simplilearn | PURDUE UNIVERSITY

Natural Language Processing and Speech Recognition

Signal Processing and Speech Recognition Models

Learning Objectives

By the end of this lesson, you will be able to:

- Relate audio signal and its processing
- Outline the Hidden Markov Model
- Interpret different models for speech recognition

DATA AND
ARTIFICIAL INTELLIGENCE

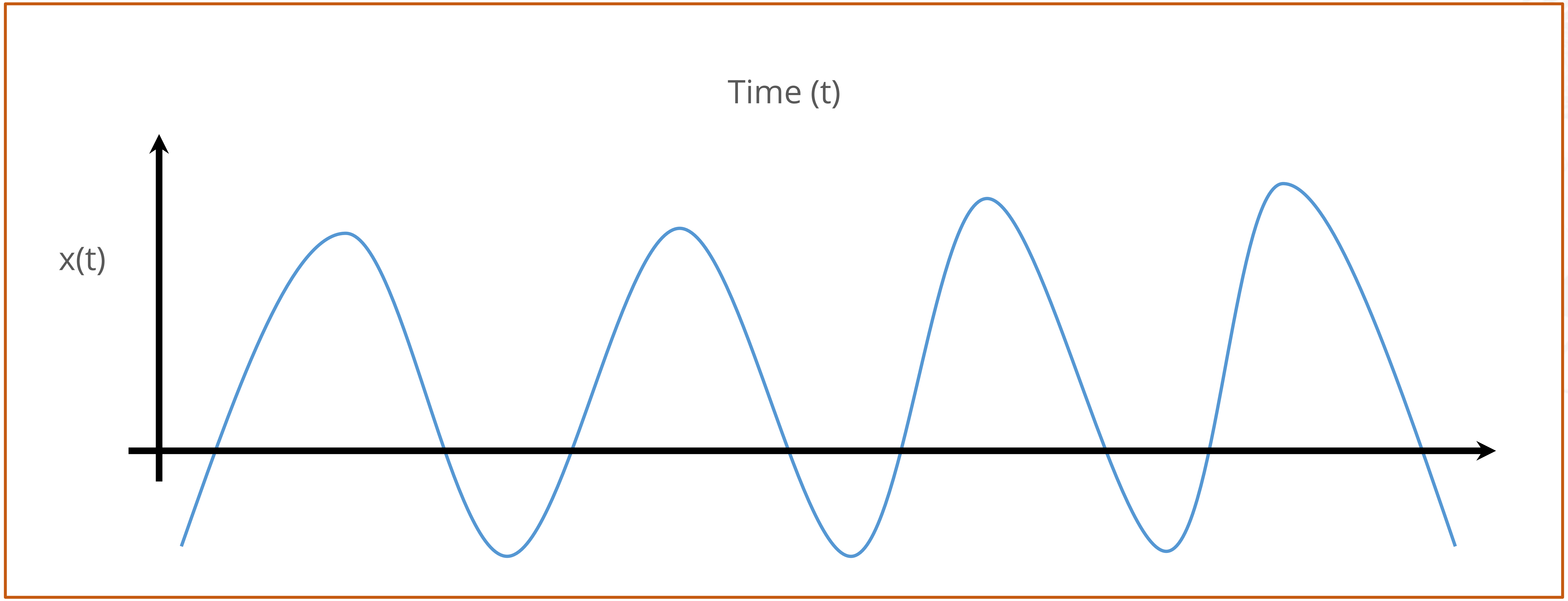
Audio Signal Processing

What Is Audio Signal Processing?

> Audio signal processing is a subfield of signal processing that is concerned with the electronic manipulation of audio signals.

What Is Signal?

Signal is a time varying physical phenomenon which is intended to convey information.



What Is Audio?

Is an electric signal that represents sound

Comes out of microphones, electrical instruments, processors, music players, etc.

Audio

Is not same as sound

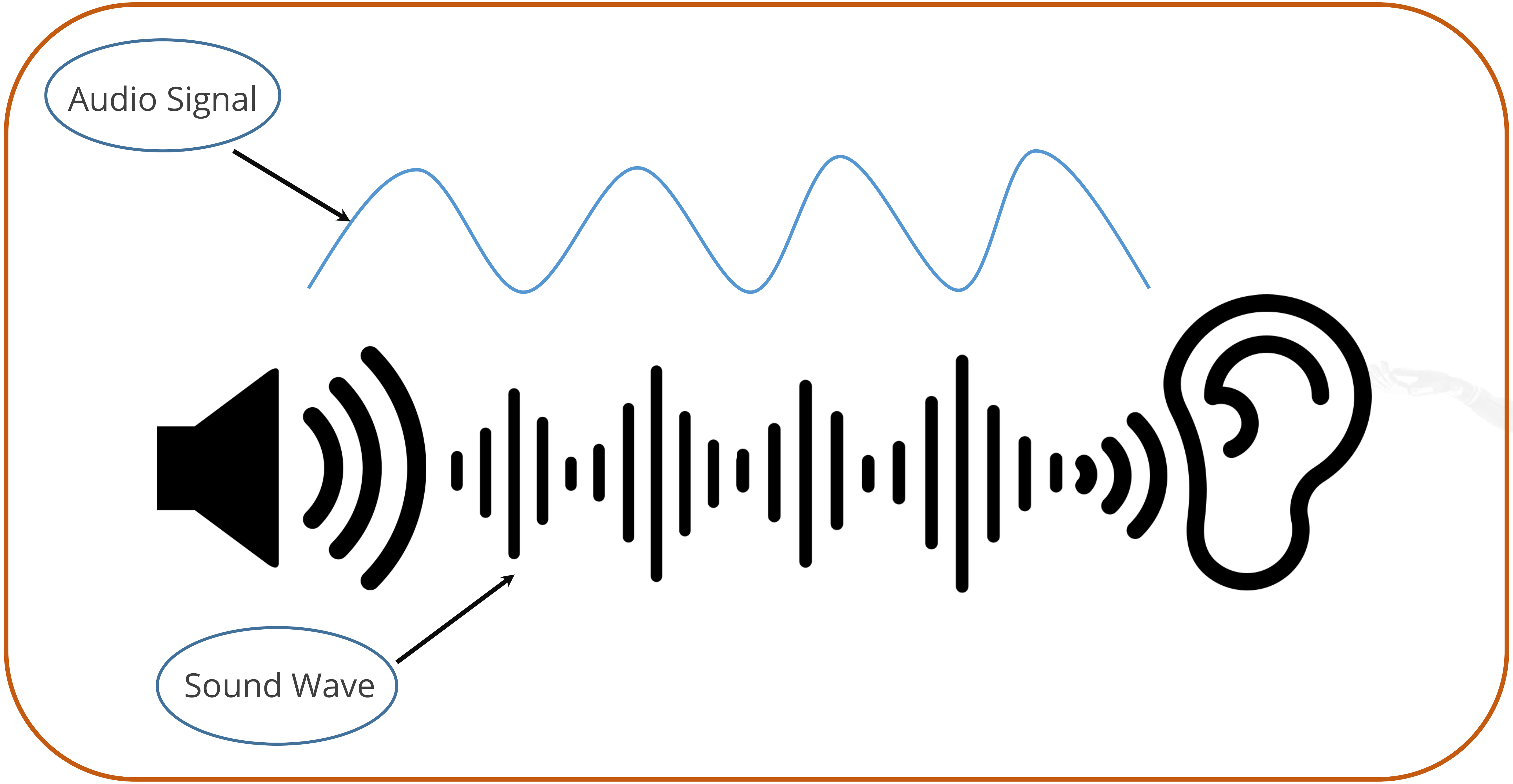
What Is Audio Signal?

> An audio signal is a representation of sound in terms of electrical variations such as voltage.
> Audio signals cannot be heard.

What Is Sound Wave?

Electrical signal travels through a transducer and gets converted into pressure variations in the air. These variations are called sound waves.

