Introduction to intelligent systems

Audio processing

Mikkel N. Schmidt

Technical University of Denmark,
DTU Compute, Department of Applied Mathematics and Computer Science.

Overview

1. Machine learning systems
2. Feature transformations
3. Audio
4. Tasks

Feedback group

- Selma Bundgaard Langvik
- Andreas Holm Matthiassen
- Jacob Danvad Nalholm
- Mikkel Nielsen Broch-Lips

Learning objectives

I Frequency spectrum and spectrogram.

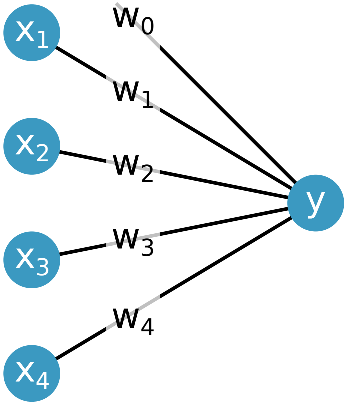
II Feature transformations and basis change.

I Understand the concepts and definitions, and know their application. Reason about the concepts in the context of an example. Use correct technical terminology.

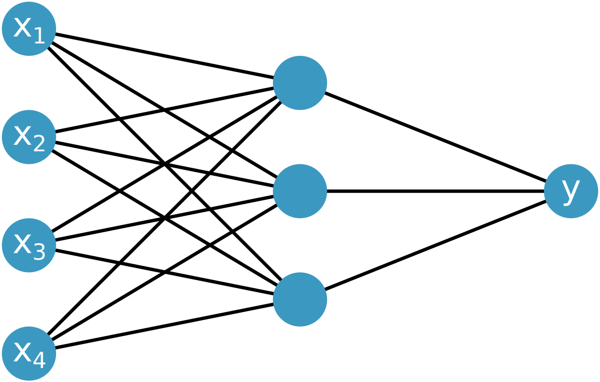
II As above plus: Read, manipulate, and work with technical definitions and expressions (mathematical and Python code). Carry out practical computations. Interpret and evaluate results.

Machine learning systems

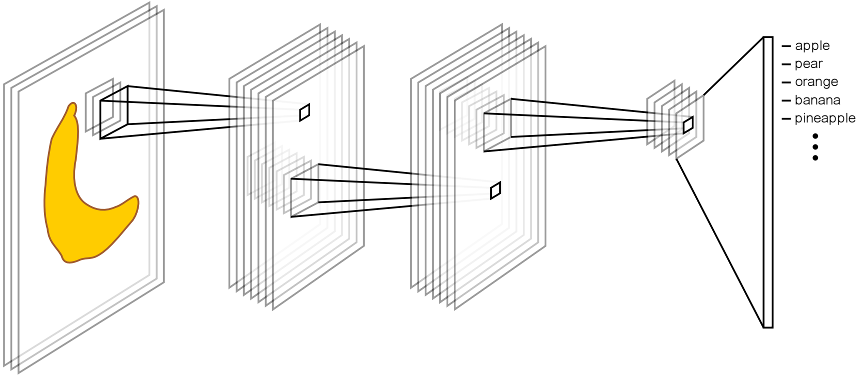
Linear models



Neural networks

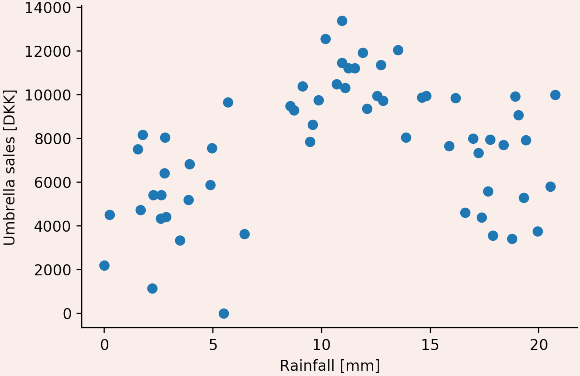


Convolutive neural network



Feature transformations

Clustering umbrella sales



- We want to examine if there are any clusters in the umbrella sales data
- We decide to use the k-means algorithm with the Euclidean distance
- What will go wrong, and how can we fix it?

Feature transformations

Mapping a set of data to a new set of values

Reasons to do feature transformations:

- To make representation more suitable for some particular algorithm
- To focus on relevant aspects of the data
- To make the data easier to process
- To remove unwanted noise
- To reduce dimensionality

Feature scaling

- Some machine learning methods are sensitive to the range of variables
- Standardize the range of a variable

Min-max normalization Rescale the range to $[0, 1]$

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Standardization Rescale to have zero mean and unit variance

$$x' = \frac{x - \bar{x}}{\sigma_x}$$

($\bar{x}$: mean, σ_x: standard deviation)

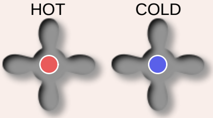
Change of basis

One way to transform a set of features is to change basis

- Represent data as a set of linear combinations of existing features
- Improve interpretation / more meaningful features

Hot and cold tap

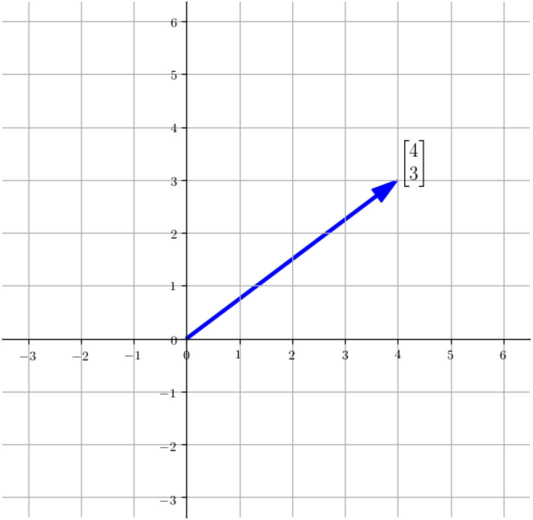
- In the good old days, a shower just had a *hot* and a *cold* tap



