A Probability Model for Interaural Phase Difference

Michael Mandel Dan Ellis
mim@ee.columbia.edu

Columbia University Department of Electrical Engineering

2006 Statistical and Perceptual Audio Workshop

Outline

Introduction

Probability model
- Observations

Experiments
- Results

Summary

Laboratory for the Recognition and Organization of Speech and Audio

Outline

Introduction

Probability model

Observations

Experiments

Results

Summary

Laboratory for the Recognition and Organization of Speech and Audio

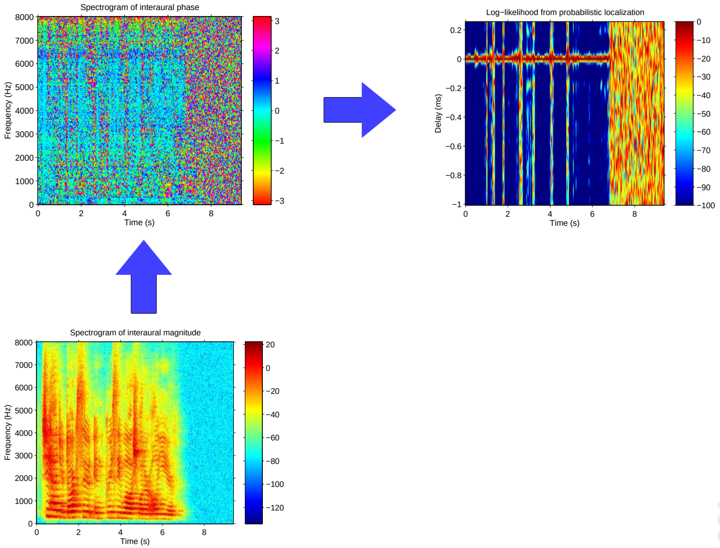
Vision

- Computer programs that can hear as well as people can
- In a room with two mics and multiple people talking
  - Locate the sound sources in space
  - Characterize the room
  - Determine what is said

Overview

- Problem: localization, time delay of arrival estimation
  - Stereo recording
  - Single source
  - Reverberation or additive noise
- Solution: probability model for interaural phase difference in reverberation
  - Considers time-frequency points individually, IID
  - No assumptions about source characteristics
- Performance: estimates delay more accurately than generalized cross-correlation in simulations

Probabilistic localization



Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

Previous work

- Knapp and Carter (1976)
  - Generalized cross correlation and PHAse Transform
  - Maximum likelihood estimator for stationary sources in uncorrelated noise
- Stern and Trahiotis (1995)
  - Based more on human perception
  - Cross-correlations in many frequency bands
  - “Straightness” measure for combining bands
- Rennie (2005)
  - EM algorithm for localizing multiple sources based on whole STFT frames

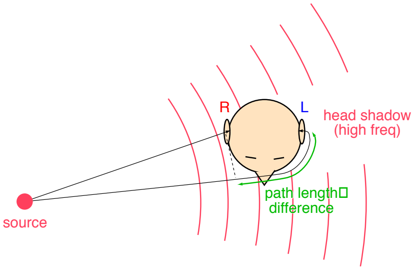
Outline

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

Data

- Speech from TIMIT database
  - Phonetically balanced sentences from 630 native English speakers
  - "She had your dark suit in greasy wash water..."
- Binaural room impulse responses from Shinn-Cunningham et al. (2005)
  - Recorded with KEMAR manikin in classroom
  - 4 locations, 7 directions, 3 distances, 3 repetitions
  - "She had your dark suit in greasy wash water..."
- Binaural anechoic impulse responses from Algazi et al. (2001)
  - Recorded with KEMAR manikin in anechoic chamber
  - 25 azimuths, 50 elevations
  - Speech-shaped noise added at 10 dB, direct to reverb ratio
  - "She had your dark suit in greasy wash water..."

