SOUND LOCALIZATION AND RECOGNITION

Fahim Zaman
7224528

Agenda

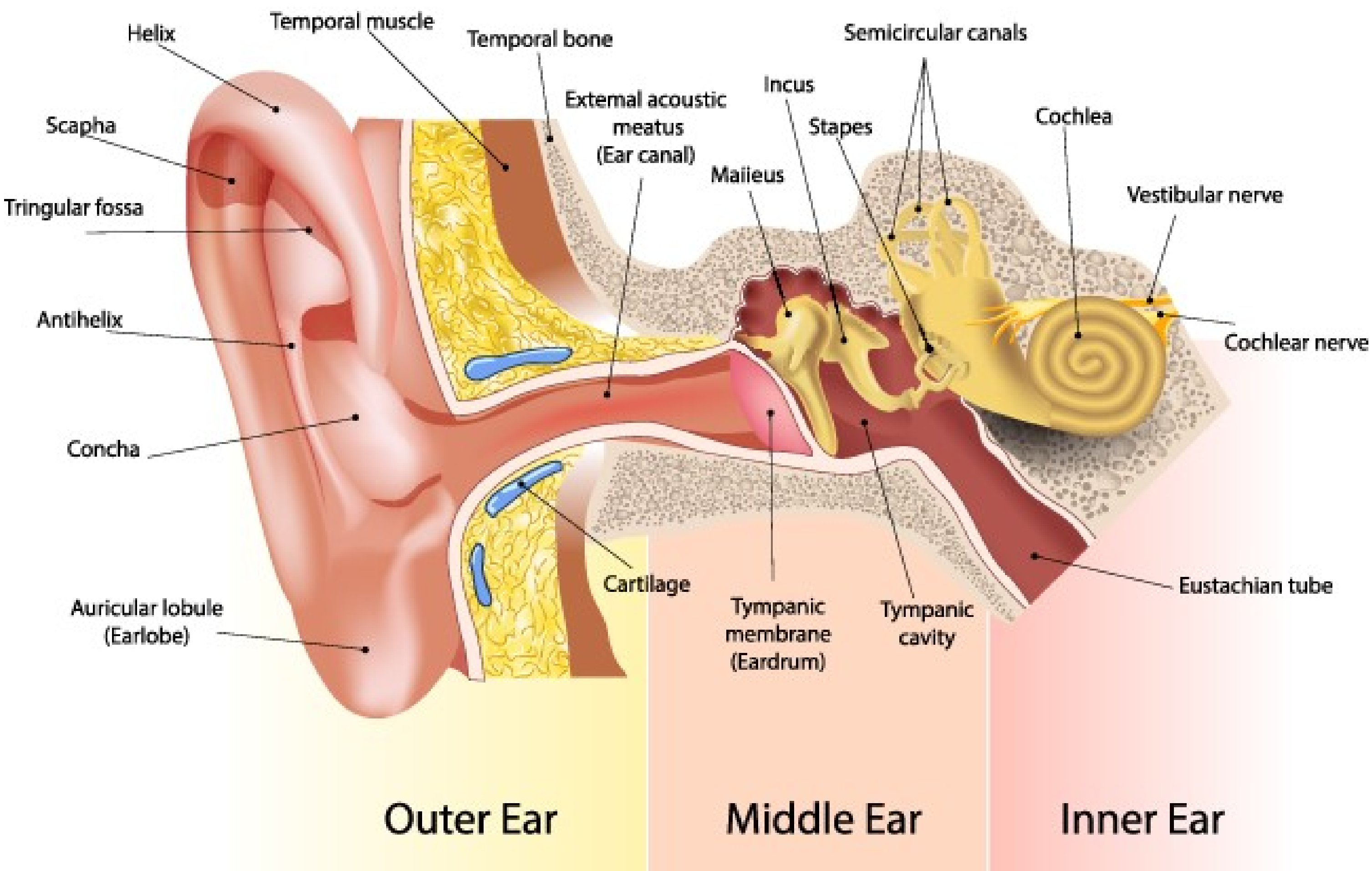
- Sound localization in humans
- Binaural methods
- Array methods
- Humanoid robot approach

Sound Localization in Humans

Sound perception in humans

- The outer ears collect sound and amplify them
- The inner ears use the changes in air pressure to convert them into neural signals
- Fluid elements in the ear used to detect frequency