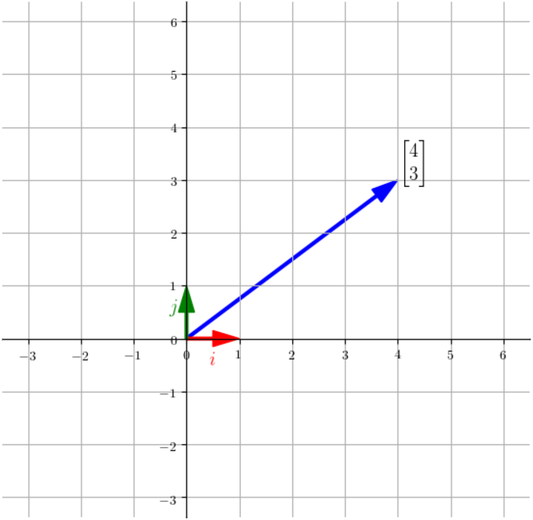
- To make the water more hot while keeping the samme pressure, you would
  1. Turn up the hot tap a bit
  2. Turn down the cold tap a bit
  3. Adjust by turning the hot a little down again
  4. Hmm. Now turn up the cold a little ...
  5. Aargh... Too cold now...
- Ideally, we would like a *temperature* and a *pressure* tap
- How can you make a linear combination of *hot* and *cold* to accieve this? (fill in the missing numbers)

$$\text{temperature} = \_\_\_ \cdot \text{cold} + \_\_\_ \cdot \text{hot}$$
$$\text{pressure} = \_\_\_ \cdot \text{cold} + \_\_\_ \cdot \text{hot}$$

