Robustness, Separation & Pitch

or

Morgan, Me & Pitch

Dan Ellis
Columbia / ICSI

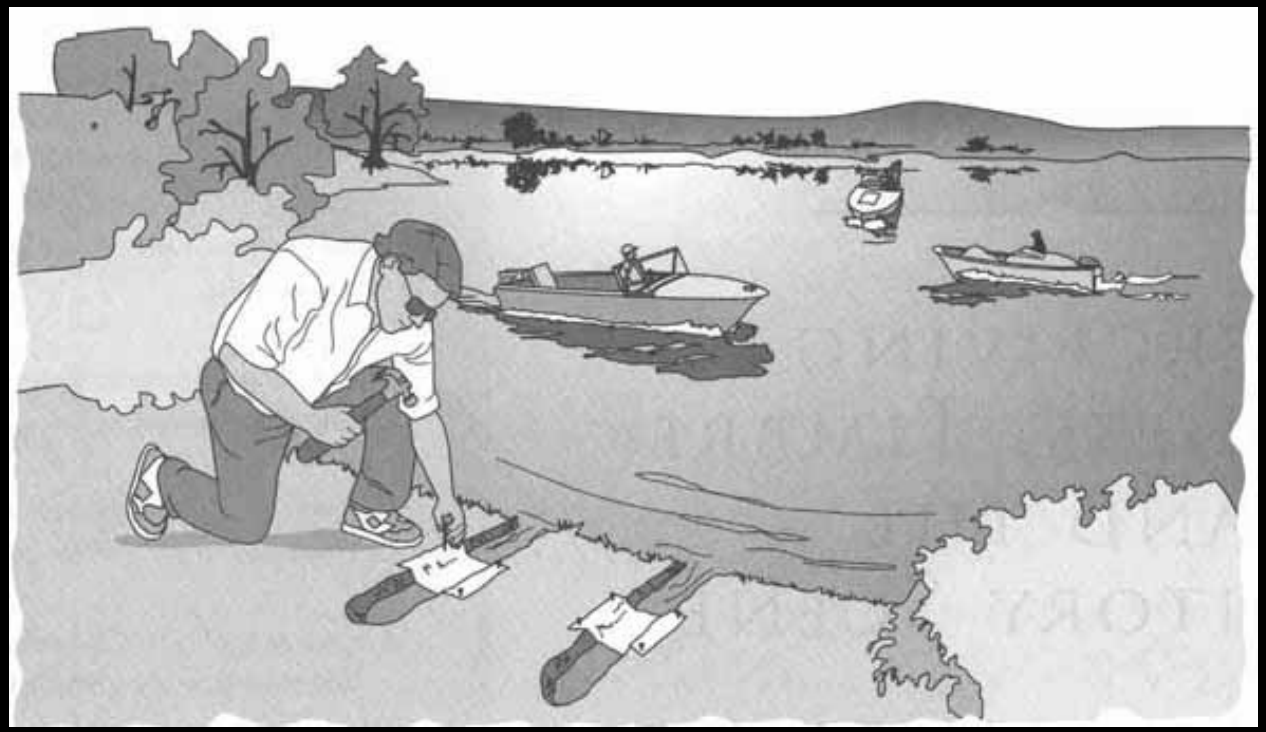
dpwe@ee.columbia.edu http://labrosa.ee.columbia.edu/

1. Robustness and Separation
2. An Academic Journey
3. Future

1953: How To Separate Speech?

- **The "Cocktail Party Problem"** *[Cherry '53]*
  - Spatial information: ATC over a single speaker
  - Pitch differences via gender differences