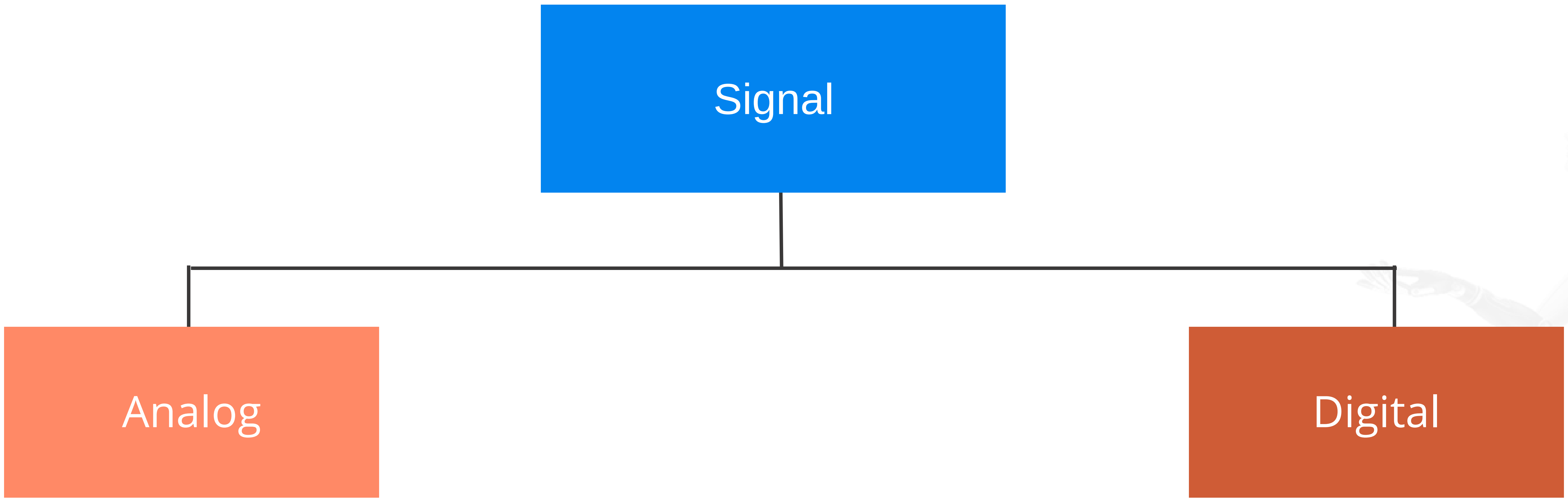
Audio Signal and Sound Wave



Parameters of an Audio Signal

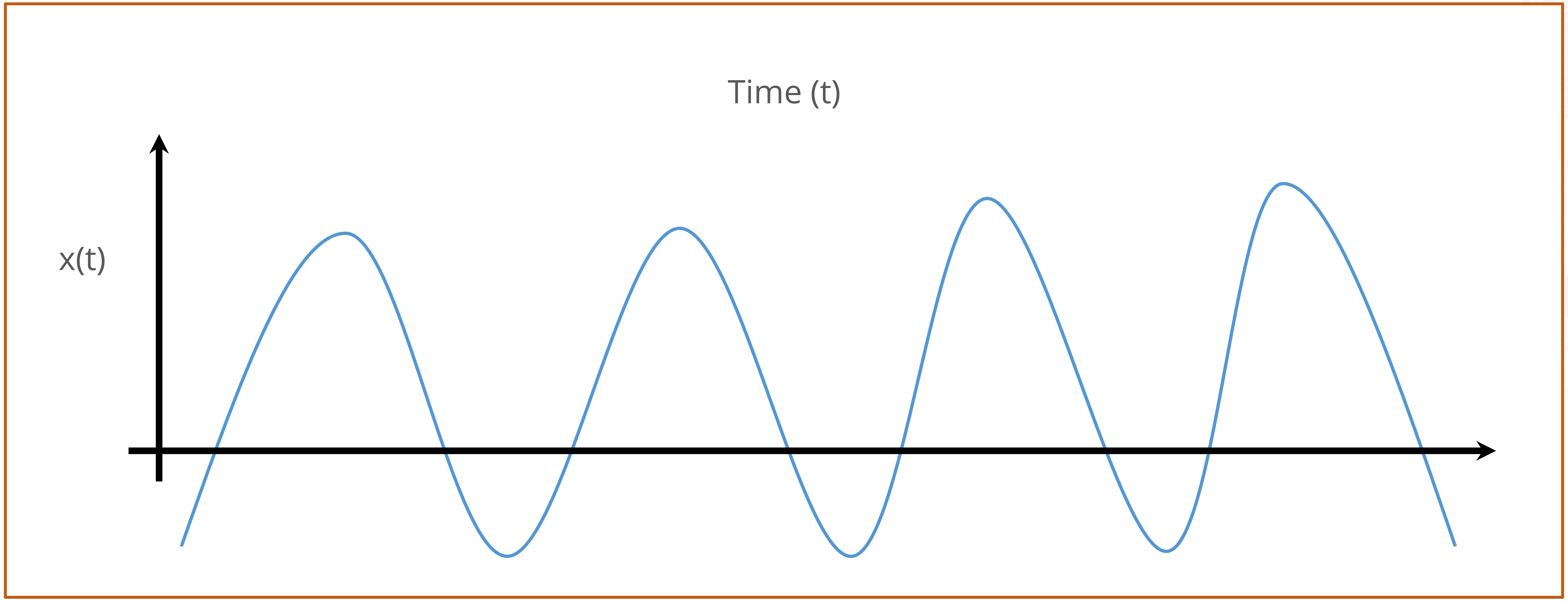
Amplitude	Refers to the peak displacement of the air molecules from the rest location
Crest	Is the highest point in the wave
Trough	Is the lowest point in the wave
Wavelength	Is the range between two consecutive crests or troughs
Cycle	Refers to a total upward motion and downward movement of a signal
Frequency	Refers to how rapidly a signal changes over time

Types of Signal



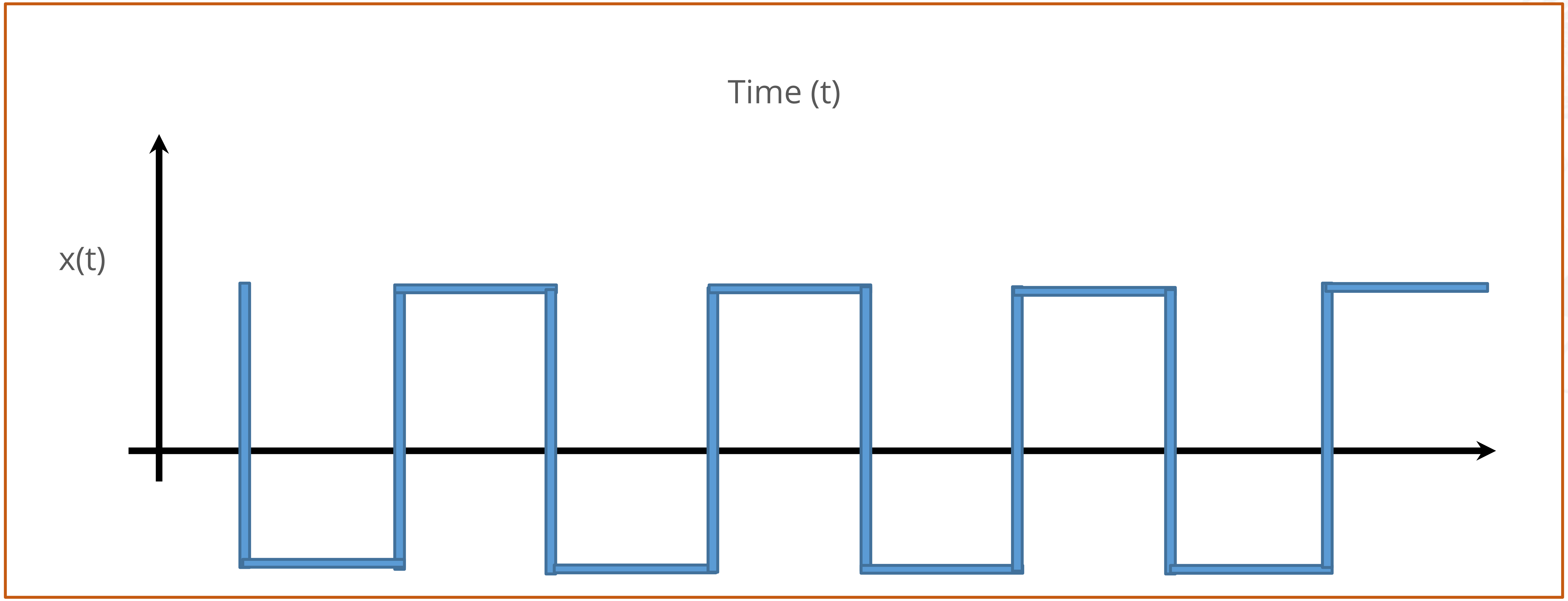
Analog Signal

An analog signal is a continuous signal that carries time-varying quantities.