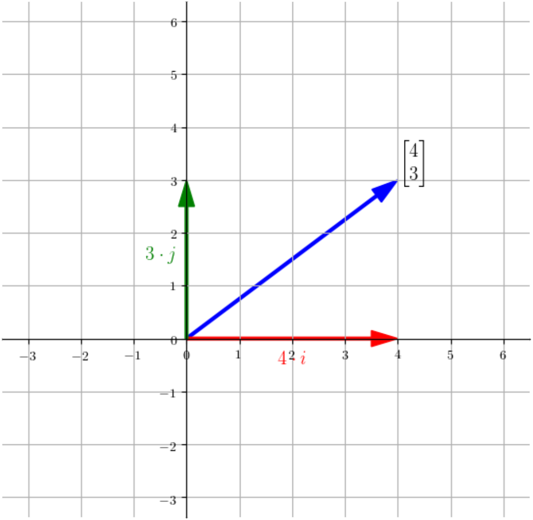
Basis of a vector space



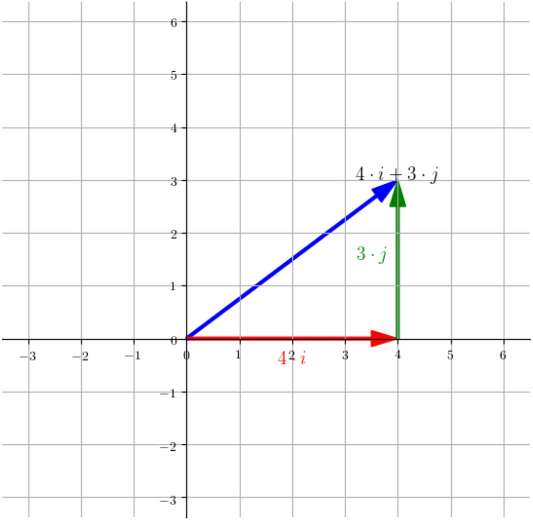
Basis of a vector space



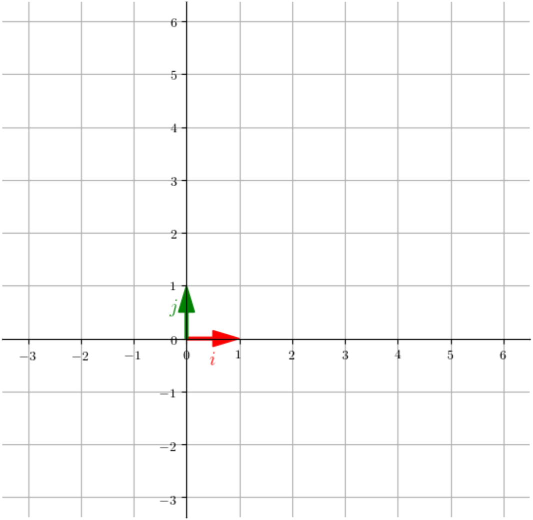
Basis of a vector space



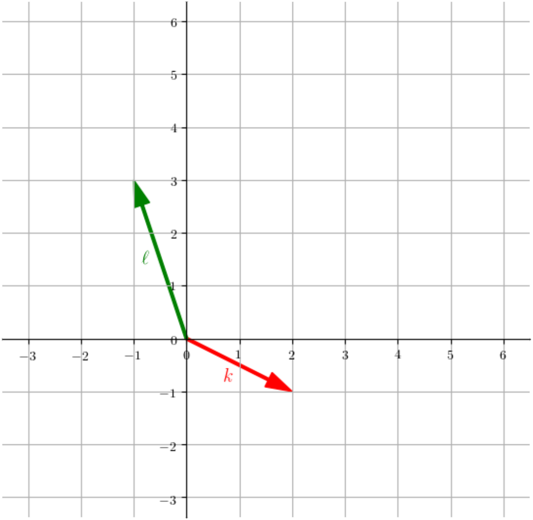
Basis of a vector space



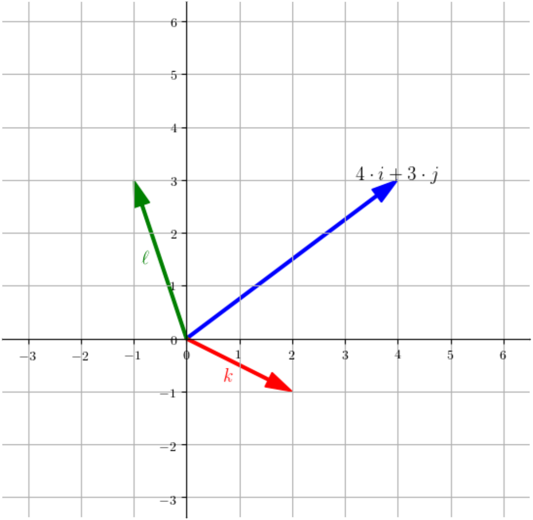
Basis of a vector space



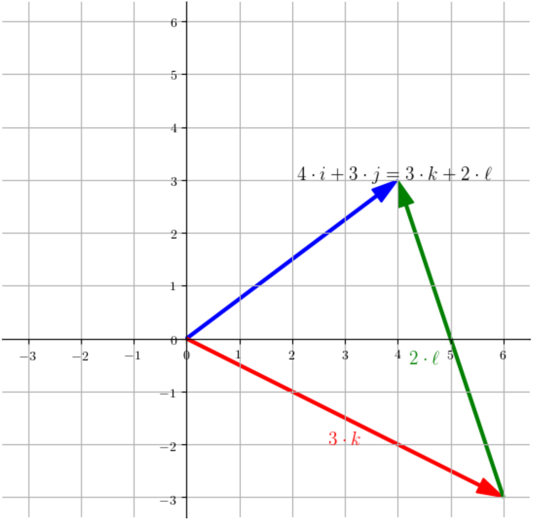
Basis of a vector space



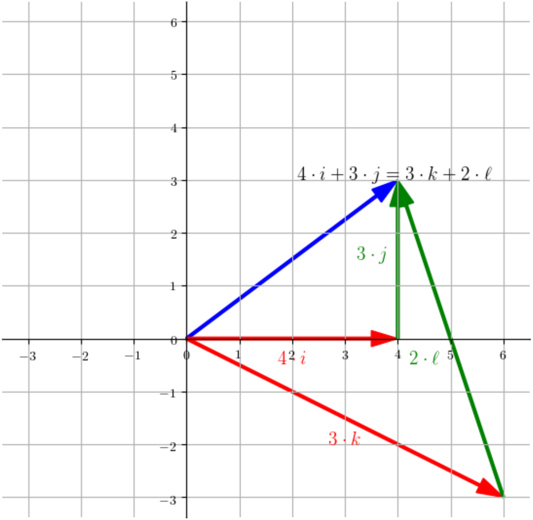
Basis of a vector space



Basis of a vector space



Basis of a vector space



Changing basis

- In the standard basis, the point is $\begin{bmatrix} 4 \\ 3 \end{bmatrix} = 4 \cdot i + 3 \cdot j$
- In the new basis, the point is $\begin{bmatrix} 3 \\ 2 \end{bmatrix} = 3 \cdot k + 2 \cdot \ell$
- The new basis is defined by

$$k = 2 \cdot i - 1 \cdot j \qquad \ell = -1 \cdot i + 3 \cdot j$$

- We can transform the point from the new basis to the standard basis

$$\begin{aligned} \begin{bmatrix} 3 \\ 2 \end{bmatrix} &= 3 \cdot k + 2 \cdot \ell \\ &= 3(2 \cdot i - 1 \cdot j) + 2(-1 \cdot i + 3 \cdot j) \\ &= 4 \cdot i + 3 \cdot j = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \end{aligned}$$

Excercise: Changing basis

- In the standard basis, a point is given by $\begin{bmatrix} 5 \\ 7 \end{bmatrix} = 5 \cdot i + 7 \cdot j$
- The new basis is defined by

$$k = 2 \cdot i - 1 \cdot j \qquad \ell = -1 \cdot i + 3 \cdot j$$

- Express the point as a coordinate in the new basis
  Hint: Solve for i and j in terms of k and ℓ and insert the result

Excercise: Changing basis

- In the standard basis, a point is given by $\begin{bmatrix} 5 \\ 7 \end{bmatrix} = 5 \cdot i + 7 \cdot j$
- The new basis is defined by

$$k = 2 \cdot i - 1 \cdot j \qquad \ell = -1 \cdot i + 3 \cdot j$$

- Express the point as a coordinate in the new basis
  Hint: Solve for i and j in terms of k and ℓ and insert the result

Solution

Solving for i and j yields

$$i = 0.6k + 0.2\ell \qquad j = 0.2k + 0.4\ell$$

$$\begin{aligned} \begin{bmatrix} 5 \\ 7 \end{bmatrix} &= 5i + 7j \\ &= 5(0.6k + 0.2\ell) + 7(0.2k + 0.4\ell) \\ &= 4.4k + 3.8\ell = \begin{bmatrix} 4.4 \\ 3.8 \end{bmatrix} \end{aligned}$$

Basis matrix

- New basis, defined in the standard basis

$$k = 2 \cdot i - 1 \cdot j \qquad \ell = -1 \cdot i + 3 \cdot j$$

- Standard basis, defined in the new basis

$$i = 0.6k + 0.2\ell \qquad j = 0.2k + 0.4\ell$$

We can write the bases as matrices where each column is a basis vector

$$\begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.4 \end{bmatrix} \tag{1}$$

Basis change as matrix multiplication

- In the standard basis, the point is $\begin{bmatrix}5\\7\end{bmatrix} = 5 \cdot i + 7 \cdot j$
- The basis is defined by

$$i = 0.6k + 0.2\ell \qquad j = 0.2k + 0.4\ell$$

- We transform the point to the new basis as

$$\begin{aligned}
\begin{bmatrix}5\\7\end{bmatrix} &= 5i + 7j \\
&= 5(0.6k + 0.2\ell) + 7(0.2k + 0.4\ell) \\
&= 5\begin{bmatrix}0.6\\0.2\end{bmatrix} + 7\begin{bmatrix}0.2\\0.4\end{bmatrix} \\
&= \begin{bmatrix}0.6 & 0.2\\0.2 & 0.4\end{bmatrix}\begin{bmatrix}5\\7\end{bmatrix} = \begin{bmatrix}4.4\\3.8\end{bmatrix}
\end{aligned}$$

Basis change as matrix multiplication

- General formula for basis change

$$\boldsymbol{y} = \underbrace{\boldsymbol{T}^{-1}\boldsymbol{x}}_{\text{matrix multiplication}}$$

$\boldsymbol{x}$ Vector in the original coordinate system
$\boldsymbol{T}$ Matrix where each column is a basis vector of the new coordinate system expressed in the old coordinate system
$\boldsymbol{y}$ Vector expressed in the new coordinate system

- We can map the other way as

$$\boldsymbol{x} = \boldsymbol{T}\boldsymbol{y}$$

- Orthonormal basis: $T^{-1} = T^{\top}$

Audio

What is sound?

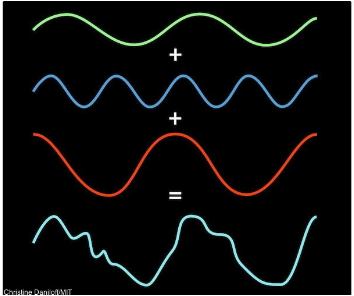
In physics

- Vibrations that propagate as a pressure wave through a transmission medium (such as air)

In psychology

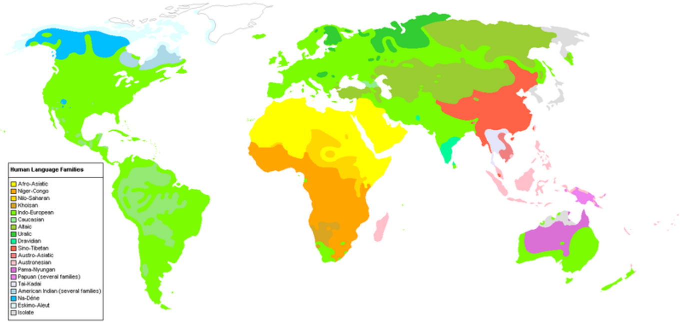
- The reception of a sound wave and its perception by the brain

