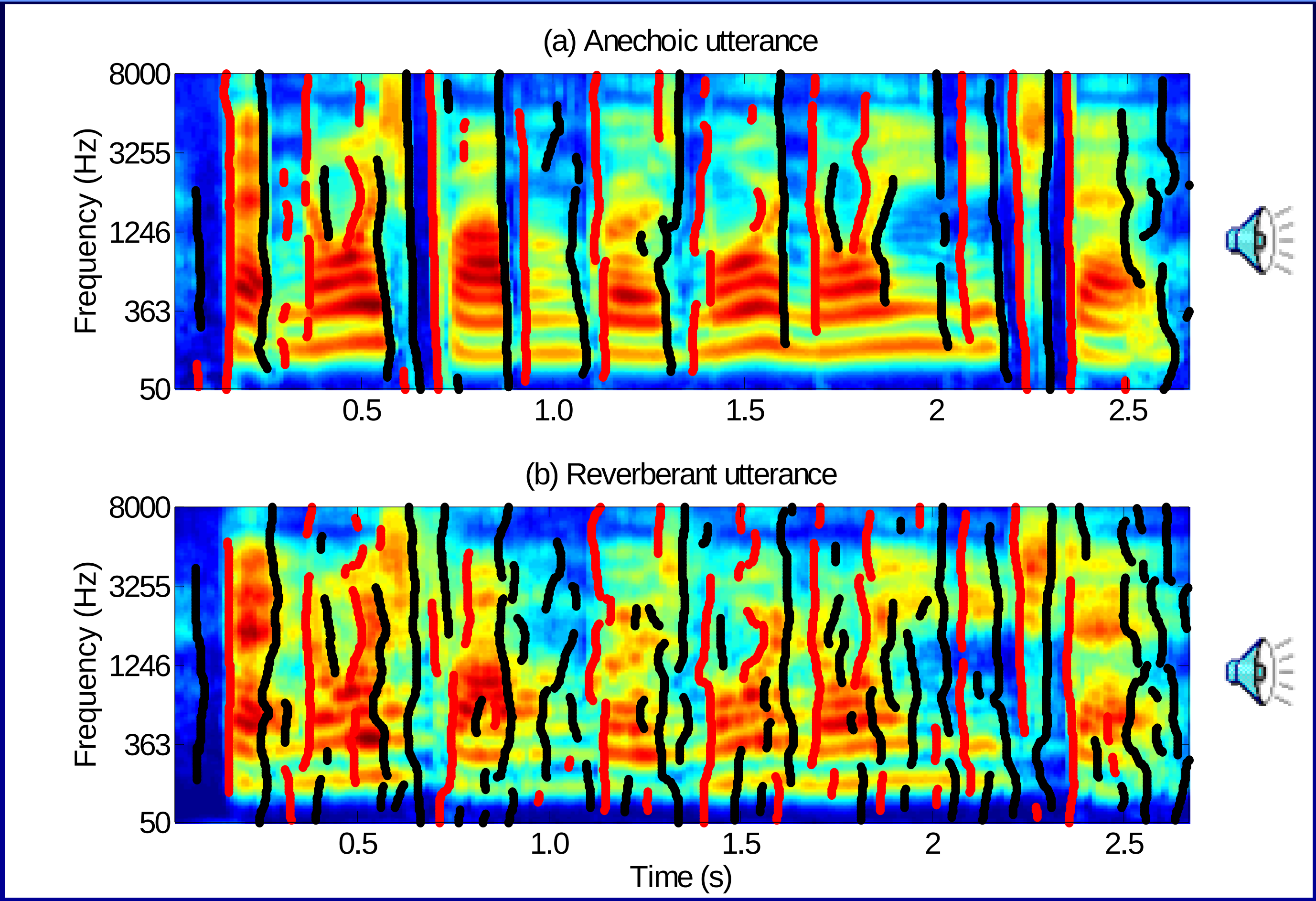
Smearing of harmonic structure is worse in the high-frequency range. The figure shows the histogram of peak positions that are nearest to the detected pitch periods for frequencies greater than 800 Hz. This smearing effect would degrade the performance of pitch-based grouping.

