Binaural Hearing for Robots

Sound-Source Localization

Radu Horaud

Binaural Hearing for Robots

1. Introduction to Robot Hearing
2. Methodological Foundations
3. **Sound-Source Localization**
4. Machine Learning and Binaural Hearing
5. Fusion of Audio and Vision

Radu Horaud

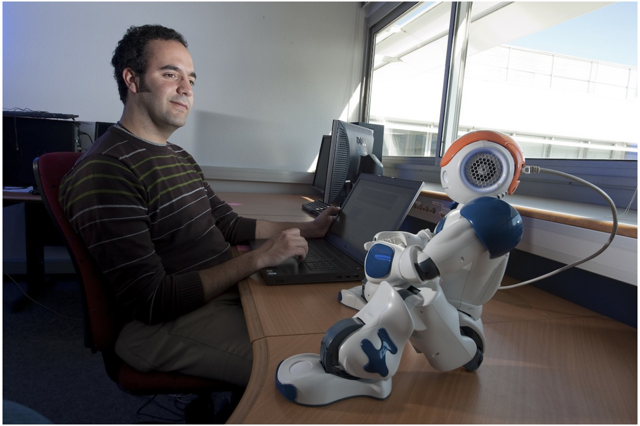
3. Sound-Source Localization

1. Time difference of arrival (TDOA)
2. Estimation of TDOA by cross-correlation
3. Estimation of TDOA in the spectral domain
4. The geometry of two microphones
5. Direction of arrival
6. Using more than two microphones
7. Embedding the microphones in a robot head
8. Learning a sound propagation model
9. Predicting direction of a sound with a robot head
10. Example of sound direction estimation

3. Sound-Source Localization

1. **Time difference of arrival (TDOA)**
2. Estimation of TDOA by cross-correlation
3. Estimation of TDOA in the spectral domain
4. The geometry of two microphones
5. Direction of arrival
6. Using more than two microphones
7. Embedding the microphones in a robot head
8. Learning a sound propagation model
9. Predicting direction of a sound with a robot head
10. Example of sound direction estimation

The Basic Setup



Notations and Definitions

We start by considering a simple situation: two microphones and one sound source:

- $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$: 3D **known** positions of microphones m_1 and m_2.
- $\boldsymbol{S}$: 3D **unknown** position of the sound source
- $s(t)$: **unknown** acoustic signal emitted by the source
- $x_1(t), x_2(t)$: **observed** acoustic signals recorded with the two microphones

Microphone Recordings

$$x_1(t) = s(t - \tau_1) + n_1(t)$$
$$x_2(t) = s(t - \tau_2) + n_2(t)$$
$$\tau = \tau_1 - \tau_2$$

- τ_1 and τ_2: times of arrival from the source to the microphone(s)
- $n_1(t)$ and $n_2(t)$: additive noise
- τ: time difference of arrival (TDOA) between the two microphones

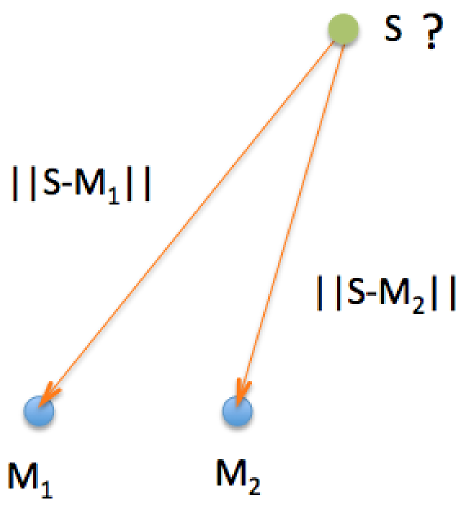
Direct Sound Propagation

We assume a direct propagation path:

$$\tau_i = \frac{\|\boldsymbol{S} - \boldsymbol{M}_i\|}{\nu}$$

- $\| \cdot \|$ is the Euclidean norm
- $\nu = 344$ m/s is the speed of sound

Geometry



Time Difference of Arrival

- The signal reaches the two microphones with a delay:

$$\tau = \tau_1 - \tau_2 = \frac{\|\boldsymbol{S} - \boldsymbol{M}_1\|}{\nu} - \frac{\|\boldsymbol{S} - \boldsymbol{M}_2\|}{\nu}$$

- τ is the:
- TDOA (**time difference of arrival**) or
- ITD (**interaural time difference**).

Session Summary

- Basic setup: two microphones and one source.
- Binaural recording with the direct propagation model.
- Geometric representation.
- Time difference of arrival (TDOA).

