Precedence Effect

Beamforming
Demo of the Franssen effect

• Demonstrates precedence
Introduction to 3D Audio (capture)

- Directivity of microphone.
  - Omni-directional
    - Advantages are that microphones capture all sound including that of interest
  - Directional
    - Capture sound from a preferred direction
Beamforming

- Given $N$ microphones combine their signals in a way that some desired result occurs.
- Word arises from the use of parabolic reflectors to form pencil “beams” for broadcast and reception.
- Alternate word: “spatial filtering”
- Towed array and fixed array sonars.
Delay and Sum Beamforming

- If the source location is known, delays relative to the microphone can be obtained.
- Signal $x$ at location $s$ arrives at microphone $m_i$ as
  
  \[ x \left( t - \frac{|s - m_i|}{c} \right) \]

- Signals at microphones can be appropriately delayed and weighted.
- Output signal is
  
  \[ y(k) = \frac{1}{N} \sum_{l=1}^{N} w_l^* x_l(k - \Delta_l) \]

\[ \Delta_l = \frac{|s - x_l|}{c} \quad w_l = \frac{1}{|s - x_l|} \]
Behavior of simple beamformer

• Usually source is assumed to be far away.
  – Weights are approximately the same in this case
• Signal from source direction adds in phase
  – So the signal is amplified $N$ times
• Signals from other directions will add up with random phase and the power will decrease by a factor of $1/N$
• Directivity index is a measure of the gain of the array in the look direction (location of the delays) in decibels
  – For $N$ microphones $10 \log_{10} (N)$
• Requires an ability to store the signal (at least for max $\{\Delta_i\}$
• Jargon: “taps” number of samples in time that are stored
• Data independent beamforming:
  – Weights are fixed

• Data dependent (adaptive)
  – Weights change according to the data

• Simple example:
  – Fixed: Delay and sum looking at a particular point (direction)
  – Adaptive: Delay and sum looking at a particular moving source
More general beamforming

- Suppose we want to take advantage of the stored data
- Write the beamformer output as

\[ y(k) = \sum_{l=1}^{N} \sum_{m=k-M}^{k} w_{lm}^* x_l(k - m) \]

- Can be written as \( y = w^H x \)
- Take Fourier transform of the weights and the signal
Speech and Audio Processing

Microphone Array Processing
Slides adapted from those of Marc Moonen/Simon Doclo
Dept. E.E./ESAT, K.U.Leuven
www.esat.kuleuven.ac.be/~moonen/
Introduction

- Each microphone is characterized by a `directivity pattern’ which specifies the gain (& phase shift) that the microphone gives to a signal coming from a certain direction (`angle-of-arrival’).
- Directivity pattern is a function of angle-of-arrival and frequency.
- Directivity pattern is a (physical) microphone design issue.
Introduction

• By weighting/filtering and summing signals from different microphones, a ‘virtual’ directivity pattern may be produced

\[
\begin{align*}
&\Sigma \\
&f_1[k] \quad y_1[k] \\
&f_2[k] \quad y_2[k] \\
&\vdots \\
&f_M[k] \quad y_M[k]
\end{align*}
\]

• This is ‘spatial filtering’ and ‘spatial filter design’, based on given microphone characteristics (with correspondences to traditional (spectral) filter design)

• Applications: teleconferencing, hands-free telephony, hearing aids, voice-controlled systems, …
Introduction

- An important aspect is that different microphones in a microphone array are in different positions/locations, hence receive different signals.
- Example: linear array, with uniform inter-microphone distances, under far-field (plane waveforms) conditions. Each microphone receives the same signal, but with different delays.
- Hence ‘spatial filter design’ based on microphone characteristics + microphone array configuration. Often simple assumptions are made, e.g. microphone gain = 1 for all frequencies and all angles.
Introduction

- **Background/history:** ideas borrowed from antenna array design/processing for RADAR & (later) wireless comms.

