Indoor Sound Localization

Fares Abawi

Universität Hamburg
Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Informatik
Technische Aspekte Multimodaler Systeme

Monday, 12-12-2016

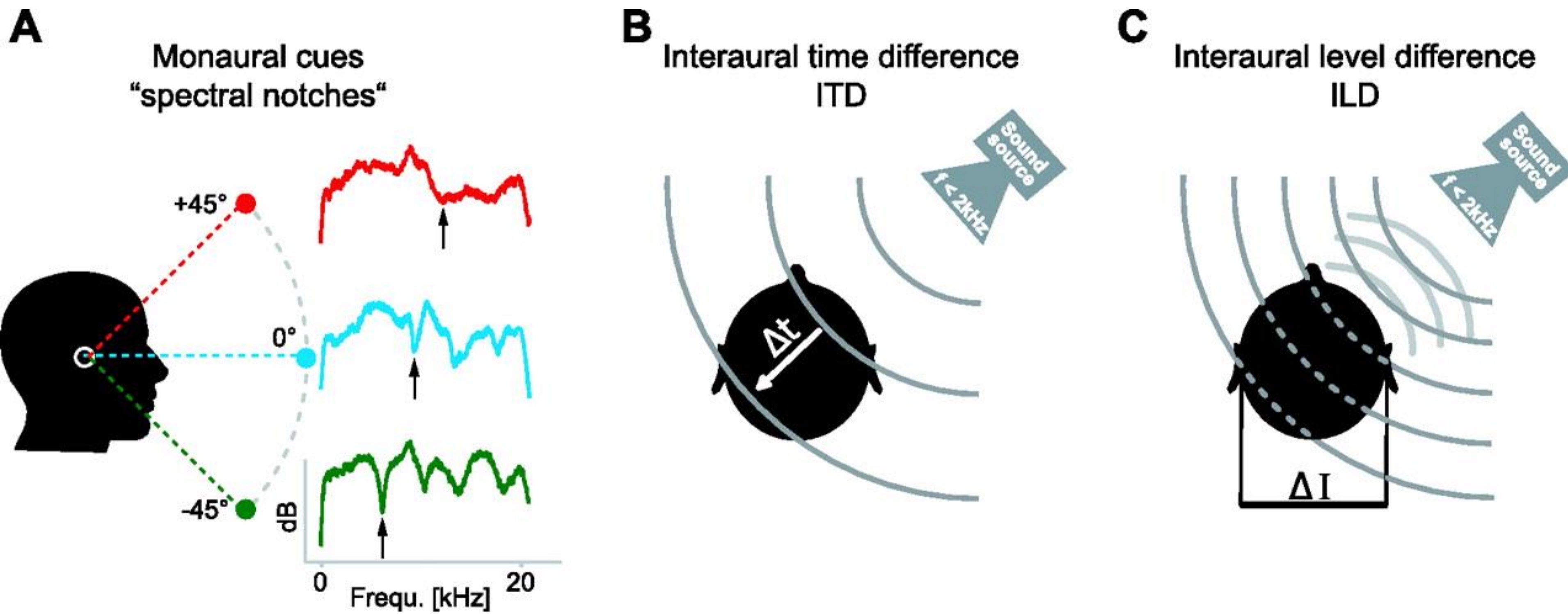
Contents

Introduction

Definition

Sound localization is …

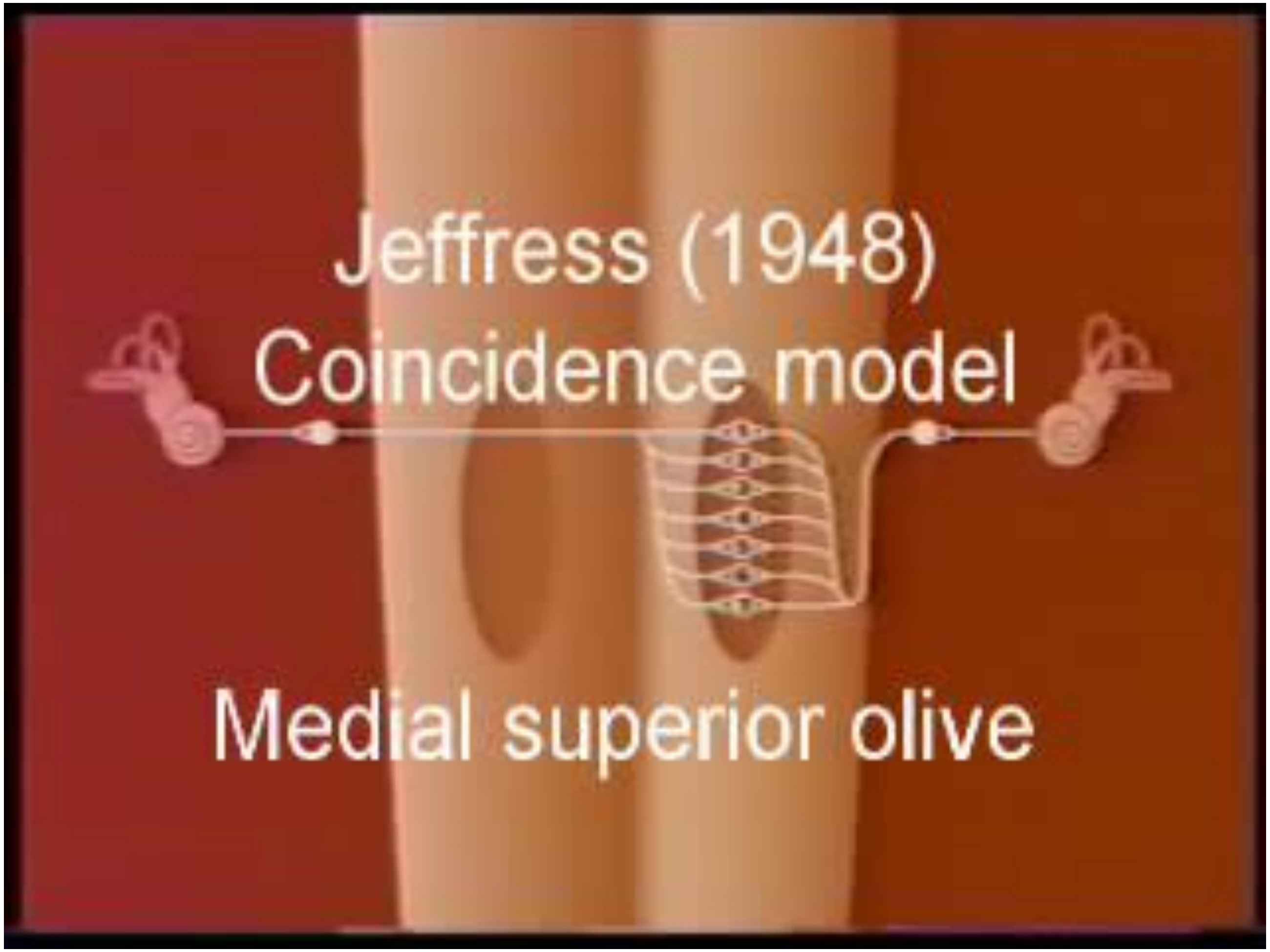
Introduction



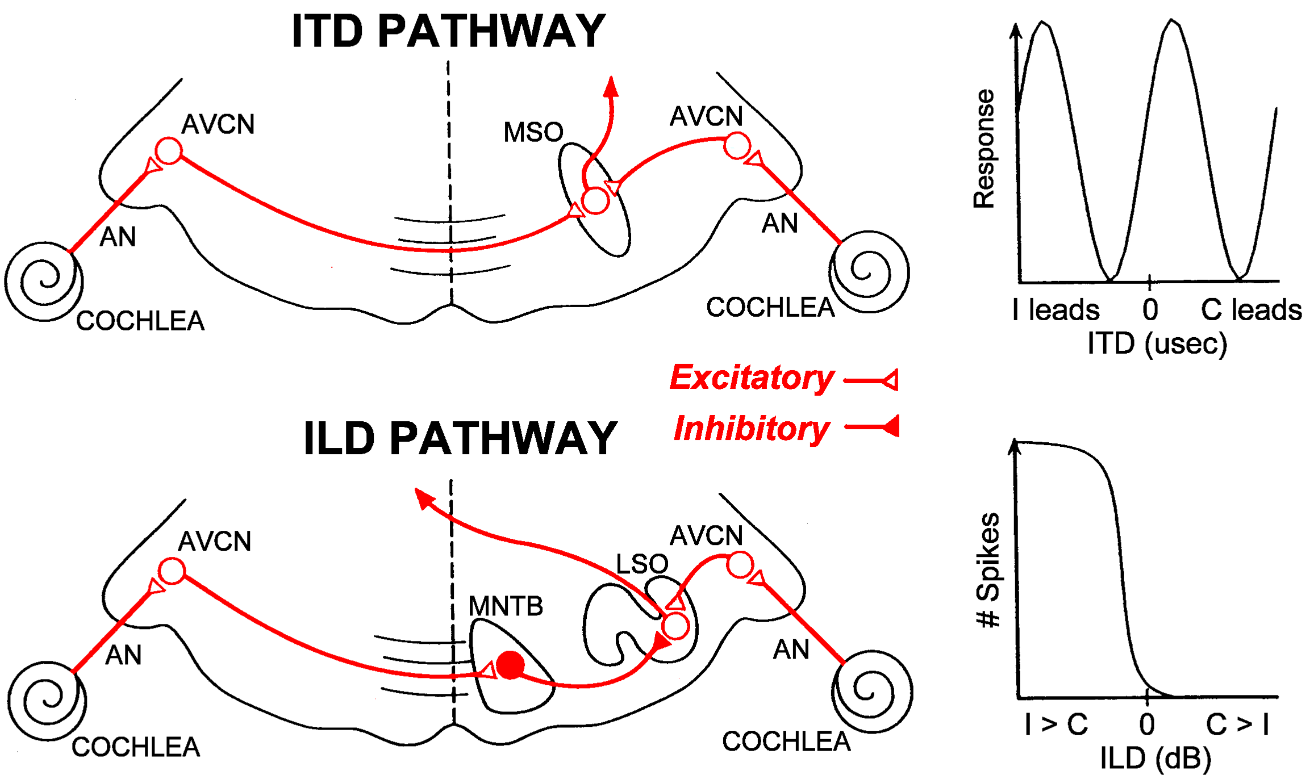
[4]

Introduction

The Jeffess Model – Oversimplified model of the mammalian MSO [VIDEO] [6]



Introduction



[4]

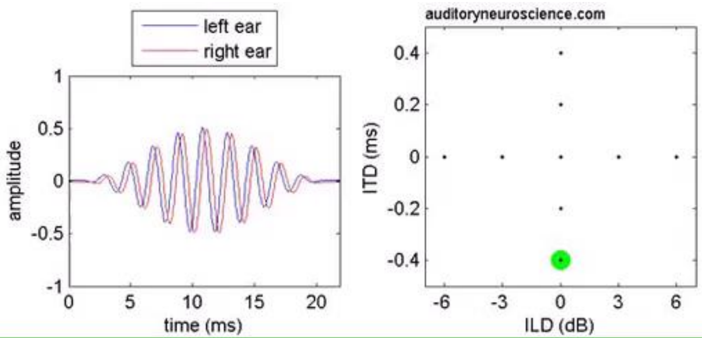
Lateral Superior Olive : ILD is performed

Medial Superior Olive : ITD is performed