Digital Signal

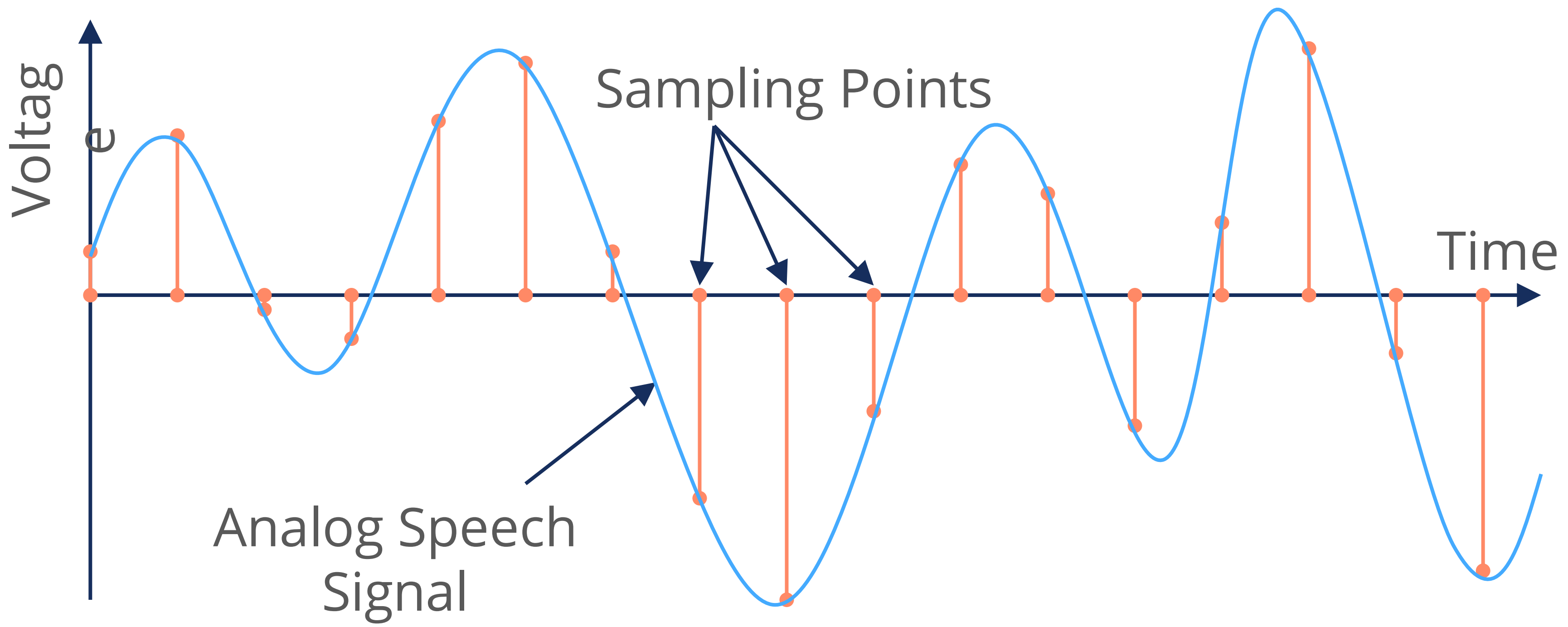
A digital signal represents data as a sequence of discrete values, at any given time it can only take one of a finite number of values.



Digitization of Speech Signals

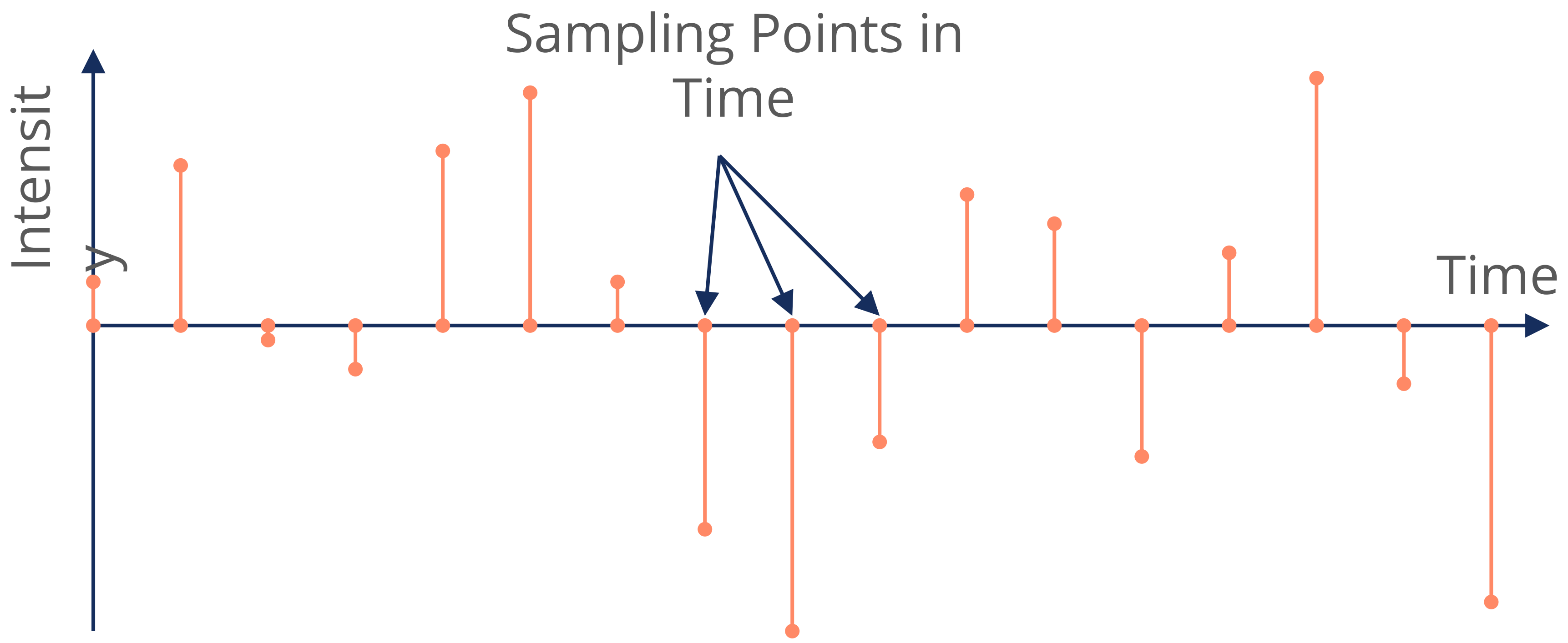
Sampling

- The analog speech signal captures pressure variations in air that are produced by the speaker.
- The analog speech input signal from the microphone is sampled periodically at a fixed sampling rate.



Sampling

- The value of the analog signal at discrete time points remains the same after sampling.
- This is the discrete-time signal.