Listening model

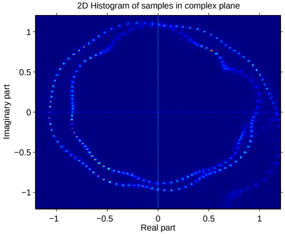


$$\ell(t) = a_\ell s(t - \tau_\ell) * n_\ell(t)$$
$$r(t) = a_r s(t - \tau_r) * n_r(t)$$

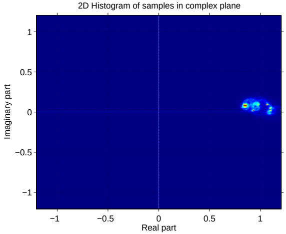
$$N(j\omega) = e^{-a + j\omega\tau} \frac{L(j\omega)}{R(j\omega)}$$

- Delay between channels because of path difference
- Gain between channels because of head shadowing
- Convolutive noise to model reverberation
- Noise assumed to be compact

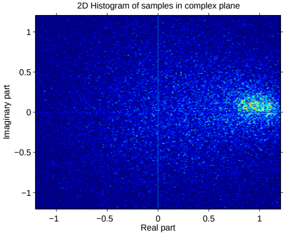
Noise distributions in the complex plane



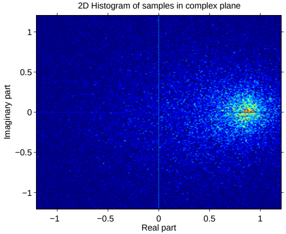
Incorrect τ



Anechoic



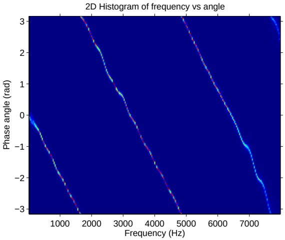
Additive



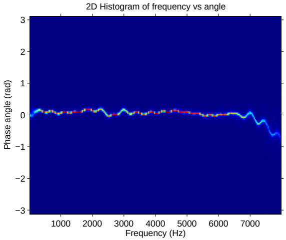
Reverberant

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

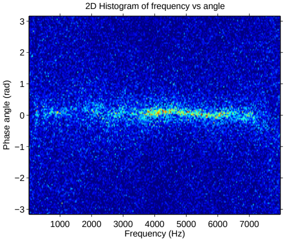
Phase of noise vs frequency



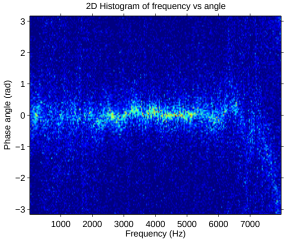
Incorrect τ



Anechoic

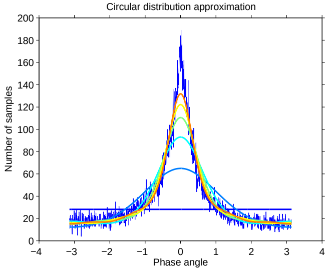


Additive



Reverberant

Probability model



$$p(\Delta\phi \,|\, \omega\tau) = \exp\left(\sum_k a_k \cos^k(\Delta\phi - \omega\tau)\right)$$

- Circular maximum entropy probability distribution
- Fourier series gives desired moments

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

Outline

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

Experiments

- Compare PHAT to probabilistic method
- Compare sensitivity of probabilistic method to learned noise distribution
- Test in simulated reverberation and additive noise, at varying azimuths
- Pool probabilistic estimates over frequency to compare to PHAT
- Measure average error in delay estimate as a function of pooling over time

PHAse Transform (PHAT)

Classical Time-Delay of Arrival (TDOA) estimator

Cross-correlation:

$$\hat{\tau} = \arg\max_{\tau} \ell(t) * r(-t)$$

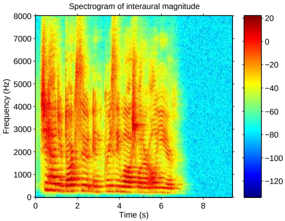
Generalized cross-correlation:

$$\hat{\tau} = \arg\max_{\tau} \sum_{\omega} \psi(j\omega) L(j\omega) R^*(j\omega) e^{j\omega\tau}$$

