Speech and Language Processing

Lecture 1
Introduction and Preparation

Information and Communications Engineering Course

Takahiro Shinozaki

Manabu Okumura

2022/10/4

Lecture Plan (Shinozaki's part)

I gives the first 6 lectures about speech recognition. Through these lectures, the backbone of the latest speech recognition techniques is explained.

1. 10/4 (remote)
   Introduction and Preparation
2. 10/4 (remote)
   Probability Distributions, Markov Models, Samplings
3. 10/6 (remote)
   Maximum Likelihood Estimation and EM Algorithm
4. 10/6 (remote)
   Bayesian Networks and Bayesian Inference
5. 10/7 (remote)
   Neural networks
6. 10/7 (remote)
   Reinforcement Learning

Handouts

- All the materials are available at my home page:

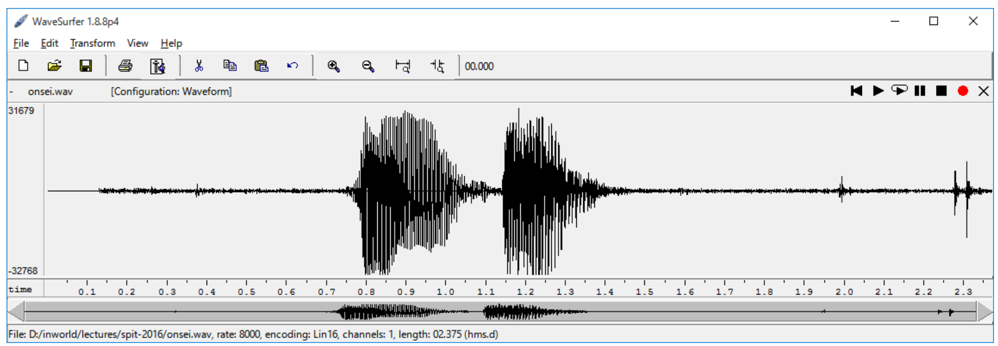
http://www.ts.ip.titech.ac.jp/shinot/lectures/asrintro/

Perception of Speech Sound

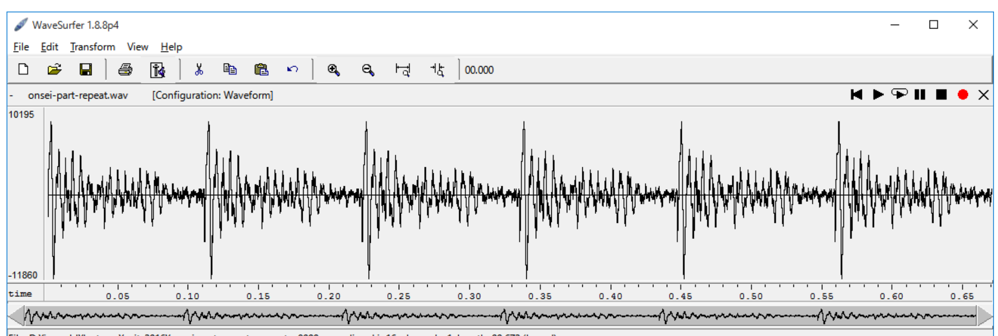
- Speech is a sound made by a physical process
- Frequency and time pattern are important