Anatomy of the Ear



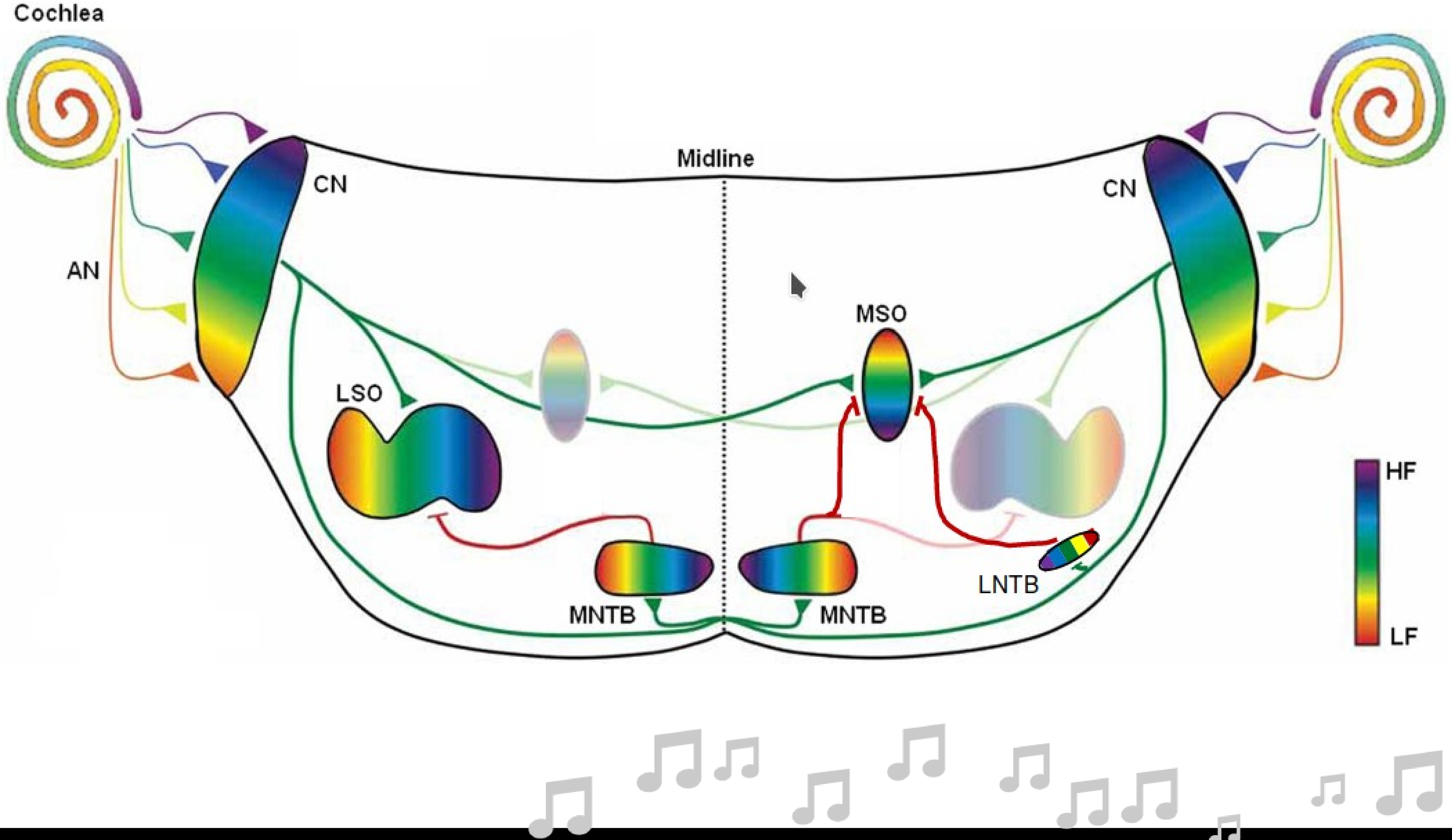
https://static1.squarespace.com/static/5b24224b297114f42b946c38/t/5b3fd94df950b709fea38d18/1530911058308/EarAnatomy.jpg

Medial Superior Olive

- Receives input from both ears AVCN’s(anteroventral cochlear nucleus)
- Calculates low frequency sounds
- Interaural time difference

Lateral Superior Olive

- Receives input from both ears AVCN’s(anteroventral cochlear nucleus)
- Calculates high frequency sounds
- Interaural level difference


Cochlea
Midline
CN
CN
AN
MSO
LSO
LNTB
MNTB
MNTB
HF
LF

Sound Localization

- Monaural
- Vertical processing
- Binaural
- Processed in the MNTB(Medial nucleus of the trapezoid body)
- Horizontal processing

Interaural Time Difference

- Differences in travel time from a sound source to closer and farther ear
- Mathematically works well for frequencies >=2kHz
- In humans highest sensitivity between 0.7 and 1 kHz, insensitive after ~1.4 kHz

Interaural Level Difference

- For smaller wavelengths (higher frequencies)
- Physical presence of the head makes a shadow
- Louder in one ear than the other

Sound Localization in Robotics

Potential Constraints

- Geometric
- Embeddability
- Real time
- Broadband
- Environmental

Approach

- Frequency decompositions modelled by FFT, bandpass filters or gammatone filters
- Binaural processing
- Array processing

