Binaural Localization Inference

Michael Mandel

12/6/05

Binaural Localization

- Determine directions sounds are coming from
- Hard because of noise, simultaneity
- But only need reliable “glimpses” of single sources, even in dense mixtures
- Useful for
  - pointing directional mics, video cameras
  - classifying auditory scene (number of sources, etc)
  - cuing source separation

System

- Filterbank the two channels
- Break each band into small frames
- Assume one source dominates each frame
- $L_\omega(t) \approx a_{\omega,t} R_\omega(t - \tau_{\omega,t})$
- Compute likelihood of direction $p(\theta_\omega(t) | L_\omega(t), R_\omega(t))$
- Use sequential Monte Carlo to combine estimates over time and across frequencies

Likelihood Model

$$p(\theta|L,R)=\int_{a,\tau} p(\theta|a,\tau,L,R)\,p(a,\tau|L,R)\,da\,d\tau$$

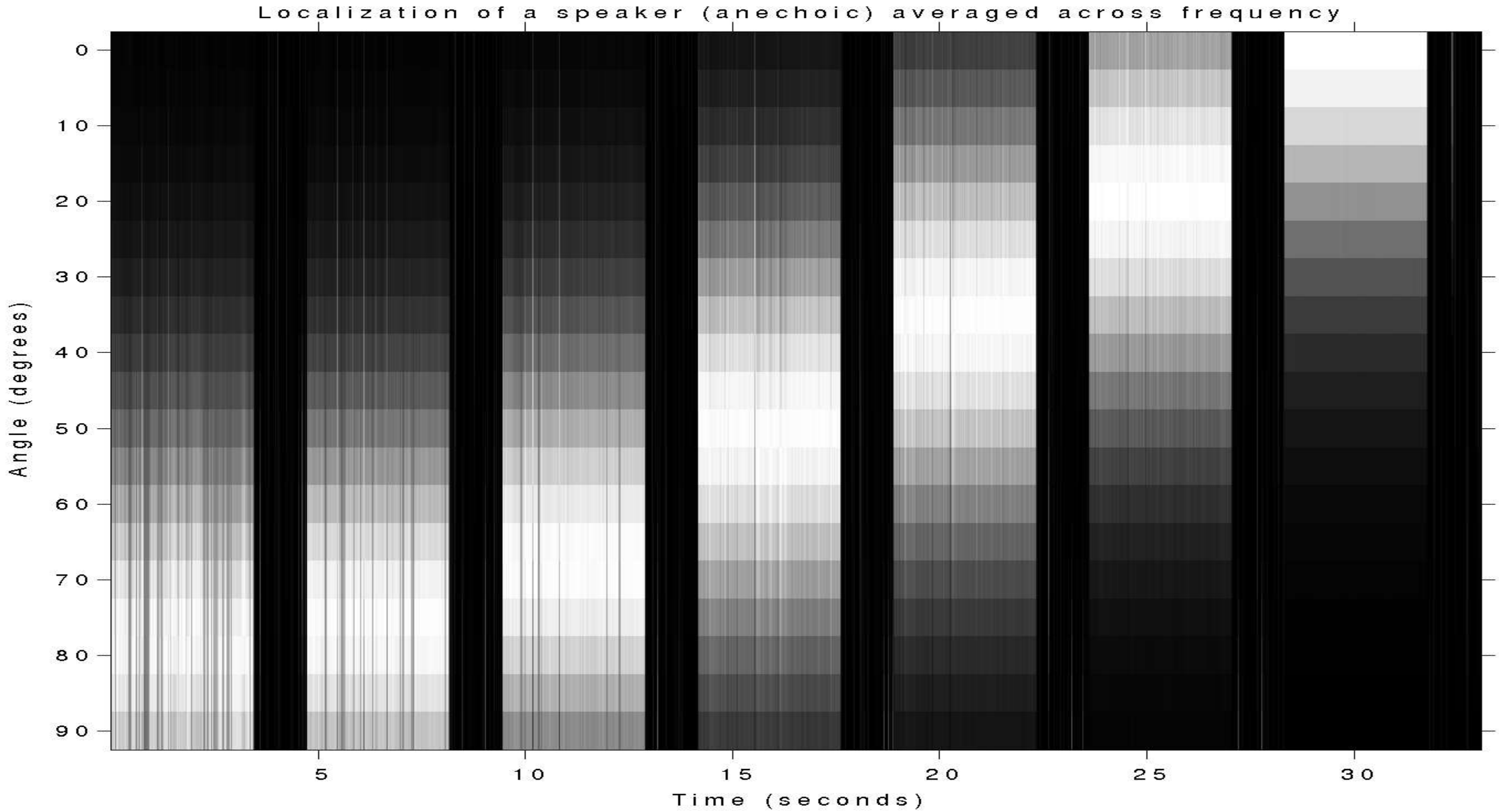
$$=\int_{a,\tau} p(\theta|a,\tau)\,p(a,\tau|L,R)\,da\,d\tau$$

$$=E_{a,\tau|L,R}[p(\theta|a,\tau)]$$

$$p(\theta|L,R)\approx\frac{1}{N}\sum_{a_i,\tau_i} p(\theta|a_i,\tau_i) \qquad (a_i,\tau_i)\sim p(a,\tau|L,R)$$