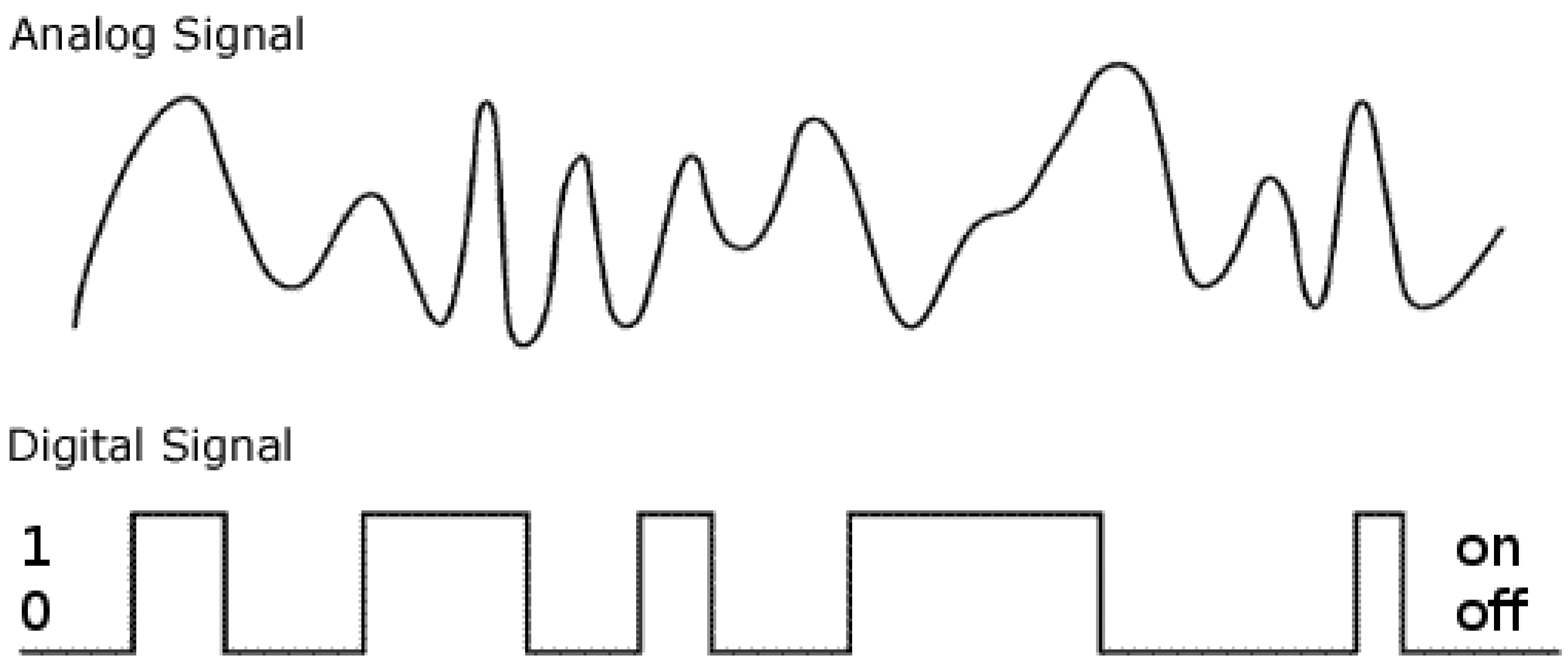


Sampling Rules

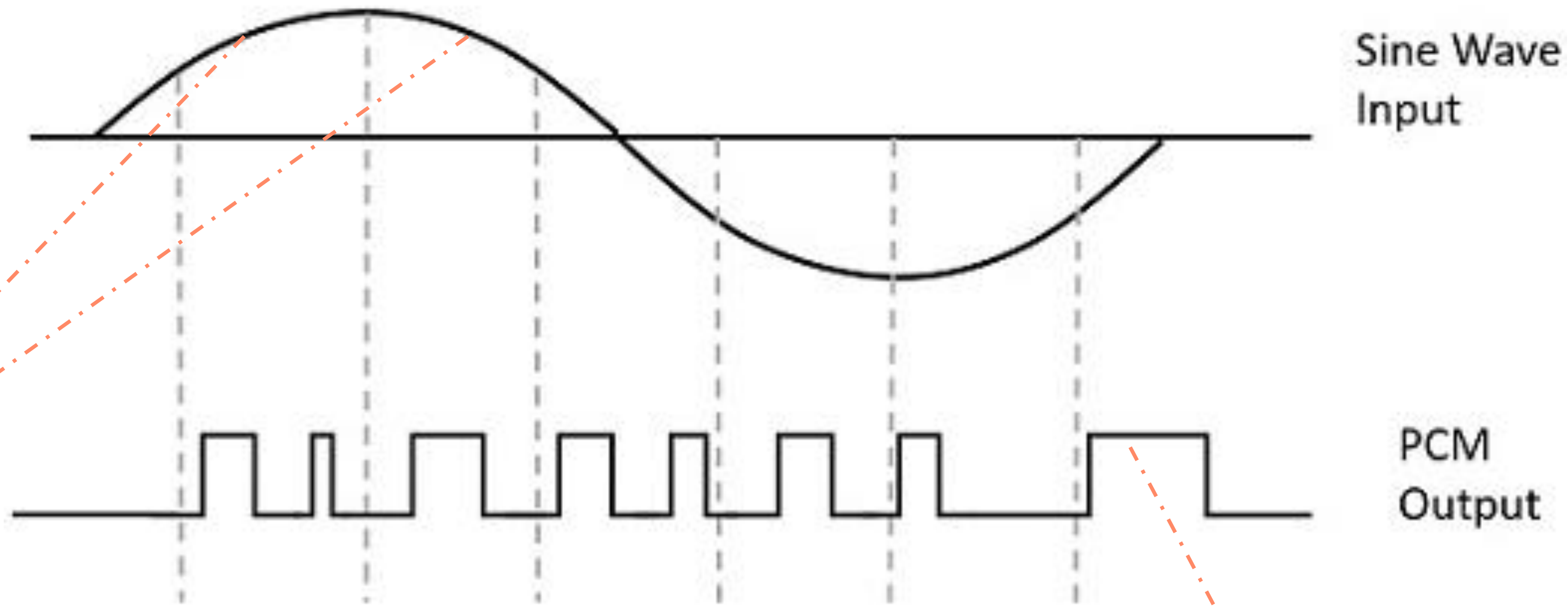
- A signal that is digitized at N samples/sec can represent frequencies up to N/2 Hz only (Nyquist Theorem).
- One can sample the speech signal at a sufficiently high rate to retain all perceivable components in the signal.
- For practical reasons, lower sampling rates are often used.
- A signal that is sampled at N samples per second must first be low-pass filtered at N/2 Hz to avoid distortions from “aliasing”.

Digitization

- Each sampled value is digitized (or quantized or encoded) into one of a set of fixed discrete levels.
- Digitization can be linear (uniform) or nonlinear (nonuniform).