Binaural Processing

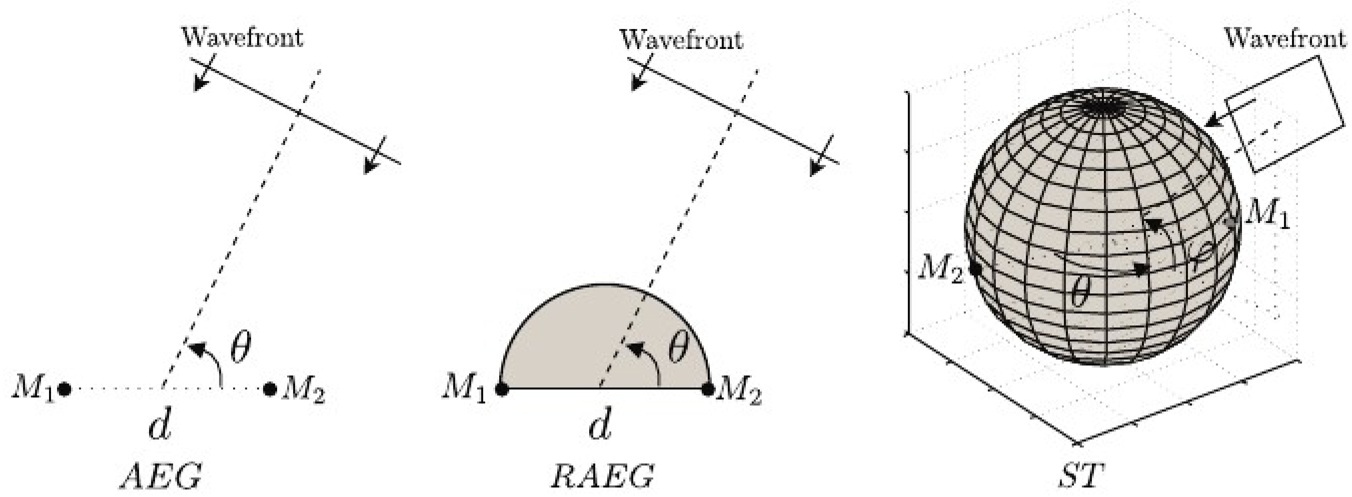
- Simpler computations
- Simpler geometry
- Built in irregularity

Head Related Transfer Function

The HRTF captures the relationship between the signal originating from a sound source and captured at a certain arbitrary reference position in space

- Auditory Epipolar Geometry
- Revised Auditory Epipolar Geometry
- Scattering Theory

$$\begin{cases} L(f) = H_L(r_s, \theta_s, \varphi_s, f)S(f), \\ R(f) = H_R(r_s, \theta_s, \varphi_s, f)S(f), \end{cases}$$



Adaptive Filters

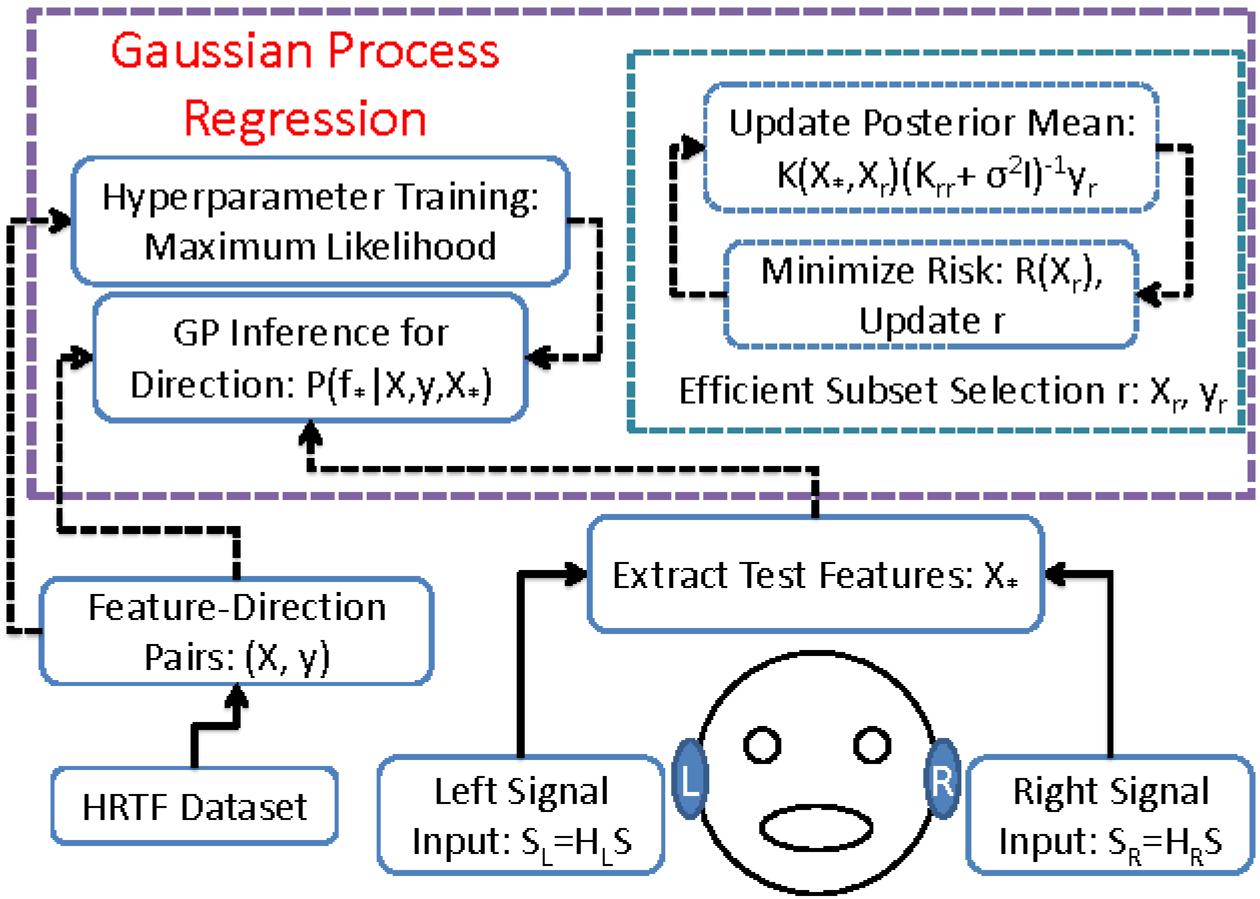
- Robots own noise
- Environmental reflection
- Unexpected environmental elements

