From Azimuth's cocktail party to accident prevention

Binaural Sound Localization

Christoph Quinten

christoph.quinten@rwth-aachen.de
AICES RWTH Aachen University

Aachen, 22.06.2016

Outline

Importance

Importance

- remarkable how we hear and separate sound

Importance

- remarkable how we hear and separate sound
- cocktail party processing becomes more relevant

Importance

- remarkable how we hear and separate sound
- cocktail party processing becomes more relevant
- understand the physical cues

Importance

- remarkable how we hear and separate sound
- cocktail party processing becomes more relevant
- understand the physical cues
- human safety and survival system
    - prevent traffic accidents

Background and Approach

Background and Approach

- Lord Rayleigh's duplex theory (1876)
  - binaural and spectral cues
  - single source localization

Background and Approach

- Lord Rayleigh's duplex theory (1876)
  - binaural and spectral cues
  - single source localization
- Proposed warning system to optimize siren localization (2011)
  - in-cabin warning system
  - UK emergency services

Background and Approach

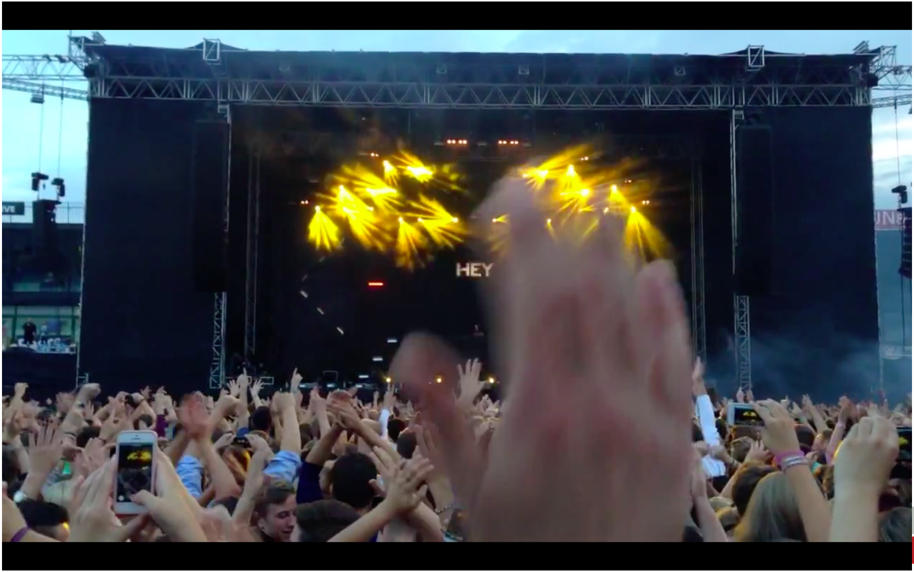
- Lord Rayleigh's duplex theory (1876)
    - binaural and spectral cues
    - single source localization
- Proposed warning system to optimize siren localization (2011)
    - in-cabin warning system
    - UK emergency services
- Uniform linear microphone array by Carter and Abraham (1980)

Example

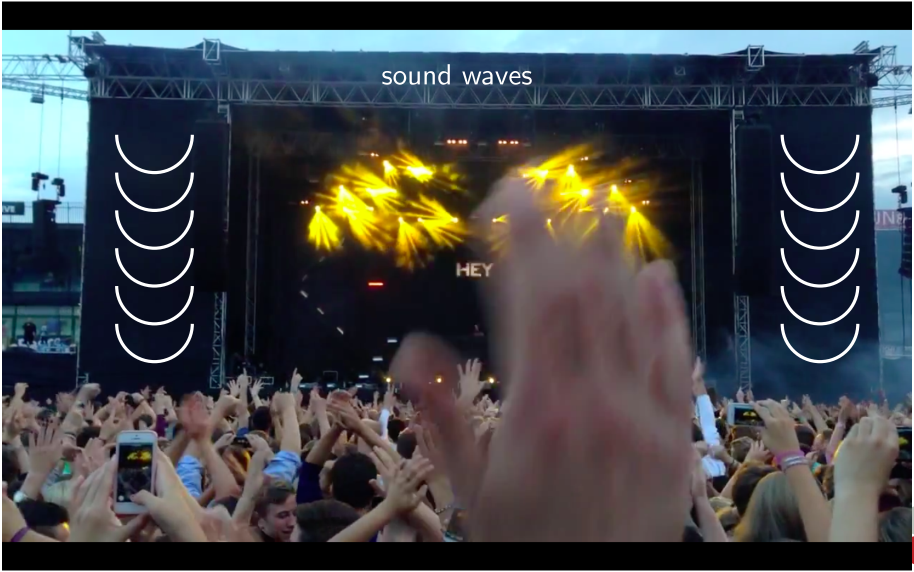
https://www.youtube.com/watch?v=b0lrGXF2Y3w