Sound is available in web version

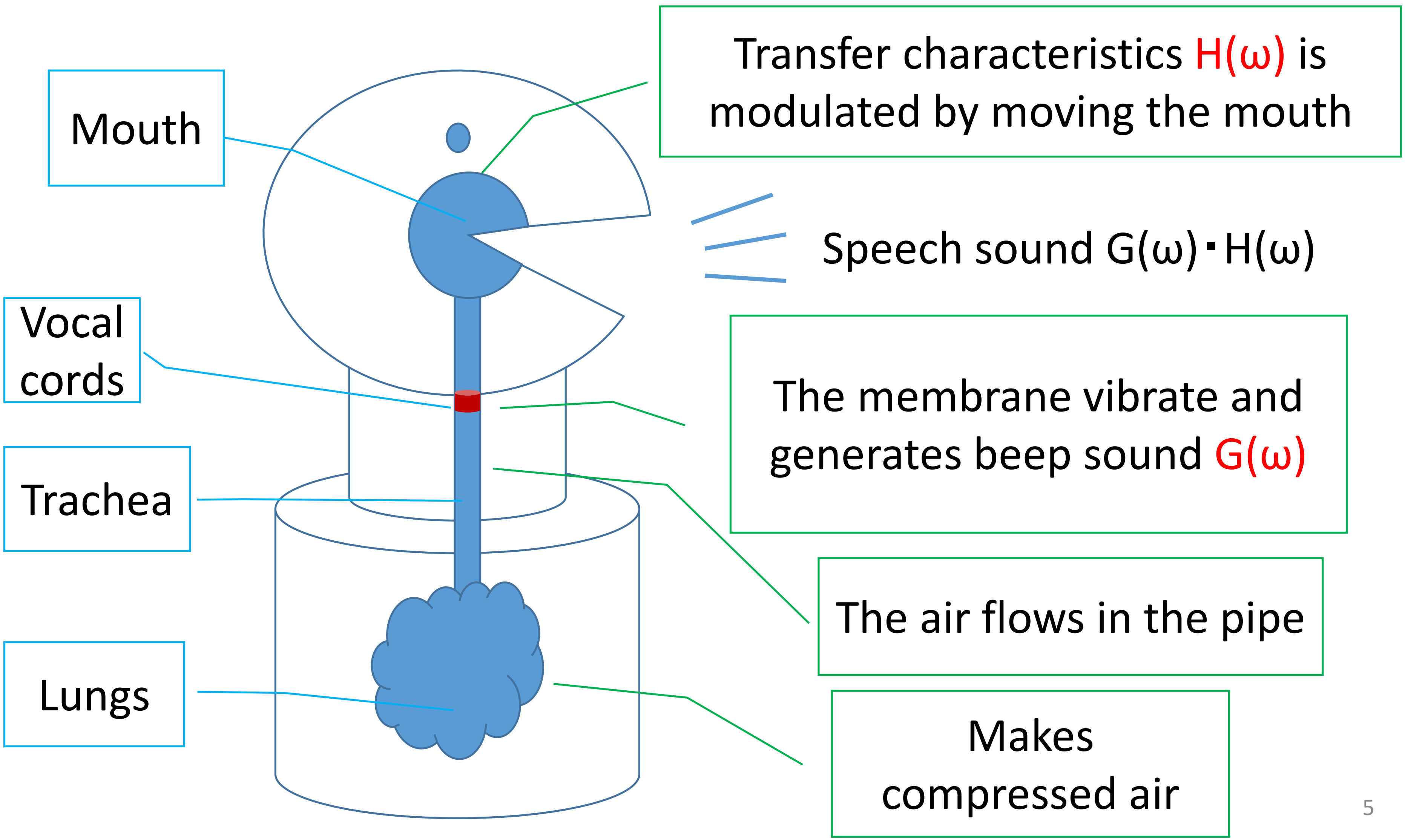
Speech sound pronunciating "Onsei"



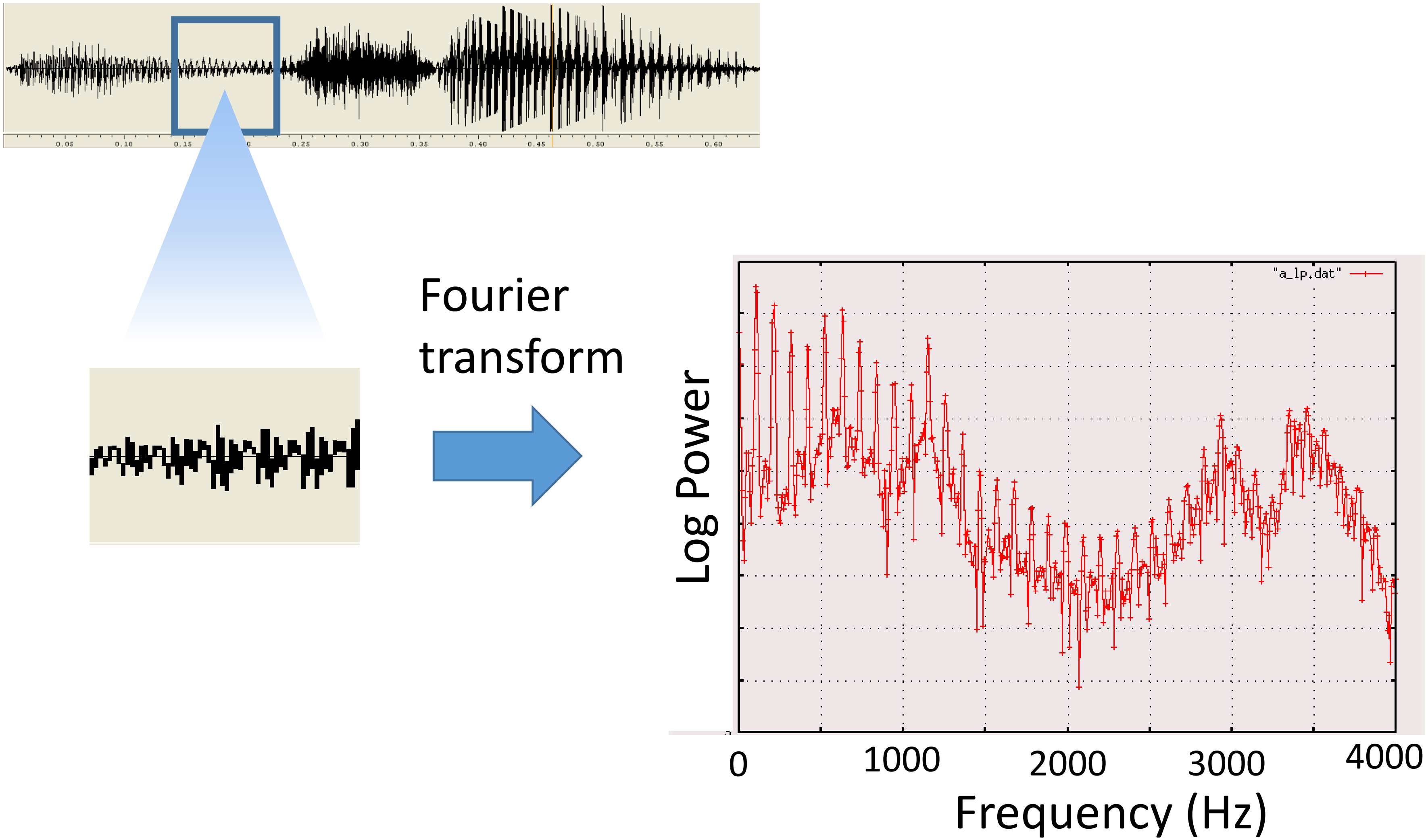
Cut the region near "n" and repeat it 6 times.