3. Sound-Source Localization

1. Time difference of arrival (TDOA)
2. **Estimation of TDOA by cross-correlation**
3. Estimation of TDOA in the spectral domain
4. The geometry of two microphones
5. Direction of arrival
6. Using more than two microphones
7. Embedding the microphones in a robot head
8. Learning a sound propagation model
9. Predicting direction of a sound with a robot head
10. Example of sound direction estimation

Estimation of TDOA

The TDOA can be estimated via **cross-correlation** (CC) between the two microphone signals:

$$\hat{\tau} = \underset{t'}{\operatorname{argmax}} \sum_{t=t_1}^{t_2} x_1(t) x_2(t - t')$$

$\hat{\tau}$ is an estimation of the **true value** τ.

Robust Estimation of TDOA

The normalized cross-correlation (NCC) offers robustness with respect to a gain difference between the two signals:

$$\hat{\tau} = \underset{t'}{\operatorname{argmax}} \frac{\sum_{t=t_1}^{t_2} x_1(t)x_2(t-t')}{(\sum_{t=t_1}^{t_2} x_1^2(t))^{1/2}(\sum_{t=t_1}^{t_2} x_2^2(t))^{1/2}}$$

The TDOA can also be estimated with the **generalized cross-correlation** (GCC) that offers more flexibility.

Estimation of TDOA Using GCC

- The short-time Fourier transform (STFT) is applied to the two signals.
- The two signals may (or may not) be filtered.
- The cross power spectral density (cross-PSD) is estimated at each frequency.
- The TDOA corresponds to the phase offset that maximizes the cross-PSD over all the frequencies.

Short-time Fourier Transform (STFT)

- STFT of signal $x(t)$ at **frame** l has a complex value for each **frequency** f:

$$X(1,l), \ldots, X(f,l), \ldots, X(F,l)$$

- This is a complex-valued vector of dimension F.
- For each frequency f and frame l, $X(f,l)$ is a complex number:

$$X(f,l) = \mathcal{X}(f,l)e^{\alpha} = \mathcal{X}(f,l)\,(\cos\alpha + j\sin\alpha)$$

- $\mathcal{X}(f,l)$ is the **magnitude** and
- α is the **phase**.
- $j^2 = -1$.

Microphone Signals in the Spectral Domain

- STFT vector of the left-microphone signal $x_1(t)$:

$$X_1(f, l), f = 1 \dots F$$

- STFT vector of the right-microphone signal $x_2(t - t')$:

$$X_2(f, l)e^{\alpha(t')}, f = 1 \dots F$$

- This results from the time-shifting theorem.
- t' corresponds to the unknown TDOA.

Cross Power Spectral Density

$$\Phi_{x_1,x_2}(1) = \frac{1}{L}\sum_{l=1}^{L} X_1(1,l)X_2^*(1,l)$$

$$\vdots$$

$$\Phi_{x_1,x_2}(f) = \frac{1}{L}\sum_{l=1}^{L} X_1(f,l)X_2^*(f,l)$$

$$\vdots$$

$$\Phi_{x_1,x_2}(F) = \frac{1}{L}\sum_{l=1}^{L} X_1(F,l)X_2^*(f,l)$$

Session Summary

A number of important concepts were addressed:

- Cross-correlation in the temporal domain
- The short-time Fourier transform of a microphone signal
- The TDOA in the spectral domain
- Cross-power spectral density of two signals

3. Sound-Source Localization

1. Time difference of arrival (TDOA)
2. Estimation of TDOA by cross-correlation
3. **Estimation of TDOA in the spectral domain**
4. The geometry of two microphones
5. Direction of arrival
6. Using more than two microphones
7. Embedding the microphones in a robot head
8. Learning a sound propagation model
9. Predicting direction of a sound with a robot head
10. Example of sound direction estimation

Cross-Correlation in the Frequency Domain

A time delay corresponds to a phase shift in the spectral domain:

$$R_{x_1,x_2}(t') = \sum_{f=1}^{F} \Phi_{x_1,x_2}(f) e^{\alpha(t')}$$

TDOA estimation:

$$\hat{\tau} = \underset{t'}{\operatorname{argmax}}\, R_{x_1,x_2}(t')$$

Prefiltering the Microphone Signals

Convolution between signals and filters:

$$y_1(t) = h_1(t) \star x_1(t)$$
$$y_2(t) = h_2(t) \star x_2(t)$$

In the spectral domain, for each frequency f and frame l:

$$Y_1(f, l) = H_1(f)X_1(f, l)$$
$$Y_2(f, l) = H_2(f)X_2(f, l)$$

Generalized Cross-Correlation

$$\Phi_{y_1,y_2}(f) = \frac{1}{L}\sum_{l=1}^{L} H_1(f)H_2^*(f)\Phi_{x_1,x_2}(f)$$

$$R_{y_1,y_2}(t') = \sum_{f=1}^{F} \Phi_{y_1,y_2}(f,t')e^{\alpha(t')}$$

TDOA estimation:

$$\hat{\tau} = \underset{t'}{\text{argmax}}\, R_{y_1,y_2}(t')$$

Filter Design

$\Psi(f) = H_1(f)H_2^*(f)$ is the frequency weighting. We have:

- The *Smooth COherent Transform* (SCOT):

$$\Psi(f) = (\Phi_{x_1,x_1}\Phi_{x_2,x_2})^{-1/2}$$

- The *PHAse Transform* (PHAT):

$$\Psi(f) = |\Phi_{x_1,x_1}|^{-1}$$

TDOA Estimation with PHAT

The PHAT cross-correlation:

$$R_{y_1,y_2}(t') = \sum_{f=1}^{F} \frac{\Phi_{x_1,x_2}(f)}{|\Phi_{x_1,x_1}(f)|} e^{\alpha(t')}$$

The TDOA estimation:

$$\hat{\tau} = \underset{t'}{\operatorname{argmax}}\, R_{y_1,y_2}(t')$$

Session Summary

- Cross-correlation in the frequency domain
- Signal filtering
- Generalized cross-correlation
- Choosing a filter
- TDOA estimation in the spectral domain

3. Sound-Source Localization

1. Time difference of arrival (TDOA)
2. Estimation of TDOA by cross-correlation
3. Estimation of TDOA in the spectral domain
4. **The geometry of two microphones**
5. Direction of arrival
6. Using more than two microphones
7. Embedding the microphones in a robot head
8. Learning a sound propagation model
9. Predicting direction of a sound with a robot head
10. Example of sound direction estimation

Finding the Position of the Sound Source

The position of the sound source is given by solving the equation:

$$\hat{\tau} = \tau(\boldsymbol{S})$$

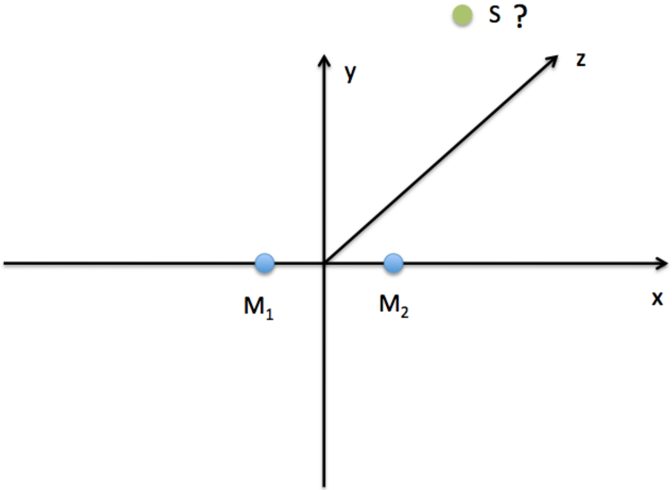
In more detail:

$$\hat{\tau} = \frac{\|\boldsymbol{S} - \boldsymbol{M}_1\| - \|\boldsymbol{S} - \boldsymbol{M}_2\|}{\nu}$$

Let $M_1 = (-m, 0, 0)$ and $M_2 = (m, 0, 0)$.

$S = (x, y, z)$ is the unknown source location.

Finding the Position of the Sound Source



Finding the Position of the Sound Source

The equation develops as:

$$\nu\hat{\tau} = \left((x+m)^2 + y^2 + z^2\right)^{1/2} - \left((x-m)^2 + y^2 + z^2\right)^{1/2}$$

which be written as:

$$\frac{x^2}{\frac{\nu^2\hat{\tau}^2}{4}} - \frac{y^2}{\frac{4m^2-\nu^2\hat{\tau}^2}{4}} - \frac{z^2}{\frac{4m^2-\nu^2\hat{\tau}^2}{4}} = 1$$

Two-Sheet Hyperboloid $\mathcal{H}$

This equation has the following structure:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{b^2} = 1$$

This is a *two-sheet hyperboloid*, $\mathcal{H}$, with foci in $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$.

The sound source $S = (x, y, z)$ should lie on $\mathcal{H}$.

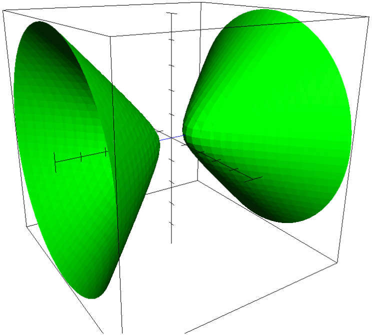
Where is the Sound Source?

An estimated time difference of arrival $\hat{\tau}$ defines a surface $\mathcal{H}_{\hat{\tau}}$.

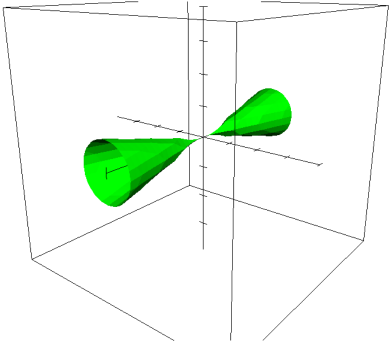
The surface exists only if:

$$0 < |\hat{\tau}| < \frac{\|M_1 - M_2\|}{\nu}$$

Two-Sheet Hyperboloids



small TDOA



large TDOA

Choosing a Sheet

- The hyperboloid has to sheets (branches), one with $x < 0$ and one with $x > 0$.
- The source can be on either one of these two branches. From the above equations, we can write:

$$x = \frac{\nu\hat{\tau}\left(2\|\boldsymbol{S} - \boldsymbol{M}_1\| - \nu\hat{\tau}\right)}{4m}$$

The sign of x, that depends on $\hat{\tau}$, selects one of the two sheets.