Neural Networks

- Learning head effect
- Learn different environments

Gaussian Models

- Auditory events with ITD values in histograms
- Approximated with gaussian models
- Results in model with peaks from sound sources
- Can compute multiple sound sources



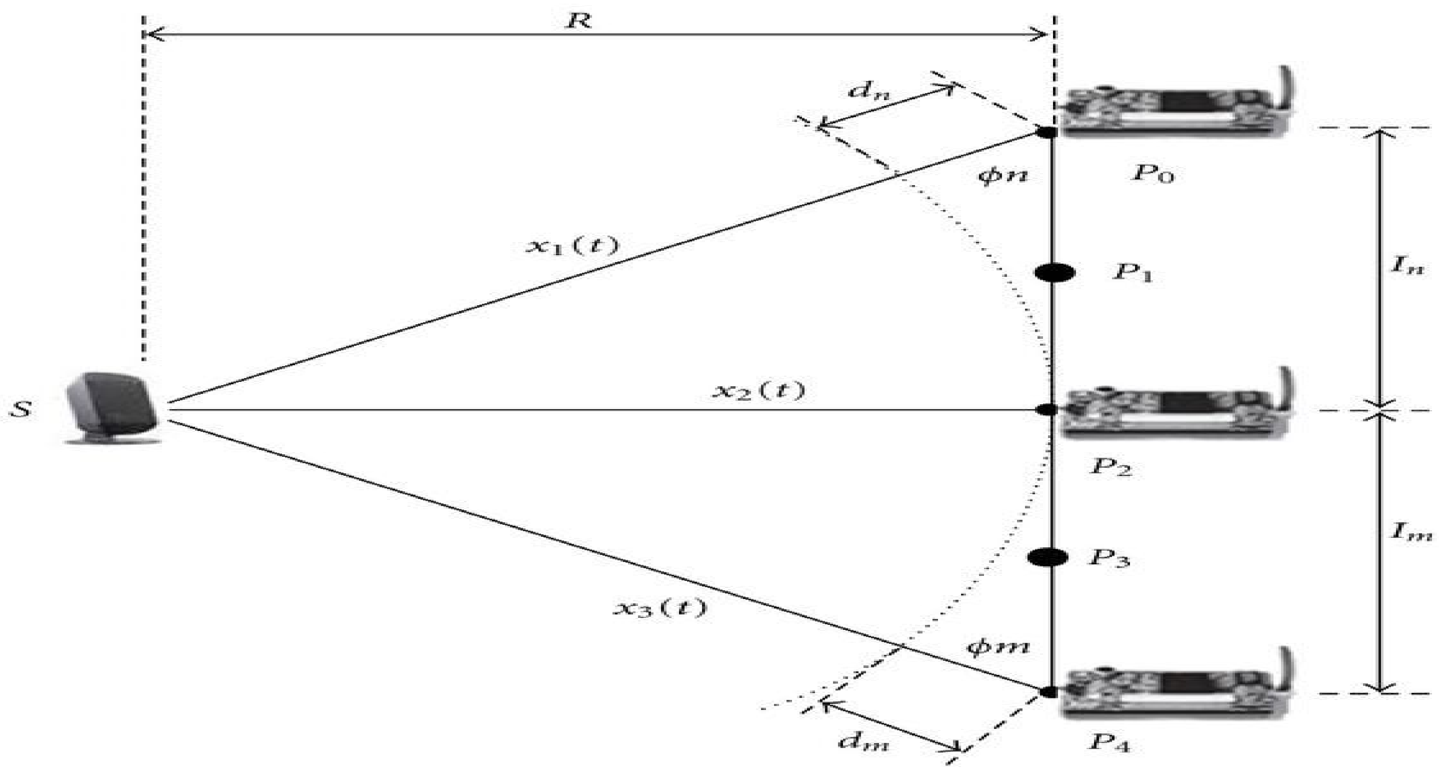
https://www.semanticscholar.org/paper/Gaussian-process-models-for-HRTF-based-3D-sound-Luo-Zotkin/24f37465aae38112c3ad5969d786e47a07b4824d

Spectral Cues

- Elevation dependent on positions of notches in the spectrum of sound signals
- Acoustic reflections from head and outer ear
- Scattering patterns from shape of pinnae

Distance Localization

- Triangular theory
- Theoretically only for static sources
- Also not too accurate in humans