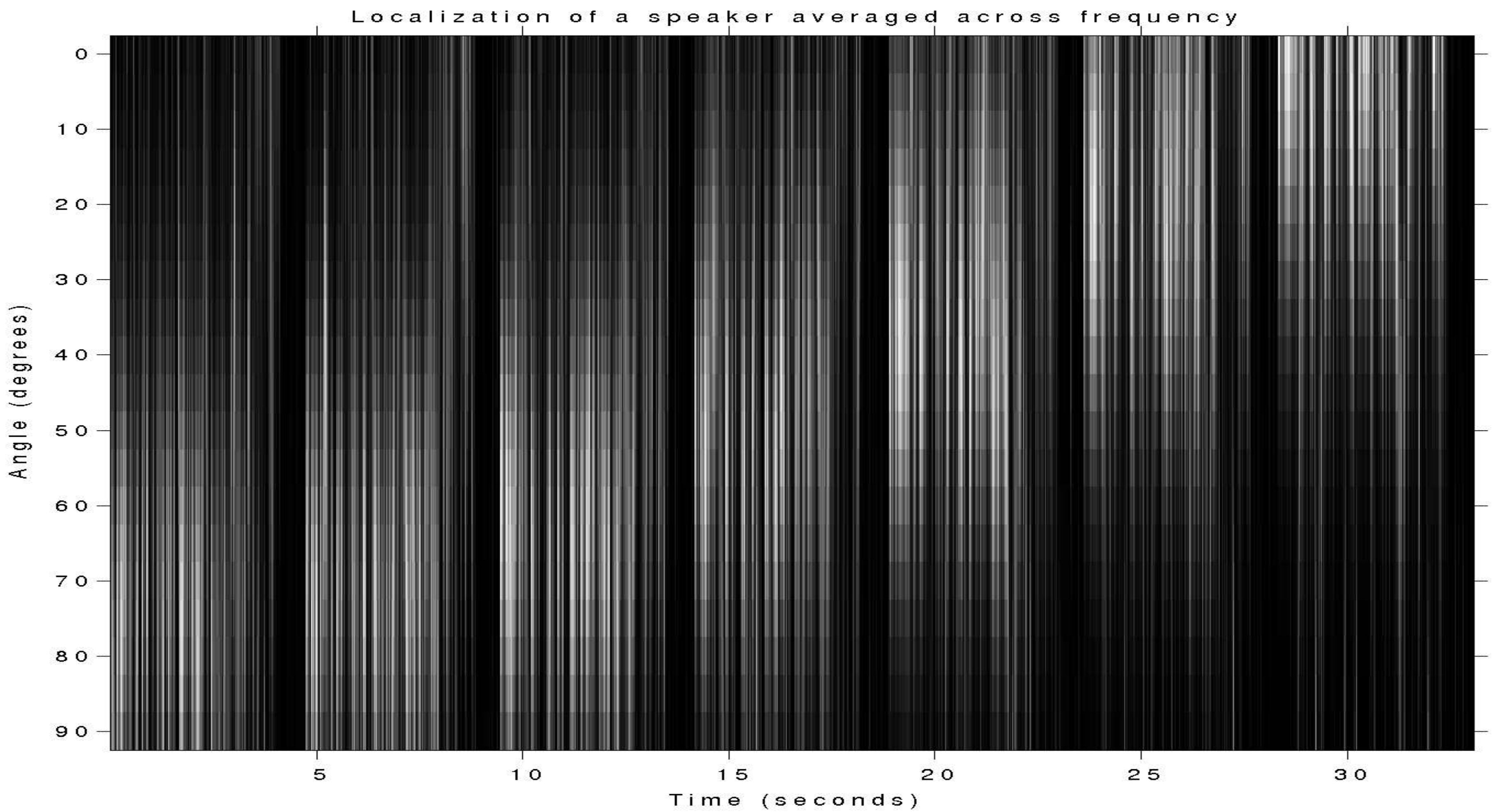
$$p(\theta|L,R)\approx p(\theta|a_{ml}\tau_{ml})$$