Session Summary

- The TDOA is combined with the geometry of two microphone and an emitting source
- For any two given microphones, the TDOA has lower and upper bounds
- The source is “somewhere" onto a two-sheet hyberboloid

3. Sound-Source Localization

1. Time difference of arrival (TDOA)
2. Estimation of TDOA by cross-correlation
3. Estimation of TDOA in the spectral domain
4. The geometry of two microphones
5. **Direction of arrival**
6. Using more than two microphones
7. Embedding the microphones in a robot head
8. Learning a sound propagation model
9. Predicting direction of a sound with a robot head
10. Example of sound direction estimation

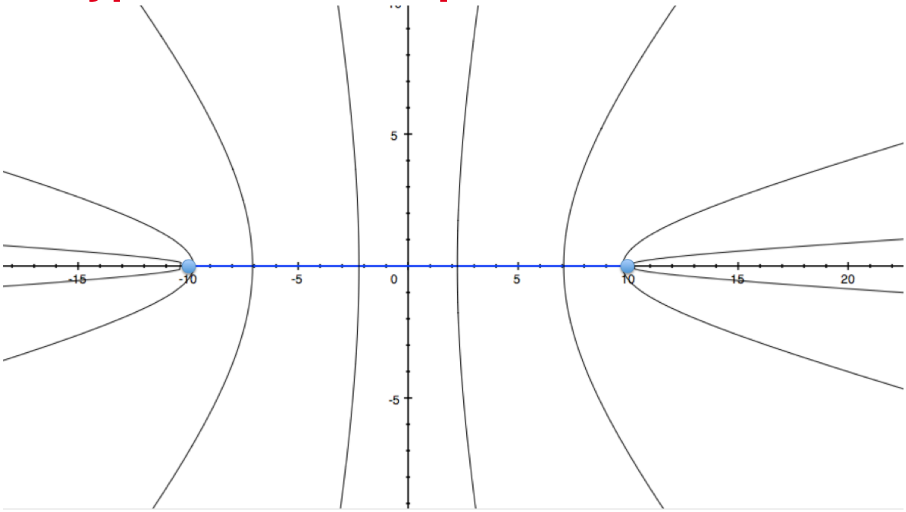
The Azimuthal Plane

We consider an *azimuthal* plane, namely a plane that contains the two microphones, namely the plane $y = 0$:

$$\frac{x^2}{a^2} - \frac{z^2}{b^2} = 1$$

There is a hyperbola for each value of the TDOA

Each Hyperbola Corresponds to a TDOA



Direction of Arrival

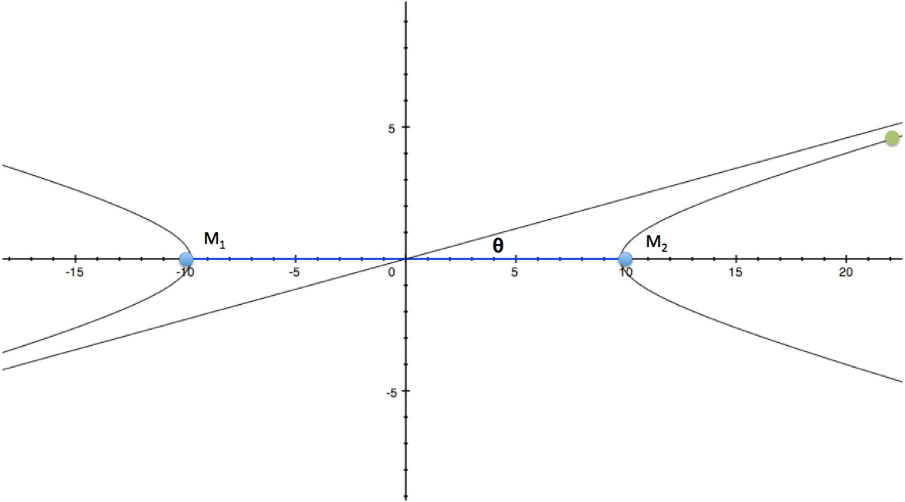
The slopes of the two *asymptotes* of the hyperbole are

$$\tan\theta = \pm\frac{b}{a} = \pm\frac{\sqrt{4m^2 - (\nu\hat{\tau})^2}}{\nu\hat{\tau}}$$

The value of $\hat{\tau}$ selects one of the branches ($x > 0$ or $x < 0$).