Frequency content



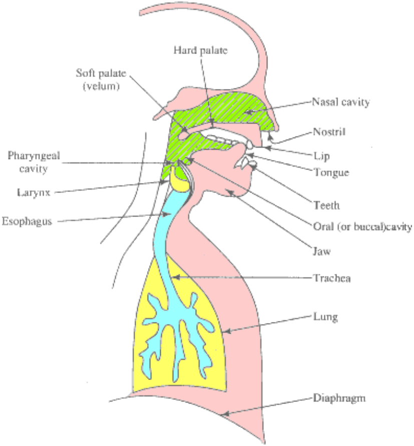
Speech

- Speech signals are sequences of sounds
- The basic sounds and the transitions between them serve as symbolic representation of information: *semantics*
- The arrangement of these sounds (symbols) is governed by the rules of language
- The study of these rules and their implications in human cmmunication is called *linguistics*
- The study of and classification of the sounds of speech is called *phonetics*



Speech production

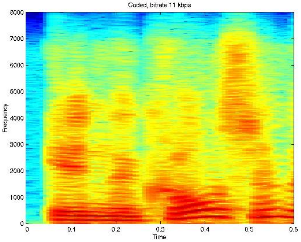
- Speech is produced by the human vocal tract
- The vocal tract is excited either by short burst of *periodic* signal or by *"noise"* from the flow
- Periodic signals (voiced sounds) are produced by air flow through tight and vibrating vocal cords
- Noise (unvoiced sounds) are produced by turbulent flow with relaxed vocal cords



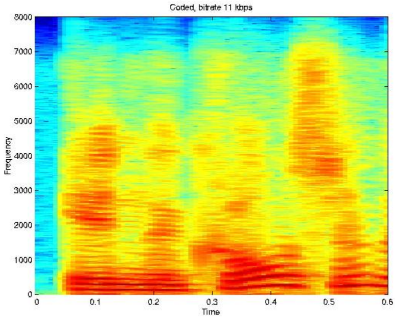
Speech production

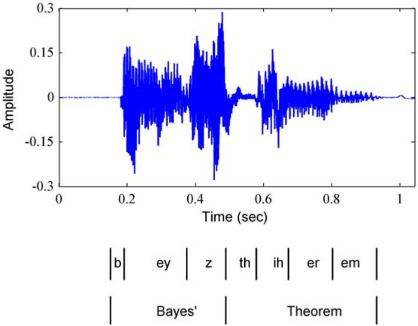
- The vocal tract (nasal and oral cavities) transforms the voiced or unvoiced sounds to phonemes (speech sounds)
- Speech sounds are classified broadly into phonemes classes
  - Vowels (voiced)
  - Consonants (unvoiced)
- Phonemes corresponds to *formants*—peaks in the power spectrum modulation (red areas in figure)
- Formant frequencies in the range

F1	270-730 Hz
F2	840-2290 Hz
F3	1690-3010 Hz

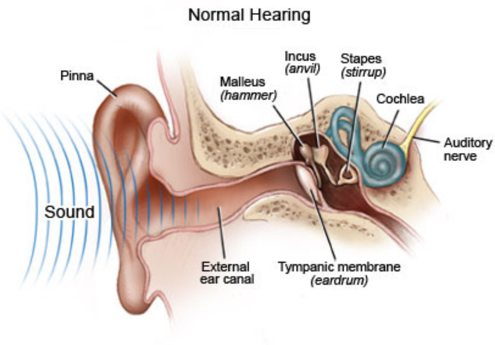


Speech spectrogram





Human hearing performs frequency analysis



- Cochlea is filled with a watery liquid
- Liquid moves in response to the vibrations coming from the middle ear via the oval window
- Hair cells sense the motion—convert motion to electrical signals
- Communicated via neurotransmitters to many thousands of nerve cells

Human hearing performs frequency analysis

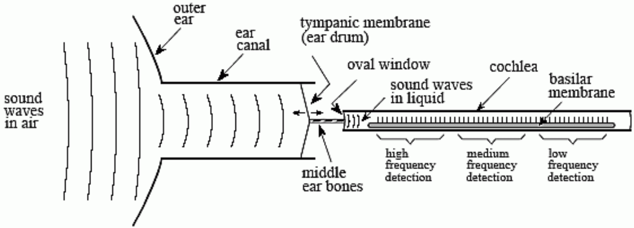


FIGURE 22-1
Functional diagram of the human ear. The outer ear collects sound waves from the environment and channels them to the tympanic membrane (ear drum), a thin sheet of tissue that vibrates in synchronization with the air waveform. The middle ear bones (hammer, anvil and stirrup) transmit these vibrations to the oval window, a flexible membrane in the fluid filled cochlea. Contained within the cochlea is the basilar membrane, the supporting structure for about 12,000 nerve cells that form the cochlear nerve. Due to the varying stiffness of the basilar membrane, each nerve cell only responses to a narrow range of audio frequencies, making the ear a frequency spectrum analyzer.

http://www.dspguide.com/ch22/1.htm

http://www.cochlea.eu/en/cochlea/function

Sound intensity

TABLE 22-1
Units of sound intensity. Sound intensity is expressed as power per unit area (such as watts/cm²), or more commonly on a logarithmic scale called *decibels SPL*. As this table shows, human hearing is the most sensitive between 1 kHz and 4 kHz.

Softer Louder

Watts/cm^2	Decibels SPL	Example sound
10^{-2}	140 dB	Pain
10^{-3}	130 dB	
10^{-4}	120 dB	Discomfort
10^{-5}	110 dB	Jack hammers and rock concerts
10^{-6}	100 dB	
10^{-7}	90 dB	OSHA limit for industrial noise
10^{-8}	80 dB	
10^{-9}	70 dB	
10^{-10}	60 dB	Normal conversation
10^{-11}	50 dB	
10^{-12}	40 dB	Weakest audible at 100 hertz
10^{-13}	30 dB	
10^{-14}	20 dB	Weakest audible at 10kHz
10^{-15}	10 dB	
10^{-16}	0 dB	Weakest audible at 3 kHz
10^{-17}	-10 dB	
10^{-18}	-20 dB	

http://www.dspguide.com/ch22/1.htm
http://www.cochlea.eu/en/cochlea/function

Sampling and waveforms

The following is done in order to process audio in a computer

Low pass filtering Frequency content above some upper level is discarded

Sampling The signal is measured at discrete time intervals

Quantization The signal amplitudes are represented as (usually discrete) numbers

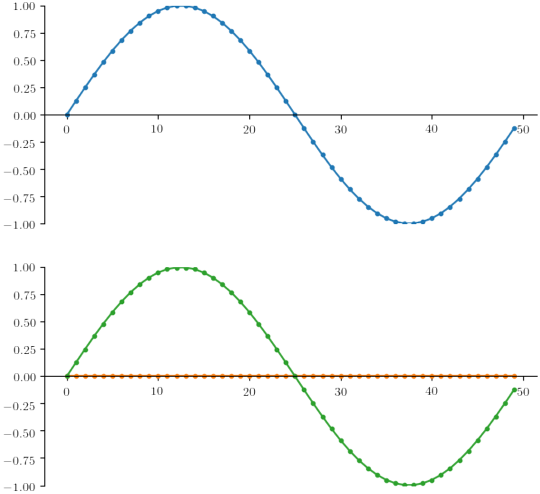
Exercise: Audio as a point in a vector space