- **Microphone array processing considerably more difficult than antenna array processing:**
  - narrowband radio signals versus broadband audio signals
  - far-field (plane wavefronts) versus near-field (spherical wavefronts)
  - pure-delay environment versus multi-path reverberant environment

- **Classification:**
  - **fixed beamforming:** data-independent, fixed filters $f_m[k]$  
    *e.g.* delay-and-sum, weighted-sum, filter-and-sum
  - **adaptive beamforming:** data-dependent, adaptive filters $f_m[k]$  
    *e.g.* LCMV-beamformer,
Beamforming basics

General form: filter-and-sum beamformer

- linear microphone array with $M$ microphones and inter-micr. distance $d_m$
- Microphone gains are assumed to be equal to 1 for all freqs./angles
  (otherwise, this characteristic is to be included in the steering vector, see next page)
- source $S(\omega)$ at angle $\theta$ (far-field, no multipath)
- filters $f_m[k]$ with filter length $L$

Terminology: `Broadside' direction: $\theta = 90^\circ$  `End-fire' direction: $\theta = 0^\circ$

\[
F_m(\omega) = \sum_{k=0}^{L-1} f_m[k]e^{-jk\omega}
\]
Near-field beamforming

- Far-field assumptions not valid for sources close to microphone array
  - spherical wavefronts instead of planar waveforms
  - include attenuation of signals
  - 3 spherical coordinates $\theta, \phi, r$ (=position $q$) instead of 1 coordinate $\theta$

- Different steering vector:

\[
d(\omega, \theta) \rightarrow \mathbf{d}(\omega, q) = \begin{bmatrix} a_1 e^{-j\omega \tau_1(q)} & a_2 e^{-j\omega \tau_2(q)} & \ldots & a_M e^{-j\omega \tau_M(q)} \end{bmatrix}^T
\]

\[
a_m = \frac{\|q - p_{\text{ref}}\|}{\|q - p_m\|}, \quad \tau_m(q) = \frac{\|q - p_{\text{ref}}\| - \|q - p_m\|}{c} f_s
\]

with $q$ position of source
$p_{\text{ref}}$ position of reference microphone
$p_m$ position of $m^{th}$ microphone
Beamforming basics

Data model:

- **Microphone signals** are delayed versions of $S(\omega)$

$$Y_m(\omega, \theta) = e^{-j\omega \tau_m(\theta)} . S(\omega)$$

$$y_m[k] = s[k - \tau_m(\theta)]$$

$$\tau_m(\theta) = \frac{d_m \cos \theta}{c} f_s$$

- Stack all microphone signals in a vector

$$Y(\omega, \theta) = d(\omega, \theta) . S(\omega)$$

$$d(\omega, \theta) = \begin{bmatrix} 1 & e^{-j\omega \tau_2(\theta)} & \ldots & e^{-j\omega \tau_M(\theta)} \end{bmatrix}^T$$

d is `steering vector’

- **Output signal** $Z(\omega, \theta)$ is

$$Z(\omega, \theta) = \sum_{m=1}^{M} F_m^*(\omega) Y_m(\omega, \theta) = F^H(\omega) \cdot Y(\omega, \theta)$$
Beamforming basics

Data model:

• **Microphone signals** are corrupted by additive noise

\[ y_m[k] = s[k - \tau_m(\theta)] + n_m[k] \]

• **Stack** all noise signals in a vector

\[ \mathbf{N}(\omega) = \begin{bmatrix} N_1(\omega) & N_2(\omega) & \ldots & N_M(\omega) \end{bmatrix}^T \]

• Define **noise correlation matrix** as

\[ \Phi_{NN}(\omega) = \mathbb{E}\{\mathbf{N}(\omega)\mathbf{N}(\omega)^H\} \]

• **We assume** noise field is homogeneous, i.e. diagonal elements of

\[ \Phi_{ii}(\omega) = \Phi_{\text{noise}}(\omega) \quad \forall i \]