AVICII live @ HockeyPark Mönchengladbach

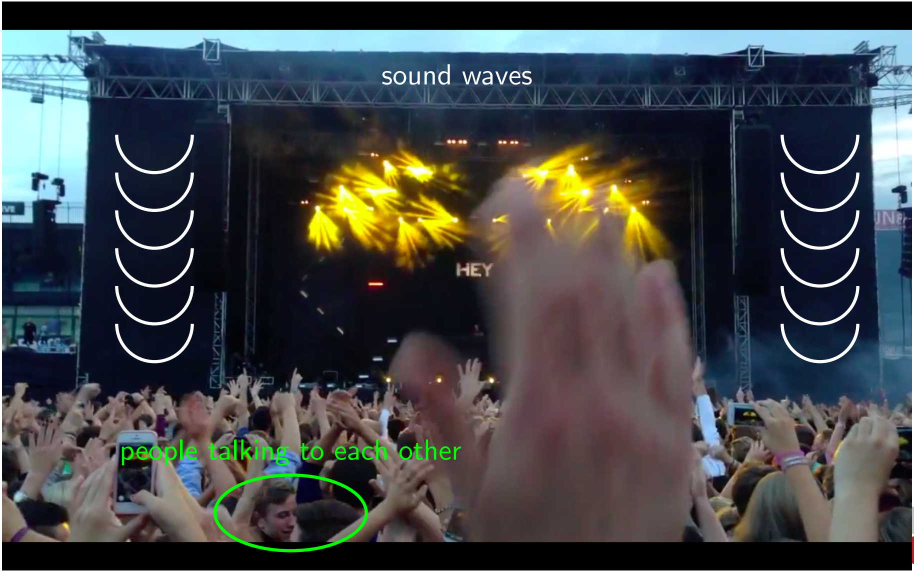
Example



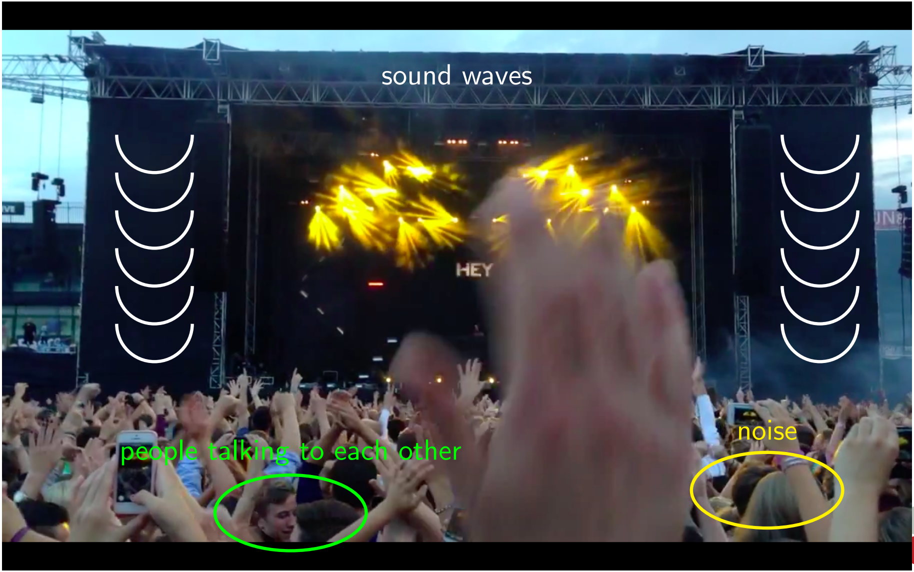
Example



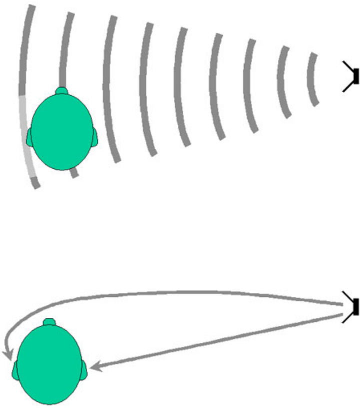
Example



Example



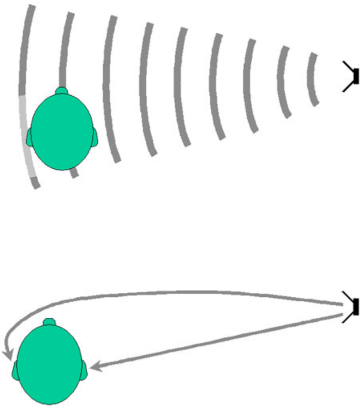
Lord Rayleigh's duplex theory



Duplex Theory [5]

Lord Rayleigh's duplex theory

- sound arrives earlier at the ear closer to the source
- this sound has higher intensity than at the other ear
- only for simple single sound sources in free field



Duplex Theory [5]

Binaural cues

Binaural cues

Interaural Differences

- **ITD** = *Interaural Time Difference*
- **ILD** = *Interaural Level Difference (aka IID = Intensity Difference)*
- **IPD** = *Interaural Phase Difference*

Binaural cues

Interaural Differences

- **ITD** = *Interaural Time Difference*
- **ILD** = *Interaural Level Difference (aka IID = Intensity Difference)*
- **IPD** = *Interaural Phase Difference*

- ITD
  - it takes longer for the sound to arrive at the farer ear
  - maximum is about $600\mu s$
  - depends on the speed of sound, geometry of head and ears
  - not effective at high frequencies

Binaural cues

Interaural Differences

- **ITD** = *Interaural Time Difference*
- **ILD** = *Interaural Level Difference (aka IID = Intensity Difference)*