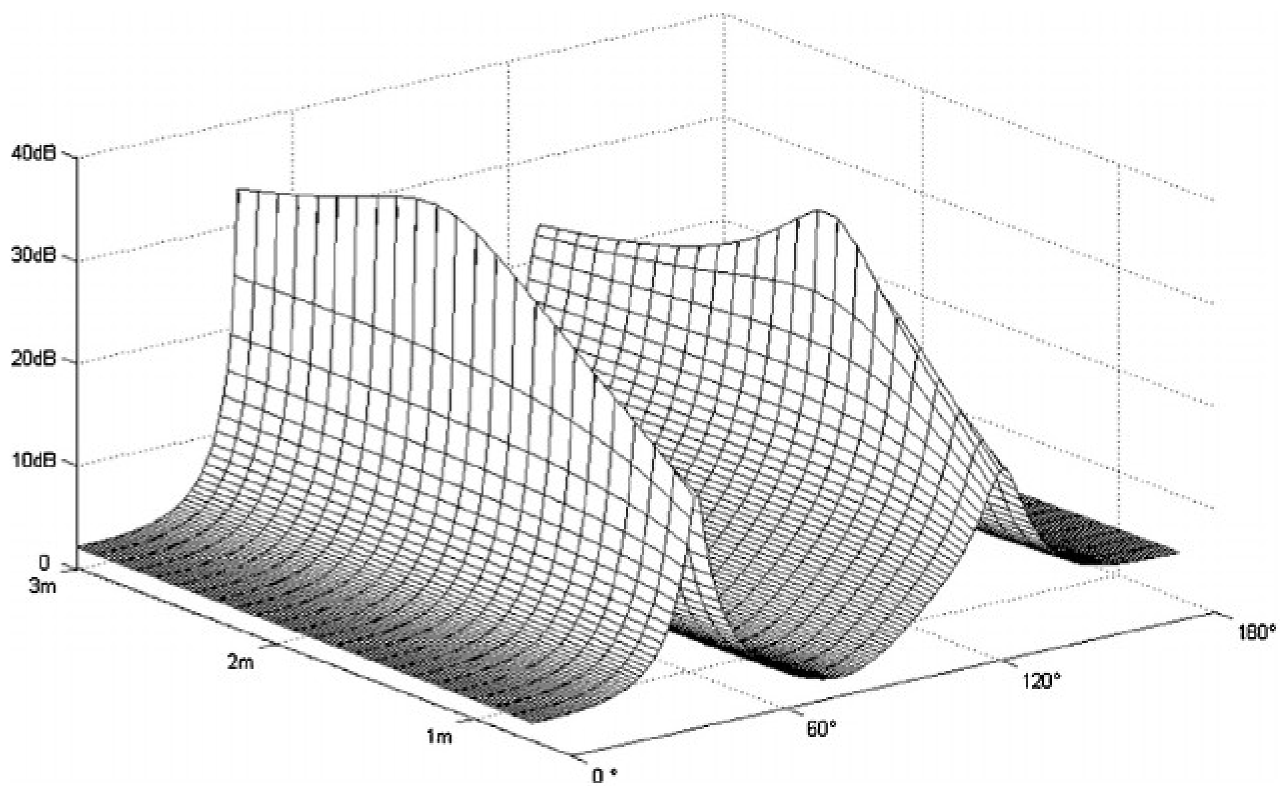
https://www.researchgate.net/figure/Triangular-configuration-of-sensor-nodes-sound-source-for-the-localization-where-Pi_fig5_236006629

Array processing

- Using array of microphones
- Computation heavy
- Mostly not broadband capable

MUSIC

- MUltiple SIgnal Classification
- Assumes sound source is independent, zero mean stationary and of a single frequency
- Can compute multiple frequencies discretely