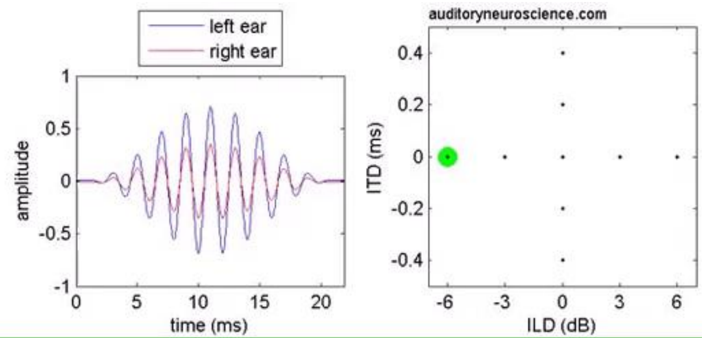
Introduction

Binaural cues
[VIDEOS] [7]

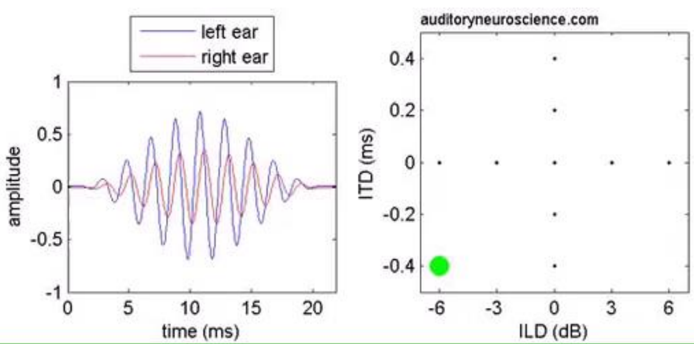
Varying ITD



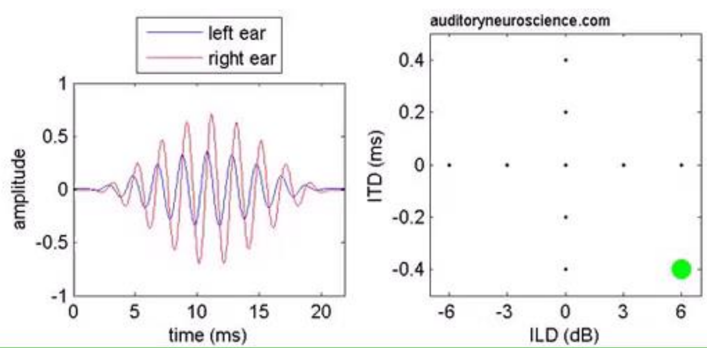
Varying ILD



Varying ITD & ILD



Trading ITD off against ILD



Checkpoint

- Introduction
- Cross-Correlation
- Quality Effecting Factors
- Sound Localization:
  - Time Difference of Arrival
  - Steered Beamforming
  - Bio-Inspired Sound Localization
- Comparison
- Summary
- References

Cross-Correlation

$$(s_A \star s_B)[n] = \sum_{m=-\infty}^{\infty} {s_A}^*[n] s_B[n+m]$$

Get the delay between two signals by shifting one against the other

Multiply-> Sum-> Shift-> Repeat !