Reverberation effects on temporal envelope



- Response envelope of a gammatone filter centered near 1 kHz to the utterance "That noise problem grows more annoying each day." (a) T60 = 0 and (b) T60 = 0.3 s
- Amplitude modulation (AM) depth is reduced, but the AM pattern is reasonably maintained

Onset and offset detection



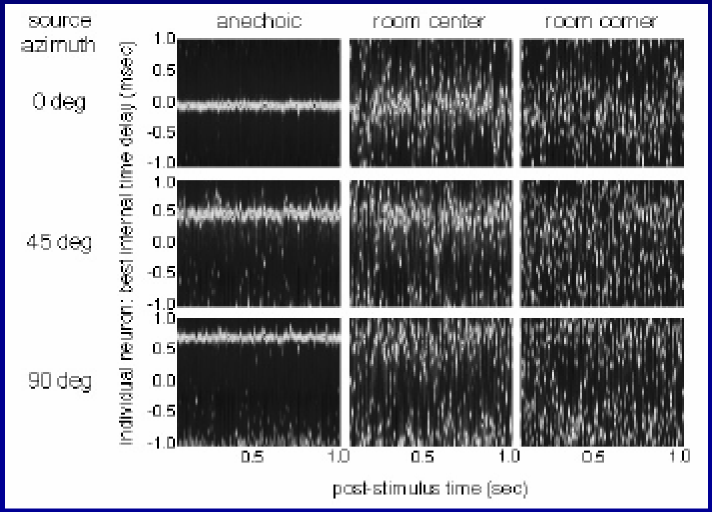
Cochleogram representation. Red/black marks indicate detected onsets/offsets. The utterance: “That noise problem grows more annoying each day.”

Reverberation effects on onset/offset detection

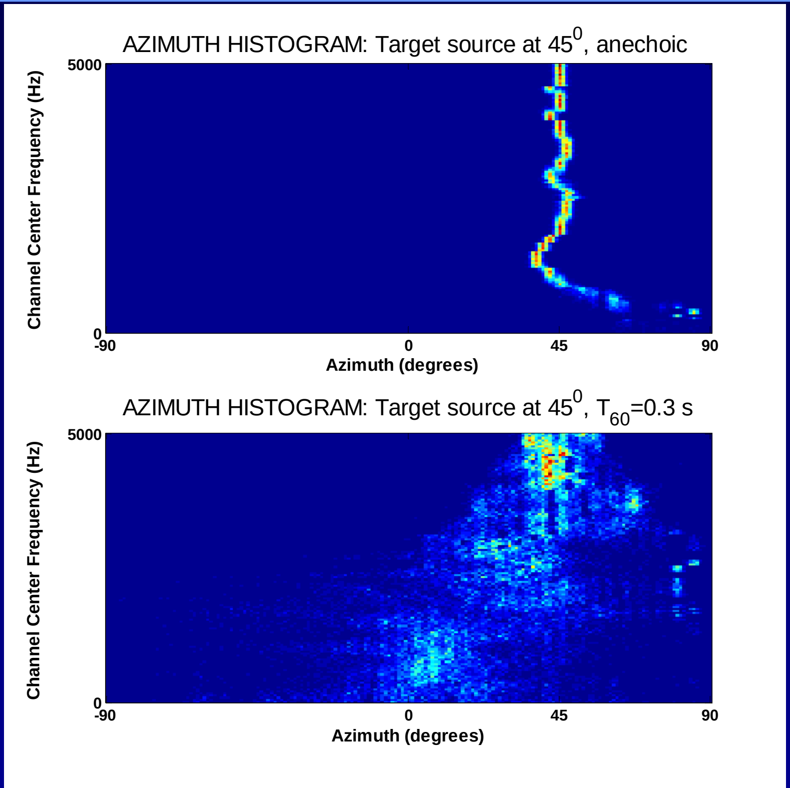
- **Both the times and strengths of onsets and offsets are affected**
  - Onset times are slightly shifted
  - Onsets of weak phones (e.g. unvoiced stops) are smeared
  - Offset times are shifted forward (delayed)
- **Reverberation introduces spurious offsets**