An audio signal of length N can be thought of as a point in an N-dimensional vector space, $\mathbb{R}^N$

- What is the standard basis of this vector space?
- How can we construct any possible audio signal by a linear combination of such basis vectors?
- How do you think each of these basis vectors sounds
- Is this a good basis for representing sound? Can you come up with a better basis, perhaps inspired by the human auditory system?

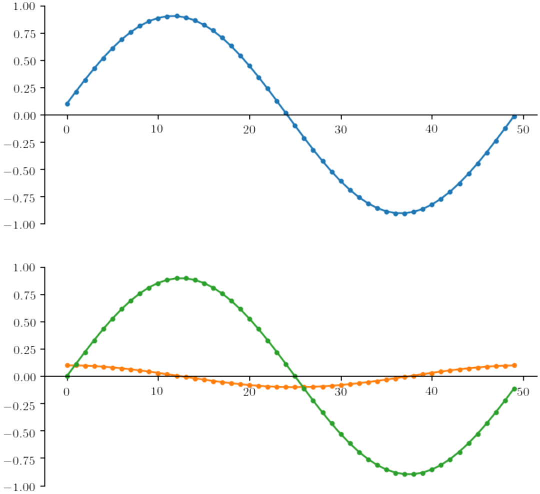
Sum of sinusoids

A weighted sum of a sine and cosine function can give a sinusoid with arbitrary phase-shift



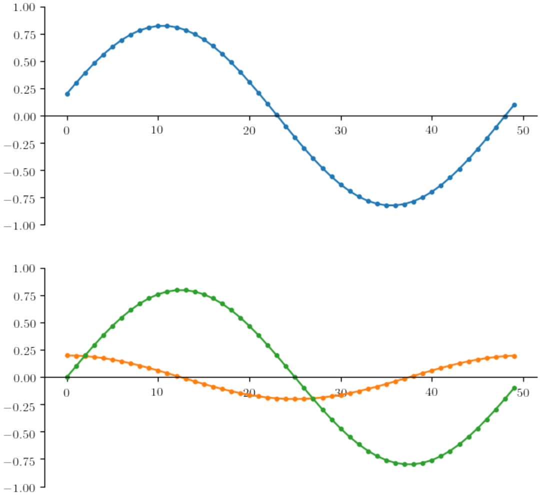
Sum of sinusoids

A weighted sum of a sine and cosine function can give a sinusoid with arbitrary phase-shift



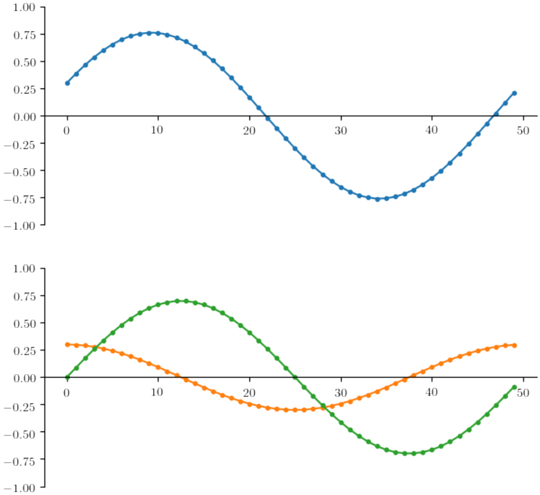
Sum of sinusoids

A weighted sum of a sine and cosine function can give a sinusoid with arbitrary phase-shift



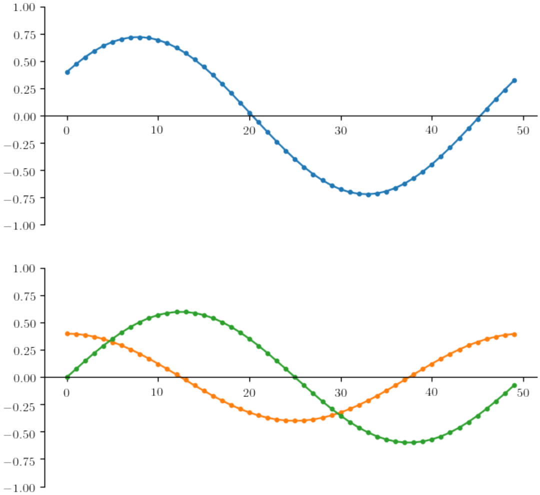
Sum of sinusoids

A weighted sum of a sine and cosine function can give a sinusoid with arbitrary phase-shift



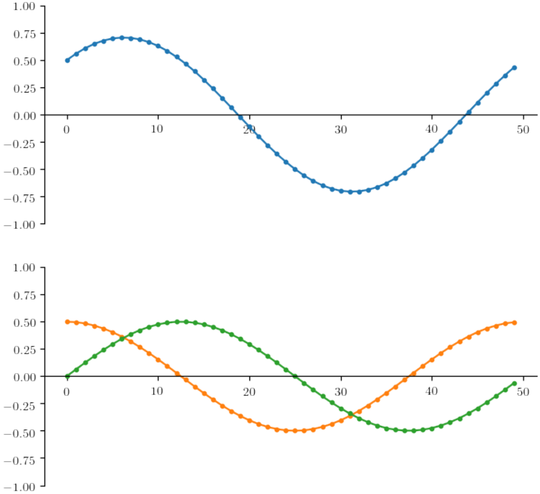
Sum of sinusoids

A weighted sum of a sine and cosine function can give a sinusoid with arbitrary phase-shift