The sign of the slope selects front of back side.

Direction of Arrival



Session Summary

- The space is projected onto the azimuthal plane
- In this plane the source lies on a two-branch hyperbola
- The sound direction (azimuth) can be estimated from the asymptote
- One has to solve the left-right and front-back ambiguities.

3. Sound-Source Localization

1. Time difference of arrival (TDOA)
2. Estimation of TDOA by cross-correlation
3. Estimation of TDOA in the spectral domain
4. The geometry of two microphones
5. Direction of arrival
6. **Using more than two microphones**
7. Embedding the microphones in a robot head
8. Learning a sound propagation model
9. Predicting direction of a sound with a robot head
10. Example of sound direction estimation

Increasing the Number of Microphones

- The time difference of arrival (TDOA) must be robustly estimated
- The TDOA is estimated using GCC (generalized cross-correlation)
- The geometry determines the sound direction in the azimuthal plane
- What about increasing the number of microphones?
- Can we find the sound location in 3D?

Microphone Arrays

- Consider M non coplanar microphones ($M \geq 4$). There is one equation for each microphone:

$$x_i(t) = s(t - \tau_i) + n_i(t), \quad \forall i, 1 \leq i \leq M$$

- as well as a TDOA between any microphone pair:

$$\tau_{i,j} = \tau_i - \tau_j$$

How Many Hyperboloids?

- For any three microphones i, j, k:

$$\tau_{i,j} = \tau_{i,k} + \tau_{k,j}$$

- Consequence: Only $M - 1$ pairs are needed among all the $M(M-1)/2$ microphone pairs.

M Microphones in General Position

$$\nu\hat{\tau}_{1,2} = \|\boldsymbol{S} - \boldsymbol{M}_1\| - \|\boldsymbol{S} - \boldsymbol{M}_2\|$$
$$\vdots$$
$$\nu\hat{\tau}_{1,M} = \|\boldsymbol{S} - \boldsymbol{M}_1\| - \|\boldsymbol{S} - \boldsymbol{M}_M\|$$

This system of equations has a unique solution $\boldsymbol{S}$ for a set of TDOAs $\hat{\tau}_{1,2}, \ldots, \hat{\tau}_{1,M}$

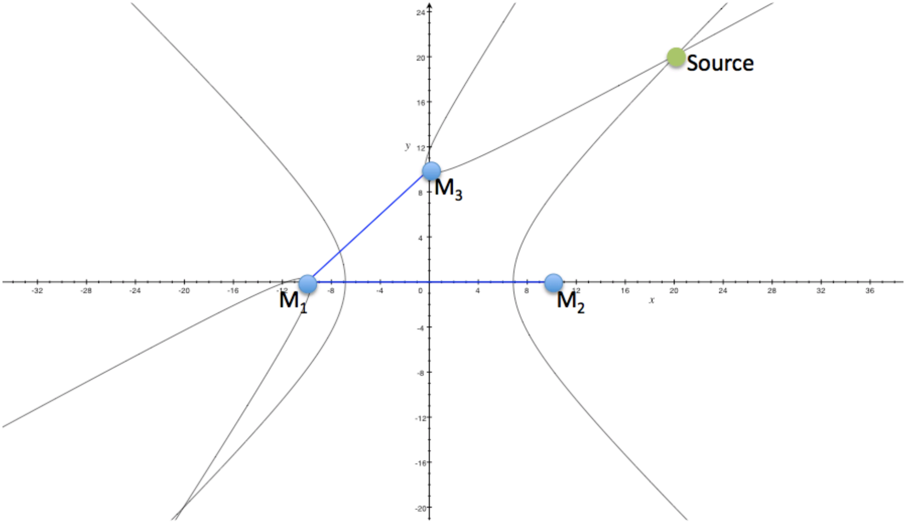
Example: Three *Planar* Microphones

- $\boldsymbol{M}_1 = (-m, 0, 0)$, $\boldsymbol{M}_2 = (m, 0, 0)$, $\boldsymbol{M}_3 = (0, m, 0)$
- We obtain two hyperboles:

$$\nu\hat{\tau}_{1,2} = \|\boldsymbol{S} - \boldsymbol{M}_1\| - \|\boldsymbol{S} - \boldsymbol{M}_2\|$$
$$\nu\hat{\tau}_{1,3} = \|\boldsymbol{S} - \boldsymbol{M}_1\| - \|\boldsymbol{S} - \boldsymbol{M}_3\|$$

- The time delays are estimated using GCC (generalized cross-correlation) applied to the microphone pairs $(1, 2)$ and $(1, 3)$.