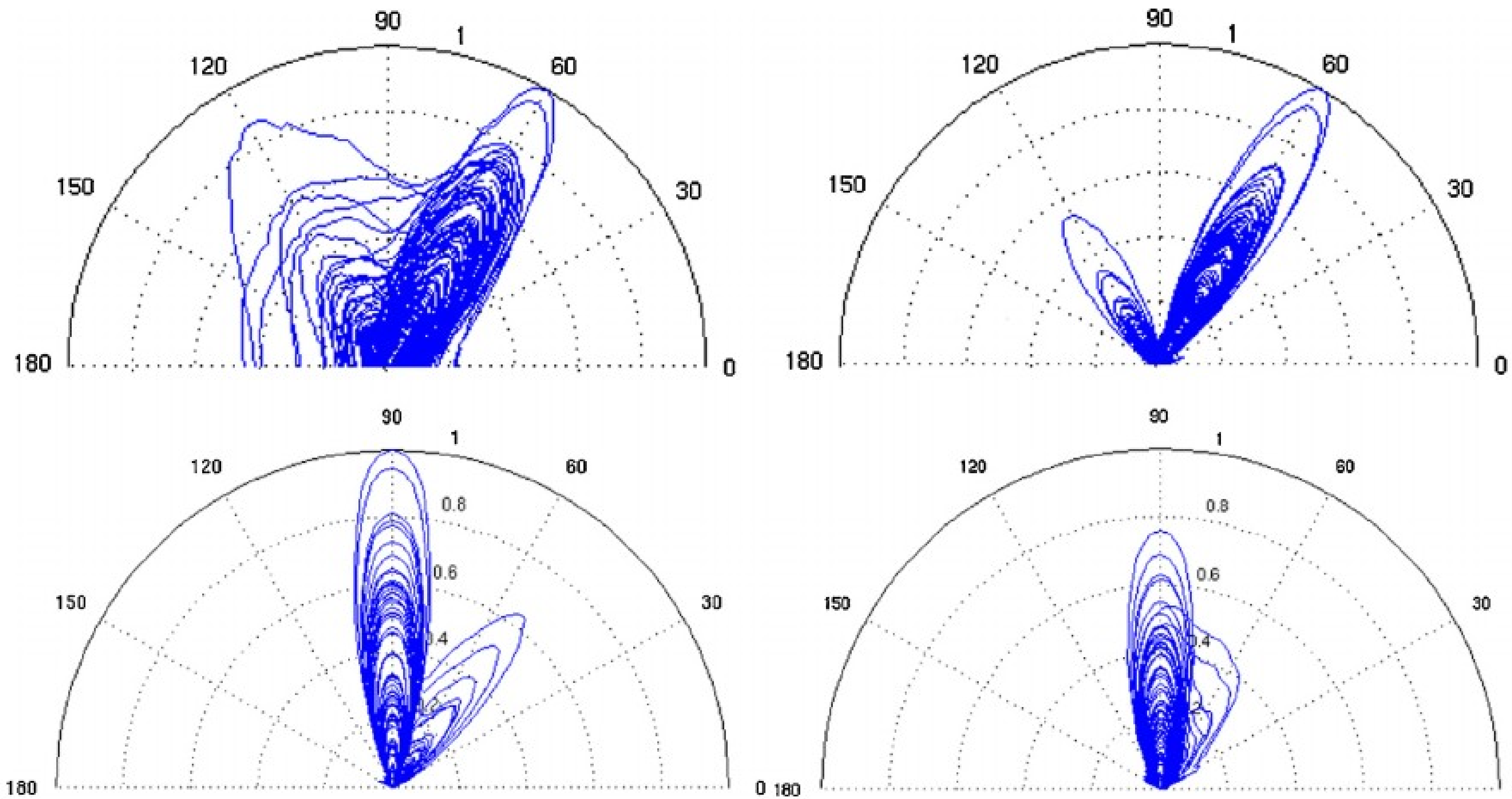

40dB
30dB
20dB
10dB
0
3m
2m
1m
0°
60°
120°
180°

TDOA

- Time Delay(s) of Arrival
- Calculate delay between two microphones of array
- Similar notion to ITD, but different computation

Beamforming

- Most used in robotics
- Low computation cost
- Performance dependent on number microphones
- Computes energy map



90
1
120
60
150
30
180
0
0.8
0.6
0.4

Humanoid Robots

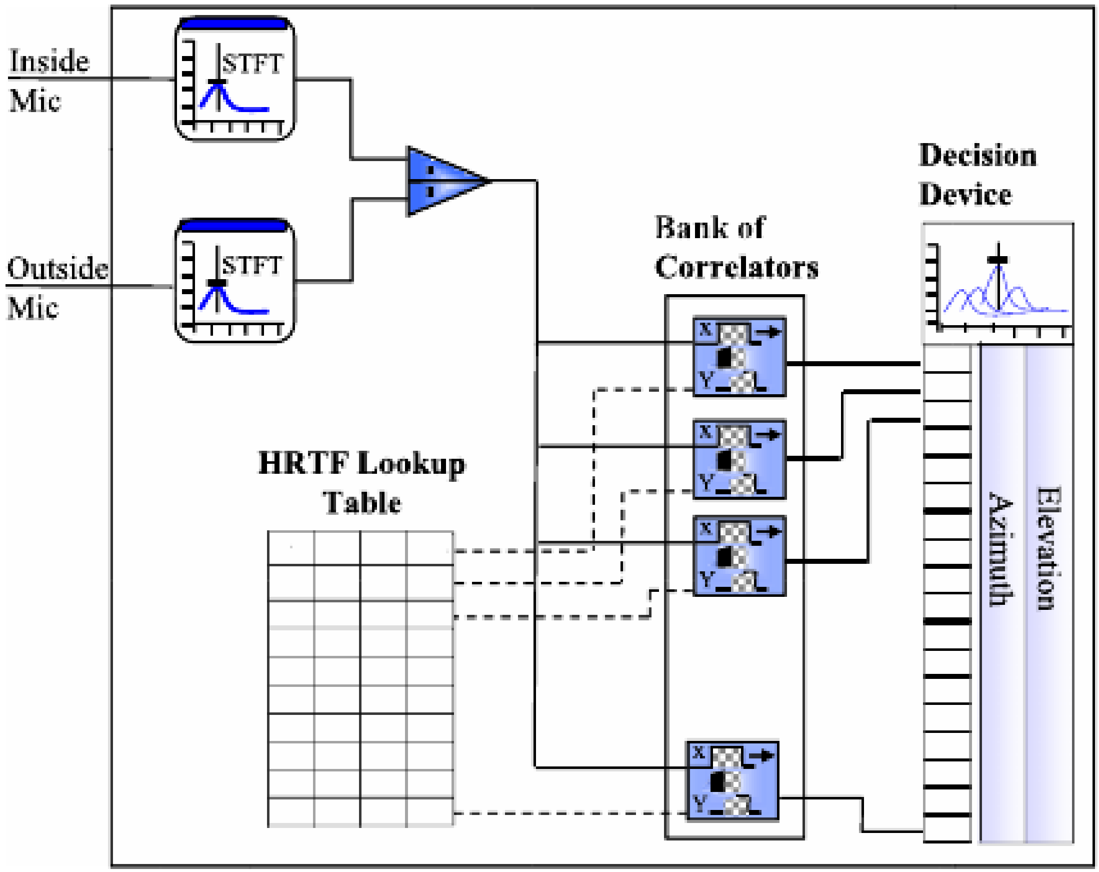
- We look at one proposed algorithm for sound localization in humanoid robots
- Proposed to handle reverberance
- Localize sound in 3d space
- Potential real time calulations