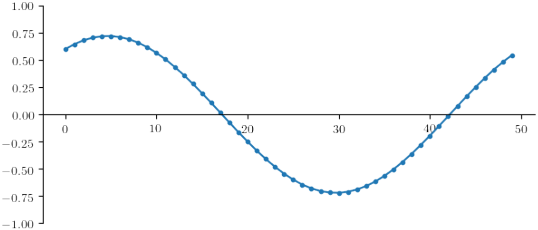
Sum of sinusoids

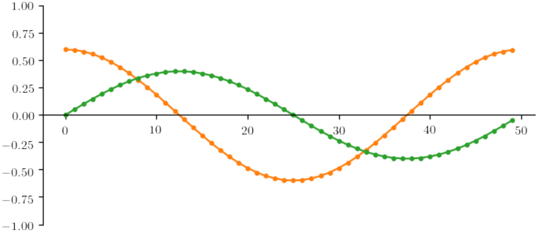
A weighted sum of a sine and cosine function can give a sinusoid with arbitrary phase-shift



Sum of sinusoids

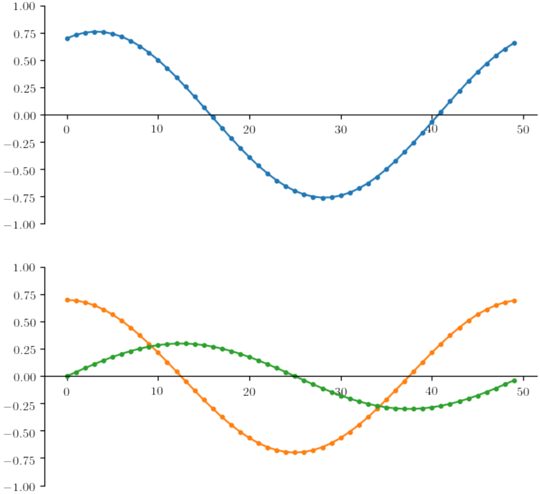
A weighted sum of a sine and cosine function can give a sinusoid with arbitrary phase-shift





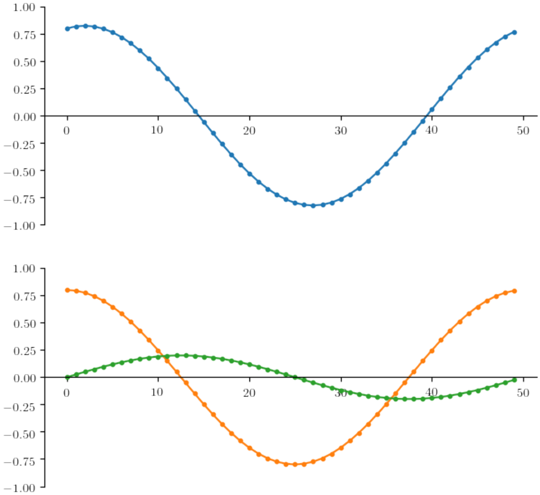
Sum of sinusoids

A weighted sum of a sine and cosine function can give a sinusoid with arbitrary phase-shift



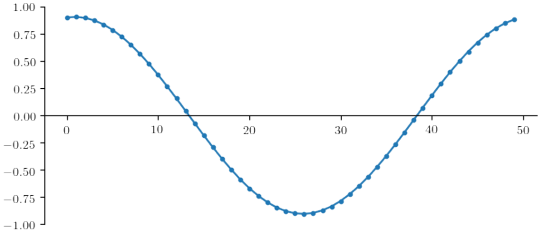
Sum of sinusoids

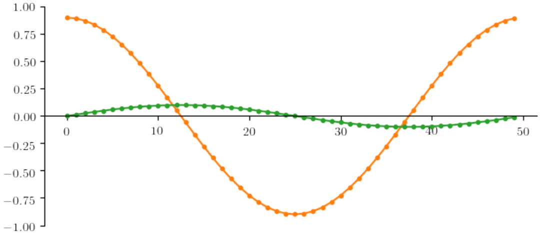
A weighted sum of a sine and cosine function can give a sinusoid with arbitrary phase-shift



Sum of sinusoids

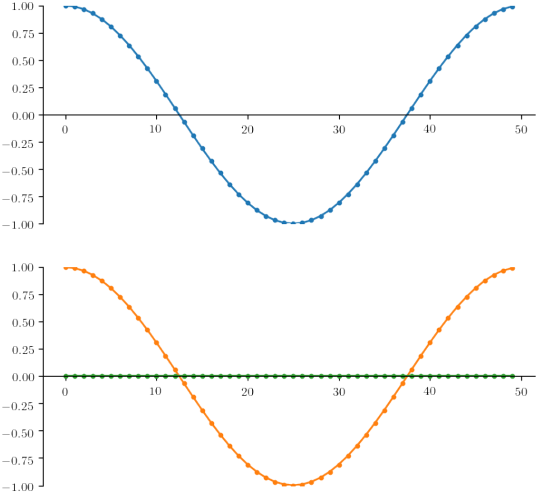
A weighted sum of a sine and cosine function can give a sinusoid with arbitrary phase-shift





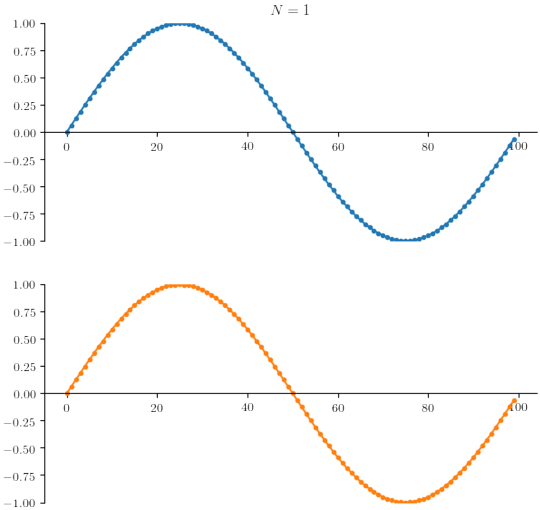
Sum of sinusoids

A weighted sum of a sine and cosine function can give a sinusoid with arbitrary phase-shift