Linear Coding

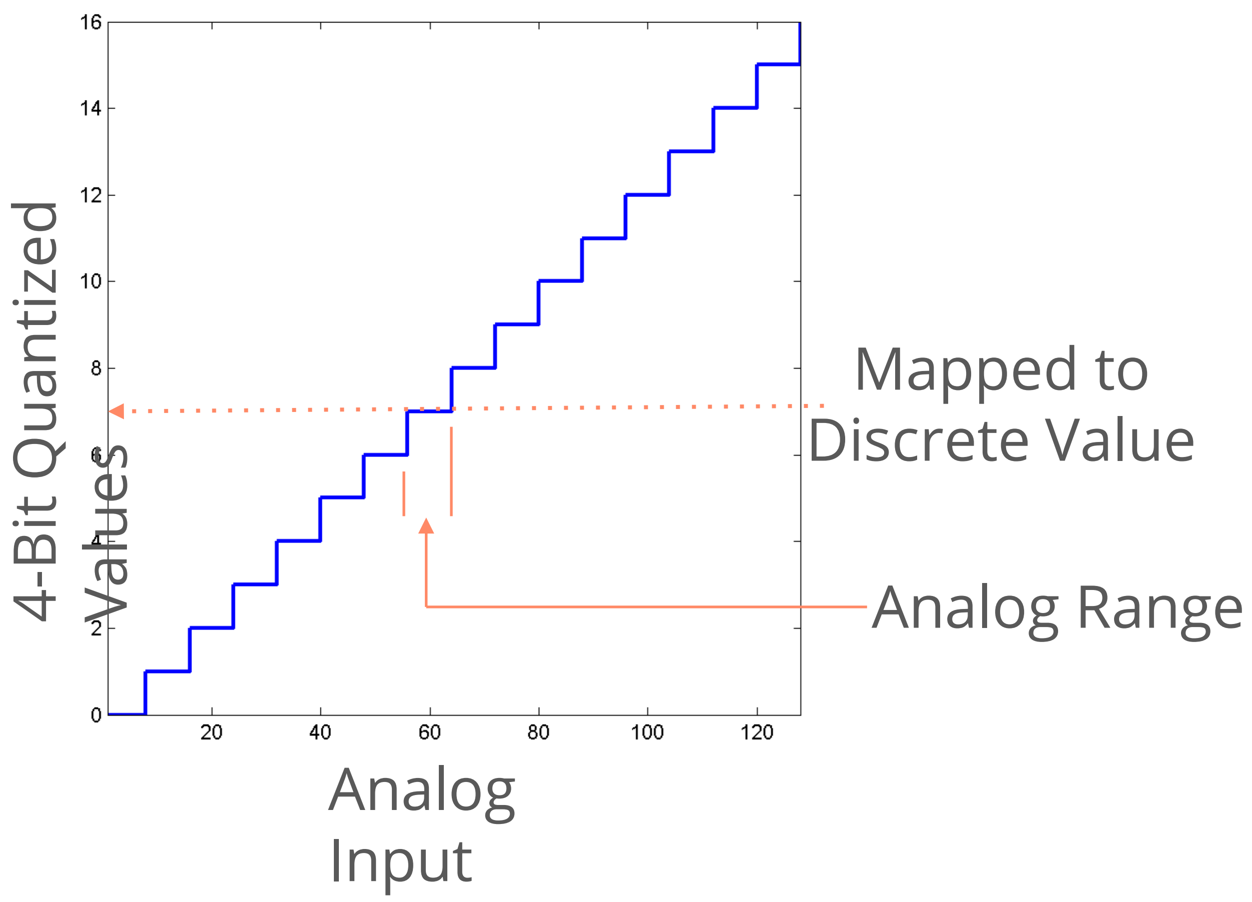


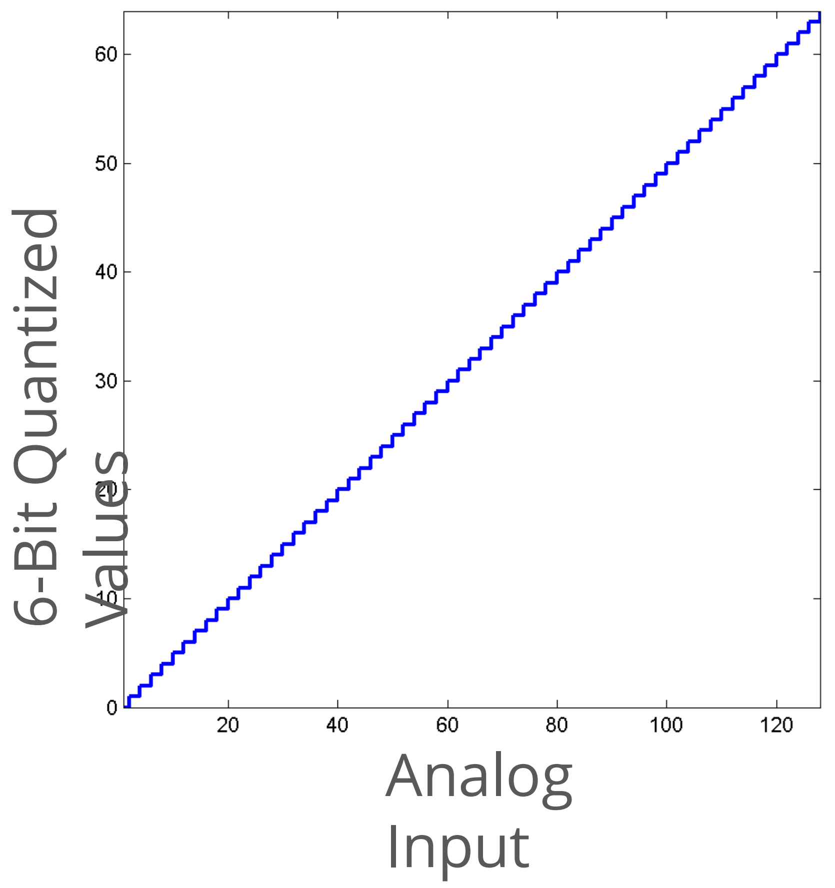
Input analog range split into uniformly spaced levels

Number of discrete levels determine number of bits needed to represent a quantized signal

Linear Coding

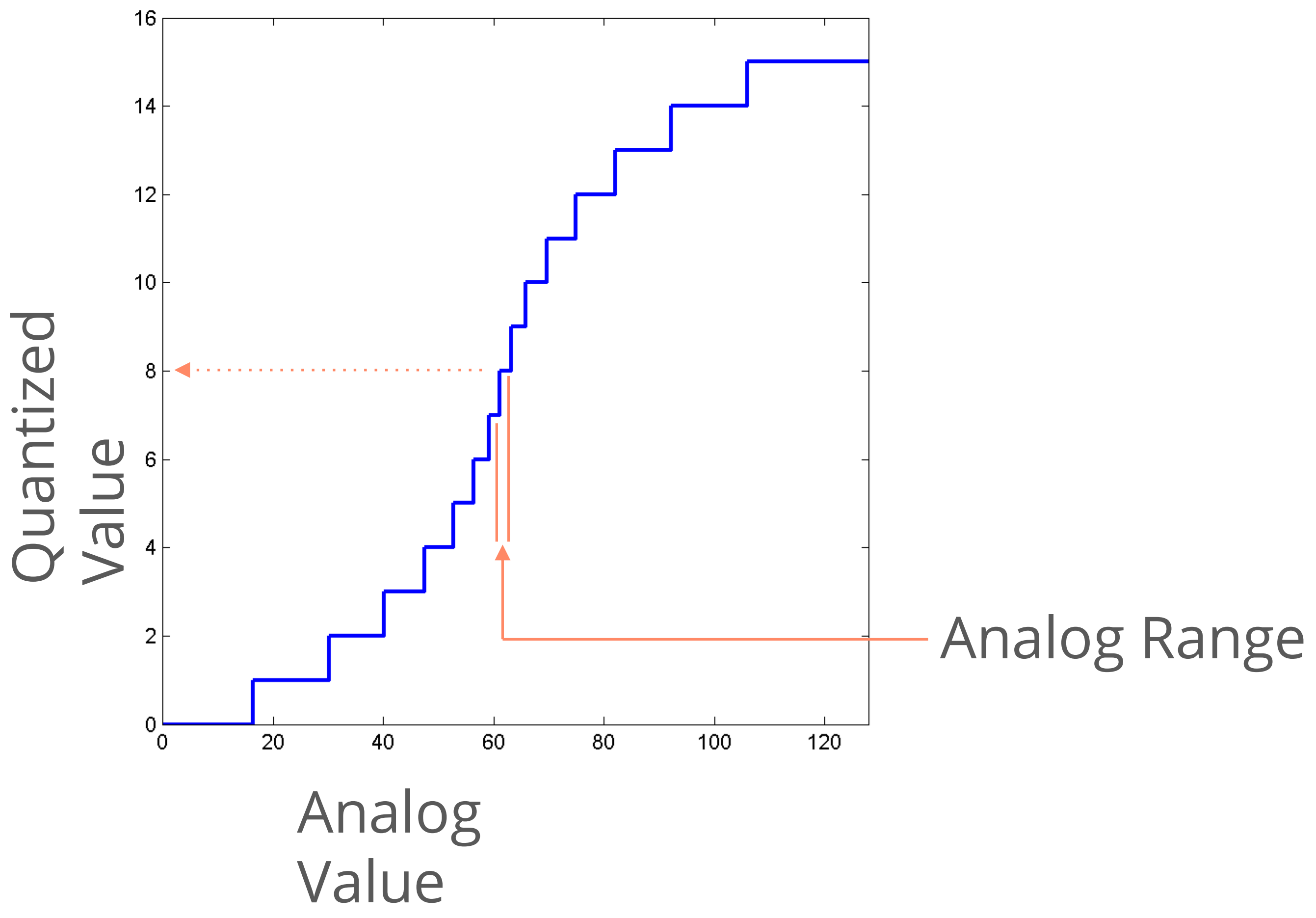
Below plots show PCM quantizations into 16 and 64 levels:





Nonlinear Coding

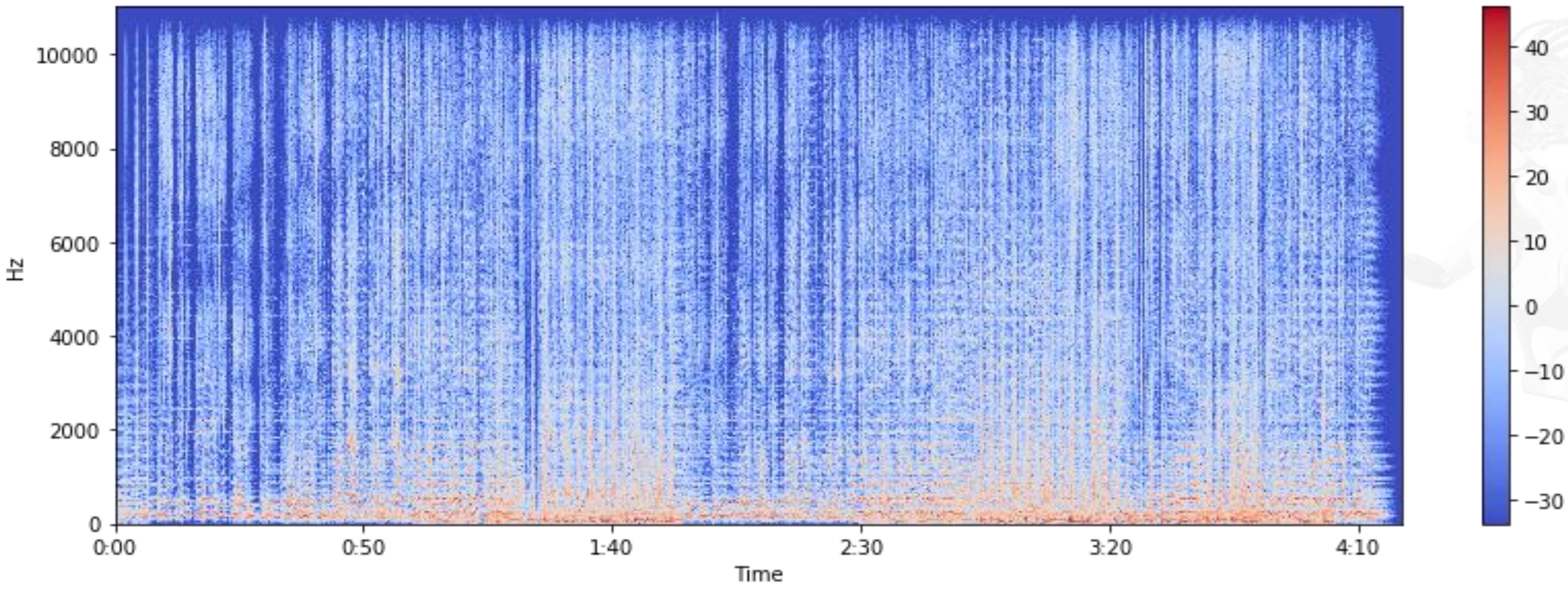
Converts nonuniform segments of the analog axis to uniform segments of the quantized axis



Spectrogram

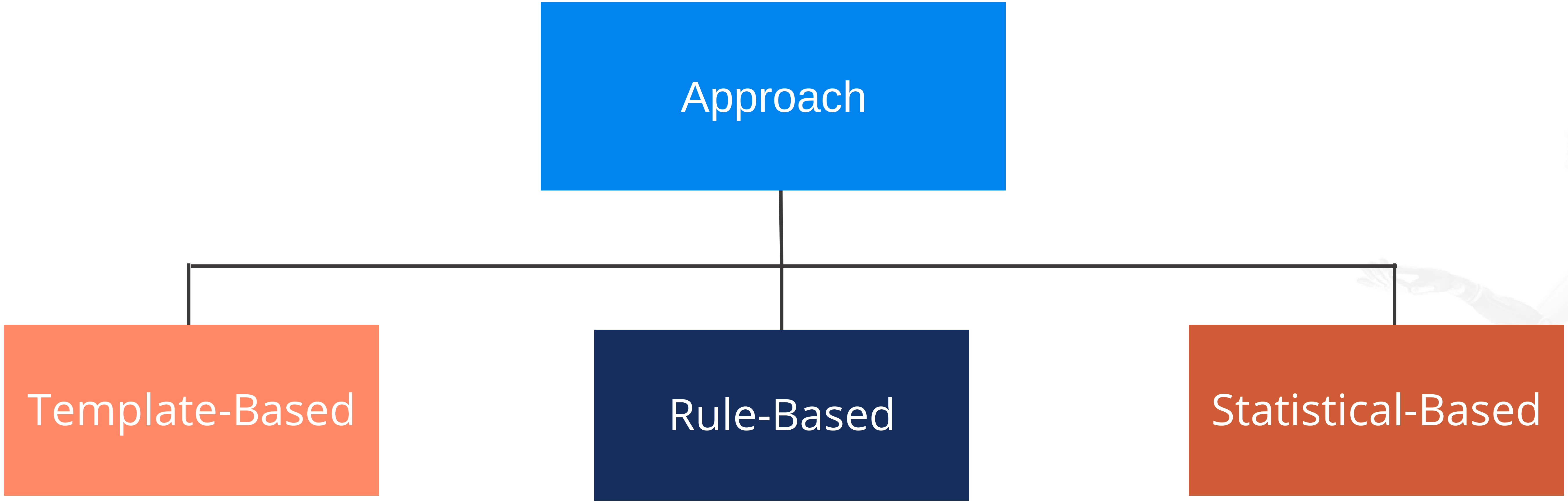
- Is a visual representation of the audio or other signal frequency range as it varies over time
- Translates signal in such a way that at a given time, you can know the amplitude of the given frequency
- Helps you determine the amplitude of an audio signal of different frequencies playing at a given time

Diagram of a Spectrogram



Approaches for Speech Recognition

Types of Approaches



Template-Based Approach

- Stores examples of units (words, phonemes) then finds the example that most closely fits the input
- Extracts features from speech signal, using solutions developed for all sorts of applications
- Works for discrete utterances and a single user

Template-Based Approach

- Hard to distinguish as very similar templates
- Result quickly degrades when input differs from templates
- Needs techniques to mitigate this degradation:
    - More subtle matching techniques
    - Multiple templates which are aggregated

Rule-Based Approach

- Uses knowledge of phonetics and linguistics to guide the search process

- Replaces templates with rules expressing anything and everything that help to decode:
  - Phonetics, phonology
  - Syntax
  - Pragmatics

Rule-Based Approach

- Uses “blackboard” architecture which:
  - At each decision point, lays out the possibilities
  - Applies rules to determine which sequences are permitted

Statistics-Based Approach

- Seen as an extension of template-based approach
- Uses more powerful mathematical and statistical tools than the template-based approach
- Is an “anti-linguistic” approach

Statistics-Based Approach

- Collects a large corpus of transcribed speech recordings
- Trains the computer to learn the correspondences
- Applies statistical processes to search through the space of all possible solutions at run time
- Selects the statistically and the most likely one

DATA AND
ARTIFICIAL INTELLIGENCE

Acoustic Modeling

Acoustic Modeling

> Acoustic modeling refers to the process of statistical representations of the computed feature vector sequences from the speech waveform.