- **IPD** = *Interaural Phase Difference*

- ILD
  - shadowing effect of the head as consequence of absorption and reflection
  - maximum above approximately 1.5kHz
  - may not exhibit a relationship to azimuth
  - not effective at low frequencies

Binaural cues

Interaural Differences

- **ITD** = *Interaural Time Difference*
- **ILD** = *Interaural Level Difference (aka IID = Intensity Difference)*
- **IPD** = *Interaural Phase Difference*

- IPD
  - depends on the frequency and the ITD
  - underlies localization performance with low frequency sounds
  - intensity differences responsible for performance

Cone of confusion

Cone of confusion

- ITD and ILD cannot uniquely locate sound

Cone of confusion

- ITD and ILD cannot uniquely locate sound
- sound sources in the median plane have no interaural differences

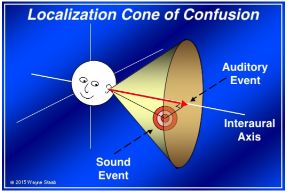
Cone of confusion

- ITD and ILD cannot uniquely locate sound
- sound sources in the median plane have no interaural differences
- differences specify a cone of directions

Cone of confusion

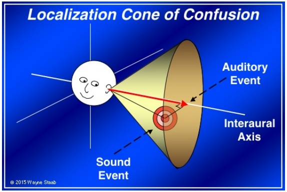
- ITD and ILD cannot uniquely locate sound
- sound sources in the median plane have no interaural differences
- differences specify a cone of directions
- similar differences (front from back or up from down)