Example With Three Microphones



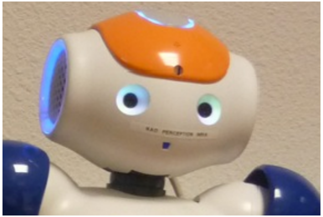
Session Summary

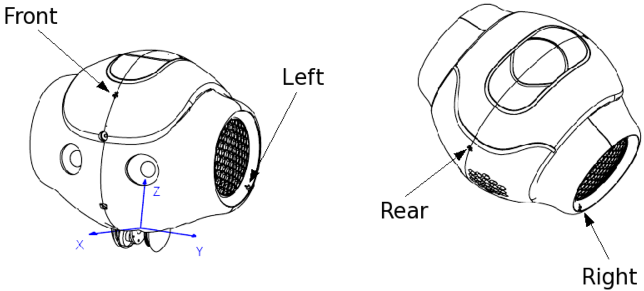
- A microphone array is composed of four non-coplanar or three planar microphones
- Each microphone and TDOA forms a two-branch hyperboloid
- The sound source is at the intersection of these hyperboloids
- There are two solutions: positive and negative

3. Sound-Source Localization

1. Time difference of arrival (TDOA)
2. Estimation of TDOA by cross-correlation
3. Estimation of TDOA in the spectral domain
4. The geometry of two microphones
5. Direction of arrival
6. Using more than two microphones
7. **Embedding the microphones in a robot head**
8. Learning a sound propagation model
9. Predicting direction of a sound with a robot head
10. Example of sound direction estimation

Microphones Embedded in a Robot Head





Sound Localization with the NAO Robot

NAO has four non-coplanar microphones

The microphones signals are corrupted by various sources of noise:

- environmental noise
- reverberations
- robot noise (motors, fans, etc.)

A robust localization method is needed.

Far-field Assumption

The sound source, $\boldsymbol{S}$ is parameterized by its direction with respect to the robot head:

- Euclidean coordinates: x, y, z.
- Spherical coordinates: $r \sin\alpha \sin\beta,\; r \sin\alpha \cos\beta,\; r \cos\alpha$
- We make a **far field** assumption: r is large with respect to the head size:

We only estimate α (azimuth) and β (elevation), the source direction.

TDOA Estimation Between Microphone Pairs

- The TDOA between two microphone pairs can be estimated using cross-correlation (CC), normalized cross-correlation (NCC), generalized cross-correlation (GCC), for example:

$$Q_{ij}(t') = \text{NCC}(x_i(t), x_j(t - t'))$$

- For a **known** sound direction (α, β), one can estimate the corresponding TDOA:

$$\tau_{i,j} = \underset{t'}{\text{argmax}}\, Q_{ij}(t')$$

A More General Sound Propagation Model

These $(\alpha, \beta) \leftrightarrow \tau$ input-output values do not necessarily verify the direct-path propagation model:

$$\tau_{i,j} \stackrel{?}{=} \frac{1}{\nu} \left(\|\boldsymbol{S}(\alpha, \beta) - \boldsymbol{M}_i\| - \|\boldsymbol{S}(\alpha, \beta) - \boldsymbol{M}_j\| \right)$$

More generally, we seek a function f that **verifies** such a relationship between known input and known output:

$$\tau = f(\alpha, \beta)$$

Session Summary

- Nao has four microphones
- The microphones capture noise coming form the electronics inside the head
- The head plays the role of an acoustic filter
- The propagation model is approximated by a linear regression function

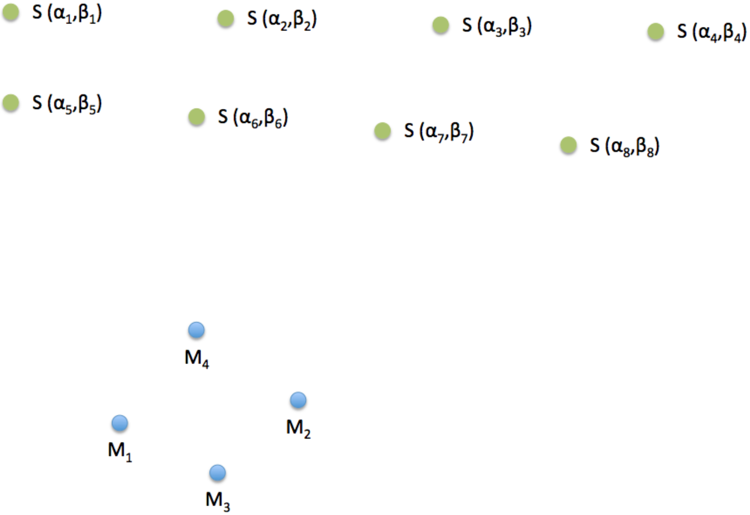
3. Sound-Source Localization

1. Time difference of arrival (TDOA)
2. Estimation of TDOA by cross-correlation
3. Estimation of TDOA in the spectral domain
4. The geometry of two microphones
5. Direction of arrival
6. Using more than two microphones
7. Embedding the microphones in a robot head
8. **Learning a sound propagation model**
9. Predicting direction of a sound with a robot head
10. Example of sound direction estimation

Learning a Regression Function

- This problem of finding f in
$$\tau = f(\alpha, \beta)$$
falls in the category of **supervised learning.**
- $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ can be **learned** by **linear regression**, for example:
$$\tau = a\alpha^2 + b\beta^2 + c\alpha + d\beta + e$$
- The parameters (a, b, c, d, e) characterize the regression and they must be estimated.