STATE-OF-THE-ART

- Uses spatial likelihood functions of several microphone arrays
- These are then integrated with a weighted average of each function
- Uses multiple possible results to find the best one
- Can be done probabilistically or with neural networks

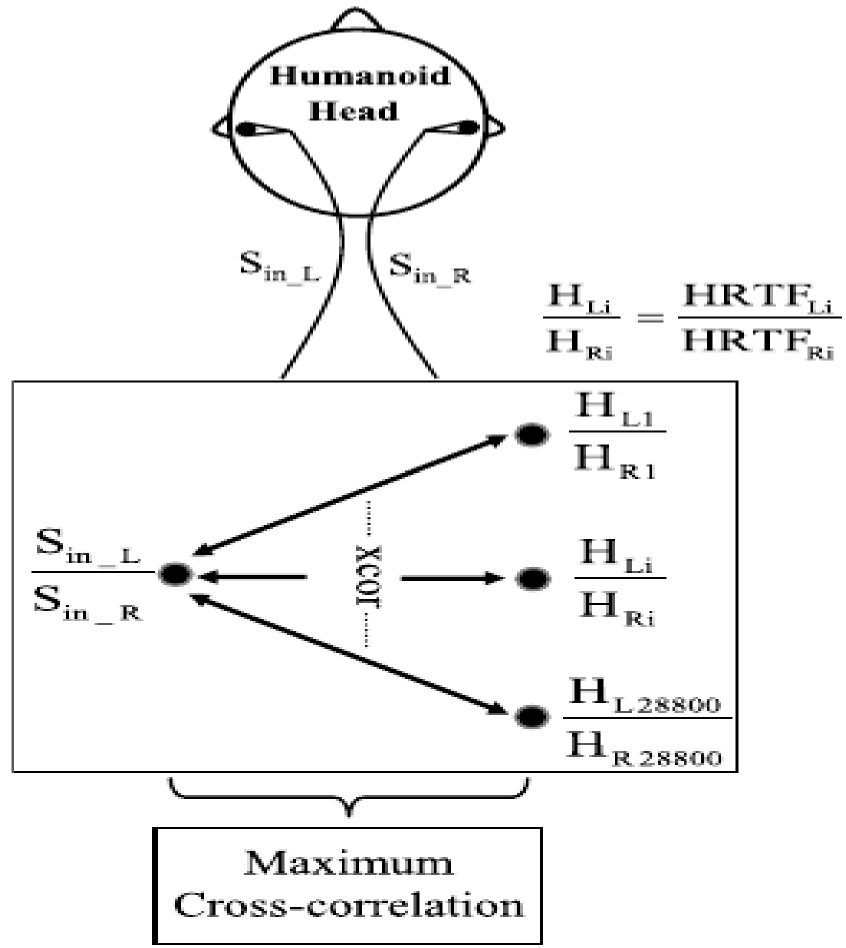
Monaural Algorithm

- Each ear has two microphones, one inside the ear and one outside
- A spatially modified acoustic signal received inside the ear
- Echoes and noise are modeled for
- Position identified by filter response on the noise from look up table


Inside
Mic
Outside
Mic
STFT
STFT
Decision
Device
Bank of
Correlators
HRTF Lookup
Table
Azimuth
Elevation