Cone of confusion



Cone Of Confusion [6]

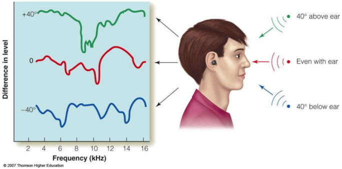
Cone of confusion



Cone Of Confusion [6]

- sound localization along a cone of confusion
- higher frequencies appear higher in elevation
- static sound sources spectral cues stay fixed over duration of the sound
- outside of the free field cues evolve as echoes and reverberation

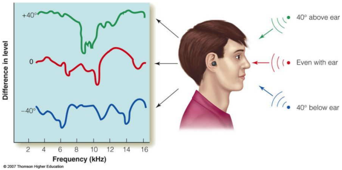
Spatial Localization



HRTF [8]

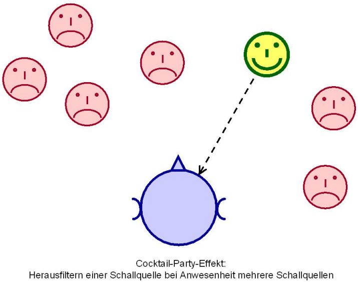
Spatial Localization

- higher complexity of locating moving sounds
- *Head related transfer function* (**HRTF**):
  frequency response as function of azimuth and elevation
- *Precedence effect*:
  sound is reflected from various surfaces before reaching the ears
- *Azimuth*:
  head and ear related coordinate system



HRTF [8]

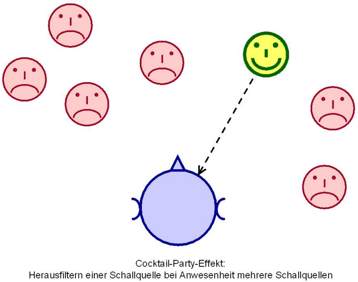
Cocktail Party Processing (CPP)



CPP example [10]

Cocktail Party Processing (CPP)

- phenomenon of selective attention
- separate single voice on a cocktail party
- other example: someone calls your name → focus on this talk



CPP example [10]

Human safety and survival system

- in everdyday life the hearing of noise is in focus
- humans must be able to locate sound sources very fast

Human safety and survival system

- in everdyday life the hearing of noise is in focus
- humans must be able to locate sound sources very fast
- enhanced road and car safety is strongly required
- approaching vehicles can be estimated from an analysis of sound sources

Human safety and survival system

- in everdyday life the hearing of noise is in focus
- humans must be able to locate sound sources very fast
- enhanced road and car safety is strongly required
- approaching vehicles can be estimated from an analysis of sound sources
- prevent traffic accidents
- saving humans (civilians, lifesaver and victims) life

Example

https://www.youtube.com/watch?v=ubwmKF7-8Og

Einsatzfahrt RTW

Siren Localization

Siren Localization

- sound = first warning of events

Siren Localization

- sound = first warning of events
- sirens can cause confusion, disorientation, possible danger

Siren Localization

- sound = first warning of events
- sirens can cause confusion, disorientation, possible danger
- todays cars improved sound system masks alerts

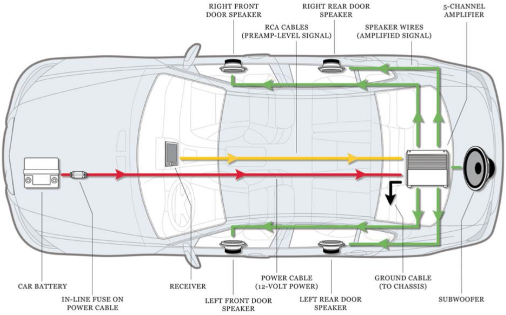
Siren Localization

- sound = first warning of events
- sirens can cause confusion, disorientation, possible danger
- todays cars improved sound system masks alerts
- either the driver don't hear the siren or the siren is not strong enough to be recognized in the noisy environment