- **Auditory Scene Analysis** *[Bregman '90]*

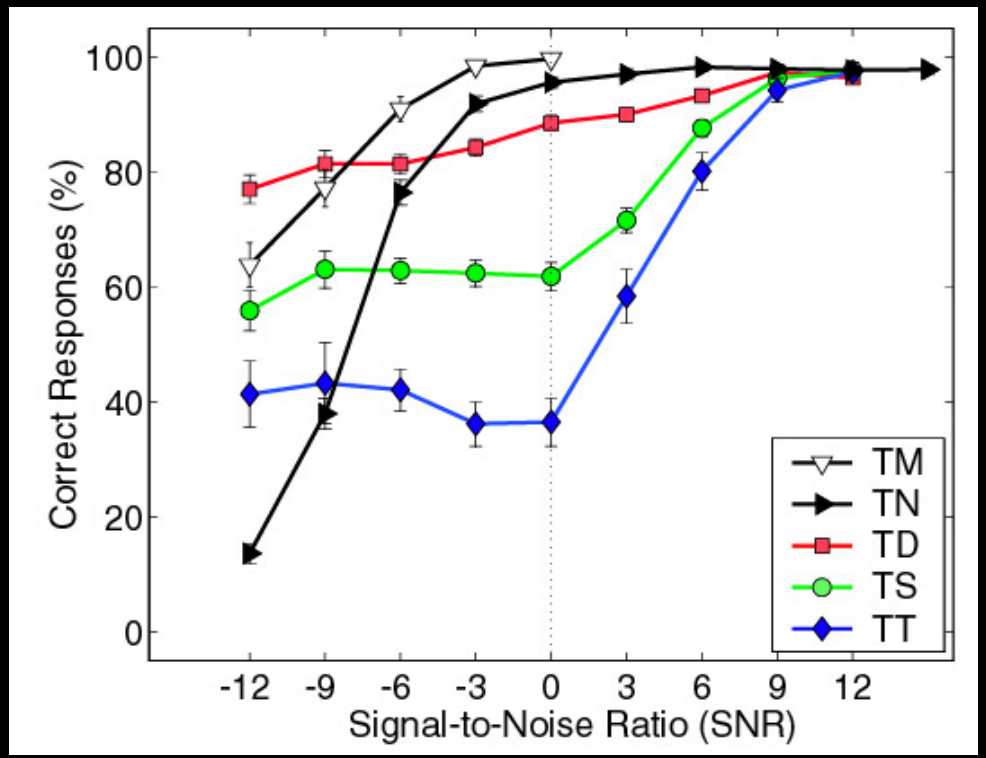


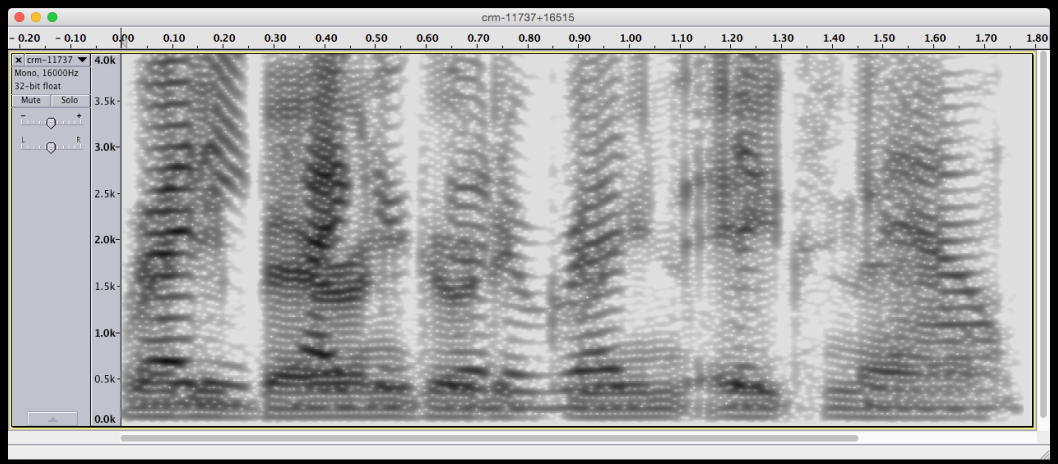
- **Grouping cues**
  - Onset
  - Harmonicity
  - Common Fate
  - "Schema"

The Usefulness of Pitch

Brungart et al.'01

- Common pitch can link energy from a single source





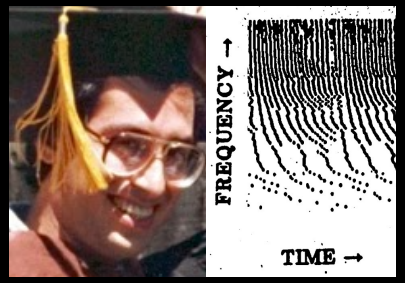
Normal mix

"Pitchless"