Convolution Theorem:

Convolution in the time domain is simply a multiplication in the frequency domain and vice versa

Cross-Correlation

Complexity:

Cooley-Tuckey FFT $= n.\log(n)$

Time-Domain xcorr $= n^2$

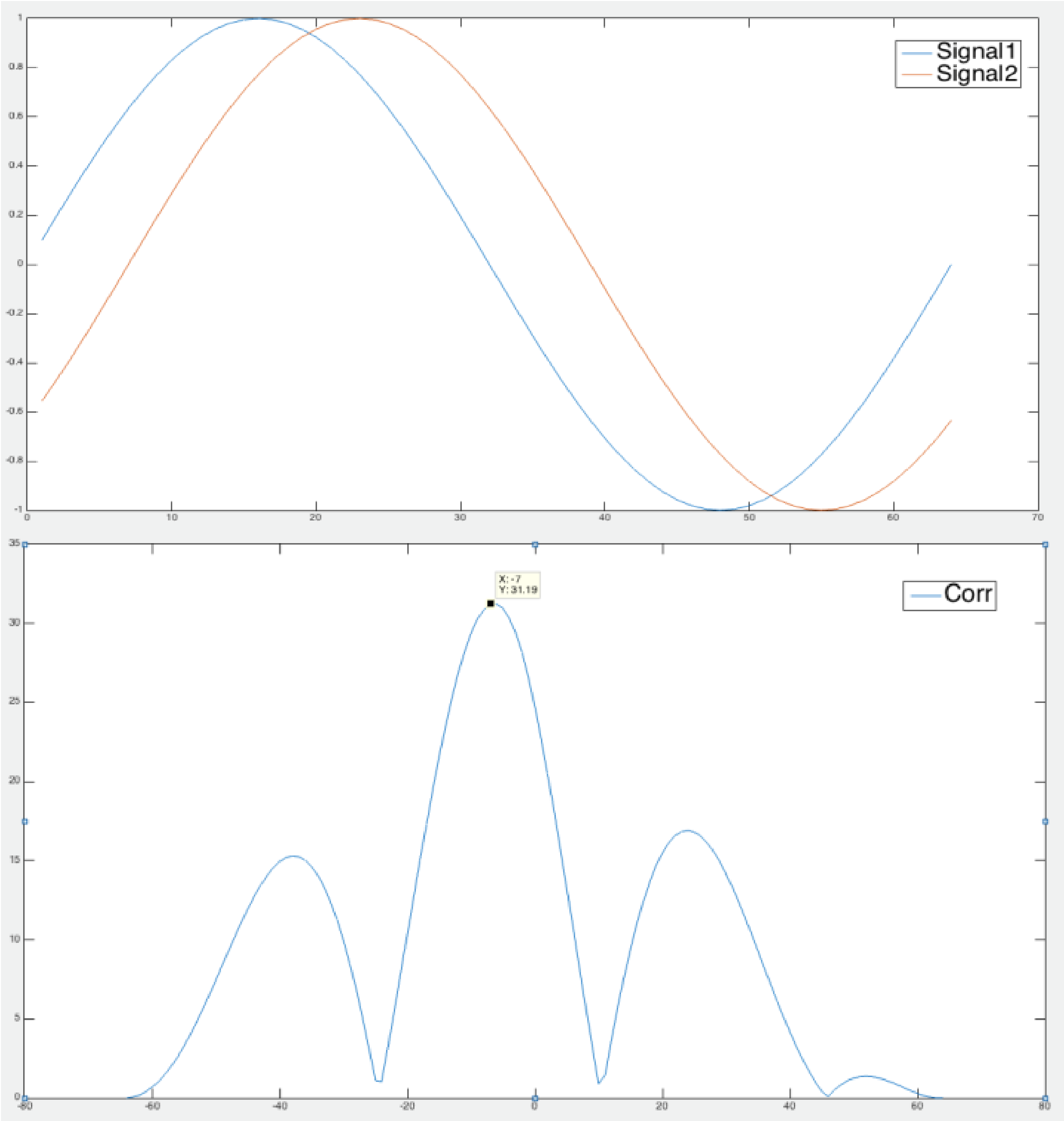
Notes on Time->Frequency Domain Transformation

- The sampling frequency must be twice the maximum frequency a system needs to acquire, according to the **Nyquist Theorem**, in order to avoid **temporal aliasing**.
- A windowing function (**Analysis window**) must be applied to signal before transformation to avoid frequency leakage and smearing. The window can be in the form of a **Hann window, Hamm window** or the like.
- **Keep in mind:** The cross-correlation of two signals produces a vector with a length of both signal lengths -1. If ignored the cross-correlation will be distorted due to circular convolution.