• **Then** noise coherence matrix is

\[ \Gamma_{NN}(\omega) = \frac{1}{\phi_{\text{noise}}(\omega)} \cdot \Phi_{NN}(\omega) \]
Beamforming basics

Definitions:

• **Spatial directivity pattern**: ‘transfer function’ for source at angle $\theta$

\[ H(\omega, \theta) = \frac{Z(\omega, \theta)}{S(\omega)} = \sum_{m=1}^{M} F_m^*(\omega) e^{-j\omega \tau_m(\theta)} = F^H(\omega) \cdot d(\omega, \theta) \]

• **Steering direction** $\theta_{\text{max}}$ = angle $\theta$ with maximum amplification (for 1 freq.)

• **Beamwidth** = region around $\theta_{\text{max}}$ with amplification $> -3\text{dB}$ (for 1 freq.)

• **Array Gain** = improvement in SNR

\[ G(\omega, \theta) = \frac{\text{SNR}_{\text{Output}}}{\text{SNR}_{\text{Input}}} = \frac{|F^H(\omega) \cdot d(\omega, \theta)|^2}{F^H(\omega) \cdot \Gamma_{NN}(\omega) \cdot F(\omega)} \]
Beamforming basics

Definitions:

- **Array Gain** = improvement in SNR

\[
G(\omega, \theta) = \frac{SNR_{Output}}{SNR_{Input}} = \frac{|F^H(\omega) \cdot d(\omega, \theta)|^2}{F^H(\omega) \cdot \Gamma_{NN}(\omega) \cdot F(\omega)}
\]

- **Directivity** = array gain for \( \theta_{max} \) and diffuse noise (=coming from all directions)

\[
DI(\omega) = \frac{|F^H(\omega) \cdot d(\omega, \theta_{max})|^2}{F^H(\omega) \cdot \Gamma_{diffuse}^{NN}(\omega) \cdot F(\omega)}
\]

- **White Noise Gain** = array gain for \( \theta_{max} \) and spatially uncorrelated noise (\( \Gamma_{NN} = I \))
  (e.g. sensor noise)

\[
WNG(\omega) = \frac{|F^H(\omega) \cdot d(\omega, \theta_{max})|^2}{F^H(\omega) \cdot F(\omega)}
\]

ps: often used as a measure for robustness
Delay-and-sum beamforming

- Microphone signals are delayed and summed together. Array can be virtually steered to angle $\psi$.

- Angular selectivity is obtained, based on constructive (for $\theta = \psi$) and destructive (for $\theta \neq \psi$) interference. For $\theta = \psi$, this is referred to as a `matched filter':

- For uniform linear array:
  \[ d_m = (m-1)d \quad \Delta_m = (m-1)\Delta \]

- PS: $H(\omega, \theta) = H(\omega, -\theta)$ (if microphone characteristics are ignored)
Delay-and-sum beamforming

- Spatial directivity pattern $H(\omega, \theta)$ for **uniform** DS-beamformer

\[
H(\omega, \theta) = \sum_{m=1}^{M} e^{-j(m-1)(\omega \frac{d(\cos \theta - \cos \psi)}{c} f_s)} = \frac{e^{-j\gamma/2} \sin(M\gamma / 2)}{e^{-j\gamma/2} \sin(\gamma / 2)}
\]

- $H(\omega, \theta)$ has sinc-like shape and is frequency-dependent

Spatial directivity pattern for $f=5000$ Hz

- $M=5$ microphones
- $d=3 \text{ cm}$ inter-microphone distance
- $\psi=60^\circ$ steering angle
- $f_s=16 \text{ kHz}$ sampling frequency

Spatial directivity pattern for $f=5000$ Hz

- $\psi=60^\circ$
- Wavelength $= 4 \text{ cm}$
Delay-and-sum beamforming

- For \( f \geq \frac{c}{d(1 + |\cos \psi|)} \) an ambiguity, called spatial aliasing, occurs.