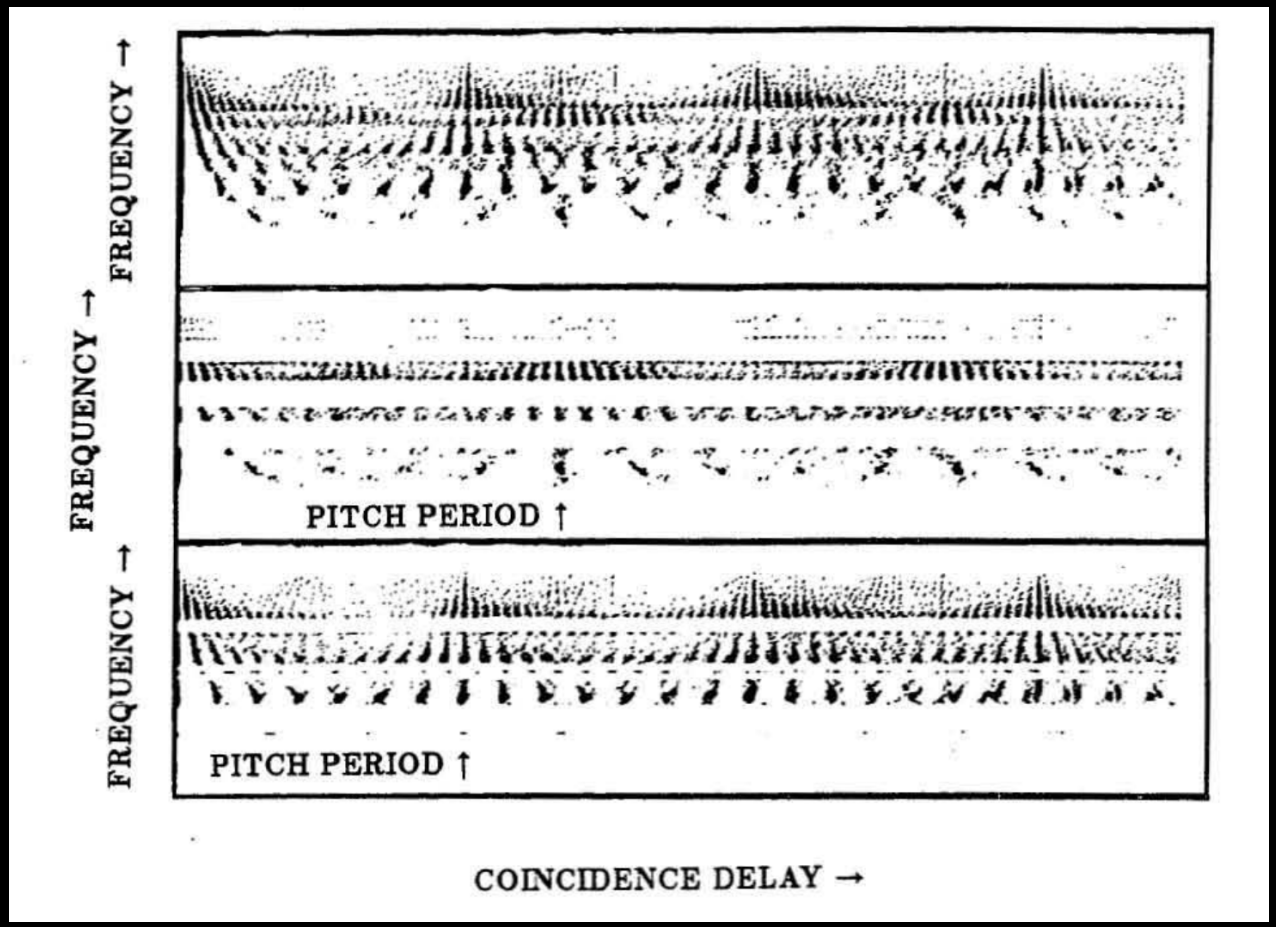
1984: Perception-Inspired Separation

- Model the periodicity information in the auditory nerve

Lyon 1984



Weintraub 1985

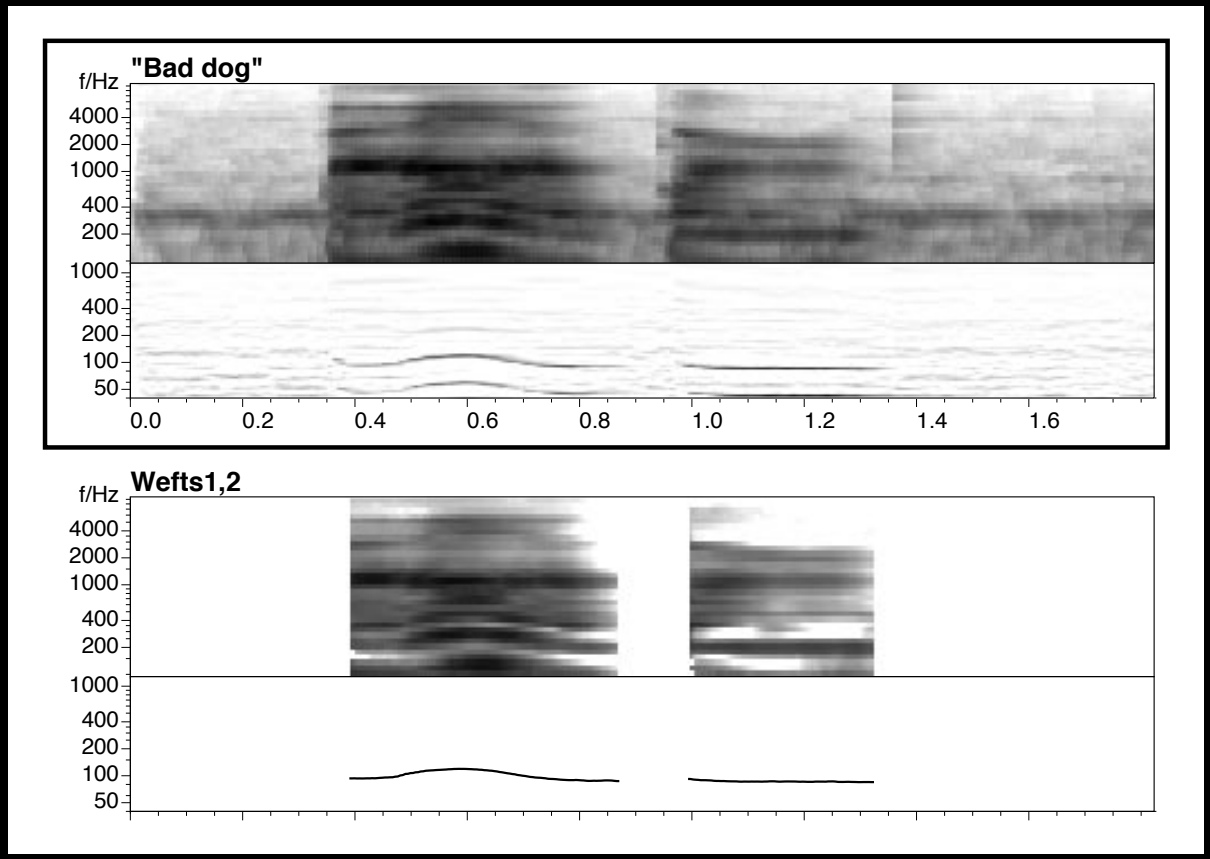


Mix

Female

1996: An Academic Journey

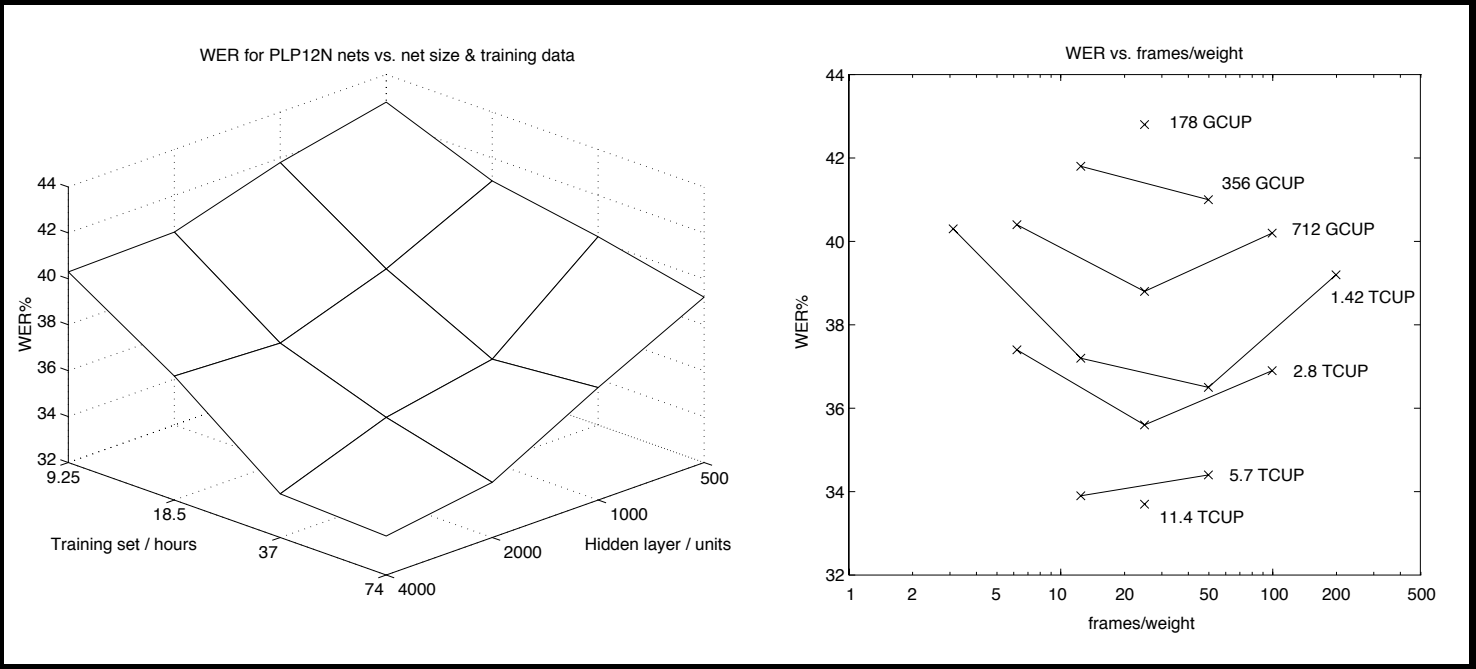
- Dan in 1996: The "weft"