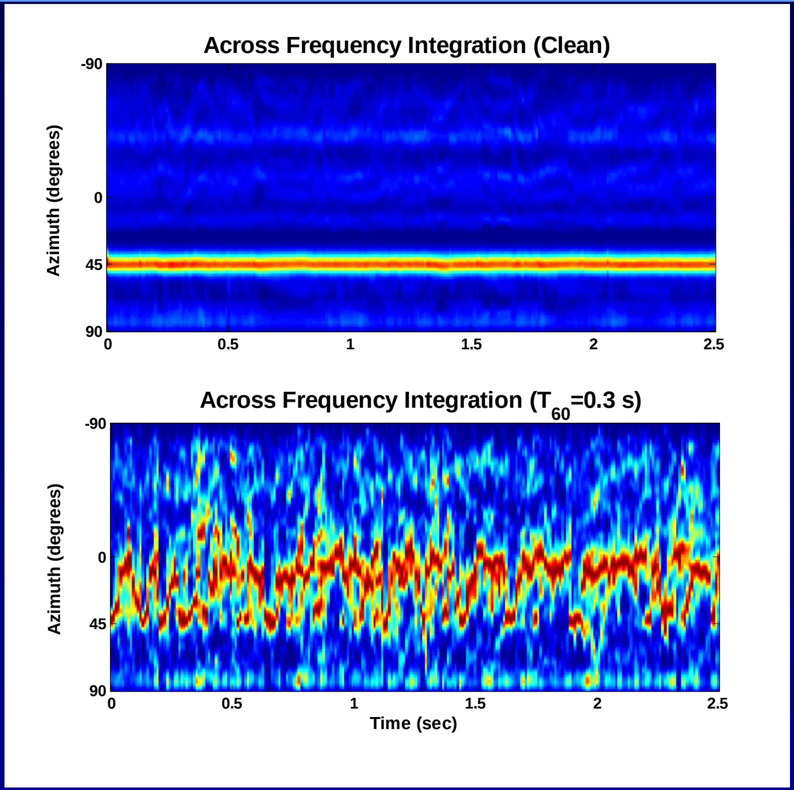
Reverberation effects on binaural cues: ITD

- Shinn-Cunningham and Kawakyu (2003) showed that the responses of a neural model to ITD (interaural time difference) are poor indicators of source azimuth in the presence of reverberation
- Integration over time enhances the estimation robustness



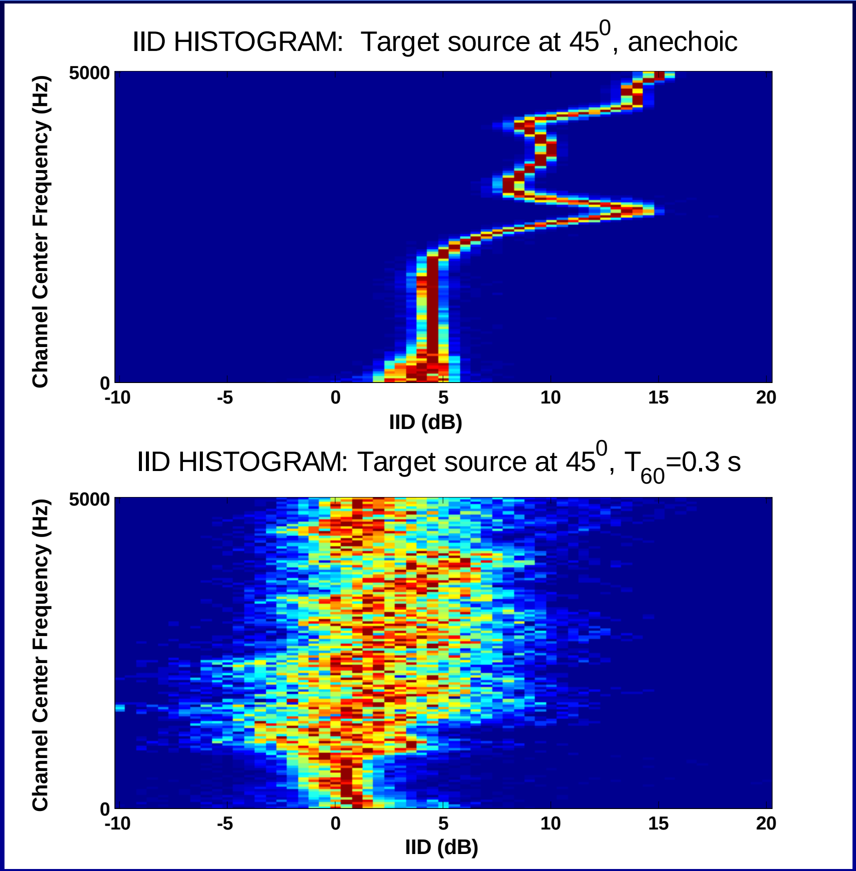
ITD estimation in time-frequency (T-F) units