Siren Localization

- sound = first warning of events
- sirens can cause confusion, disorientation, possible danger
- todays cars improved sound system masks alerts
- either the driver don't hear the siren or the siren is not strong enough to be recognized in the noisy environment
- when siren is heard, looking around trying to determine the direction

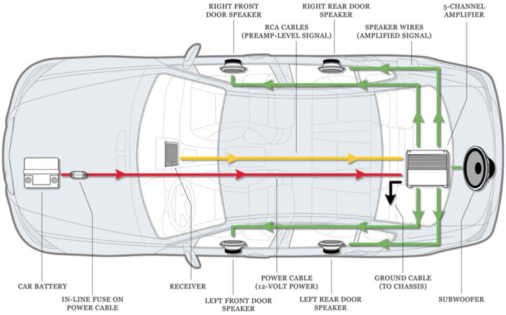
First Proposal - In-Cabin Warning System



Car loudspeaker system [7]

First Proposal - In-Cabin Warning System

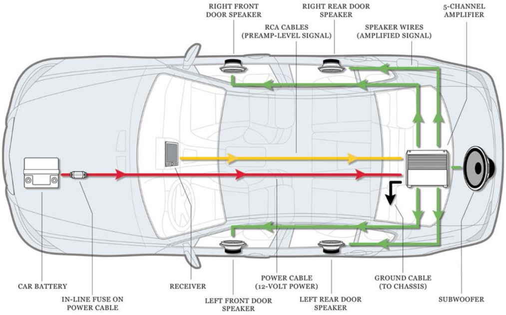
1. minimize/eliminate spatial ambiguity inherent in existing sirens
   - patterns contain broadest audible frequency and expand them



Car loudspeaker system [7]

First Proposal - In-Cabin Warning System

1. minimize/eliminate spatial ambiguity inherent in existing sirens
   - patterns contain broadest audible frequency and expand them
2. reduce problems with environmental noise
   - in-cabin auditory as visual warning system
   - relay a warning sound in the same direction as the EV



Car loudspeaker system [7]

Second Proposal - Uniform Microphone Arrays

Second Proposal - Uniform Microphone Arrays

- recent systems detect driver's status like drowsiness

Second Proposal - Uniform Microphone Arrays

- recent systems detect driver's status like drowsiness
- outer information collected by systems like camera, radar, etc.

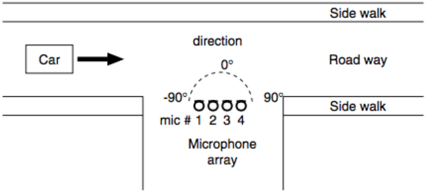
Second Proposal - Uniform Microphone Arrays

- recent systems detect driver's status like drowsiness
- outer information collected by systems like camera, radar, etc.
- environmental sound contains much information

Second Proposal - Uniform Microphone Arrays

- recent systems detect driver's status like drowsiness
- outer information collected by systems like camera, radar, etc.
- environmental sound contains much information
- use microphone array to record sounds of arriving vehicles

Sound recording

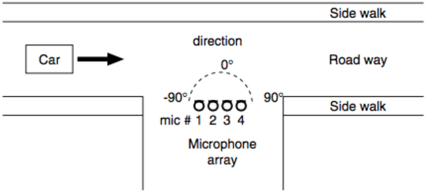


environment [4]

Sound recording

Temperature	15.9 °C
Road condition	Dry
Subject	4 cars
Measurement	5 times
Speed	20, 40, 60 km/h
Weather	Fair
Frequency response	20 - 20kHz
Sampling frequency	44.1kHz

recording conditions [4]

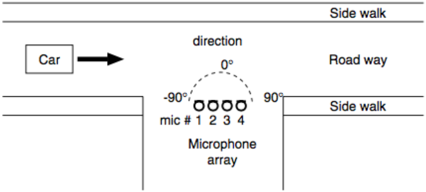


environment [4]

Sound recording

Temperature	15.9 °C
Road condition	Dry
Subject	4 cars
Measurement	5 times
Speed	20, 40, 60 km/h
Weather	Fair
Frequency response	20 - 20kHz
Sampling frequency	44.1kHz