1999: "Size Matters"

Ellis & Morgan'99

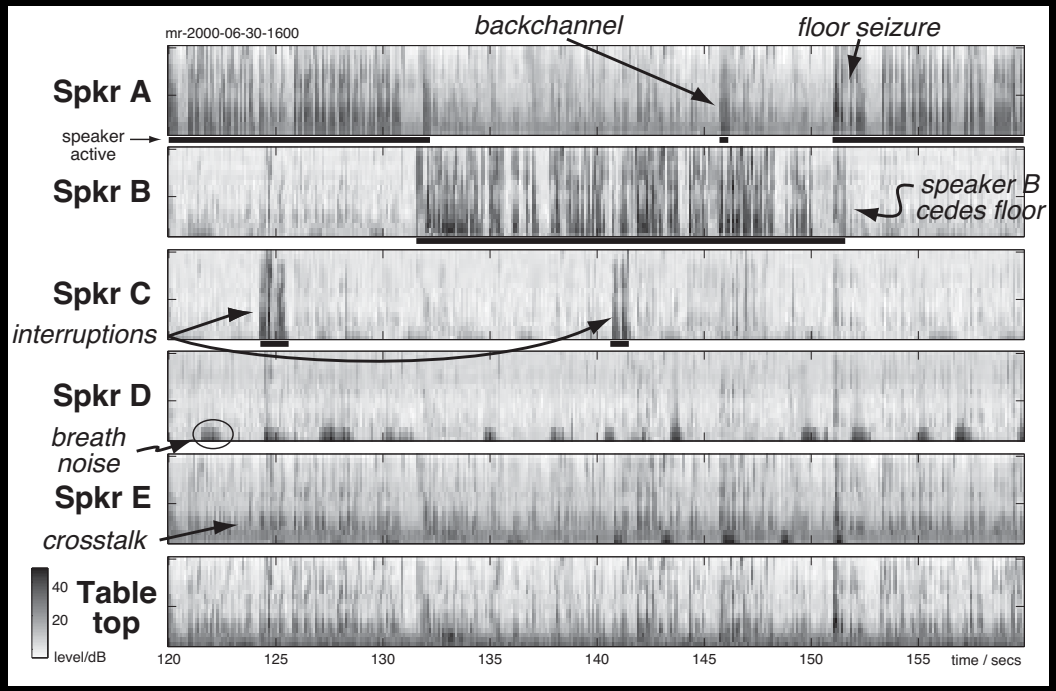
- All you need is a BDNN
  - .. and the data (and patience) to train it



2001: Overlap Remains

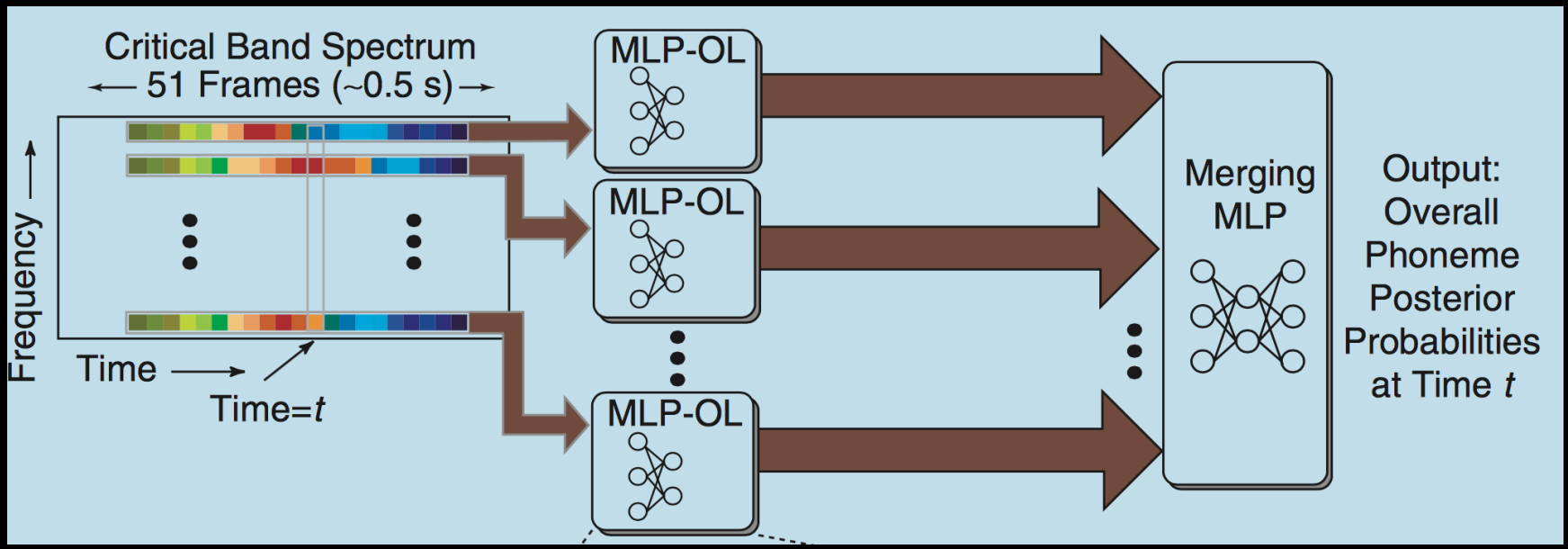
Janin, Baron, Edwards, Ellis, Gelbart, Morgan, Peskin, Pfau, Shriberg, Stolcke, Wooters '03