This is analogous to time-domain aliasing where now the spatial sampling (=d) is too large.

Aliasing does not occur (for any \( \psi \)) if

\[
d \leq \frac{c}{f_s} = \frac{c}{2f_{max}} = \frac{\lambda_{min}}{2}
\]

\( M=5, \ \psi=60^\circ, f_s=16 \text{ kHz}, \ d=8 \text{ cm} \)

\[
f = \frac{c}{d(1 + \cos \psi)}
\]

Details...

\( H(\omega, \theta) = 1 \) iff \( \gamma = 2\pi p \) for integer \( p \)

1) \( \gamma = 0 \) for \( \theta = \psi \) (for all \( \omega \))

2) if \( \psi \leq \frac{\pi}{2} \) then \( \gamma = 2\pi \) occurs for \( \theta = \pi \) and \( f = \cdots = \frac{c}{d(1 + \cos \psi)} \)

3) if \( \psi \geq \frac{\pi}{2} \) then \( \gamma = 2\pi \) occurs for \( \theta = 0 \) and \( f = \cdots = \frac{c}{d(1 - \cos \psi)} \)
Delay-and-sum beamforming

- **Beamwidth**: for a uniform delay-and-sum beamformer
  \[
  BW \approx c \frac{\sqrt{96(1-\nu)}}{\omega d M} \sec \psi
  \]
  with e.g. \( \nu = \frac{1}{\sqrt{2}} \) (-3 dB)
  hence large dependence on # microphones, distance (compare p14 & 15) and frequency (e.g. BW infinitely large at DC)

- **Array topologies**:
  - Uniformly spaced arrays
  - Nested (logarithmic) arrays (small \( d \) for high \( \omega \), large \( d \) for small \( \omega \))
  - Planar / 3D-arrays
Weighted-sum beamforming

- Sensor-dependent complex weight + delay (compare to p. 13)

- Weights added to allow for better beam shaping
- Design similar to traditional (spectral) filter design

\[ z[k] = \sum_{m=1}^{M} w_m \cdot y_m [k + \Delta_m] \]

\[ H(\omega, \theta) = \sum_{m=1}^{M} w_m \cdot e^{-j(m-1)\omega \frac{d(\cos \theta - \cos \psi)}{c} \cdot f_s} \]

Ex: Dolph-Chebyshev design:
beampattern with uniform sidelobe level (`equiripple')
Filter-and-sum beamforming

- Sensor-dependent filters implement frequency-dependent complex weights to obtain a desired response over the whole frequency/angle range of interest.

- Design strategies: desired beampattern is $P(\omega, \theta)$
  - Non-linear:
  $$\min_{f_m[k], m=1\ldots M} \int_{\theta_1}^{\theta_2} \int_{\omega_1}^{\omega_2} \left| H(\omega, \theta) - |P(\omega, \theta)| \right|^2 d\omega d\theta$$
  - Quadratic:
  $$\min_{f_m[k], m=1\ldots M} \int_{\theta_1}^{\theta_2} \int_{\omega_1}^{\omega_2} \left| H(\omega, \theta) - P(\omega, \theta) \right|^2 d\omega d\theta$$
  - Frequency sampling, i.e. design weights for sampling frequencies $\omega_l$ and then interpolate:
  $$\min_{F_m(\omega_l), m=1\ldots M} \int_{\theta_1}^{\theta_2} \left| H(\omega_l, \theta) - P(\omega_l, \theta) \right|^2 d\theta$$

$$z[k] = \sum_{m=1}^{M} f_m[-k] \otimes y_m[k]$$

$$H(\omega, \theta) = \sum_{m=1}^{M} F_m(\omega) \cdot e^{-j(m-1)\omega d \cos \theta / c f_s}$$
Filter-and-sum beamforming