Results

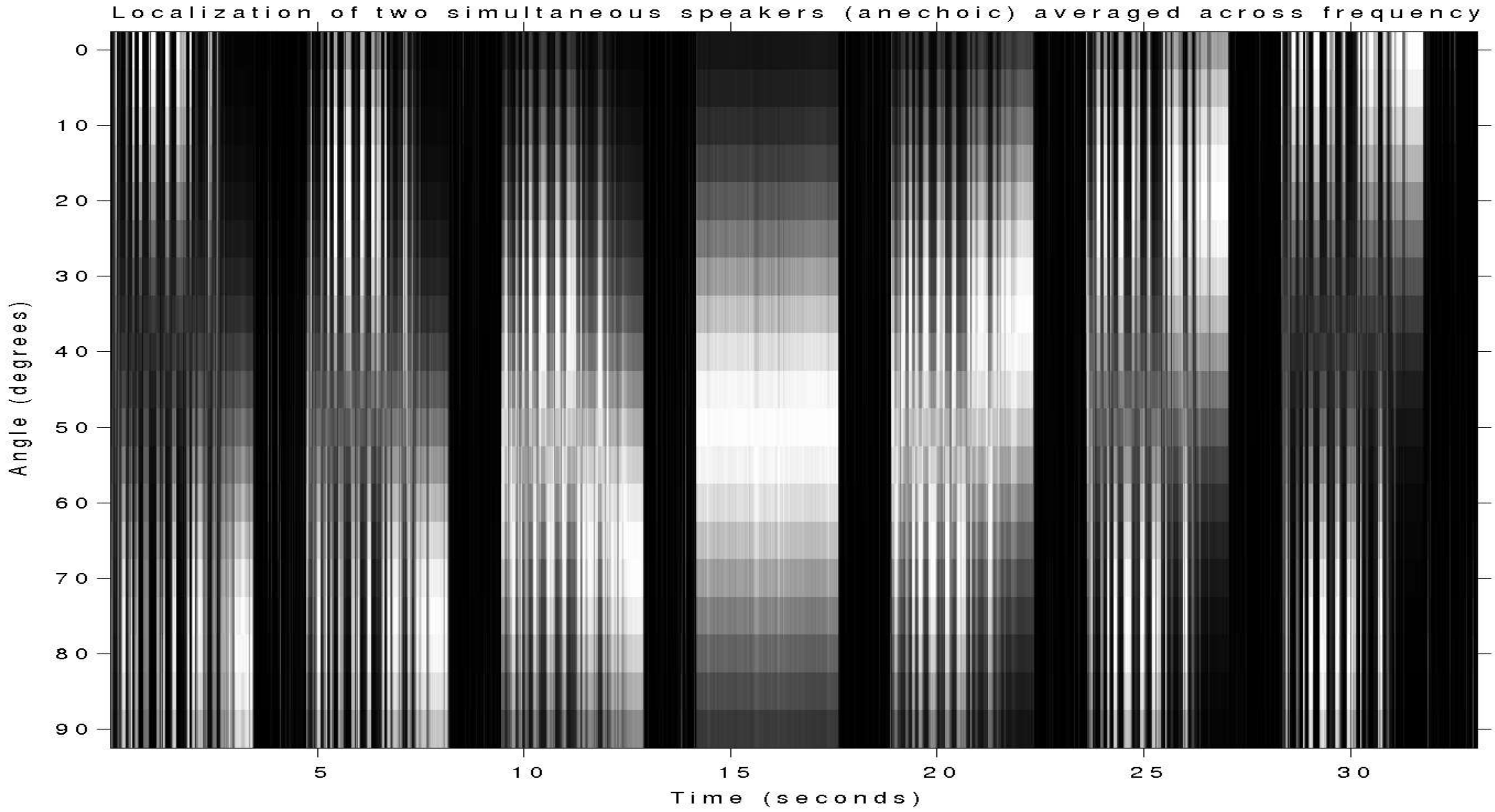


Localization of a speaker (anechoic) averaged across frequency
Angle (degrees)
0
10
20
30
40
50
60
70
80
90
5
10
15
20
25
30
Time (seconds)