Types of Acoustic Models

- Hidden Markov Model (HMM)
- Segmental Models
- Super-Segmental Models
- Maximum Entropy Models
- Conditional Random Fields (CRF)

Acoustic Modeling: Different Approaches

- Vector Quantization
- Gaussian PDFs
- Kernel Methods

Vector Quantization

- Computes the observation probabilities directly on the feature vectors
- Prefers probability density function over space

Vector Quantization: Pedagogical Steps

- Clusterize
- Get prototype vectors
- Compute distances with a metric
  - Euclidean
  - Mahanabolis

Vector Quantization: Pedagogical Steps

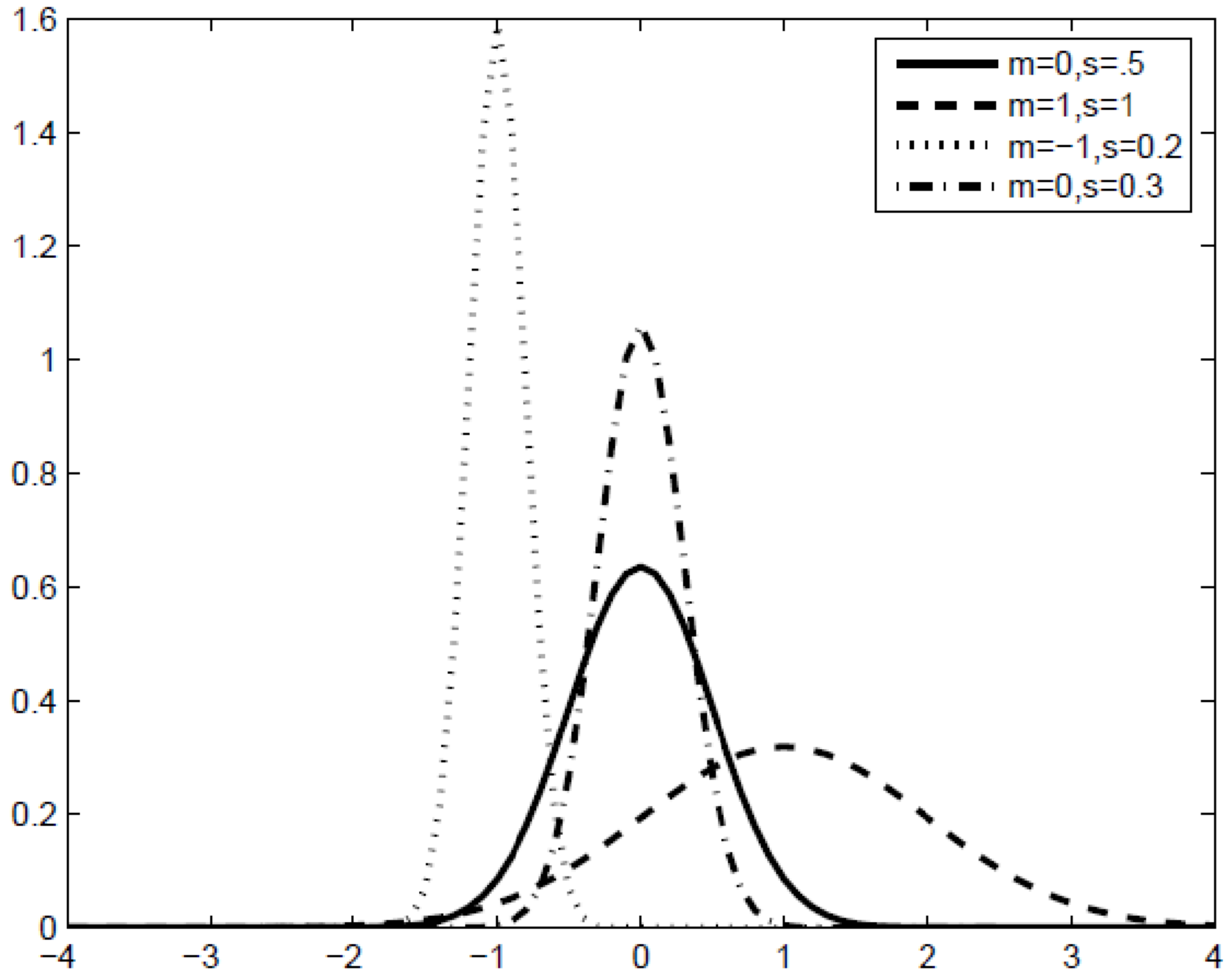
- Train with an algorithm
  - Knn
  - K-means
- Get the most probable symbol given as an observation

Gaussian PDFs

- Univariate Gaussians
  - Are the simplest usage of Gaussian probability estimator
  - Probability is equal to the area under the curve = 1

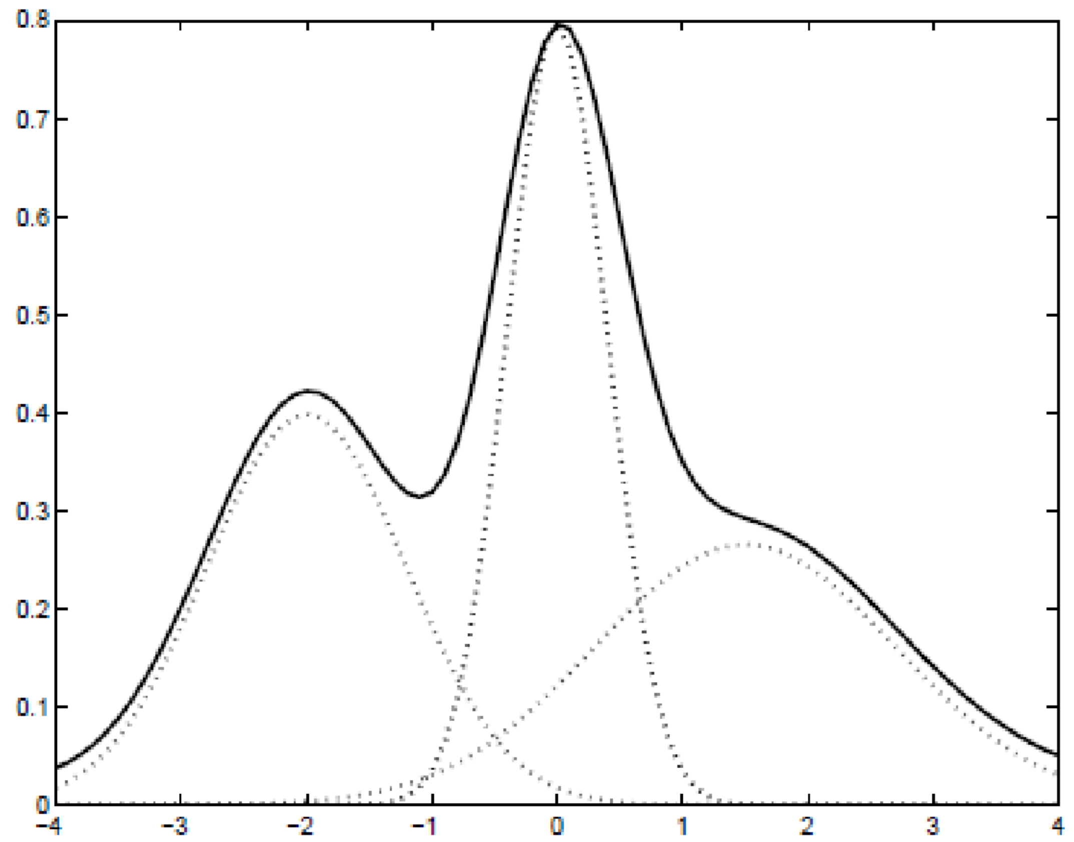
Gaussian PDFs

Univariate Gaussian shows the probable value of a feature to be generated by an HMM state.



Gaussian PDFs

- Multivariate Gaussians
  - Convert single spectral feature to a 39-dimension vector
  - Use a Gaussian rule for each feature



Gaussian PDFs

- Gaussian mixture models
  - Are weighted mixture of multivariate Gaussians
  - Are trained with the Baum-Welch algorithm

Language Modeling

Language Modeling

> A language modeling is the process of including the conceptual weaknesses of the language available to produce event probabilities.

DATA AND ARTIFICIAL INTELLIGENCE

Hidden Markov Model (HMM)

Hidden Markov Model

> A Hidden Markov Model is a statistical model in which the system being modeled is assumed to be a Markov process (memoryless process i.e. its future and past are independent) with hidden states.

Hidden Markov Model

- Has a set of states, each of which has limited number of transitions and emissions
- Each transition between states has an assigned probability
- Each model begins at **start state** and ends at **end state**

Markov Model (MM)

- Let's apply the Markov Model for the weather prediction
- Consider three types of weather:
  - Sunny
  - Rainy
  - Foggy
- Weather at day n is:

$$q_n \in \{\text{sunny}, \text{rainy}, \text{foggy}\}$$

- q_n depends on the known weathers of the past days (q_{n-1}, q_{n-2}, ...)