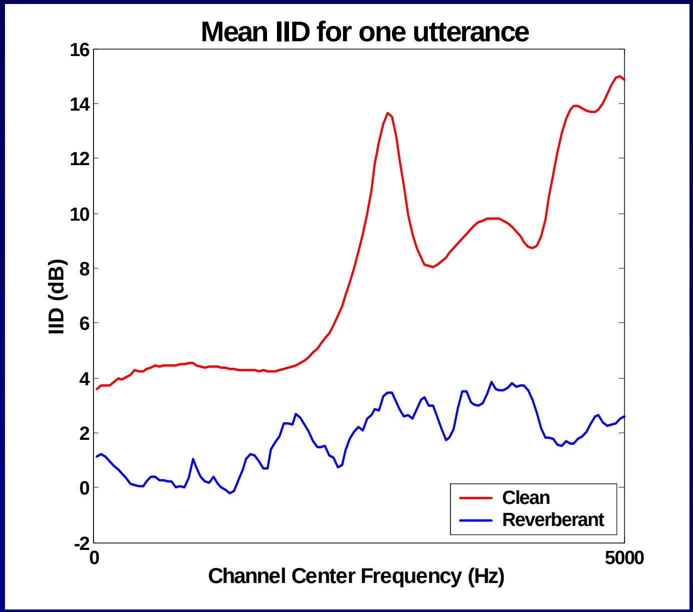




ITD estimation in individual T-F units using a cross-correlation model (Roman et al.'03). The input is natural speech. The distribution of local azimuth estimates is much noisier in the reverberant condition

Interaural intensity difference estimation in T-F units





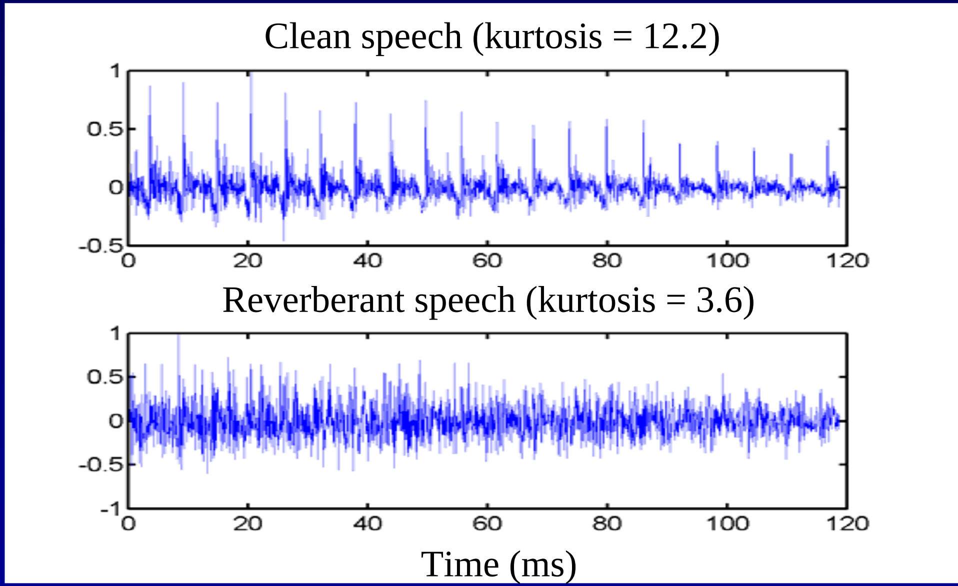
The distribution of IID (interaural intensity difference) is also much noisier in reverberation, and the mean IID values lose characteristics

Outline of presentation

- **Introduction**
- **Human performance**
- **Reverberation effects**
  - On pitch
  - On onset/offset
  - On binaural cues
- **Monaural enhancement of reverberant signal**
- **Binaural segregation of reverberant signal**
- **Discussion and summary**