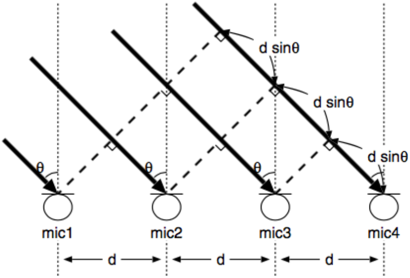
recording conditions [4]



environment [4]

- sounds captured and sampled at 44.1kHz
- vehicle sounds split into
  - frictional sound (between road and tires)
  - engine sound

Sound Localization



microphone array [4]

- $d[m]$ element spacing
- $\theta[rad]$ direction of sound wave
- $dsin\theta$ propagation distance
- time delay[s] $\tau_S = dsin\theta/c$

Sound Localization - Cross correlation method

$\phi_{ij}(\tau) = \int_{t_S - T}^{t_S + T} x_i(t) x_j(t+\tau)\,\mathrm{d}t$

Sound Localization - Cross correlation method

$\phi_{ij}(\tau) = \int_{t_S - T}^{t_S + T} x_i(t) x_j(t+\tau)\,\mathrm{d}t$

- $x_i(t)$, $x_j(t)$: input signals recorded at ith and jth microphone
- cross-correlation $\phi_{ij}(\tau)$ at time t_S with τ delay
- T length of the window that satisfies $T > 2d/c$

Sound Localization - Cross correlation method

$$\phi_{ij}(\tau) = \int_{t_S-T}^{t_S+T} x_i(t)x_j(t+\tau)\,\mathrm{d}t$$

- $x_i(t)$, $x_j(t)$: input signals recorded at ith and jth microphone
- cross-correlation $\phi_{ij}(\tau)$ at time t_S with τ delay
- T length of the window that satisfies $T > 2d/c$

$$\phi_{ij}(\tau) = \int_{-f_c}^{f_c} e^{j2\pi f\tau}\,\mathrm{d}f = 2f_c\frac{sin2\pi f_c\tau}{2\pi f_c\tau}$$

Sound Localization - Cross correlation method

$$\phi_{ij}(\tau) = \int_{t_S - T}^{t_S + T} x_i(t) x_j(t+\tau)\,\mathrm{d}t$$

- $x_i(t)$, $x_j(t)$: input signals recorded at ith and jth microphone
- cross-correlation $\phi_{ij}(\tau)$ at time t_S with τ delay
- T length of the window that satisfies $T > 2d/c$

$$\phi_{ij}(\tau) = \int_{-f_c}^{f_c} e^{j2\pi f\tau}\,\mathrm{d}f = 2f_c \frac{sin 2\pi f_c \tau}{2\pi f_c \tau}$$

- accuracy depends on the bandwith of the input signal
- $\phi(\tau)$ has a clear peak when f_c is significantly large

Proposed method

Proposed method

- time delay is computed using pairs of microphones

Proposed method

- time delay is computed using pairs of microphones
- m-microphone array: $\phi_C(\tau) = \sum_{i=0}^{m-1} \phi_{i,i+1}(\tau)$

Proposed method

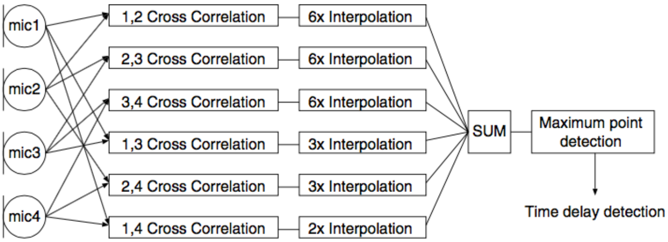
- time delay is computed using pairs of microphones
- m-microphone array: $\phi_C(\tau) = \sum_{i=0}^{m-1} \phi_{i,i+1}(\tau)$
- conventional method: $\phi_C(\tau) = \phi_{12}(\tau) + \phi_{23}(\tau) + \phi_{34}(\tau)$

