Multi-channel Noise Reduction

- When more than one microphone is available for sound pickup, the signal enhancement task may be facilitated by exploiting the multivariate deterministic and stochastic properties of the signal.

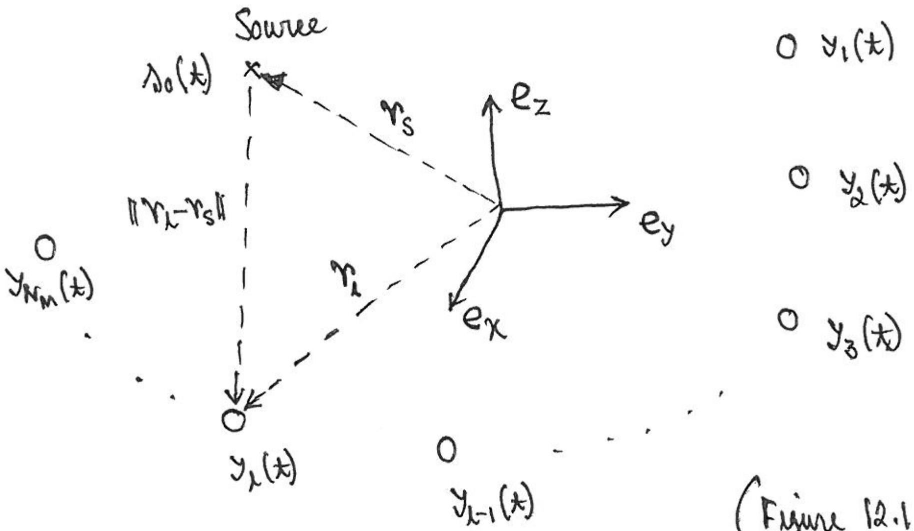
- From a deterministic viewpoint, the signals at the various microphones differ in that they arrive via different acoustic paths at the microphones and thus, also differ in their short term phase and amplitude.

- From a stochastic perspective, multi-channel methods allow the evaluation of the second-order and higher-order statistics of the spatial sound field.

- Sources which are close to array of microphones will generate mostly ~~of~~ coherent microphone signals, while distant and distributed sources lead to uncorrelated signals.

- Thus, short-term amplitude, short-term phase, and the statistics of the signals may be used to differentiate between sources and to perform source separation and enhancement.

Spatial sampling of Sound Fields



(Figure 12.1 of Reference Book)

- Reference coordinate $\{e_x, e_y, e_z\}$
- N_M microphones and single source
- w.r.t reference coordinate system, source position and "l"-th microphone are denoted by vectors r_s and r_l, $l = 1, 2, \cdots N_M$.
- In an ideal environment (no reverberation/noise), the microphone signals $y_l(t)$, $l = 1, 2, \cdots, N_M$, are delayed and attenuated version of the source signal $s_0(t)$,

$$y_l(t) = \frac{1}{\|r_l - r_s\|} s_0(t - \tau_l).$$

Here $\tau_l = \dfrac{\|r_l - r_s\|}{c}$ [c is sound speed]

- "reference point" — the origin of the reference coordinate
- Signal at reference point

$$s(t) = \frac{1}{\|r_s\|} s_0(t - \tau_0)$$

- The reference point ~~could~~ could be the geometric ~~of~~ center of the array.

- We are mainly interested in the delay of the signals relative to the signal which is received at the reference point.

- The relative signal delay $\Delta\tau_l$ is the time delay difference between the received signal at the reference point and at the "l"-th microphone.

$$\Delta\tau_l = \tau_0 - \tau_l = \frac{1}{c}\left(\|r_s\| - \|r_l - r_s\|\right)$$

- What is $y_l(t)$?

$$y_l(t) = \frac{1}{\|r_l - r_s\|}\, s_0(t - \tau_l)$$

$$= \frac{\|r_s\|}{\|r_l - r_s\|} \cdot \frac{1}{\|r_s\|} \cdot s_0\big(t - \tau_0 + \underbrace{\tau_0 - \tau_l}_{\Delta\tau_l}\big)$$

$$= \frac{\|r_s\|}{\|r_l - r_s\|}\, s(t + \Delta\tau_l) \quad \left[\because s(t) = \frac{1}{\|r_s\|}\, s_0(t - \tau_0)\right]$$

So, we represent $y_l(t)$ as a function of the signal at the reference point.

The Farfield Model

- The sound source is far away that the microphones.

~~That means~~ $\cancel{\|r_l\|}$

- Also assume that the reference point is close to microphones. That is $\|r_l\| \ll \|r_s\|$.

$$\therefore \quad \frac{\|r_s\|}{\|r_l - r_s\|} \approx 1$$

- But we <u>do not</u> assume that $\Delta\tau_l = \frac{1}{c}\left(\|r_s\| - \|r_l - r_s\|\right) = 0$. If we assume that then all $y_l(t)$ are same and of no use to bring diversity.