- **Meeting Recorder Project**
  - natural speech interactions
  - ~10% of speech frames have overlaps



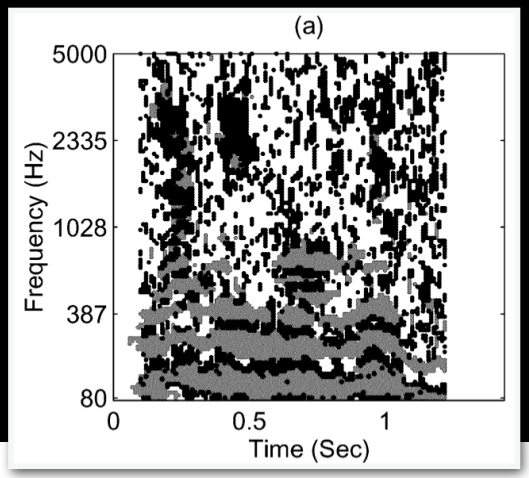
2003: EARS

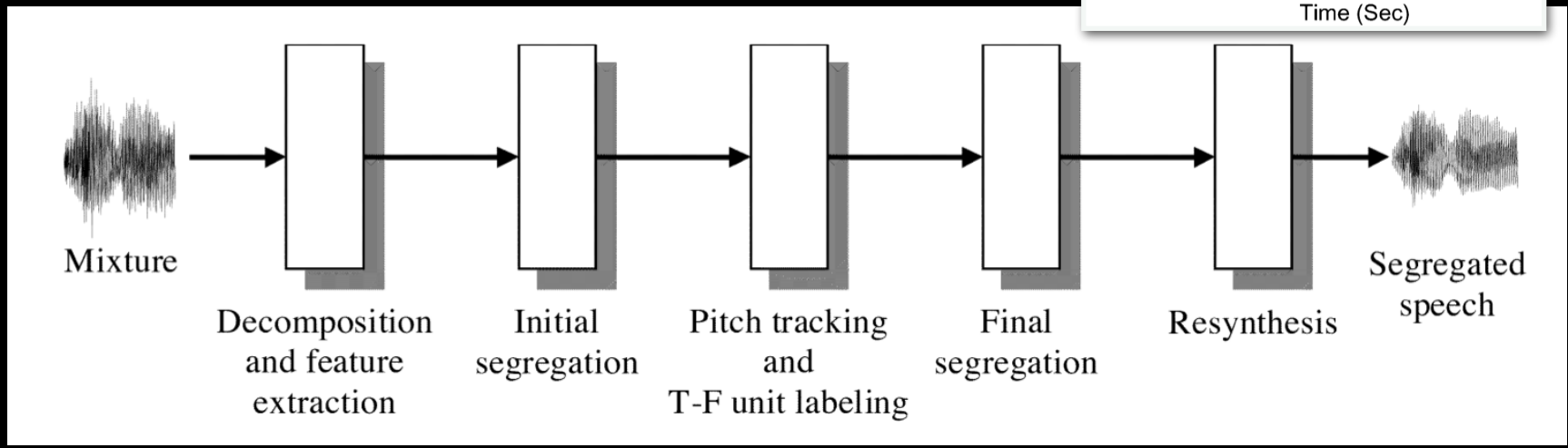
- "Pushing the envelope (aside)"



2004: Pitch Based Separation

- Literal implementations of the process described in Bregman 1990:
  - compute "regularity" cues:
    - common onset
    - gradual change
    - harmonic patterns
    - common fate