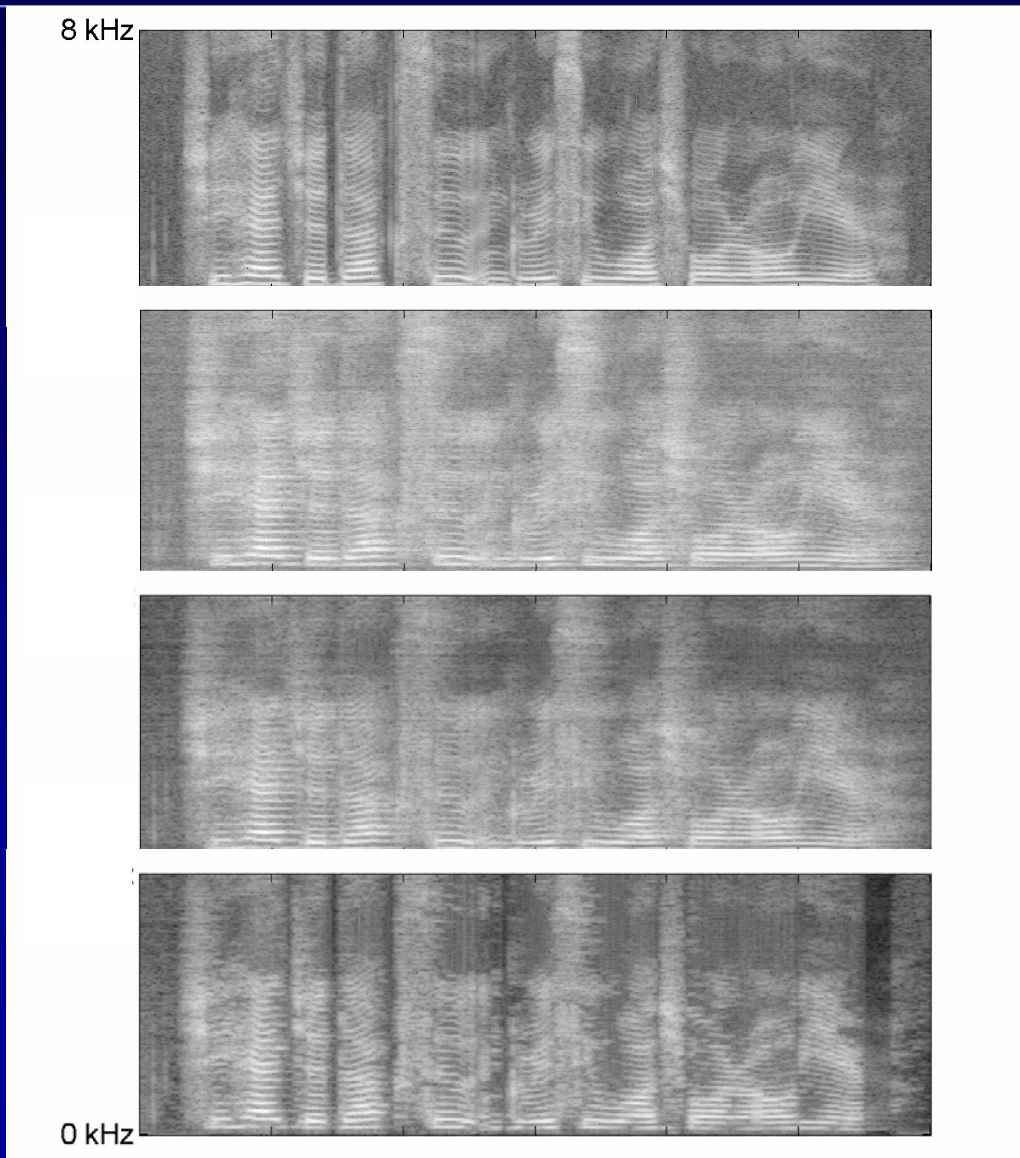
A two-stage enhancement algorithm (Wu'03)

- **Identify an inverse filter to reduce coloration distortion by maximizing kurtosis of LPC residue (Gillespie et al.'01)**