Utterance Generation

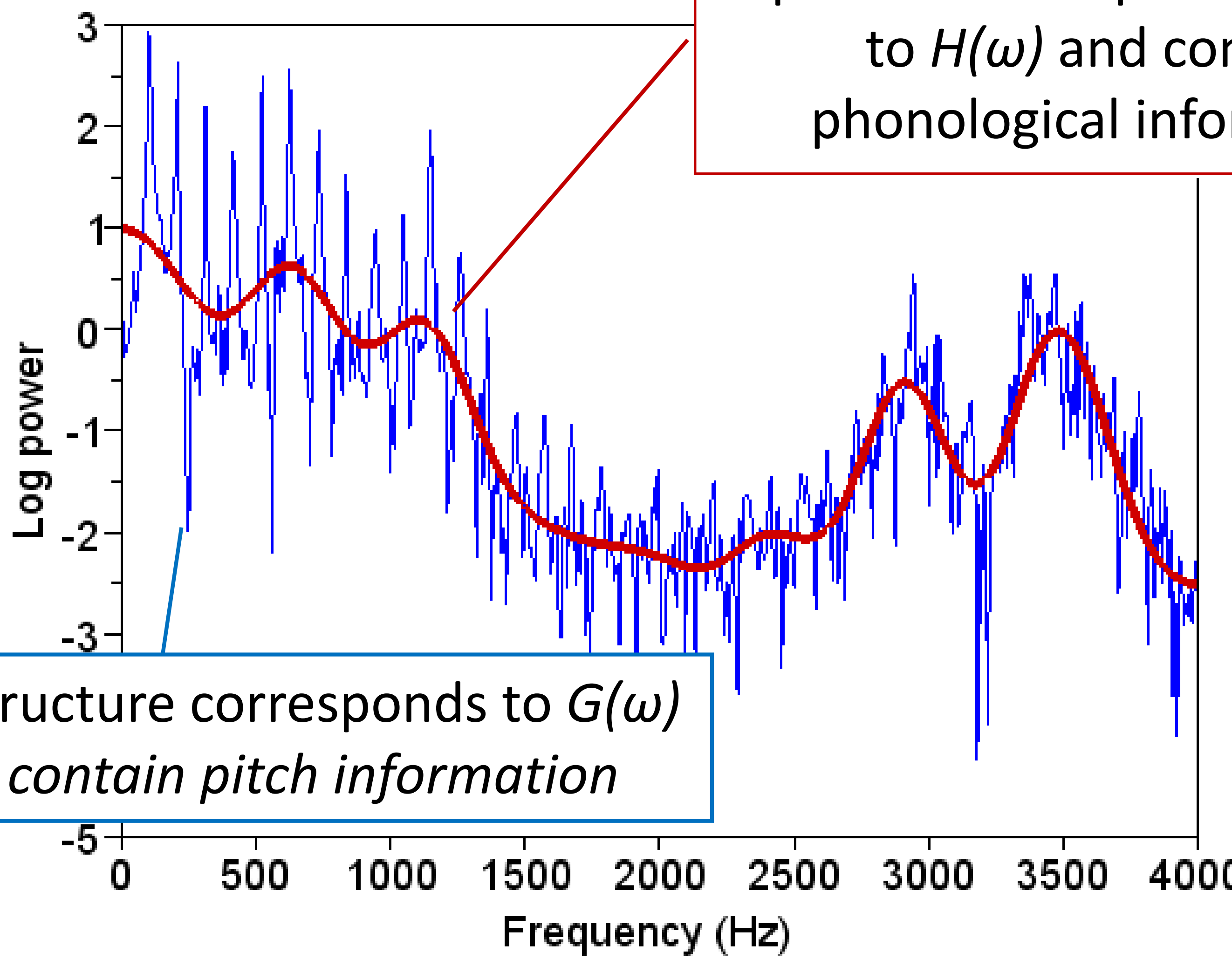


Spectral Analysis



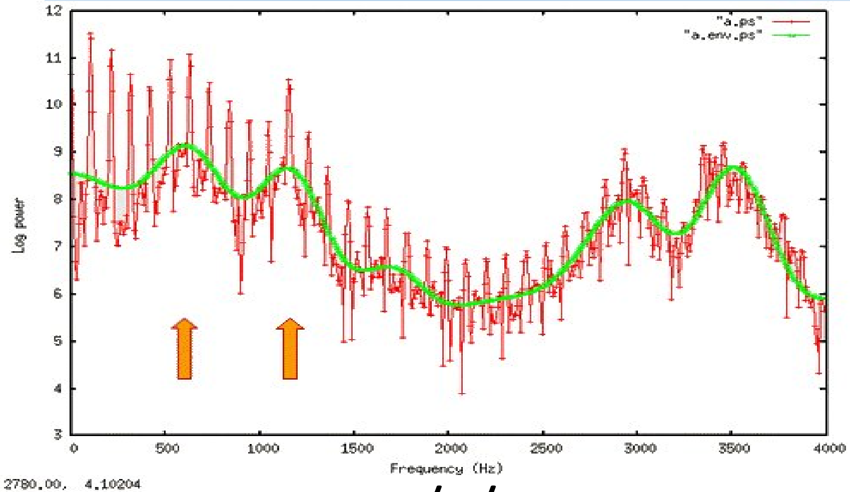
Spectral Envelope

Spectral envelope corresponds to *H(ω)* and contains phonological information

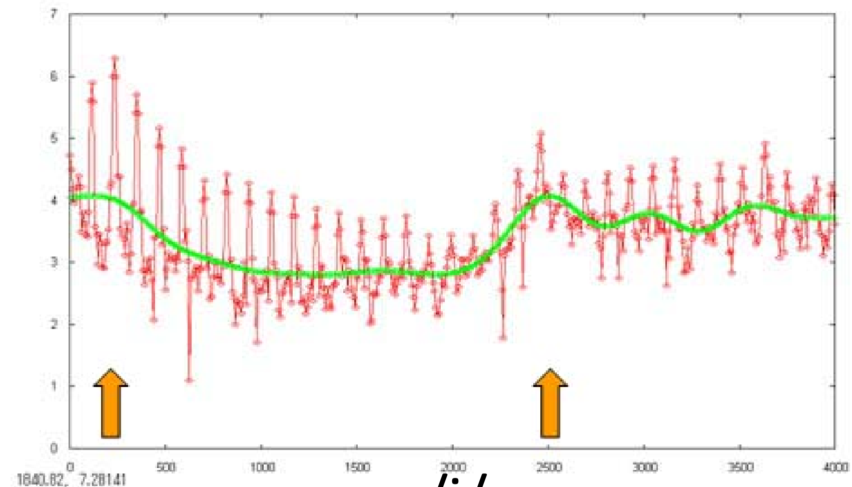


Fine structure corresponds to *G(ω)* *and contain pitch information*

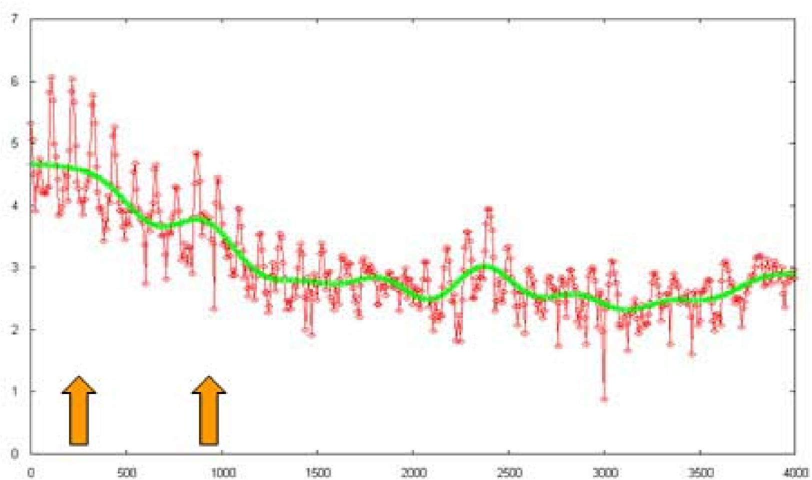