Hu & Wang 2004

Original v3n7

Brown 1992

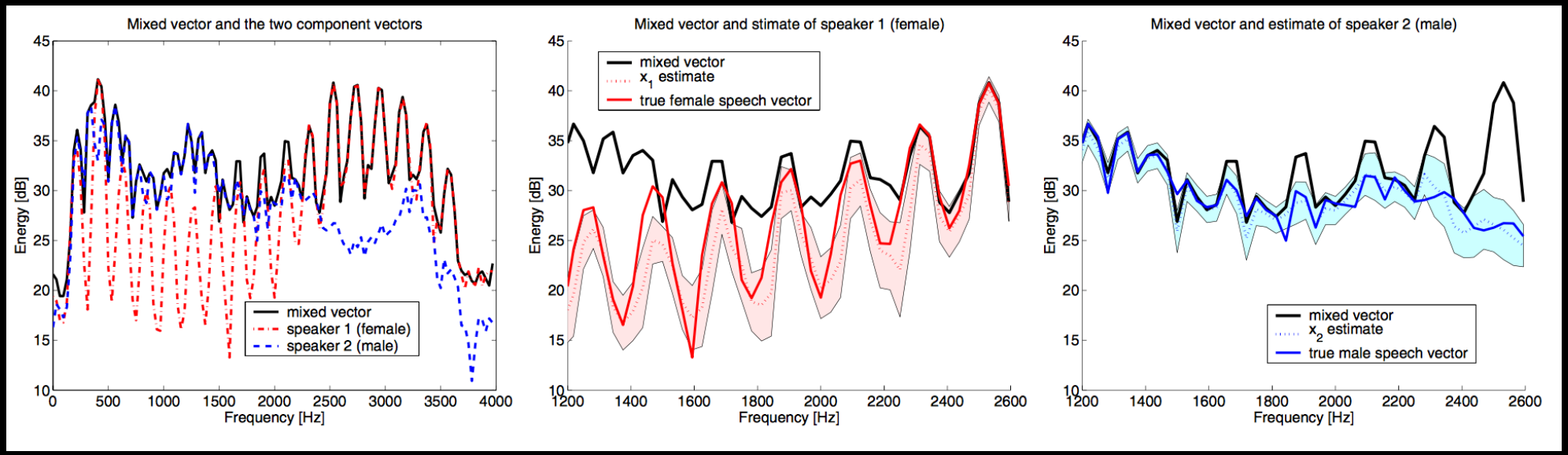
Ellis 1996

Hu & Wang 2004

2004: Model-Based Separation

Roweis '01
Kristjansson, Attias, Hershey '04

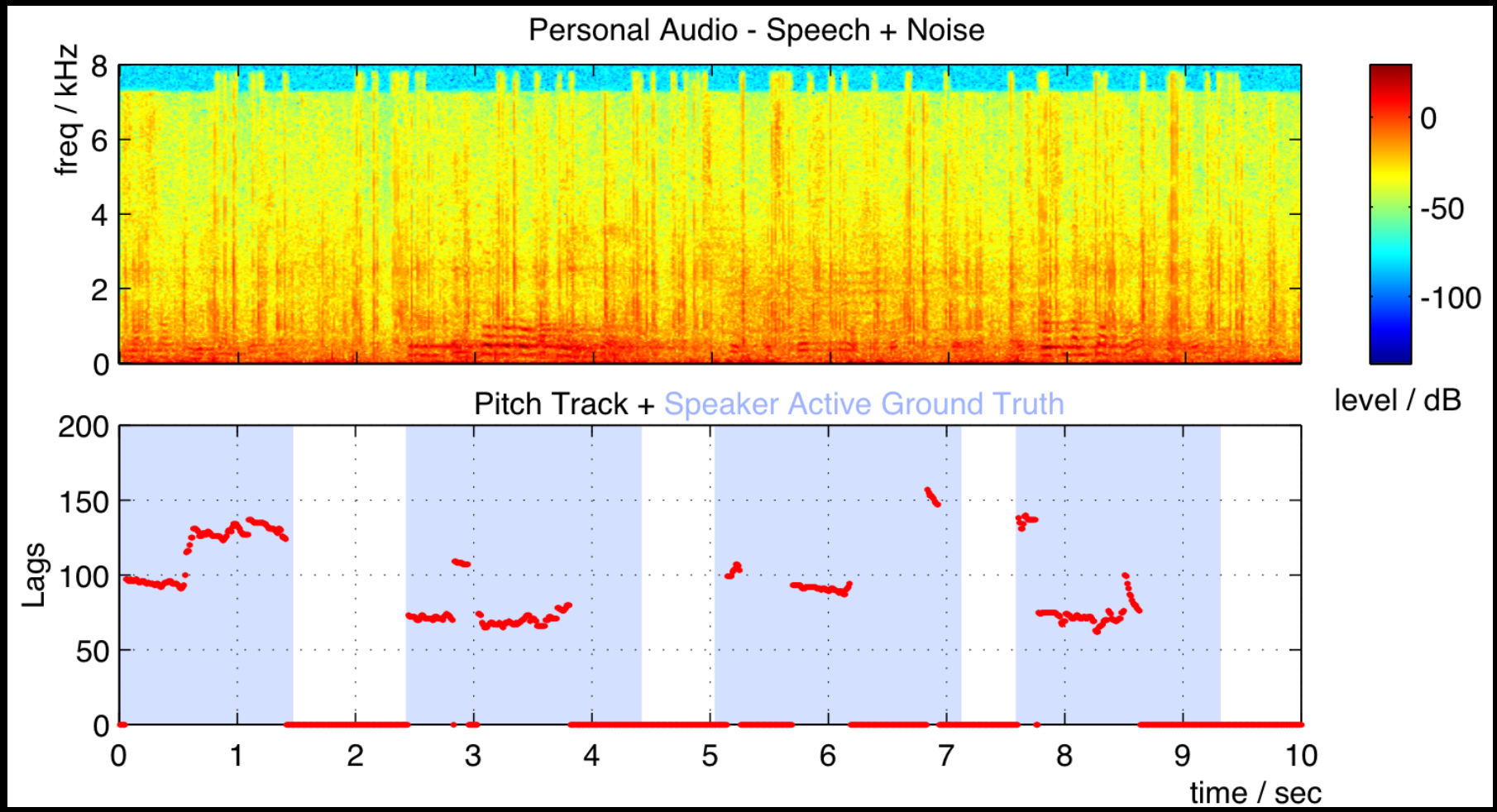
- **Data-driven separation**
  - Learn codebooks for individual speakers
  - Find best combination of sources
- **Pitch gives the "grist"**