- So,

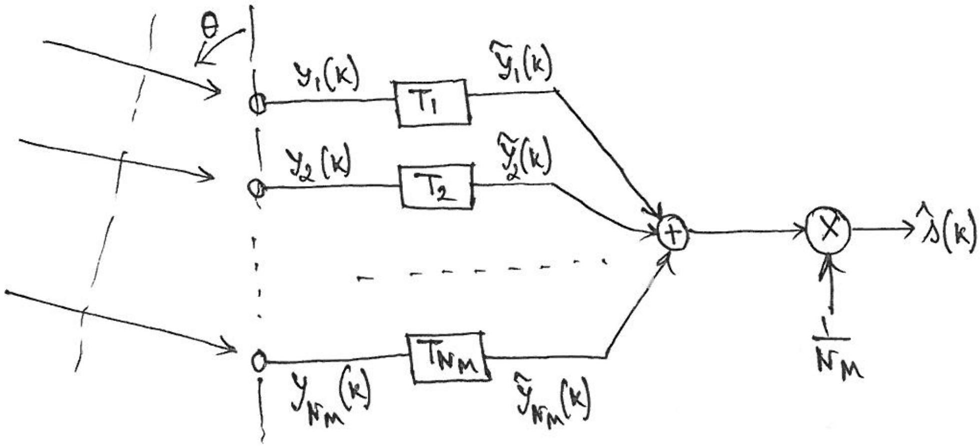
$$y_l(t) = \frac{\|r_s\|}{\|r_l - r_s\|} \, s(t + \Delta\tau_l)$$

$$\approx s(t + \Delta\tau_l) \quad \left[\text{as } \frac{\|r_s\|}{\|r_l - r_s\|} \approx 1\right]$$

Beamforming : (a) To combine the microphone signals such that a desired, spatial selectivity is achieved.
(b) In single source case, we need to form a beam of high gain in the source direction.

Delay-and-Sum Beamforming

- Assume insignificant reverberation and low noise.
- Then use "delay" for phase alignment.
- The resulting phase-aligned signals are added to form a single output signal.



- For continuous t, ↓ Farfield Model

$$y_l(t) = s_l(t) + n_l(t) = s(t + \Delta\tau_l) + n_l(t),$$

where $s(t)$ is the desired source signal at the reference point and $n_l(t)$ is the noise signal.

- Assuming farfield model, the delayed signal

$$\tilde{y}_l(t) = \tilde{s}_l(t) + \tilde{n}_l(t) = s(t + \Delta\tau_l - T_l) + n_l(t - T_l).$$

- If everything is equalized, then $T_l = T_B + \Delta\tau_l$.

- Now, considering sampling, the output of the delay-and-sum beamformer is

$$\hat{s}(kT) = \frac{1}{N_M}\sum_{l=1}^{N_M} \tilde{y}_l(kT) = \frac{1}{N_M}\sum_{l=1}^{N_M} \tilde{s}_l(kT) + \frac{1}{N_M}\sum_{l=1}^{N_M} \tilde{n}_l(kT)$$

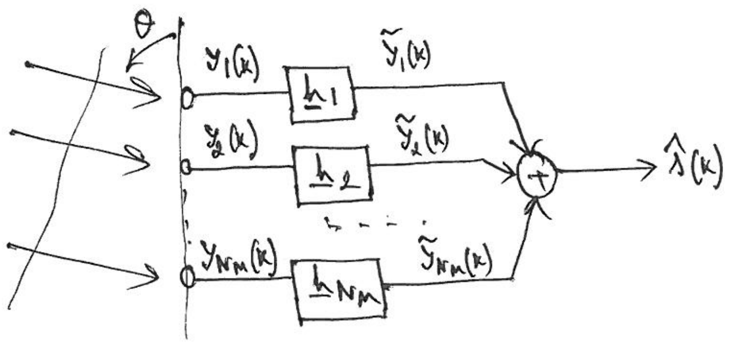
- For perfect equalized case, $T_l = T_B + \Delta\tau_l$

$$\hat{s}(k) = s(k - T_B f_s) + \frac{1}{N_M}\sum_{l=1}^{N_M} \tilde{n}_l(k).$$

Here f_s = sampling freq. $= \frac{1}{T}$.

- How to equalize?

Filter-and-Sum beamforming



$$\tilde{y}_l(k) = \sum_{m=0}^{M} h_l(m)\, y_l(k-m).$$ [M length FIR filter]

$$\underline{y}_l(k) = [\, y_l(k),\ y_l(k-1),\ \cdots,\ y_l(k-M)\,]^T$$

$$\underline{h}_l = [\, h_l(0),\ h_l(1),\ \cdots\ h_l(M)\,]^T$$

$$\therefore \quad \hat{s}(k) = \sum_{l=1}^{N_M} \underline{h}_l^T\, \underline{y}_l(k)$$

- How to find filter coefficients? Design filters.