Vowels and Spectral Envelopes



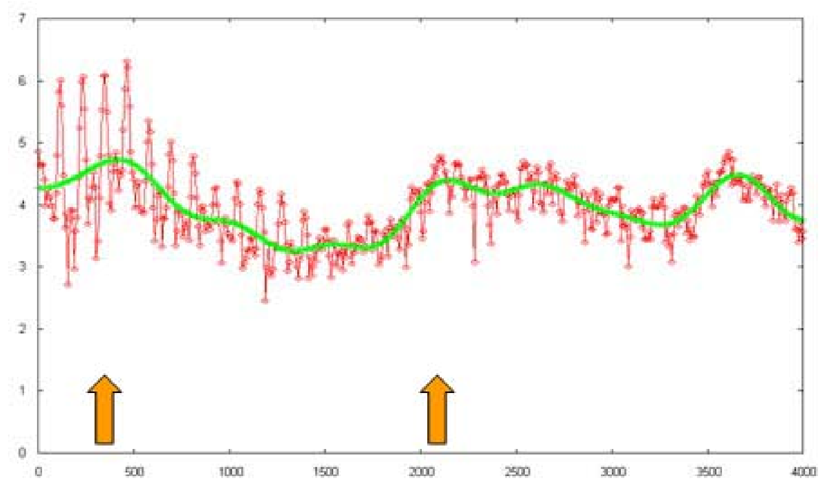
/a/



/i/



/u/



/e/

Experiment: Replacing the Sound Source ***G***

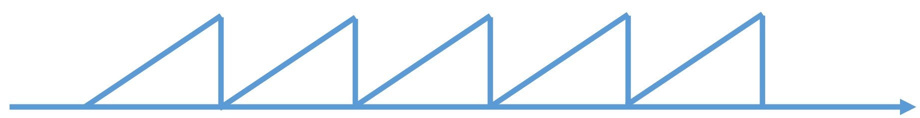
1. Record a voice

2. Analyze the voice, and decompose it to the sound source ***G*** and the transmission characteristics ***H***

X(ω) (original voice)	=	G(ω)	×	H(ω)