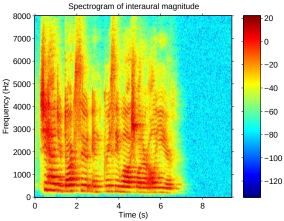
e.g. PHAse Transform:

$$\hat{\tau} = \arg\max_{\tau} \sum_{\omega} \frac{1}{|L(j\omega)||R(j\omega)|} L(j\omega) R^*(j\omega) e^{j\omega\tau}$$

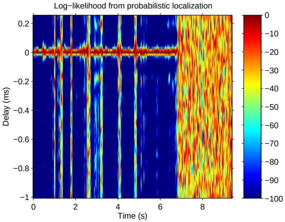
Example localizations



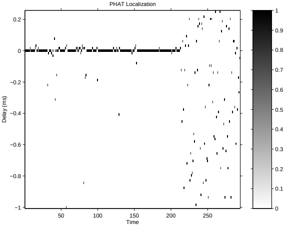
Speech signal



Speech signal



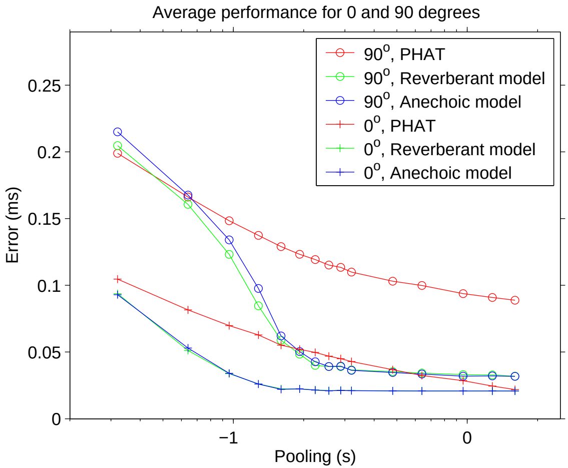
Probabilistic localization



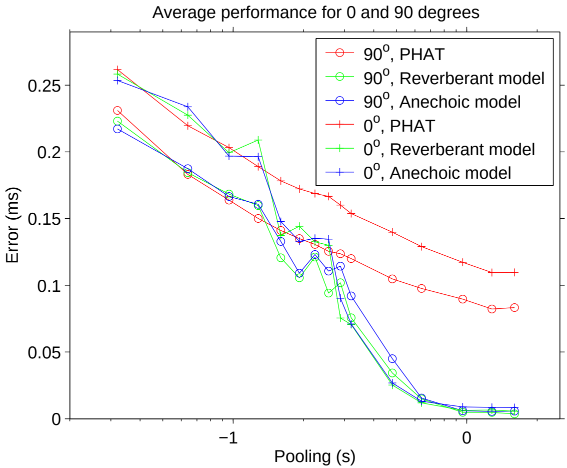
PHAT localization

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

Results in reverberation



Results in additive noise



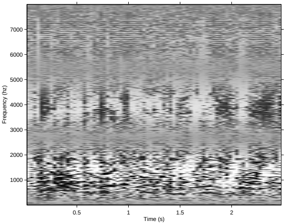
Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

Outline

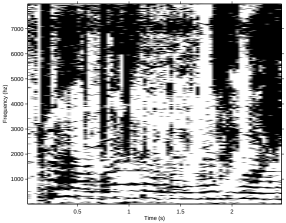
Summary

- Probabilistic model of phase at individual interaural spectrogram points
- Estimates true delays better than PHAT in reverberation
- Estimates true delays as well as, if not better than, PHAT in additive noise
- Makes no assumptions about the sources involved
- Allows powerful analyses, e.g. likelihoods of delays over arbitrary regions of spectrogram
- Lends itself to localizing multiple sources

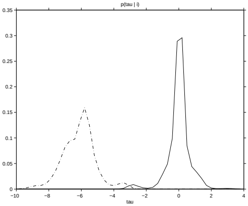
EM Algorithm



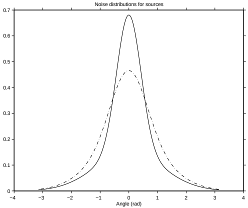
EM separation



Ground truth



$p(\tau \mid i)$



$p(\Delta\phi - \omega\tau \mid i)$

Thank you

Any questions?

Laboratory for the Recognition and Organization of Speech and Audio