- **Estimate and subtract the effects of long-term reverberation**

Results of Wu's enhancement algorithm



Original speech

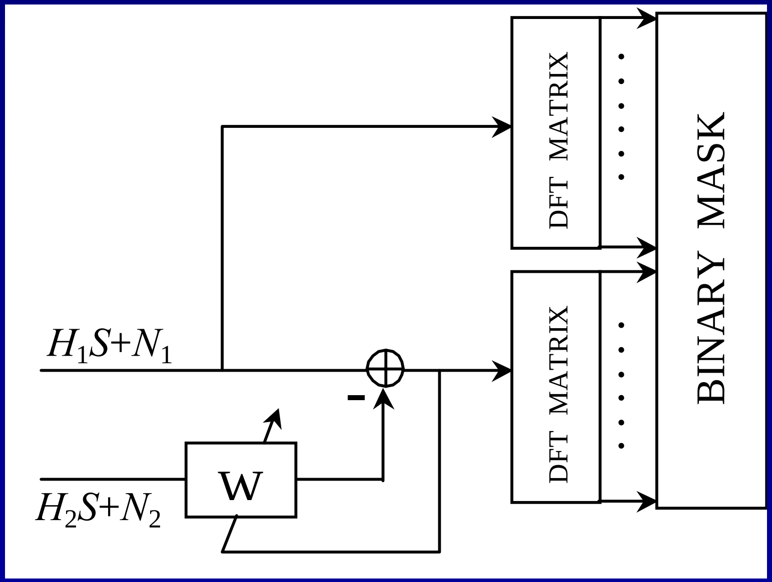
Reverberant speech

Inverse-filtered speech

Enhanced speech

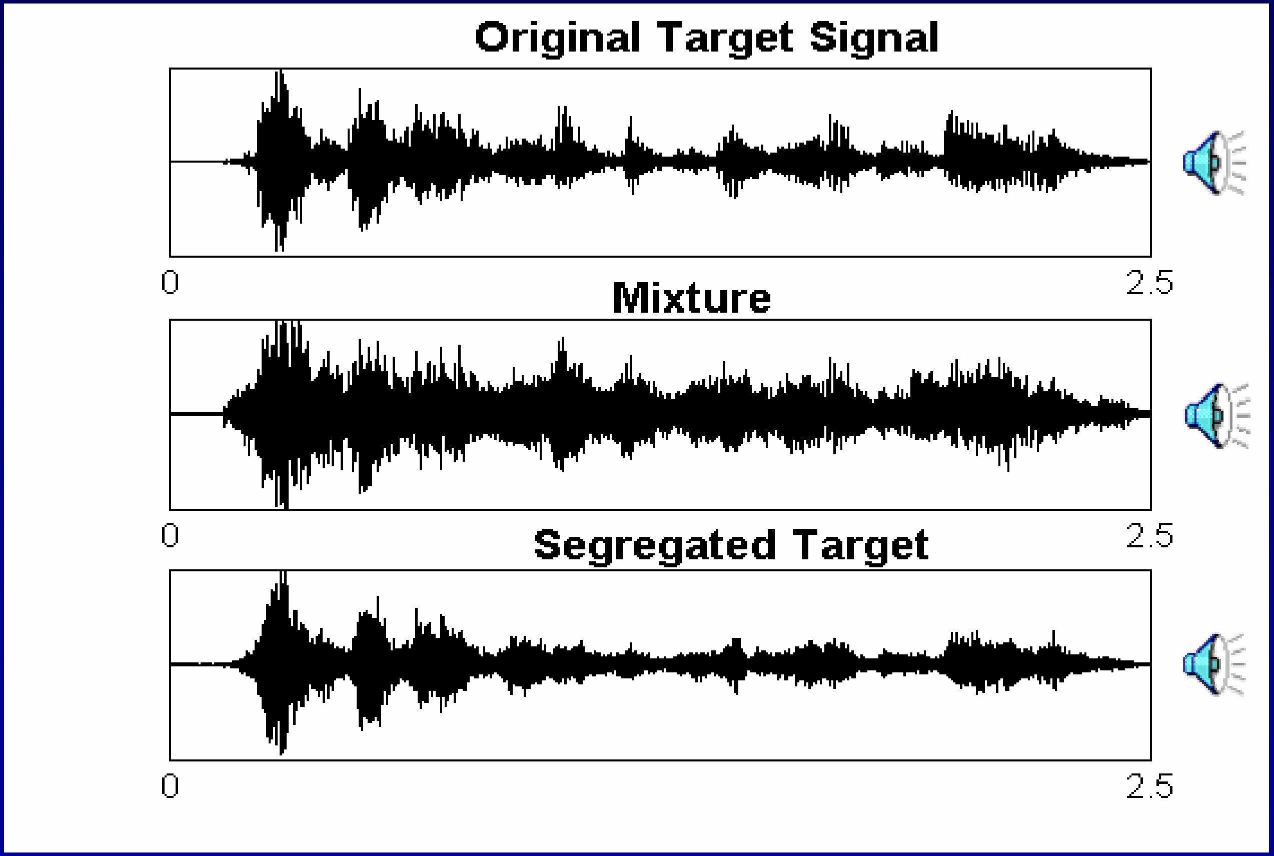
Binaural segregation of reverberant speech