Markov Model (MM)

- According to the Markov model, you need to find:

$$\boxed{P(q_n \mid q_{n-1}, q_{n-2}, q_1)}$$

- The above equation computes the probability of any possible weather of today based on the past weather data

Markov Model (MM)

- For example, if the weather for last three days was sunny, sunny, and rainy, the probability that tomorrow would be foggy is:

$$\boxed{P(q_4 \;= \text{Foggy} \,|\; q_3 \;= \text{Rainy} |\; q_4 \;= \text{Sunny} \,|\; q_4 \;= \text{Sunny})}$$

Markov Model (MM)

- Make a simplified assumption of Markov:

$$\{q_1, q_2, \ldots, q_n\}$$

$$P(q_n \mid q_{n-1}, q_{n-2}, \ldots \ldots q_1) = P(q_n \mid q_{n-1})$$

- The weather for tomorrow depends only on today's weather (first order Markov model)

Markov Model (MM)

- For example, if yesterday's weather was rainy and today's is foggy, what is the probability that tomorrow it will be sunny?

$$P(q_3 = \text{Sunny} \mid q_2 = \text{Foggy},\ q_1 = \text{Rainy}) = \boxed{P(q_3 = \text{Sunny} \mid q_2 = \text{Foggy})}$$

$$= 0.2$$

Markov Assumption

Hidden Markov Model

- Consider this scenario for the hidden Markov model:

You are locked in a room for several days and you try to predict the weather outside. The only piece of evidence you have is the presence or absence of an umbrella with the person who comes into the room bringing your daily meals.

Hidden Markov Model

- Assume the probabilities as shown in the table:

Weather	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

Probability $P(x_i \mid q_i)$ of carrying an umbrella $(x_i = \text{true})$ based on the weather q_i on day i

Hidden Markov Model

- Finding the probability of a certain weather:

$$q_n \in \{\,\text{sunny}, \text{rainy}, \text{foggy}\}$$

- The formula is based on observations, x.

Hidden Markov Model

- Using Bayes's rule:

$$P(q_i \mid x_i) = \frac{P(q_i \mid x_i)P(q_i)}{P(x_i)}$$

- For observations of n days:

$$P(q_{i\,\ldots,}\, q_n \mid x_{1\,\ldots,}\, x_n) = \frac{P(x_{1\,\ldots,}\, x_n \mid q_{i\,\ldots,}\, q_n)P(q_{i\,\ldots,}\, q_n)}{P(x_{1\,\ldots,}\, x_n)}$$

Three Fundamental Problems of HMMs

Evaluation	For a given model, compute the probability that a particular output sequence was produced by the model
Decoding	For a given model, find the most likely sequence of hidden states which could have generated a given output sequence
Learning	For a given set of output sequences, find the most likely set of state transition and output probabilities

Knowledge Check

Knowledge Check

1

Which of the following algorithms is best suited for speech recognition?

a. Hidden Markov Model

b. Markov Model

c. Besiyan Method

d. All the above

Knowledge Check 1

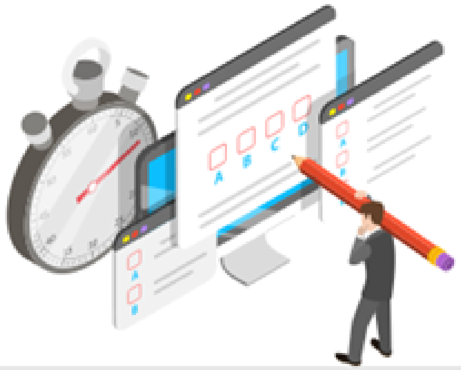
Which of the following algorithms is best suited for speech recognition?

a. Hidden Markov Model

b. Markov Model

c. Besiyan Method

d. All the above



The correct answer is **a**

Hidden Markov Model is the best algorithm for speech recognition.

Knowledge Check 2

Sound waves travel in __________ direction.

a. Longitudinal

b. Transverse

c. Both a and b

d. None of the above

Knowledge Check

2

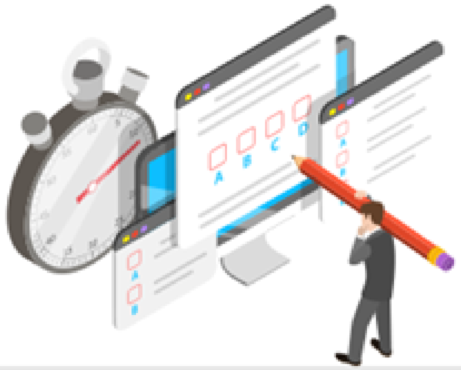
Sound waves travel in _________ direction.

a. Longitudinal

b. Transverse

c. Both a and b

d. None of the above



The correct answer is **a**

Sound waves travel in longitudinal direction.

Knowledge Check

3

It is necessary to convert analog signals to digital before feeding into a model because ______.

a. Analog signals are not understandable by model

b. Digital signals are easily interpretable by model

c. Both a and b

d. None of the above

Knowledge Check

3

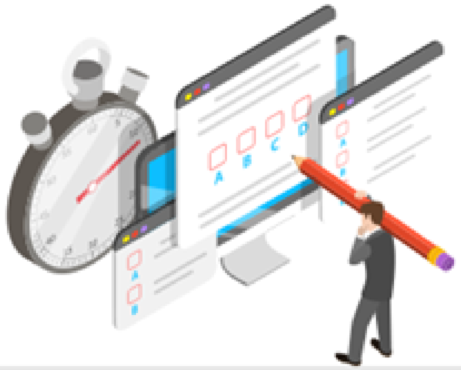
It is necessary to convert analog signals to digital before feeding into a model because ______.

a. Analog signals are not understandable by model

b. Digital signals are easily interpretable by model

c. Both a and b

d. None of the above



The correct answer is **b**

Digital signal can easily be interpreted by models because each sample of digital signal is represented with a series of bits that are either in state of 1 (on) or 0 (off).

Knowledge Check

4

Which of the following methods is used to match the unknown speech to a collection of prerecorded words or model?

a. Template Matching

b. Acoustic Modeling

c. Language Modeling

d. None of the above

Knowledge Check

4

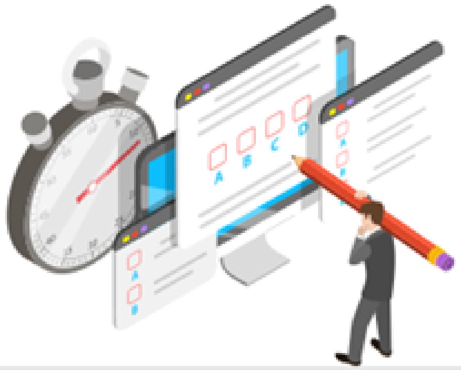
Which of the following methods is used to match the unknown speech to a collection of prerecorded words or model?

a. Template Matching

b. Acoustic Modeling

c. Language Modeling

d. None of the above



The correct answer is **a**

Template matching matches unknown speech to a collection of prerecorded words or model.

Knowledge Check

5

Which of the following models comes under acoustic models?

a. Segmental Models

b. Super-Segmental Models

c. Maximum Entropy Models

d. All the above

Knowledge Check

5

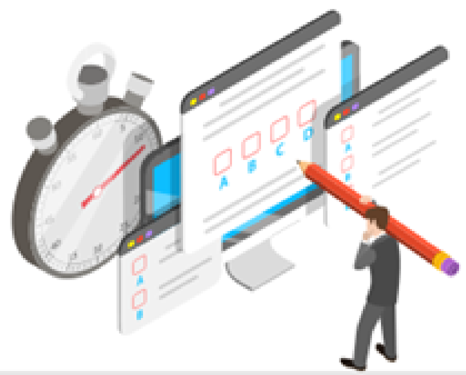
Which of the following models comes under acoustic models?

a. Segmental Models

b. Super-Segmental Models

c. Maximum Entropy Models

d. All the above



The correct answer is **d**

Segmental models, super-segmental models, and maximum entropy models come under acoustic models.

Key Takeaways

- Nonlinear coding converts nonuniform segments of the analog axis to uniform segments of the quantized axis.
- Language modeling is the process of including the conceptual weaknesses of the language available to produce event probabilities.