1. Introduction to Robot Hearing

1. Why do robots need to hear?
2. Human-robot interaction
3. Auditory scene analysis
4. **Audio signal processing in brief**
5. Audio processing in the ear
6. Audio processing in the midbrain
7. Audio processing in the brain

Using Several Microphones

For each microphone $j = 1 \dots J$ we have a different equation:

$$m_1(t) = h_{11}(t) \star s_1(t) + h_{21}(t) \star s_2(t) + \dots + h_{K1}(t) \star s_K(t) + n_1(t)$$

$$\vdots$$

$$m_j(t) = h_{1j}(t) \star s_1(t) + h_{2j}(t) \star s_2(t) + \dots + h_{Kj}(t) \star s_K(t) + n_j(t)$$

$$\vdots$$

$$m_J(t) = h_{1J}(t) \star s_1(t) + h_{2J}(t) \star s_2(t) + \dots + h_{KJ}(t) \star s_K(t) + n_J(t)$$

Binaural Hearing

The microphones are plugged into the left and right “ears" of a dummy head.

In the absence of reverberations, there is a **head-related transfer function** (HRTF) for each ear:

$$m_L(t) = h_L(t) \star s_1(t) + h_L(t) \star s_2(t) + \ldots + h_L(t) \star s_K(t) + n_L(t)$$
$$m_R(t) = h_R(t) \star s_1(t) + h_R(t) \star s_2(t) + \ldots + h_R(t) \star s_K(t) + n_L(t)$$

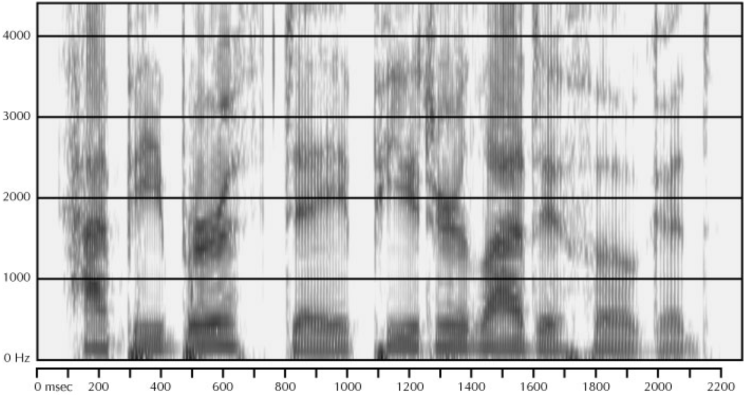
This general setting is difficult to solve in practice.

Spectral Representation

- In signal processing, it is common to perform spectral analysis, namely to apply the Fourier Transform (FT) to the microphone signals.
- In practice, one should use the discrete Fourier transform (DFT).
- The DFT applied over a short period of time is called the short-time Fourier transform (STFT).
- By applying the STFT to a signal at discrete time steps, one obtains a **spectrogram** (see next slide).
- Each spectrogram point indicates the amount of oscillation contained in the signal at time t and at frequency f

Spectrogram

A Spectrogram is an example of a time-frequency representation of a signal.



Binaural Cues for Sound Localization

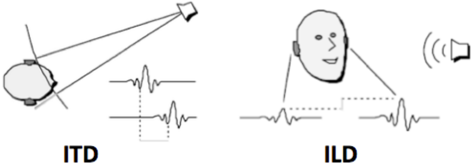
Suppose that there is a single emitting sound source.

The signals received by the two ears are **different**. Their difference is mainly characterized by two cues:

- They reach each ear at different times: interaural time difference, or ITD.
- They have different intensities (or levels): interaural intensity (or level) difference, or IID (ILD).

ITD and ILD

- ITD: interaural time difference (also TDOA or time difference of arrival)
- ILD: interaural level difference



- From these two features it is possible to localize the sound source.

Session Summary

- Audio signal processing
- Emitted and perceived signals
- Spectrograms
- Binaural localization cues