Cross-Correlation

Two sinusoids with a difference of 7 samples

Peak detected at x = -7 after performing cross-correlation

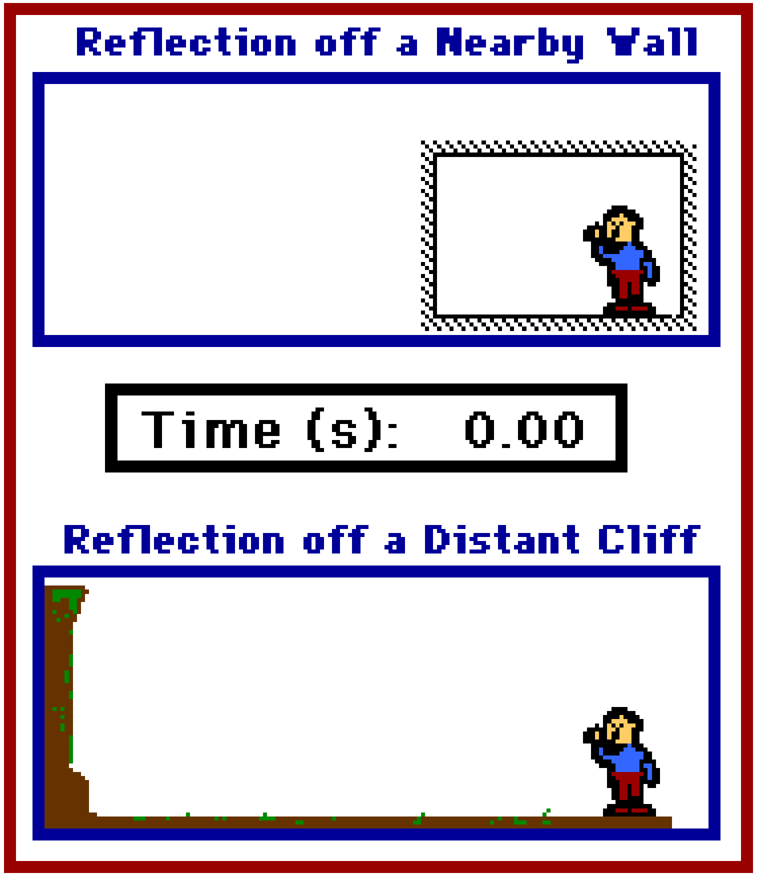


Checkpoint

- Introduction
- Cross-Correlation
- Quality Effecting Factors
- Sound Localization:
  - Time Difference of Arrival
  - Steered Beamforming
  - Bio-Inspired Sound Localization
- Comparison
- Summary
- References

Quality Effecting Factors

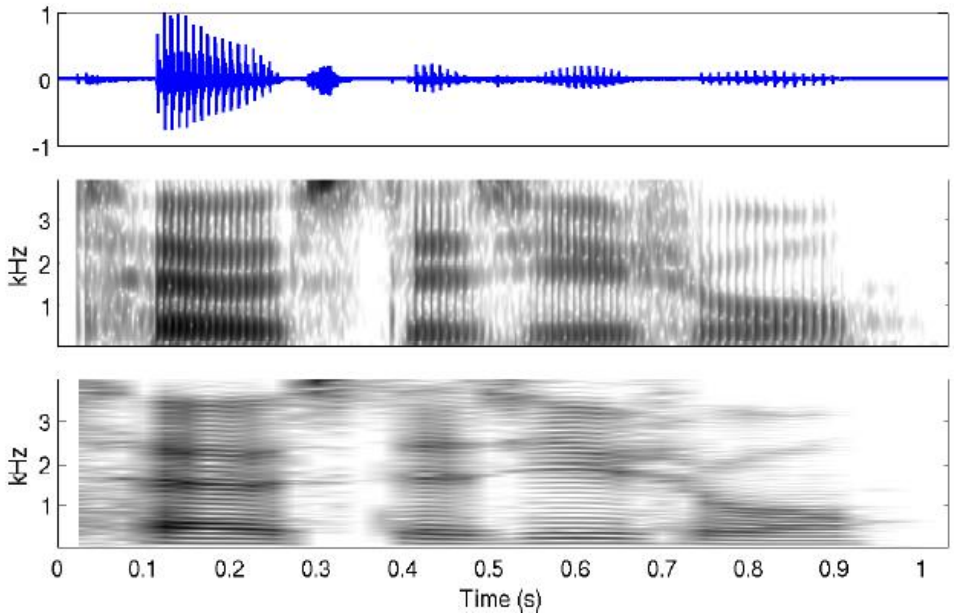
Echo and Reverb [ANIMATION] [8]



Quality Effecting Factors

Noise

Noise power spectral densities can be estimated by finding the minima from time-frequency bins that do not contain speech



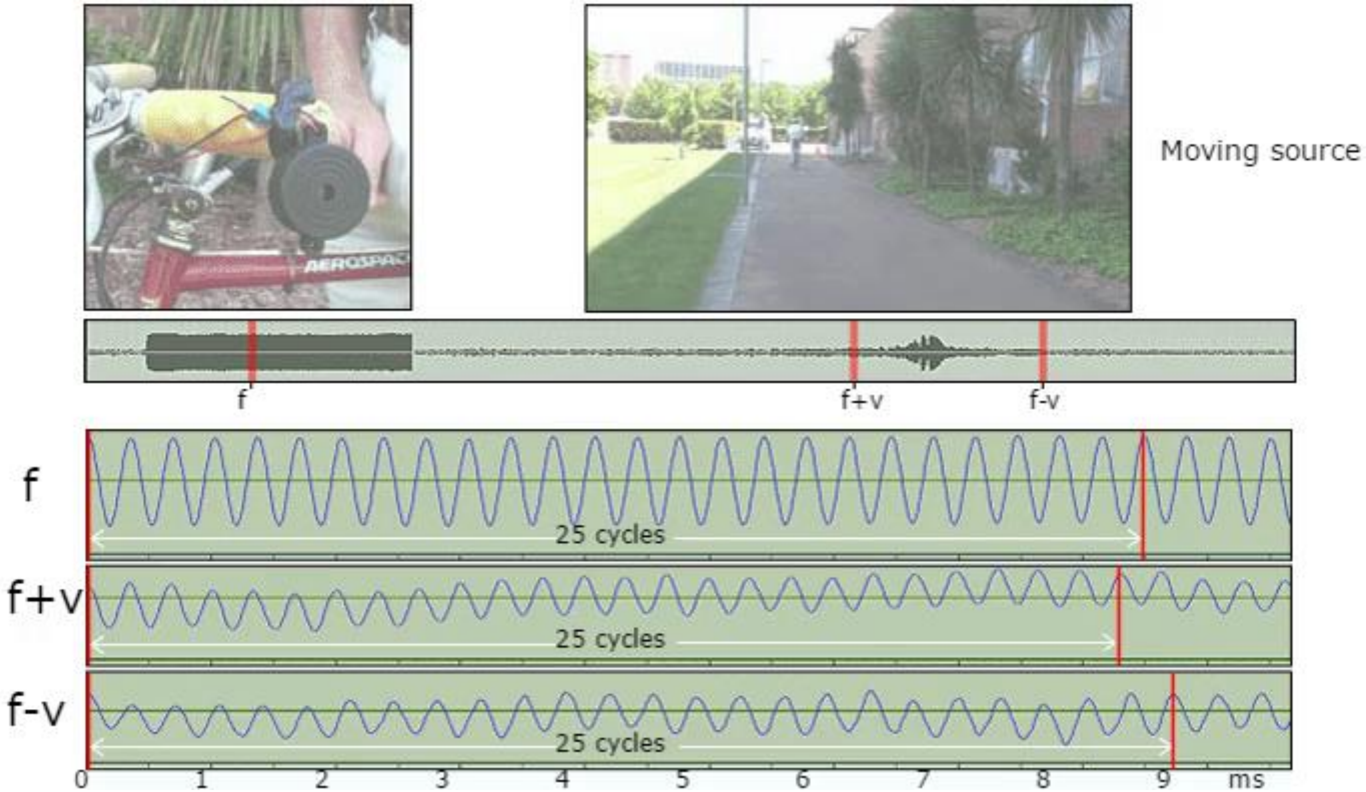
[4]

Could this work for any sound signal ?

