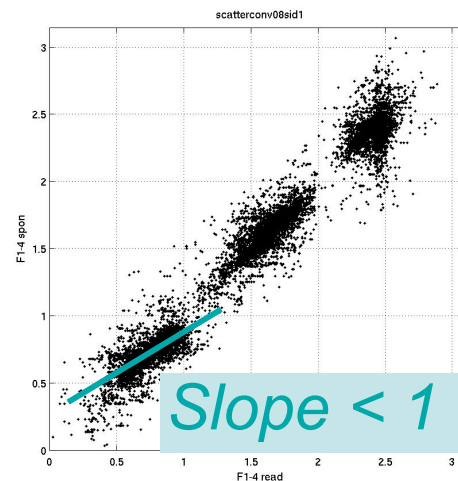
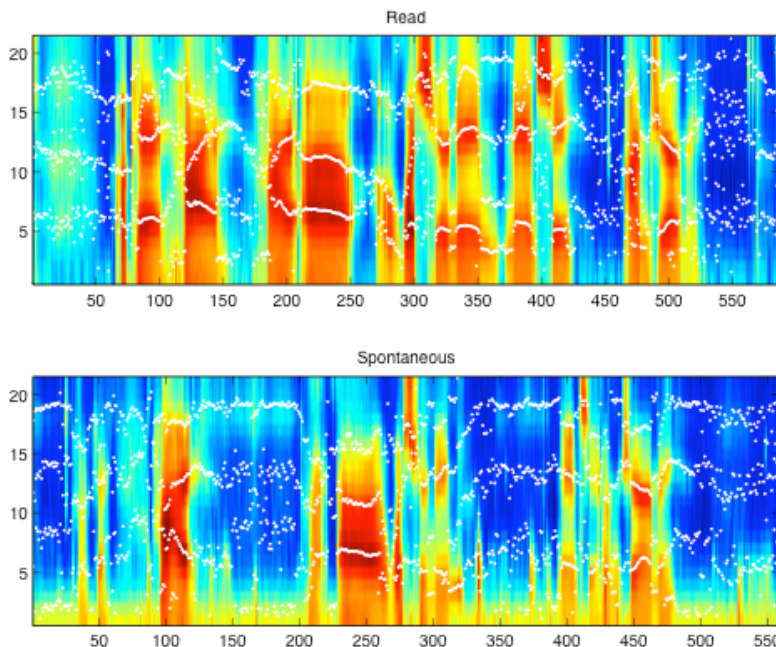


Transforming Spontaneous to Read Speech

- **Spontaneous** speaking style is a major problem in CTS
 - WER is much lower for read speech (e.g. News)
- What is the difference? Can we **transform** spont to read?
- Align and compare matched utterances from **SRI 'multistyle'**
 - speakers re-read transcripts of spontaneous discussions

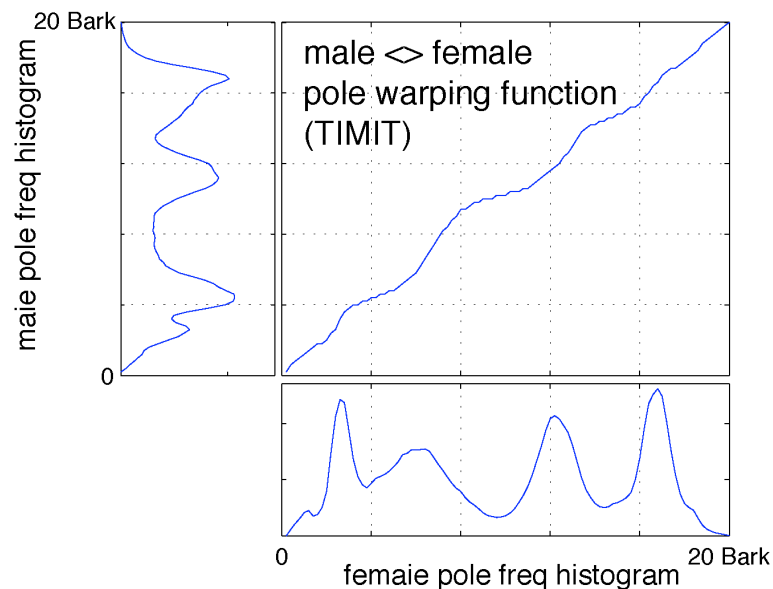


*Scatter between
PLP pole freqs in
aligned read and
spontaneous
utterances*

- **Frequency warping in
PLP pole domain?**

Pole Warping: Generalizing VTLN

- **Between-speaker variation** much greater than read/spont
- Try **speaker norm** via pole freq. histogram normalization
 - .. a single frequency warping
 - .. or modify directly in PLP pole domain?
- More **flexible** than VTLN, more **constraints** than MLLR



- Spectrogram warp, TIMIT:

<i>Condition</i>	<i>Ph.ER%</i>
M train, F test	57.9%
M train, F+warp test	55.9%
M train, F+VTLN test	48.5%

- **Next:** Adapted train, CTS