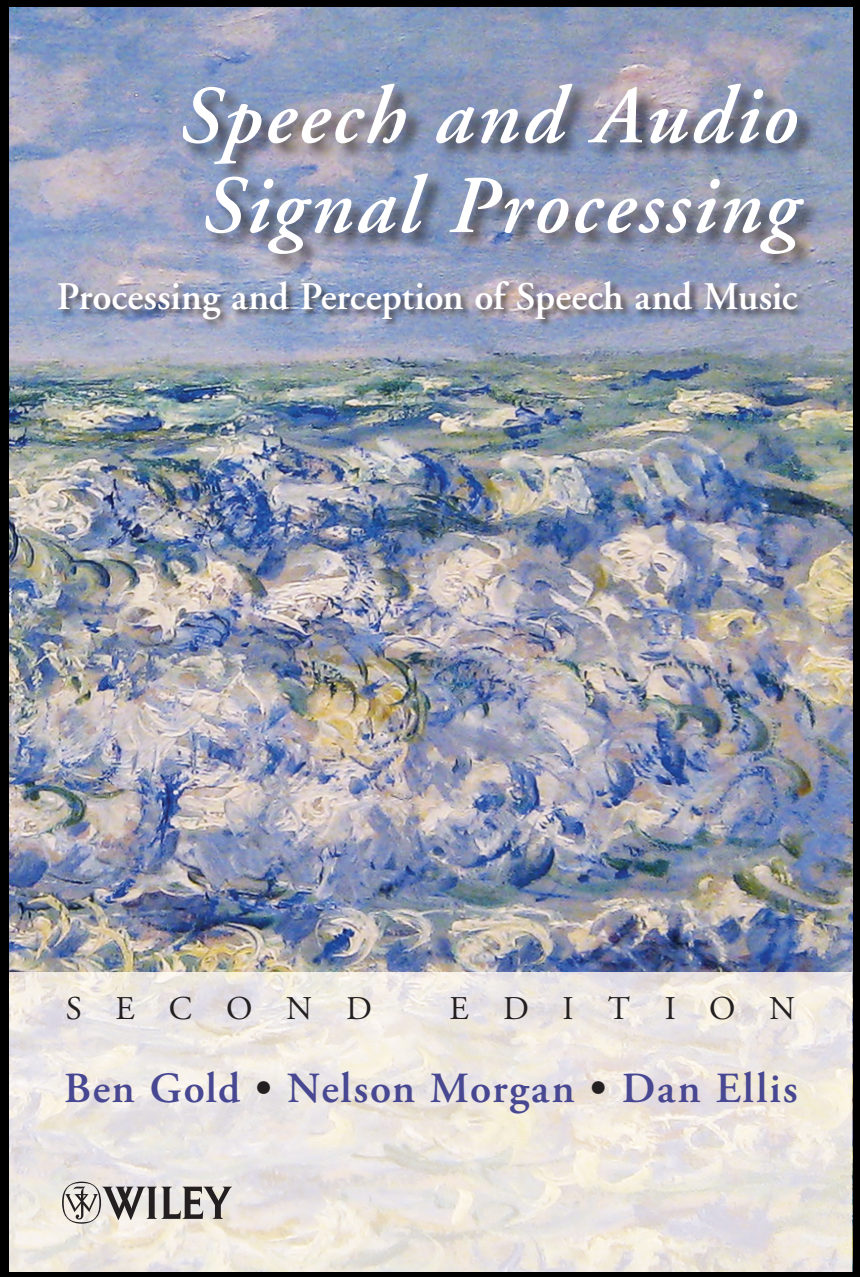
2006: Pitch for VAD

Lee & Ellis'06

- Pitch is the most robust perceptual cue to speech

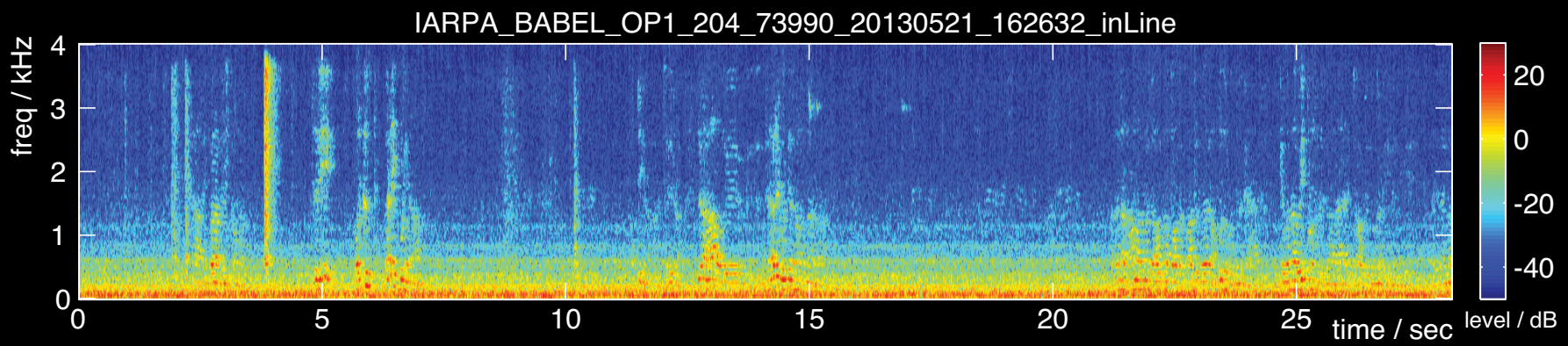


2004-2011: The Epic



2012: Project Babel

- Noisy speech is a challenge:

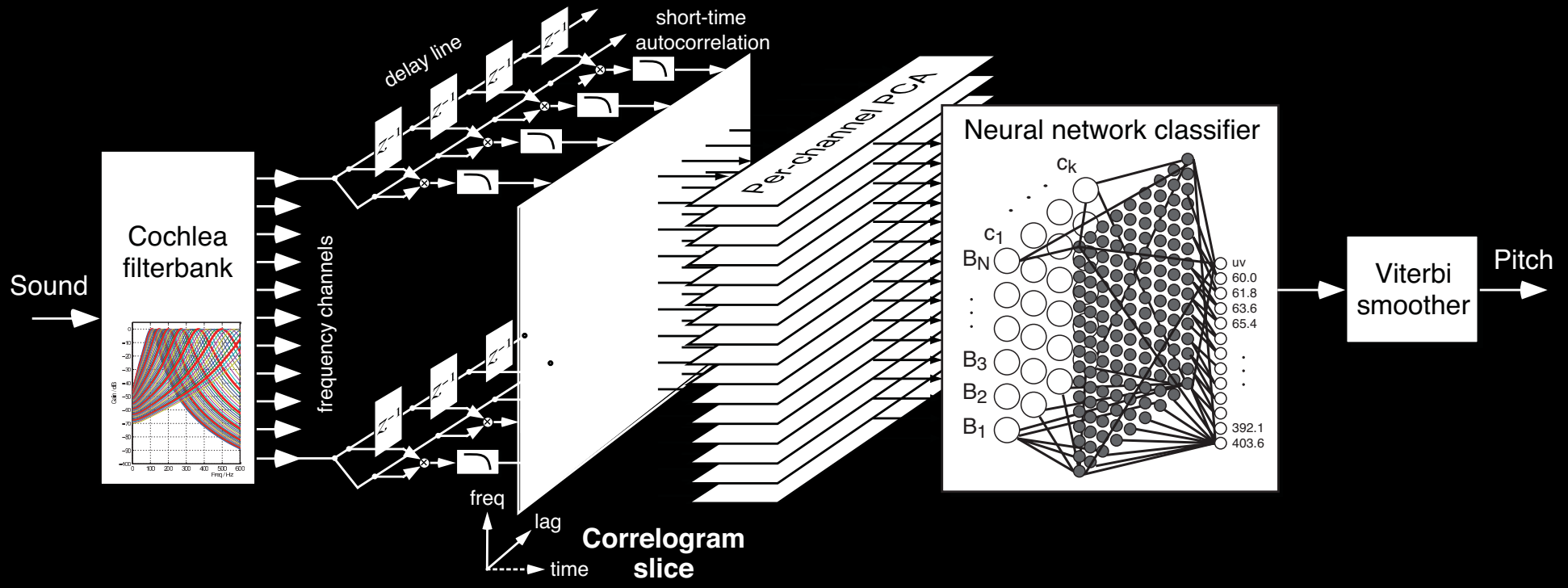


- How to disentangle speech and interference?
  - Energy peaks are speech (spectral subtraction)
  - Energy troughs are noise (Wiener, log-mmse)
  - Speech has a known form (Factorial HMM)
  - Voiced speech is periodic (Pitch-based)