- Example-1: frequency-independent beamforming (continued)

\[ M = 8 \]
Logarithmic array
\[ L = 50 \]
\[ \psi = 90^\circ \]
\[ f_s = 8 \text{ kHz} \]
Filter-and-sum beamforming

- **Example-2: `superdirective’ beamforming**
  - Maximize directivity for known (diffuse) noise fields
  - Maximum directivity $=M^2$ obtained for diffuse noise & endfire steering ($\theta=0^\circ$)

**Design:** find $F(\omega)$ that maximizes

for given steering angle theta_max

- Optimal solution is

\[ F(\omega) = \alpha \cdot \Gamma^{-1}_{NN} (\omega) \cdot d(\omega, \theta_{\text{max}}) \]

- This is equivalent to minimization of noise output power, subject to unit response for steering angle (***)

\[ \min_{F(\omega)} F^H(\omega) \cdot \Gamma_{NN}(\omega) \cdot F(\omega), \text{ s.t. } F^H(\omega) \cdot d(\omega, \theta_{\text{max}}) = 1 \]

**PS:** Delay-and-sum beamformer similarly maximizes $\text{WNG} - 1$)

\[ F(\omega) = \alpha \cdot d(\omega, \theta_{\text{max}}) \]
Filter-and-sum beamforming

- **Example-2**: `superdirective` beamforming (continued)

Directivity patterns for endfire steering:

Superdirective beamformer has **highest** DI, but very poor WNG hence problems with robustness (e.g. sensor noise)!

\[ M^2 \]

- \( M=5 \)
- \( d=3 \text{ cm} \)
- \( \theta_{\text{max}}=0^\circ \)
- \( f_S=16 \text{ kHz} \)

\[ M^2 \rightarrow 6.99=10.\log(5) \]

**PS:** diffuse noise = white noise for high frequencies
**LCMV-beamforming**

- **Adaptive filter-and-sum structure:**
  - Aim is to minimize noise output power, while maintaining a chosen frequency response in a given look direction (and/or other linear constraints, see below)
  - This corresponds to operation of a superdirective array (see (**) p25), but now noise field is unknown
  - Implemented as adaptive filter (e.g. constrained LMS algorithm)
  - Notation:

\[
z[k] = f^T y[k] = \sum_{m=1}^{M} f_m^T y_m[k]
\]

\[
y[k] = \begin{bmatrix} y_1^T[k] & y_2^T[k] & \ldots & y_M^T[k] \end{bmatrix}^T
\]

\[
y_m[k] = \begin{bmatrix} y_m[k] & y_m[k-1] & \ldots & y_m[k-L+1] \end{bmatrix}^T
\]

\[
f = \begin{bmatrix} f_1^T & f_2^T & \ldots & f_M^T \end{bmatrix}^T
\]

\[
f_m = \begin{bmatrix} f_m[0] & f_m[1] & \ldots & f_m[L-1] \end{bmatrix}^T
\]
LCMV = \textbf{Linearly Constrained Minimum Variance}

- \( f \) designed to minimize variance of output \( z[k] \):
  \[
  \min_f E\{z^2[k]\} = \min_f f^T \cdot R_{yy}[k] \cdot f
  \]

- to avoid desired signal distortion/cancellation, add linear constraints:
  \[
  C^T \cdot f = b, \quad \text{with} \quad C \in \mathbb{R}^{M \times J}, \quad b \in \mathbb{R}^J
  \]

- if noise and speech are uncorrelated, constrained output power minimization corresponds to constrained noise power minimization

- Type of constraints:
  - Frequency response in look-direction. \( \text{Ex:} \sum_{m=1}^{M} F_m(z) = 1 \) (for broadside)
  - Point, line and derivative constraints \( (=L \text{ constraints}) \)

- Solution is (obtained using Lagrange-multipliers, etc..):
  \[
  f_{\text{opt}} = R_{yy}^{-1}[k] \cdot C \cdot \left( C^T \cdot R_{yy}^{-1}[k] \cdot C \right)^{-1} b
  \]