- **Roman and Wang (2004) proposed a figure-ground segregation strategy to identify the T-F units dominated by target using spatial information, without imposing restrictions on the number, location or content of interfering sources**
- **Basic idea**
  - First perform cancellation of reverberant target (with detected target location) using adaptive filtering
  - Then label those T-F units that have been largely attenuated in the first stage since they are more likely to originate from the target location



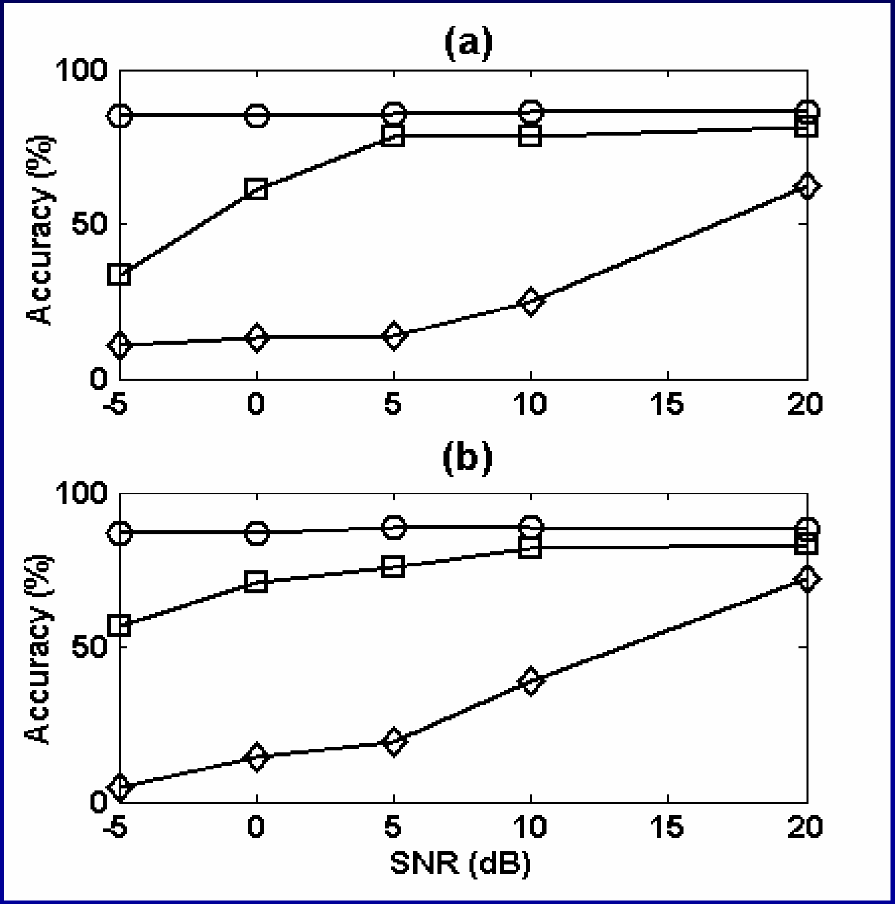
Segregation results

- An example with a target speaker at 0° and 4 other interfering speakers at (-135°, -45°, 45°, 135°) and T60 = 0.3 s



ASR results

- **The segregation output is fed to a missing data recognizer (Cooke et al.'01)**



(a) 5 speaker configuration

Baseline performance

Estimated binary mask

Ideal binary mask

(b) Nonspeech intrusion: rock music at 45º

Summary and discussion

- **Reverberation corrupts auditory cues**
  - Pitch estimation is relatively robust, but harmonic structure is smeared, particularly in high-frequency
  - AM depth is reduced but the AM pattern is reasonably maintained
  - Onset times, and especially offset times, are shifted; onset and offset synchrony is weakened
  - Binaural cues become unreliable
- **A two-stage monaural algorithm for reverberant speech enhancement**
- **A binaural algorithm for segregating reverberant speech**
- **Issues**
  - What is "ground truth" pitch for a reverberant signal?
  - Dereverberation versus enhancement
  - How to deal with both segregation and reverberation monaurally?

Acknowledgment

- **N. Roman and G. Hu for performing some computer experiments**
- **Funding by AFOSR/AFRL and NSF**