Classification-based Pitch Tracker

Lee & Ellis '12

- Subband Autocorrelation Classification (SAcC) Pitch Tracker:
  - Trained on noisy speech with true pitch targets

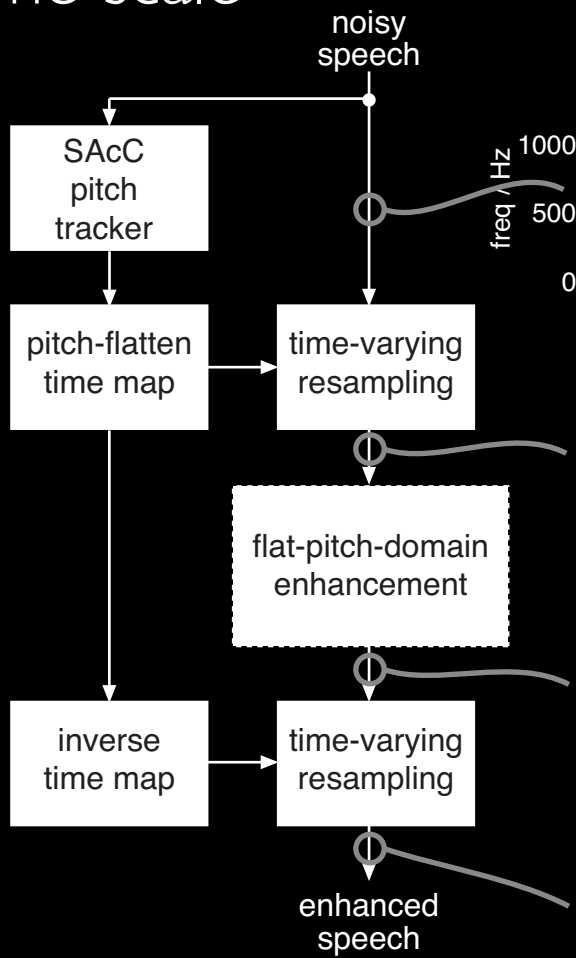


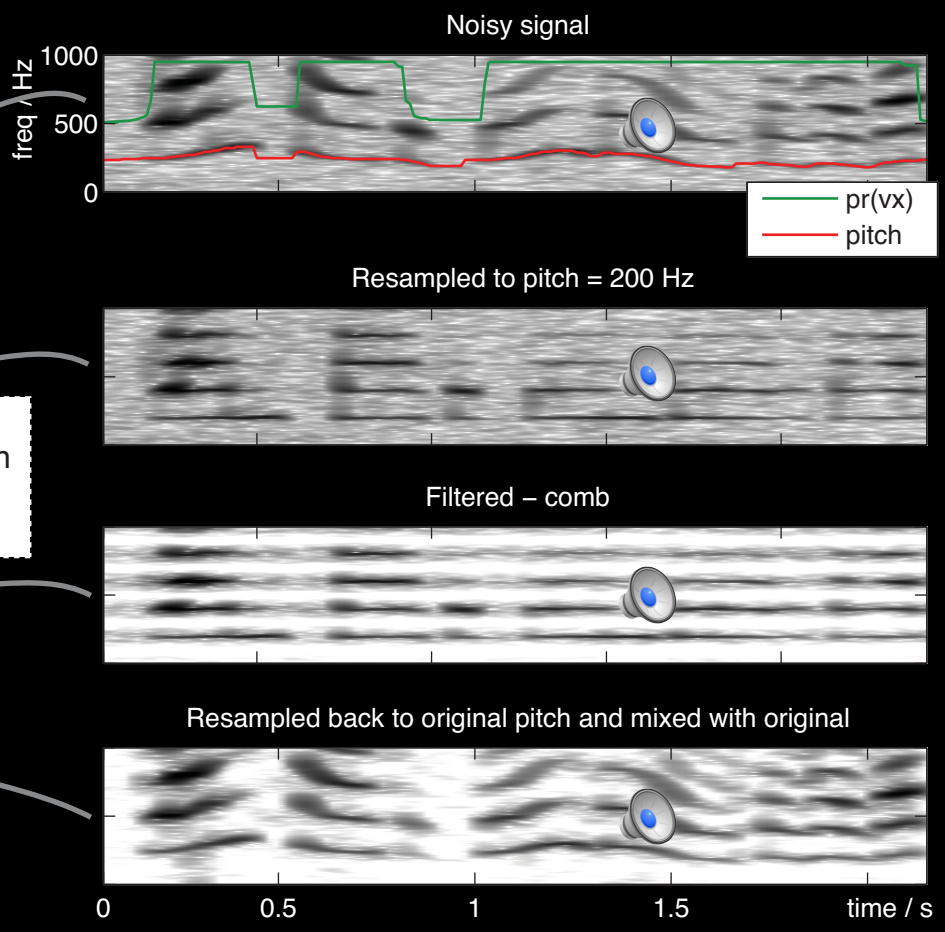
- Subband autocorrelation features
- PCA to reduce dimensions

Flat-Pitch Processing

- Time-varying filtering is tricky
  - if pitch variation and filter impulse response are on a similar time-scale

- Solution: Flatten the pitch
  - use local pitch estimate to resample
  - process constant-pitch
  - resampling is (near) invertible





Conclusions

- Pitch is a key feature for speech separation
  - "marking" signal against other speech or noise
- Some ideas don't go away
  - .. though they can change shape
- Impact from collaboration
  - you can't do good work with someone without the human connection