4. Replace the sound source ***G*** with another one ***G'***

E.g : sawtooth wave



5. Compute G'(ω) × H(ω) and re-generate waveform

G'(ω)	×	H(ω)	=	X'(ω) (synthesized sound)

Synthesized Voice Changing *G*

Sawtooth G’ (100Hz) → $G'(\omega)H(\omega)$

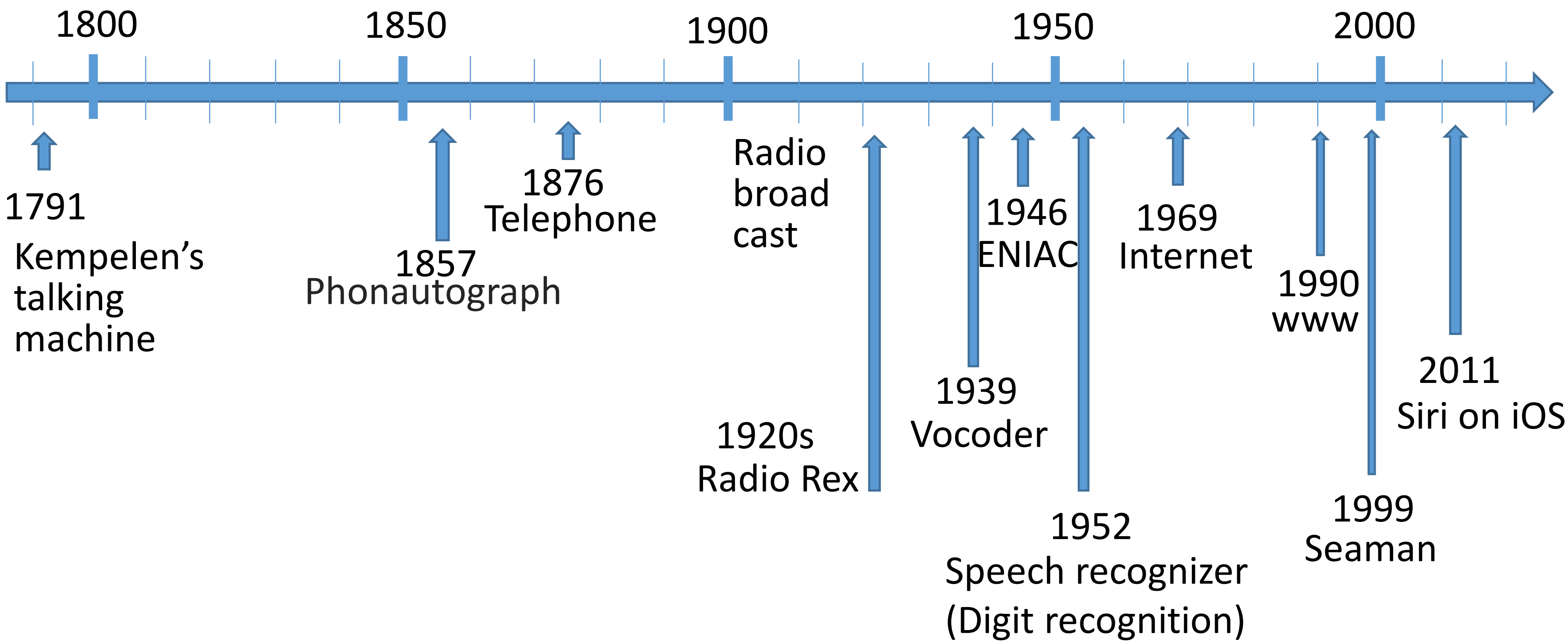
Sawtooth G’ (300Hz) → $G'(\omega)H(\omega)$