Results

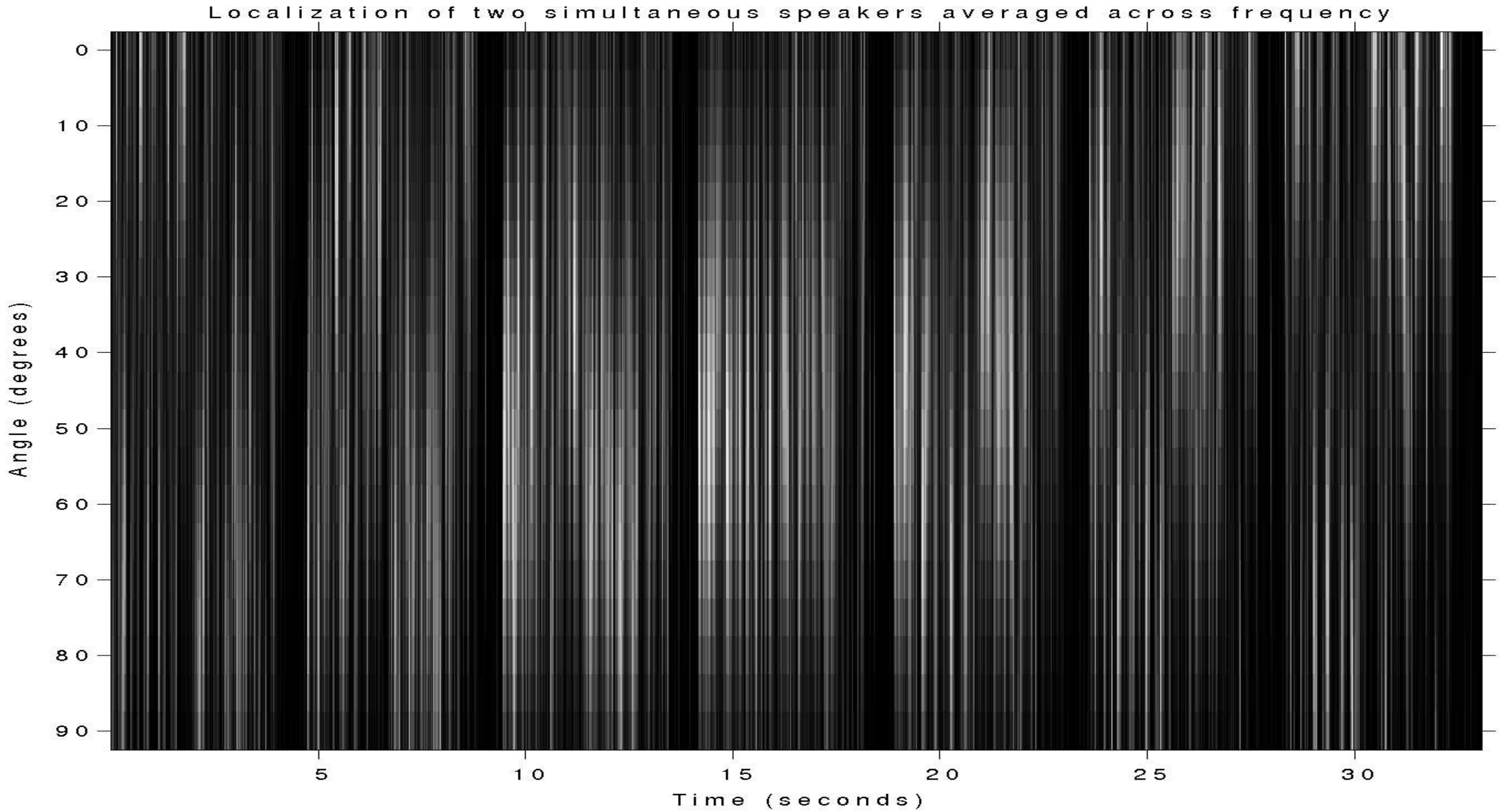

Localization of a speaker averaged across frequency
Angle (degrees)
0
10
20
30
40
50
60
70
80
90
5
10
15
20
25
30
Time (seconds)

Results



Localization of two simultaneous speakers (anechoic) averaged across frequency
0
10
20
30
40
50
60
70
80
90
Angle (degrees)
5
10
15
20
25
30
Time (seconds)

Results



Localization of two simultaneous speakers averaged across frequency
Angle (degrees)
0
10
20
30
40
50
60
70
80
90
5
10
15
20
25
30
Time (seconds)

Thank you

Questions?

Model

- a lognormal: $p(\ln a | L, R) = N(\mu, \rho)$
- tau has pseudo-likelihood from cross-correlation: $p(\tau | L, R) \propto \exp(b\, phat(\tau))$
- $phat(\tau) = \int_{2\pi} \frac{1}{|L(\omega)\bar{R}(\omega)|} L(\omega)\bar{R}(\omega) e^{j\omega\tau} d\omega$
- $p(\theta | a, \tau)$ measured empirically from ground truth, anechoic data for each frequency band