Proposed method

- time delay is computed using pairs of microphones
- m-microphone array: $\phi_C(\tau) = \sum_{i=0}^{m-1} \phi_{i,i+1}(\tau)$
- conventional method: $\phi_C(\tau) = \phi_{12}(\tau) + \phi_{23}(\tau) + \phi_{34}(\tau)$
- normalize differences with cubic spline interpolation $\rightarrow$ produce the sum of ϕ_{ij} by offsetting the time differences

Proposed method

- time delay is computed using pairs of microphones
- m-microphone array: $\phi_C(\tau) = \sum_{i=0}^{m-1} \phi_{i,i+1}(\tau)$
- conventional method: $\phi_C(\tau) = \phi_{12}(\tau) + \phi_{23}(\tau) + \phi_{34}(\tau)$
- normalize differences with cubic spline interpolation → produce the sum of ϕ_{ij} by offsetting the time differences



procedure flow [4]

Conclusions

Conclusions

- several cues in human listening

Conclusions

- several cues in human listening
- ability to locate sound is remarkable

Conclusions

- several cues in human listening
- ability to locate sound is remarkable
- cocktail party processing: focus on a talk when someone calls your name

Conclusions

- several cues in human listening
- ability to locate sound is remarkable
- cocktail party processing: focus on a talk when someone calls your name
- two systems/methods how drivers can be warned still earlier

Conclusions

- several cues in human listening
- ability to locate sound is remarkable
- cocktail party processing: focus on a talk when someone calls your name
- two systems/methods how drivers can be warned still earlier
  $\Rightarrow$ prevent more accidents and save more lifes

Questions

Thanks for your attention!

References

[1] D. Wang, G. Brown - *Binaural Sound Localization*, 2005, chapter 5 from *Computational Auditory Scene Analysis*

[2] G. Stecker, F. Gallun - *Binaural Hearing, Sound Localization, and Spatial Hearing*, 2012, chapter 14, pp. 383-433

[3] D. Moore, S. Boslem, V. Charissis - *Optimisation of Sound Localization for Emergency Vehicle Sirens through a Prototype Audio System*, 2011, Article

[4] K. Kodera, A. Itai, H. Yasukawa - *Sound Localization of Approaching Vehicles Using Uniform Microphone Array*, 2007, Conference Paper pp. 1054-1058

[5] Lord Rayleigh's duplex theory - http://www.diracdelta.co.uk/science/source/d/u/duplex%20theory%20of%20localization/source.html#.V1f0am0xaRt

[6] Cone of confusion - http://hearinghealthmatters.org/waynesworld/2015/localization-more-important-than-word-recognition/

References

[7] Car audio system -
http://images2.crutchfieldonline.com/ImageBank/v20151215083500/ImageHandler/scale/978/978/ca/learningcenter/car/amp_install_guide/system-diagram.jpg

[8] Head related transfer function -
http://images.google.de/imgres?imgurl=http%3A%2F%2Ftheheadphonelist.theheadphonelist.netdna-cdn.com%2Fwp-content%2Fuploads%2F2013%2F10%2FHRTF.jpg%253F8a4320&imgrefurl=http%3A%2F%2Ftheheadphonelist.com%2Fbrain-localize-sounds%2Fhrtf%2F&h=443&w=900&tbnid=TggfBfv16ptBYM%3A&docid=XbqbN5Gn3nKIWM&ei=CI5aV4WCNMOTaYCWoNAE&tbm=isch&client=safari&iact=rc&uact=3&dur=1194&page=1&start=0&ndsp=17&ved=0ahUKEwiFppqOmJ3NAhXDSRoKHQALCEoQMwgqKAYwBg&bih=562&biw=1280

[9] Azimuth coordinate system -
https://en.wikipedia.org/wiki/Azimuth#/media/File:Azimuth-Altitude_schematic.svg

[10] Cocktail party processing -
http://www.cocktail-party-processor.de/intro/index.html