Sawtooth G’ (500Hz) → $G'(\omega)H(\omega)$

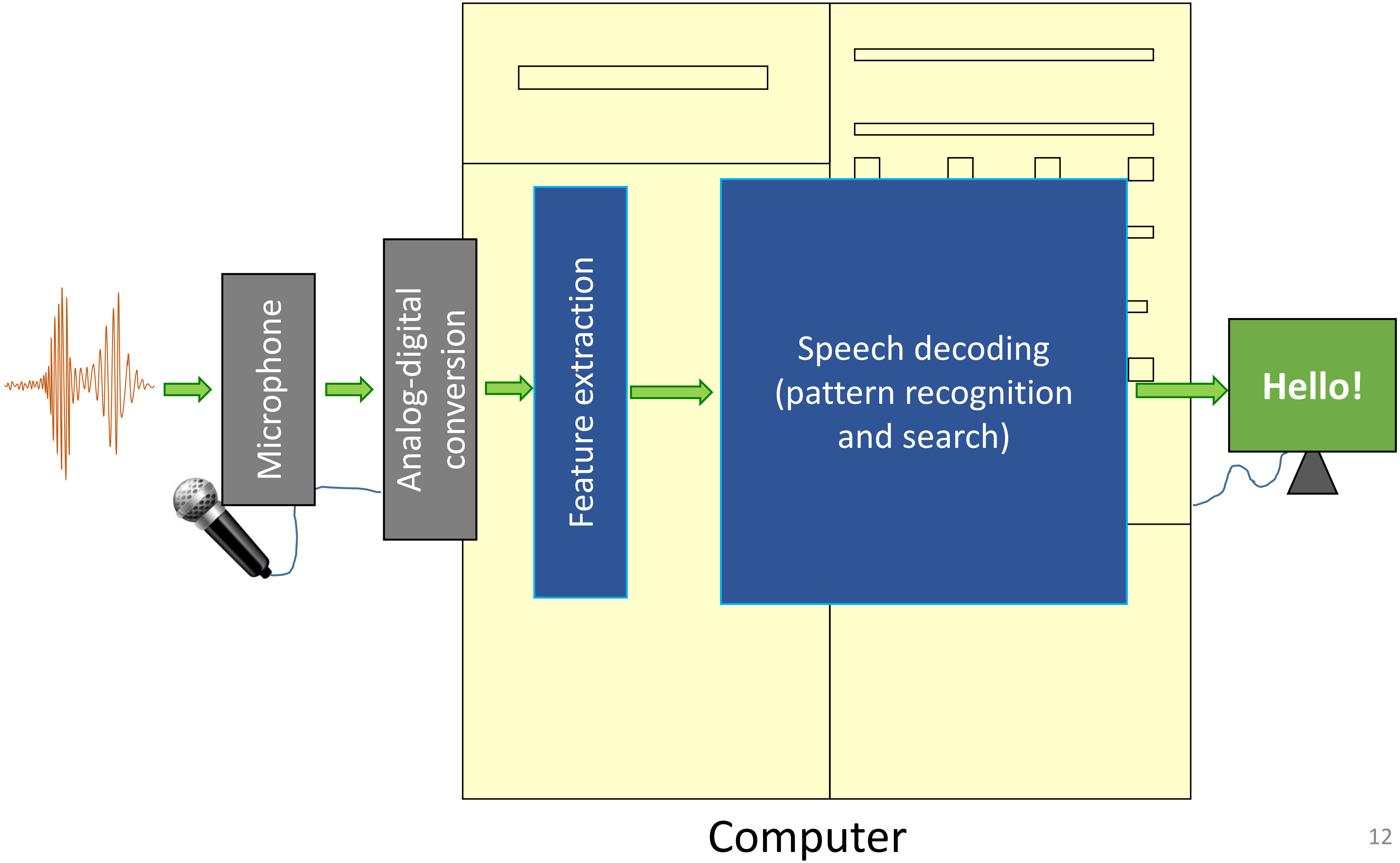
Music G’ (Acoustic11*) → $G'(\omega)H(\omega)$