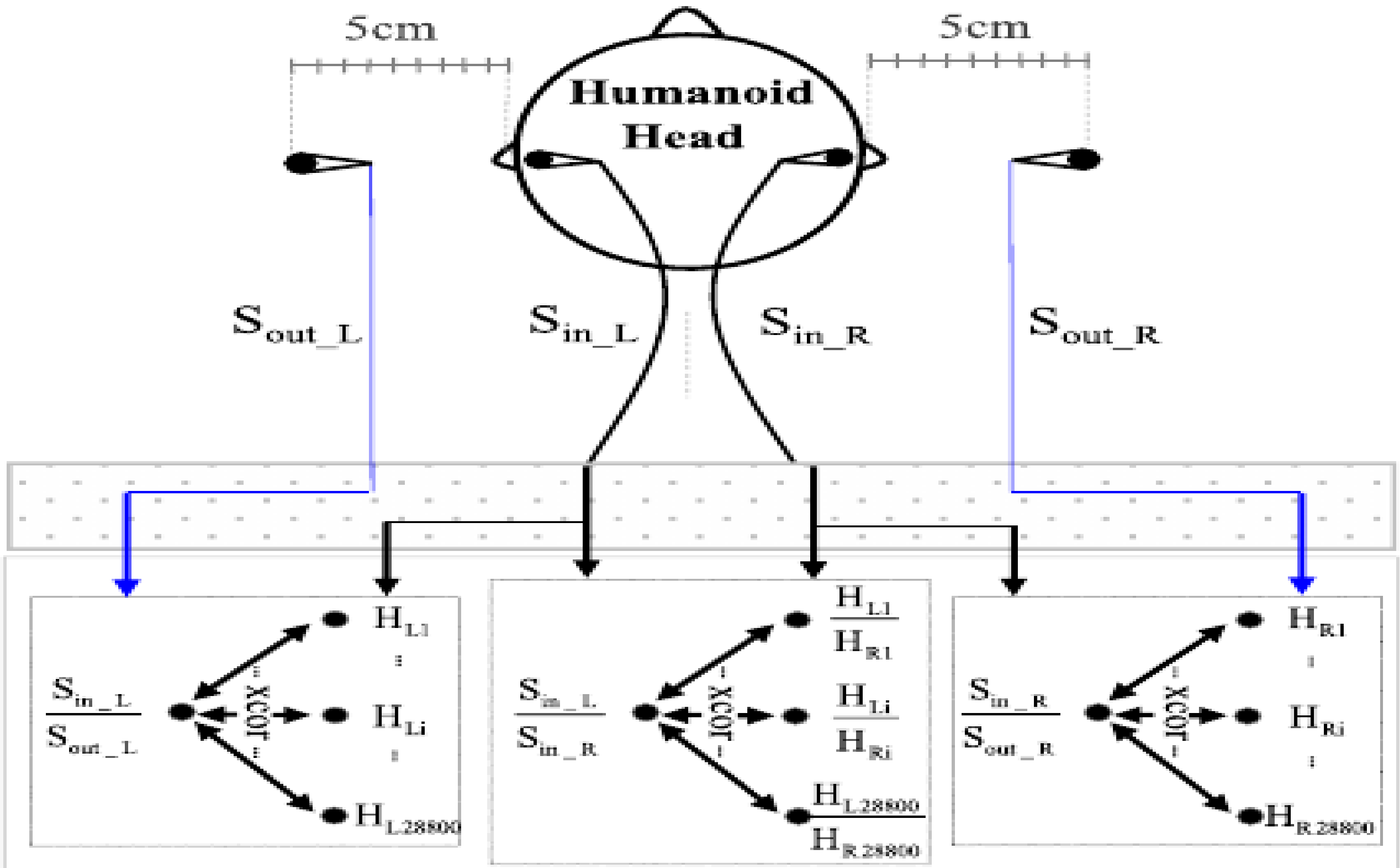
Binaural Algorithm

- Using the inner ear input of each ear
- If we divide the two inputs, to get the ratios we also cancel the source
- We match this to lookup table of HRTF pairs


Humanoid Head
S_{in_L}
S_{in_R}
$\frac{H_{Li}}{H_{Ri}} = \frac{HRTF_{Li}}{HRTF_{Ri}}$
$\frac{S_{in\_L}}{S_{in\_R}}$
$\frac{H_{L1}}{H_{R1}}$
$\frac{H_{Li}}{H_{Ri}}$
$\frac{H_{L28800}}{H_{R28800}}$
Maximum Cross-correlation

Unified Sound Localization System

- Combine results of each ear monaural algorithm and the binaural algorithm
- If they are between 5 degrees of each other they are averaged
- Else the angular error is calculated
- Bayesian network used to detect the direction
- The robot then decides the status of sound event


5cm
5cm
Humanoid
Head
S out_L
S in_L
S in_R
S out_R
S in_L / S out_L
xcor
H L1
H Li
H L28800
S in_L / S in_R
H L1 / H R1
H Li / H Ri
H L28800 / H R28800
S in_R / S out_R
H R1
H Ri
H R28800

Conclusion

- The system cannot calculate distance
- Cannot localize multiple sound sources
- Does not have auditory environment
- Potentially can use head movements or robot repositioning

References

- https://ocw.mit.edu/high-school/biology/introduction-to-cognitive-neuroscience/lecture-notes/MITHFH_cog_neuroscclass_17.pdf
- http://www2.bcs.rochester.edu/courses/crsinf/221/19.pdf
- Chapter-6---Sound-localization_2015_Handbook-of-Clinical-Neurology
- A survey on sound source localization in robotics: From binaural to array processing methods, S.Argentieria, P.Danès, P.Souères
- Advanced Binaural Sound Localization in 3-D for Humanoid Robots, Fakheredine Keyrouz

Questions