Any Environment ??

Quality Effecting Factors

Doppler shift [VIDEO] [9]

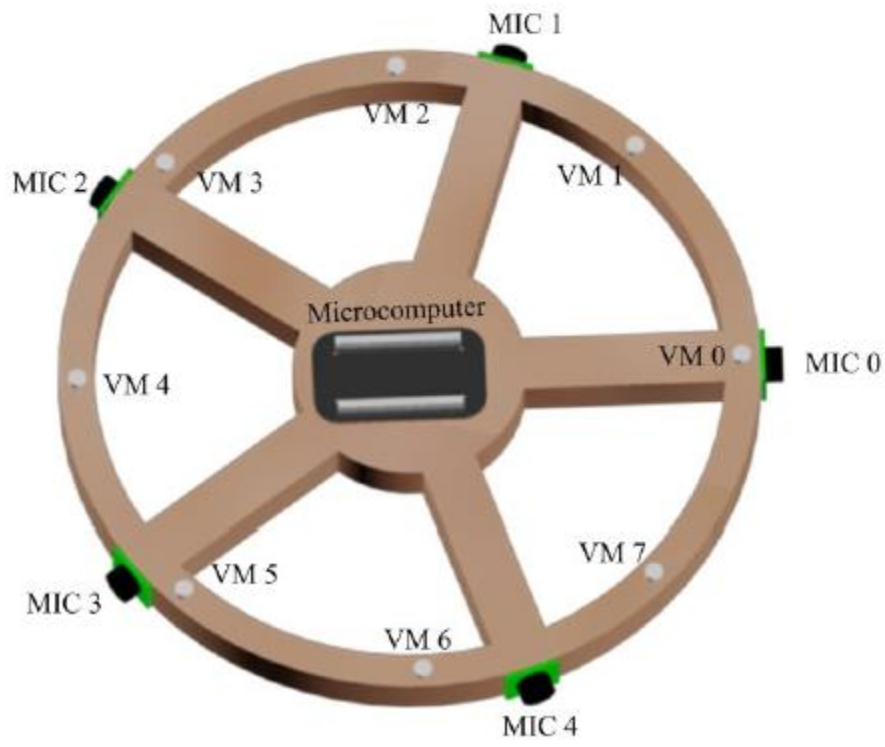


Checkpoint

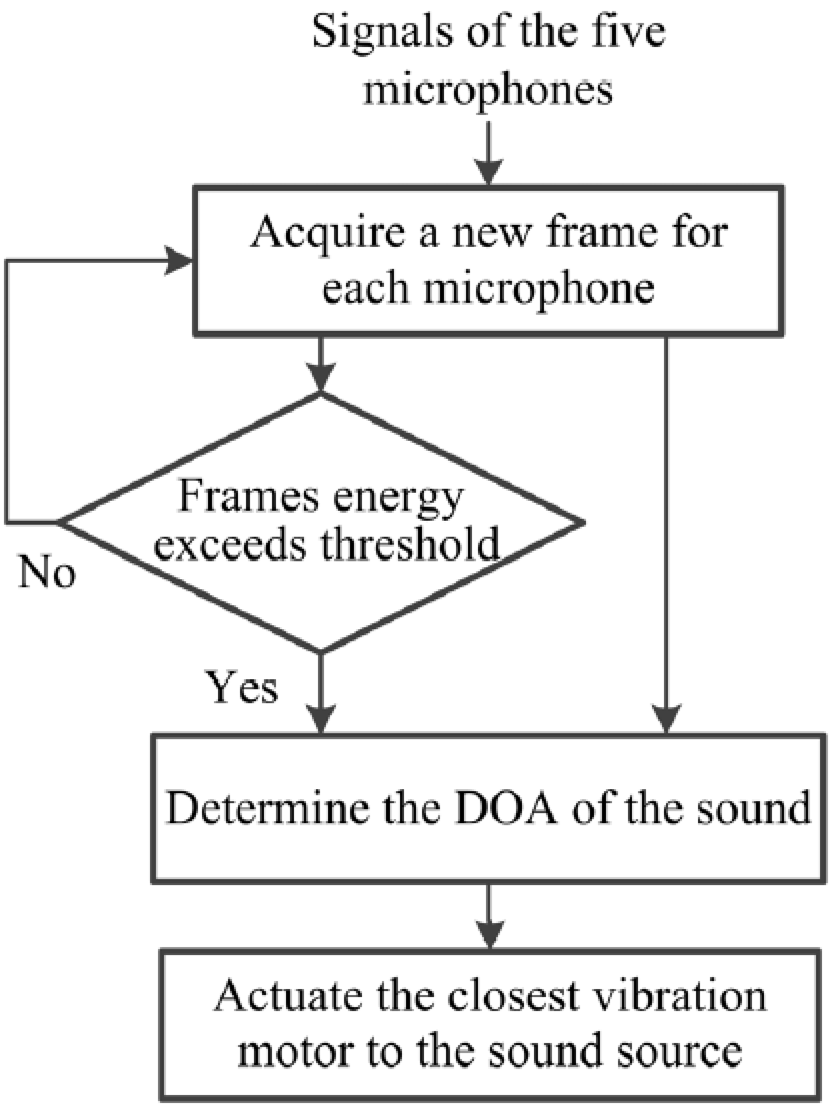
- Introduction
- Cross-Correlation
- Quality Effecting Factors
- Sound Localization:
  - Time Difference of Arrival
  - Steered Beamforming
  - Bio-Inspired Sound Localization
- Comparison
- Summary
- References

Time Difference of Arrival

In-house Alert Sounds Detection and Direction of Arrival Estimation to Assist People with Hearing Difficulties [1]