*The music is from: http://maoudamashii.jokersounds.com/

History of Speech Technology

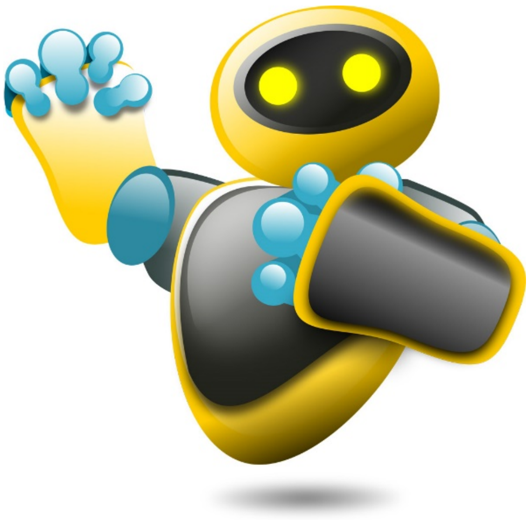


Organization of a Speech Recognition System



Applications of Speech Recognition

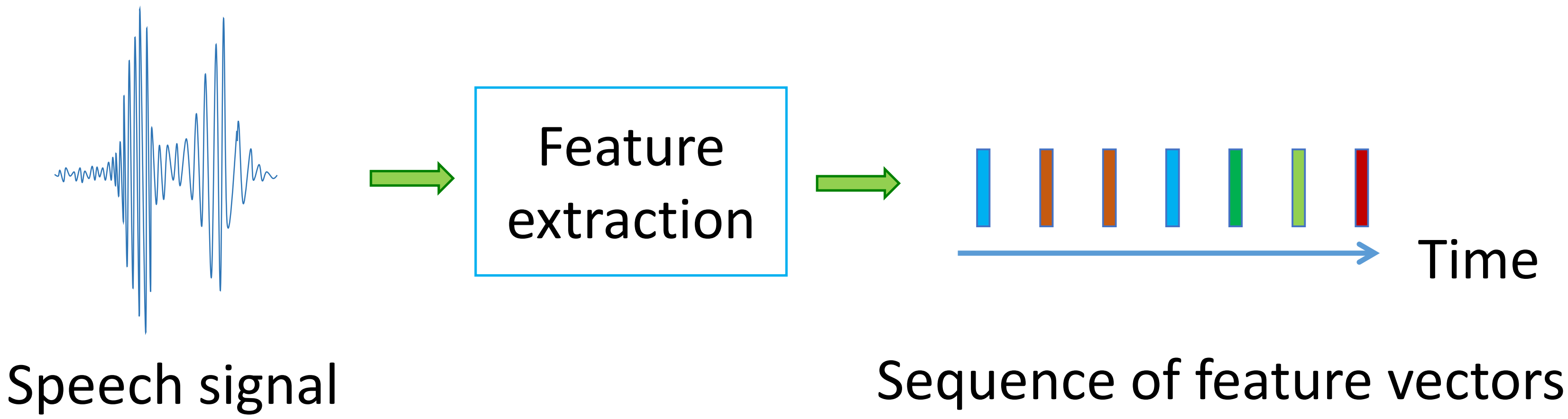
- Smartphone
  - Voice assistance
  - Speech-to-speech translation
- Judge
  - Speech retrieval system to support citizen judge
- Television
  - Automatic captioning system
- Car navigation
  - Voice commands
- Toy robots
  - Speech conversation



Feature Extraction

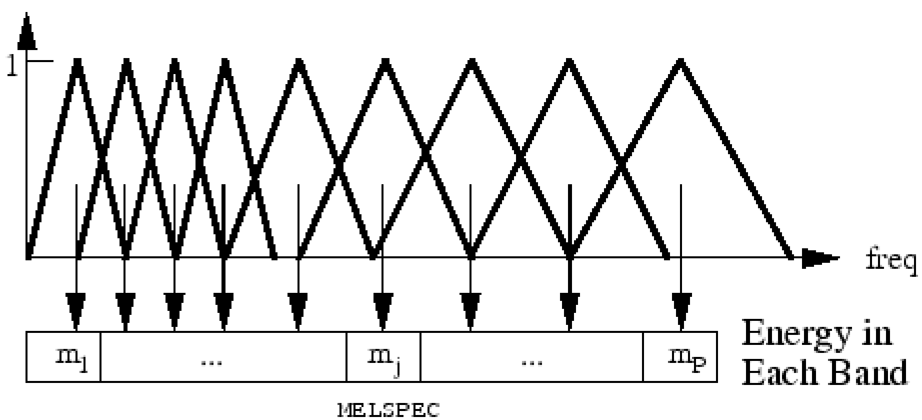
Extract useful information from the input signal in a convenient form for pattern recognition

- Help improving pattern recognition performance
- Reduce unnecessary memory and processing costs

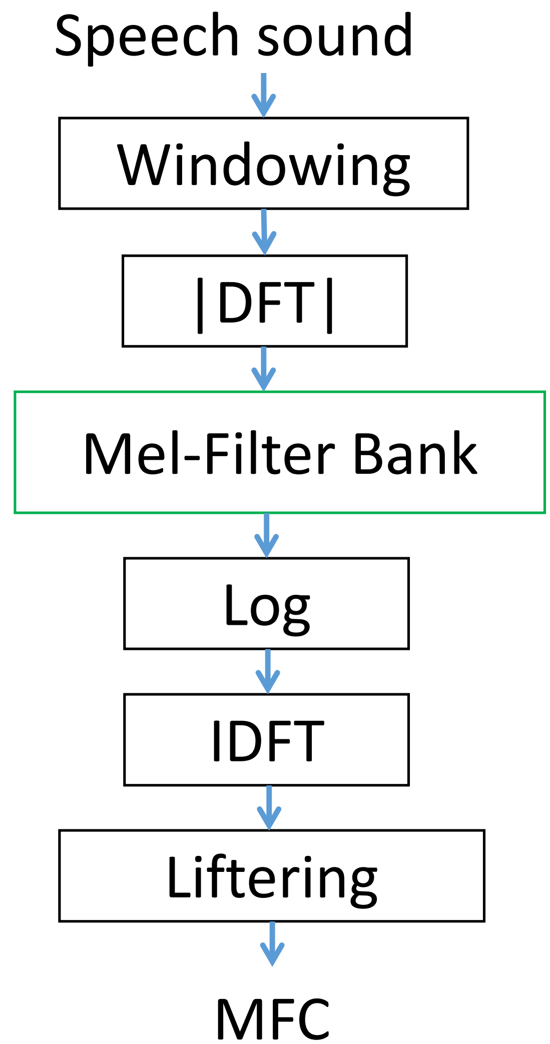


Mel-Frequency Cepstrum (MFC)