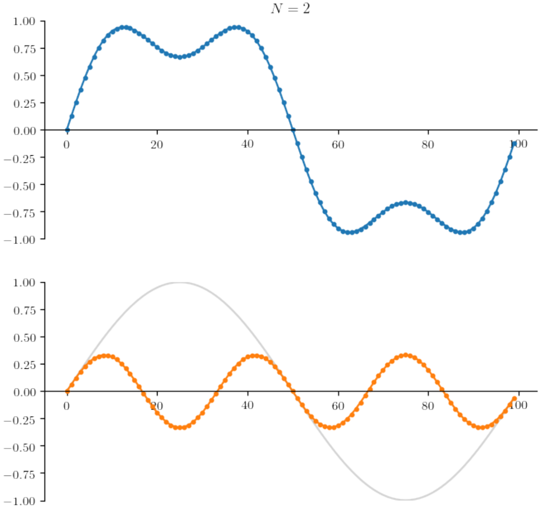
Sum of sinusoids

Any signal can be decomposed as a sum of sinusoids



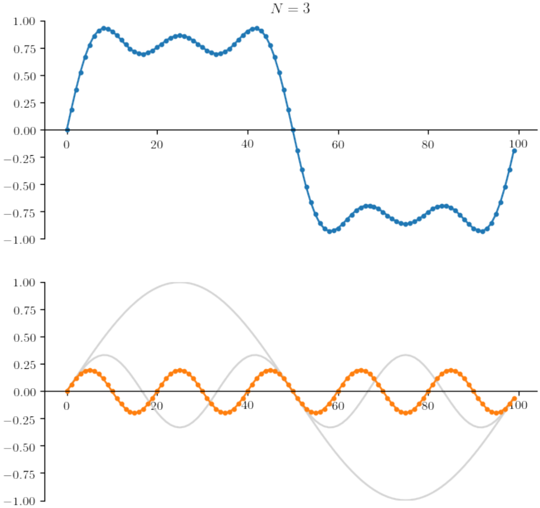
Sum of sinusoids

Any signal can be decomposed as a sum of sinusoids



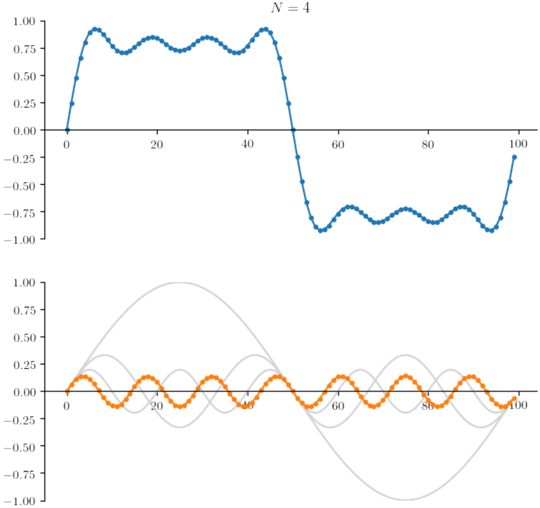
Sum of sinusoids

Any signal can be decomposed as a sum of sinusoids



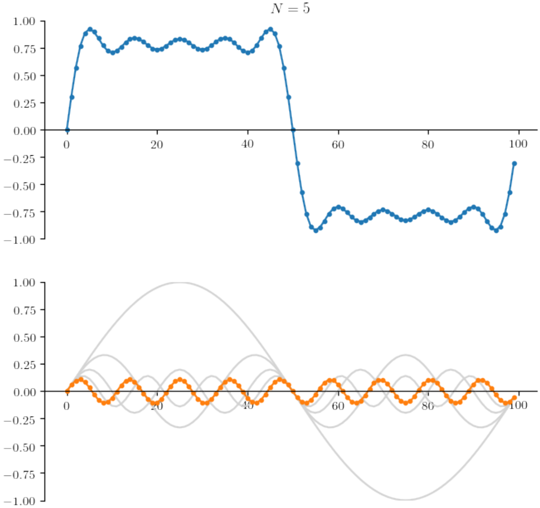
Sum of sinusoids

Any signal can be decomposed as a sum of sinusoids



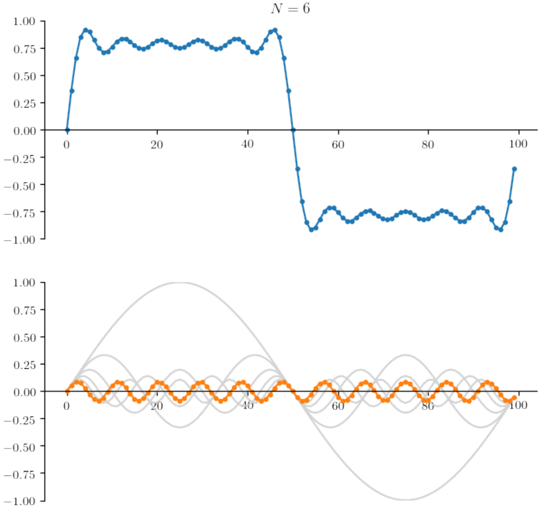
Sum of sinusoids

Any signal can be decomposed as a sum of sinusoids



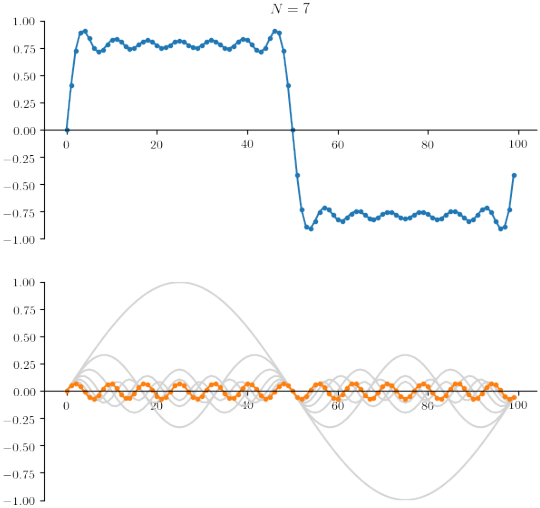
Sum of sinusoids

Any signal can be decomposed as a sum of sinusoids



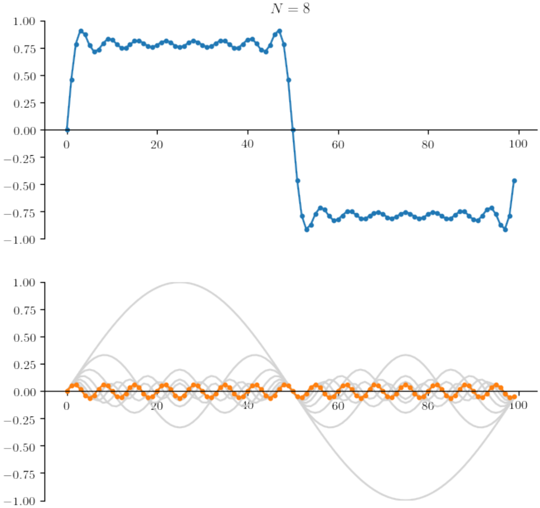
Sum of sinusoids

Any signal can be decomposed as a sum of sinusoids



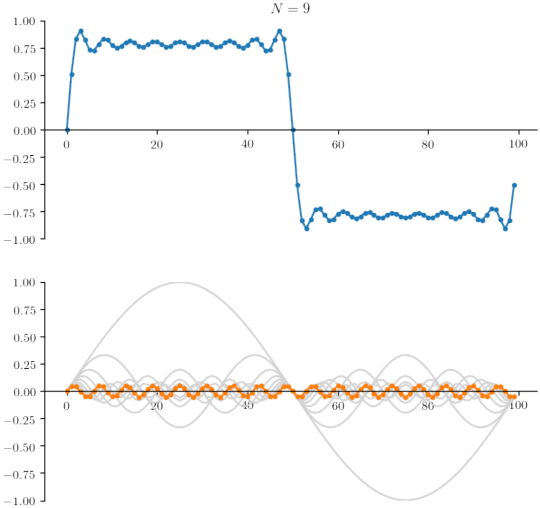
Sum of sinusoids

Any signal can be decomposed as a sum of sinusoids



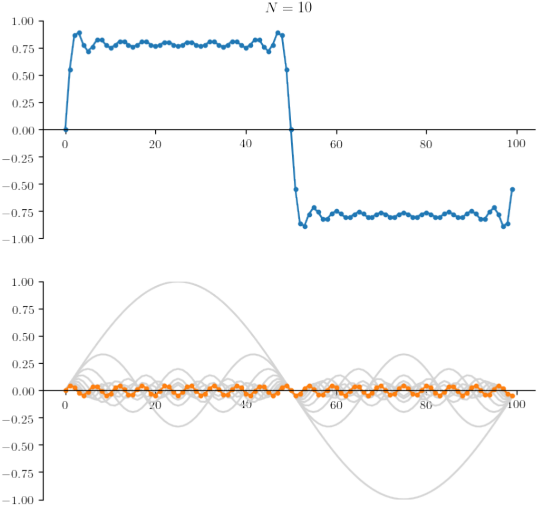
Sum of sinusoids

Any signal can be decomposed as a sum of sinusoids



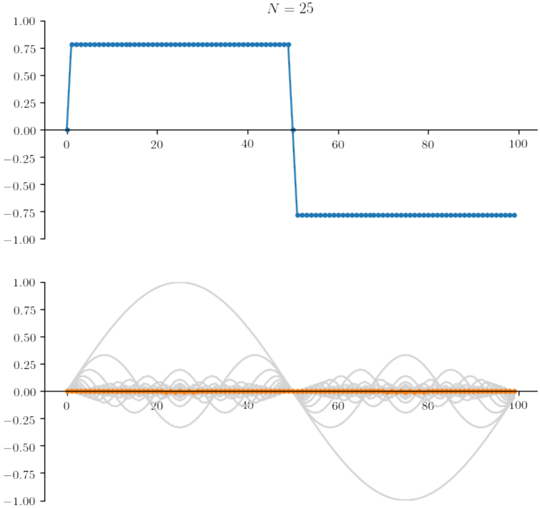
Sum of sinusoids

Any signal can be decomposed as a sum of sinusoids



Sum of sinusoids

Any signal can be decomposed as a sum of sinusoids



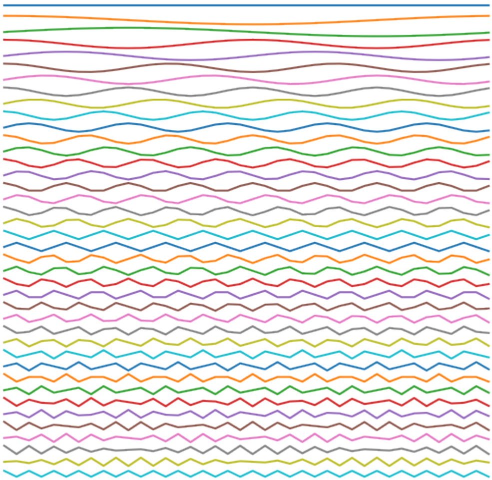
A sine+cosine basis