Time Difference of Arrival



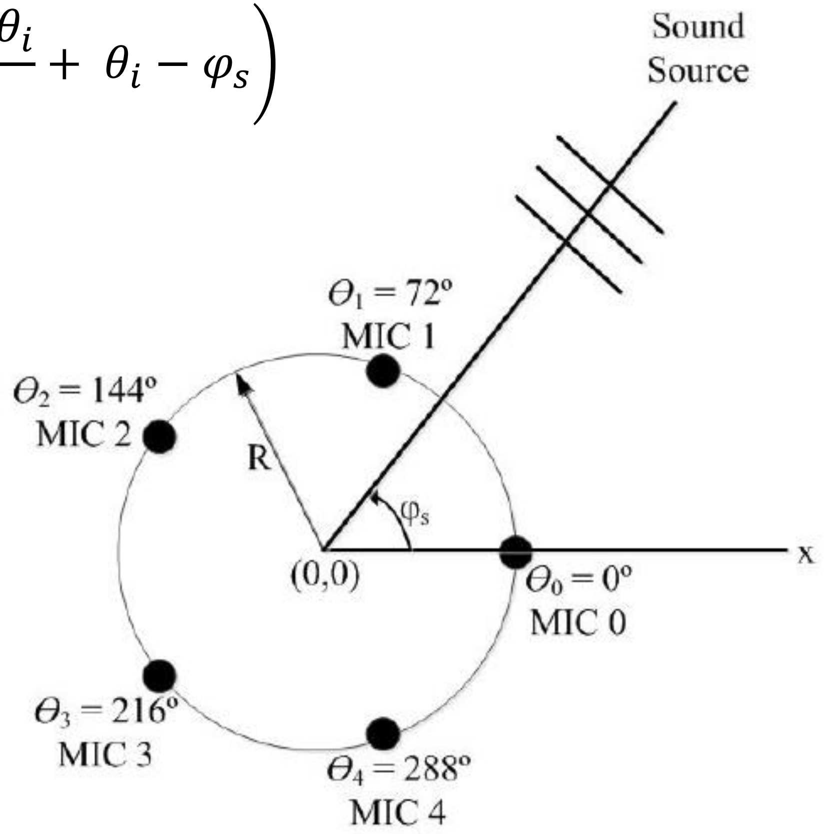
[1]

Time Difference of Arrival

Calculating the delay at which sound arrives the circular microphone array

$$\tau_{(k,i)} = 2\frac{R}{C}\sin\left(\frac{\theta_k - \theta_i}{2}\right)\sin\left(\frac{\theta_k - \theta_i}{2} + \theta_i - \varphi_s\right)$$

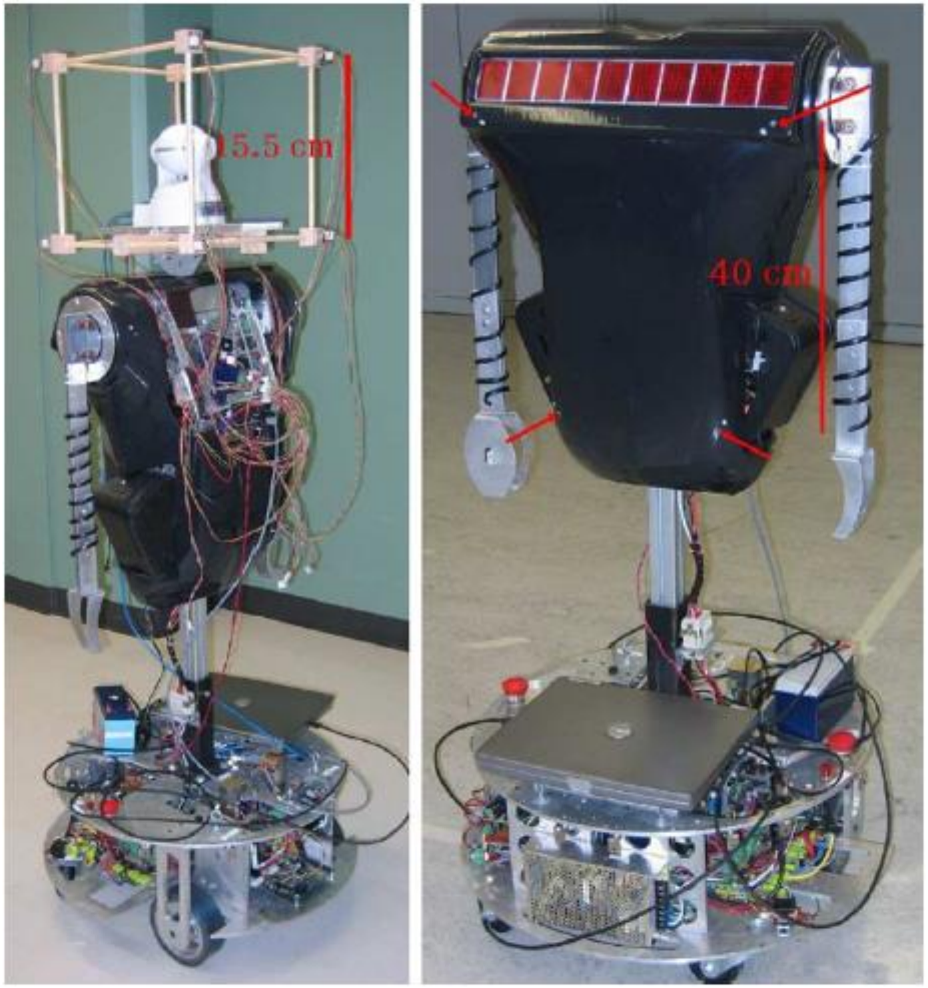
Approximating the angle by incrementing φ_s from 0° to 360° selecting the angle which reduces the difference between the analytical delay and that acquired through cross-correlation



[1]

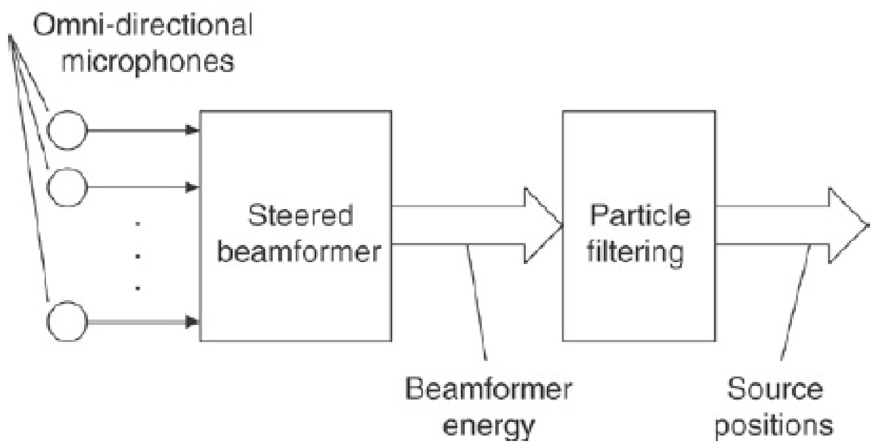
Steered Beamforming

Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering [2]