Learning a Propagation Function



Learning a Propagation Function

The source is placed in K different source directions, and for each direction the TDOA is estimated.

We collect **input-output** data: $\{\tau_k, \alpha_k, \beta_k\}_{k=1}^{k=K}$:

$$\tau_1 = a\alpha_1^2 + b\beta_1^2 + c\alpha_1 + d\beta_1 + e$$

$$\vdots$$

$$\tau_k = a\alpha_k^2 + b\beta_k^2 + c\alpha_1 + d\beta_k + e$$

$$\vdots$$

$$\tau_K = a\alpha_K^2 + b\beta_K^2 + c\alpha_K + d\beta_K + e$$

Learning a Propagation Function

Matrix notation, $\mathbf{A}\boldsymbol{y} = \boldsymbol{\tau}$ with:

$$\boldsymbol{\tau} = \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_k \\ \vdots \\ \tau_K \end{pmatrix} \quad \boldsymbol{y} = \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} \quad \mathbf{A} = \begin{pmatrix} \alpha_1^2 & \beta_1^2 & \alpha_1 & \beta_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_k^2 & \beta_k^2 & \alpha_k & \beta_k & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_K^2 & \beta_K^2 & \alpha_K & \beta_K & 1 \end{pmatrix}$$

Estimation of the Parameters

This matrix-vector equation has a least-square formulation:

$$\min_{\boldsymbol{y}} \|\mathbf{A}\boldsymbol{y} - \boldsymbol{\tau}\|^2$$

The derivative of this error function vanishes for:

$$\begin{aligned} \mathbf{A}^\top\mathbf{A}\boldsymbol{y} &= \mathbf{A}^\top\boldsymbol{\tau} \\ \boldsymbol{y} &= \left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top\boldsymbol{\tau} \end{aligned}$$

Least-square Solution

The regression parameters are obtained with:

$$\begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_k \\ \vdots \\ \tau_K \end{pmatrix}$$

Session Summary

- The regression function can be learnt from input-output data
- The source is placed in known positions (output)
- For each position a TDOA is estimated (input)
- The model parameters are estimated from these input-output pairs

3. Sound-Source Localization

1. Time difference of arrival (TDOA)
2. Estimation of TDOA by cross-correlation
3. Estimation of TDOA in the spectral domain
4. The geometry of two microphones
5. Direction of arrival
6. Using more than two microphones
7. Embedding the microphones in a robot head
8. Learning a sound propagation model
9. **Predicting direction of a sound with a robot head**
10. Example of sound direction estimation

Link Between Sound Direction and Correlation

- For each microphone pair, m_i and m_j, a propagation function is learnt:

$$\tau_{i,j} = f_{ij}(\alpha, \beta)$$

- Combine this relationship with cross-correlation:

$$Q_{ij}(t') = \text{NCC}(x_i(t), x_j(t - t'))$$

- This tells us that there is a correlation value for a sound direction:

$$Q_{ij}(\alpha, \beta) = \text{NCC}(x_i(t), x_j(t - f_{i,j}(\alpha, \beta)))$$

An Alternative to Sound-Source Direction Estimation

- **Offline learning:** Estimate the regression functions f_{ij} for each microphone pair
- **Online prediction:** Estimate the direction of the sound by combining the cross-correlation functions between all microphone pairs and for all possible directions.

Learning a Sound Propagation Model

For each microphone pair i and j:

1. Place a sound source in known directions $(\alpha_1, \beta_1), \ldots, (\alpha_k, \beta_k), \ldots, (\alpha_K, \beta_K)$.
2. For each sound source in direction (α_k, β_k) estimate the TDOA with:

$$\tau_k = \operatorname*{argmax}_{t'} \mathrm{NCC}(x_i(t), x_j(t - t'))$$

3. Collect a set of K input-output pairs $\tau_k \leftrightarrow (\alpha_k, \beta_k)$.
4. Estimate the parameters of the linear regression f_{ij}, namely $a_{ij}, b_{ij}, c_{ij}, d_{ij}, e_{ij}$.