A basis of sine and cosine functions (for even N)

$$T = \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbf{1} \\ \sqrt{2}\cos(t) \\ \sqrt{2}\sin(t) \\ \sqrt{2}\cos(2t) \\ \sqrt{2}\sin(2t) \\ \vdots \\ \sqrt{2}\cos([\frac{N}{2}-1]t) \\ \sqrt{2}\sin([\frac{N}{2}-1]t) \\ \cos(\frac{N}{2}t) \end{bmatrix}^\top$$

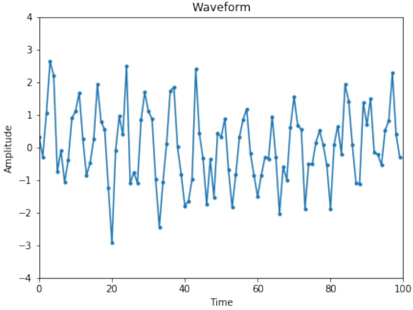
$$t = \frac{2\pi n}{N}, \qquad n = [0, 1, 2, \ldots, N-1]$$

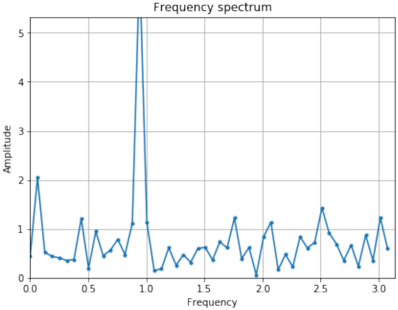
A sine+cosine basis



Frequency spectrum

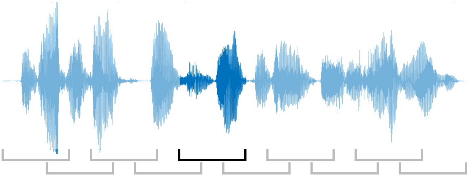
- Map waveform onto sine+cosine basis
- Combine amplitude of sine and cosine at each frequency
- Amplitude spectrum (invariant to phase shift)



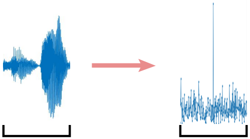


Spectrogram

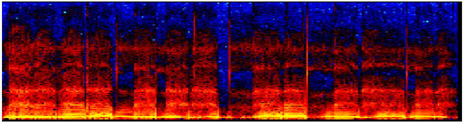
Split the signal
into blocks



For each block,
compute the spectrum



Gather spectra as columns
in matrix and plot heat map



Tasks

Tasks for today

1. Work through the three notebooks on audio analysis
   `10-SinusoidsInNoise.ipynb` `10-Spectrogram.ipynb`
   `10-AudioClassification.ipynb`
2. Today's feedback group
   - Selma Bundgaard Langvik
   - Andreas Holm Matthiassen
   - Jacob Danvad Nalholm
   - Mikkel Nielsen Broch-Lips

Lab report

- Lab 4: Neural networks (Deadline: Thursday 9 November 20:00)