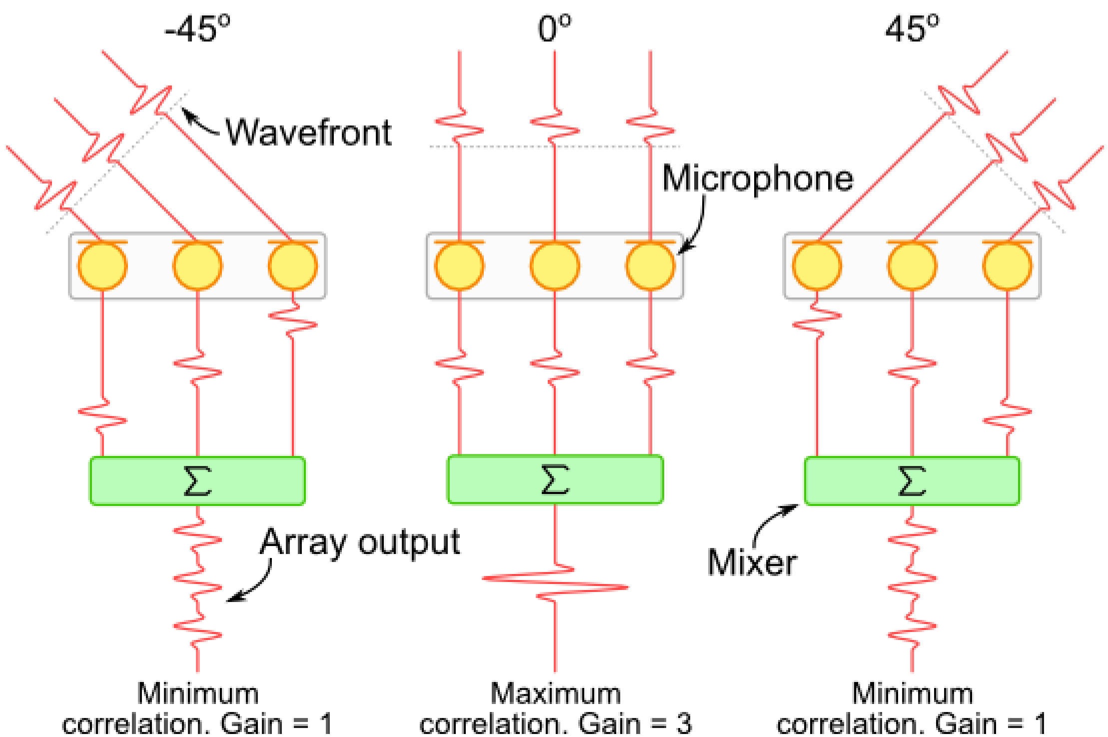
Steered Beamforming

- Detect the sound from an array of omnidirectional microphones
- Steer the beam towards all possible angles
- Use particle filtering to predict the motion of the sound source
- Can detect angle and position !



[2]

Steered Beamforming

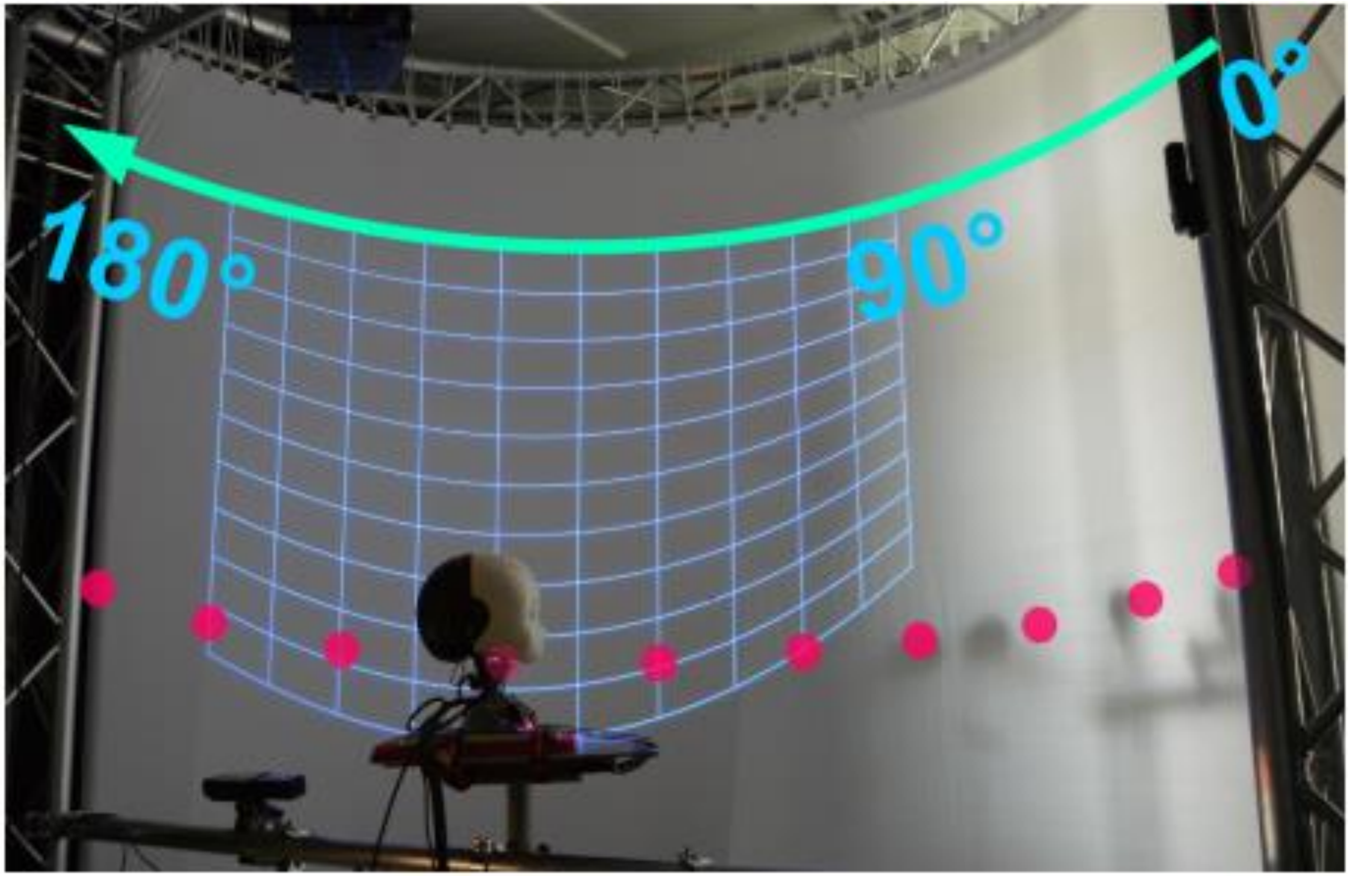


[5]