Predicting a Sound Direction

A method to find an unknown direction (azimuth α and elevation β) of a sound source:

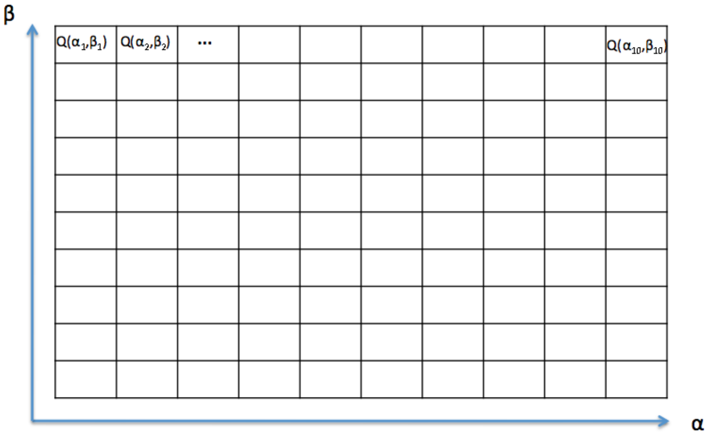
1. Discretize the space of possible directions.
2. For each direction (α, β) and for each microphone pair (m_i, m_j), compute:

$$Q_{ij}(\alpha, \beta) = \mathrm{NCC}(x_i(t), x_j(t - f_{i,j}(\alpha, \beta)))$$

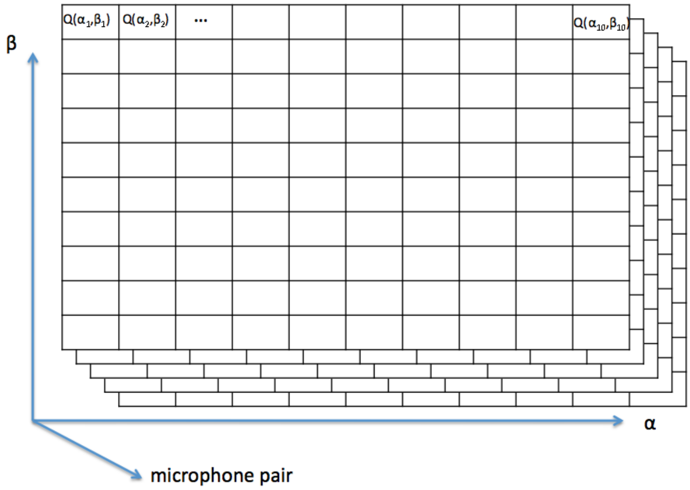
3. The (α, β) values that optimize the correlation over all the microphone pairs is the estimated sound direction:

$$\hat{\alpha}, \hat{\beta} = \underset{\alpha,\beta}{\mathrm{argmax}} \sum_{i \neq j} Q_{ij}(\alpha, \beta)$$

Correlation Map for One Microphone Pair



Correlation Map for Many Microphone Pairs



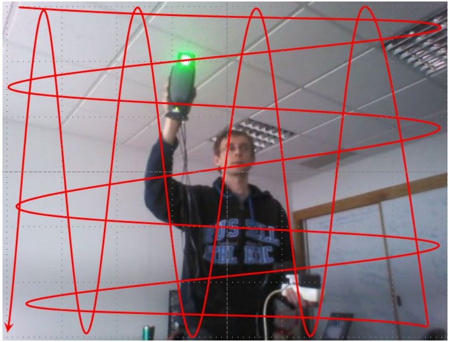
Session Summary

- Practical method for estimating the TDOA
- The linear regression parameters are estimated for each microphone pair
- The cross-correlation is estimated for each direction and for each pair
- The cross-correlations from all the pairs are summed up
- The sound-direction corresponds to the best cross-correlation

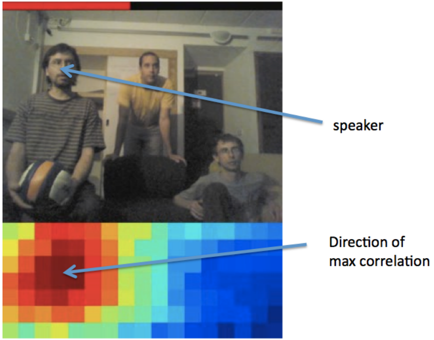
3. Sound-Source Localization

1. Time difference of arrival (TDOA)
2. Estimation of TDOA by cross-correlation
3. Estimation of TDOA in the spectral domain
4. The geometry of two microphones
5. Direction of arrival
6. Using more than two microphones
7. Embedding the microphones in a robot head
8. Learning a sound propagation model
9. Predicting direction of a sound with a robot head
10. **Example of sound direction estimation**

Example



Learning



Prediction

Nao Localizes Speech

INSERT sound-direction-correlation.avi

Session Summary

- The method was used to estimate a propagation model for NAO
- This model can then be used to estimate the direction of a speaking person
- The estimated direction corresponds to the maximum of the cross-correlation across all the microphone pairs

Week Summary

- The time difference of arrival between a microphone pair can be estimated in the temporal domain or in the frequency domain.
- General cross-correlation is a flexible tool to estimate the time difference of arrival.
- With two or more microphones it is possible to find the sound-source position in 3D.

Week Summary (Continued)

- The direct propagation equations lead to solving a non-linear optimization problem.
- It is possible to adopt a supervised learning approach that replaces the propagation model with a linear regression function.
- This yields a methodology for sound-direction estimation with a microphone array embedded in a robot head.