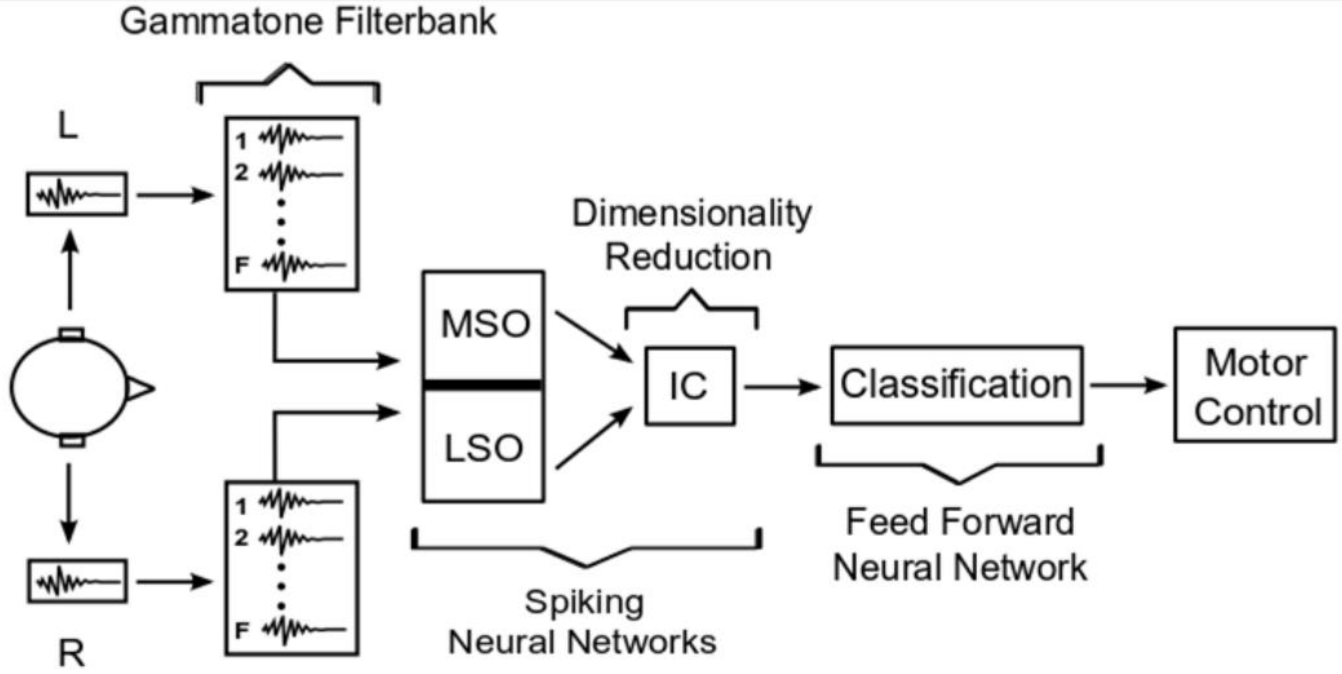
$$output = 20\log_{10}\left(\frac{1}{N}\sum_{i=0}^{N-1} e^{\frac{j\,2\pi f\,i\,l\,\sin(\theta)}{c}}\right)$$

Bio-Inspired Sound Localization

Neural and Statistical Processing of Spatial Cues for Sound Source Localisation [3]



Bio-Inspired Sound Localization



[3]

- Detect the direction of incoming sound
- Filter the sound signal (Gammatone FB)
- Detect ITD and ILD
- Reduce the dimensionality (Inferior Colliculus -> Naïve Bayes)
- Classify (FFNN)
- rotate the robot's head in the direction of the sound, aligning a single microphone with the sound source.

Checkpoint

- Introduction
- Cross-Correlation
- Quality Effecting Factors
- Sound Localization:
  - Time Difference of Arrival
  - Steered Beamforming
  - Bio-Inspired Sound Localization
- Comparison
- Summary
- References

Comparison

	TDOA	Beamforming	Bio-Inspired SSL
Steps	Cross-Correlate and measure delay	Shift, Cross-Correlate, sum and measure power	Cross-Correlate, Minimize dimensionality, feed to network and predict
Speed	Fast	Moderate	Slow
Accuracy	Lowest	Moderate	Best
Resources	Low	High	High
Training	Not Required	Not Required	Required

Checkpoint

- Introduction
- Cross-Correlation
- Quality Effecting Factors
- Sound Localization:
  - Time Difference of Arrival
  - Steered Beamforming
  - Bio-Inspired Sound Localization
- Comparison
- Summary
- References

Summary

- Mammalians Localize sound through **binaural** and **monaural cues**
- **Interaural level difference (ILD)** is the measure of sound level/loudness across two inputs
- **Interaural time difference (ITD)** is the measure of sound level/loudness across two inputs
- **The Lateral Superior Olive (LSO)** : where ILD is measured in the brain
- **The Medial Superior Olive (MSO)** : where ITD is measured in the brain
- **Cross-Correlation** measures the delay between two signal
- Cross-Correlation is performed efficiently in the Frequency domain
- Quality effecting factors:
  - **Echo**
  - **Reverb**
  - **Noise**
  - **Doppler shift**

Summary

- Computerized systems can measure the direction of sound by:
  - **Time difference of arrival or phase delay**
  - **Steered beamforming**
  - **Heuristic and statistical methods**

- Beamforming can detect more than a single sound source
- Sound can be detected by binaural or multi-microphone array systems (circular or aligned)

References

[1] M. Daoud, M. Al-Ashi, F. Abawi, and A. Khalifeh, *“In-house alert sounds detection and direction of arrival estimation to assist people with hearing difficulties,”* in IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS), pp. 297–302, Nevada, US, June 2015.

[2] J.-M. Valin, F. Michaud and J. Rouat, *“Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering,”* Robotics Autonomous Syst. J. 55, 216– 228, 2007.

[3] J. Davila-Chacon, S. Magg, J. Liu, and S. Wermter. *“Neural and statistical processing of spatial cues for sound source localization,”* in IEEE Intl. Conf. on Neural Networks (IJCNN-13), pp. 1–8, Dallas, US, 2013.

References

[4] B. Grothe, M. Pecka, and D. McAlpine, *“Mechanisms of Sound Localization in Mammals”* in Physiological Reviews Published 1 July 2010 **Vol.** 90 **no.** 3, 983-1012
http://physrev.physiology.org/content/90/3/983

[5] A. Greensted, “Delay Sum Beamforming” in The Lab Book Pages, 2012
http://www.labbookpages.co.uk/audio/beamforming/delaySum.html

[6] J. Schnupp, E. Nelken, A. King, “The Jeffress Model – Animation” in Auditory Neuroscience
https://auditoryneuroscience.com/topics/jeffress-model-animation

[7] J. Schnupp, E. Nelken, A. King, “Binaural Cues” in Auditory Neuroscience
https://auditoryneuroscience.com/topics/binaural-cue-demos

[8] “Echo and Reverb animation” in The Physics Classroom
http://www.physicsclassroom.com/mmedia/waves/er.gif

[9] “Waves and Sound: The Doppler Effect” In PHYSCLIPS ,UNSW, School of Physics, Sydney
http://www.animations.physics.unsw.edu.au/jw/doppler.htm

Further Reading

[10] B. Clénet and H. Romsdorfer, *"Circular microphone array based beamforming and source localization on reconfigurable hardware".* Diss. Master's thesis, Graz University of Technology, 2010.

[11] J. Davila-Chacon, J. Twiefel, J. Liu, and S. Wermter. *"Improving Humanoid Robot Speech Recognition with Sound Source Localisation."* International Conference on Artificial Neural Networks. Springer International Publishing, 2014.

Questions ?

Thank you !