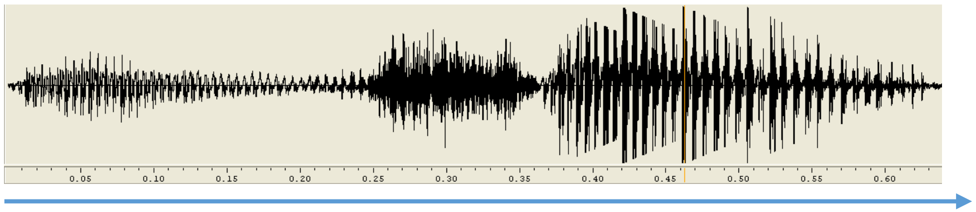
- Widely used features for speech recognition
- Emulate perceptual scale of pitches by using Mel-scale filter bank



Mel-Scale Filter Bank



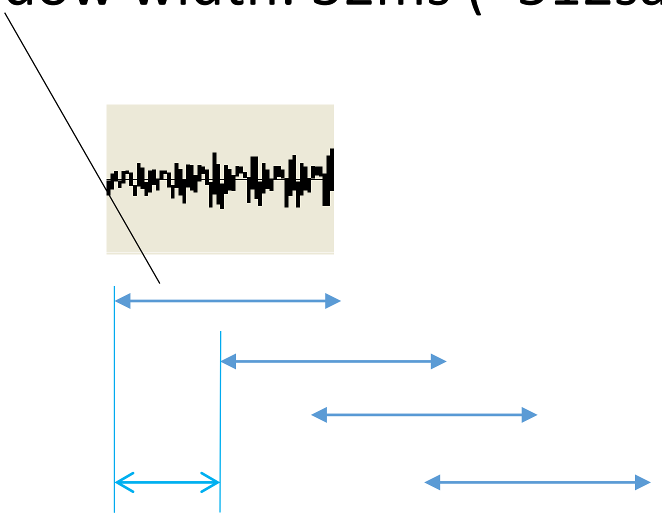
Typical Feature Extraction Process



16kHz sampling
16bit quantization

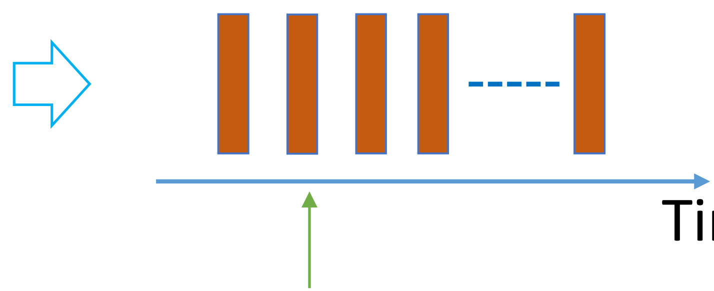
Time

Window width: 32ms (=512samples/16kHz)



shift: 10ms

Feature sequence
Sequence of real valued vectors
Rate=100Hz



Time

A vector is called a "frame"

Speech Decoding

***O**:* Input acoustic features (or a feature sequence)
***W** : A* symbol (or a symbol sequence) to recognize
e.g. phone, word, word sequence, etc.

Statistical Speech Recognition

- Use probability distribution to model speech sounds

$$\underbrace{\hat{W} = \arg\max_{W}}_{\text{Speech recognizer}} P(W \mid O)$$

Speech model

Brief Review of Probability Theory

Probability Space

- Sample space (Ω)
  - Set of all possible outcomes of an experiment
- Probability function ($f(x)$)
  - A function that maps each outcome to a probability

$$f(x) \in [0, 1] \quad \text{for all } x \in \Omega$$

$$\sum_{x \in \Omega} f(x) = 1$$

- Event (E)
  - Subset of the sample space

Probability of an event E is :

$$P(E) = \sum_{x \in E} f(x)$$

Random Variable

- A function that maps an outcome of an experiment to a value
  - Notation:
    "$P[X=x] = p$" means "the probability of a random variable X takes a value x is p"

Example

X: The value of a die

Y: Whether the value of a die is odd number or not

Z: Whether the value of a die is larger than 2 or not

Random Variable	1	2	3	4	5	6
X	1	2	3	4	5	6
Y	1	0	1	0	1	0
Z	0	0	1	1	1	1

Joint Probability

- Probability that more than one events jointly occur

Example

X : Dice 1

Y : Dice 2

$P(X = i, Y = j)$: Probability that the value of X is i and the value of Y is j

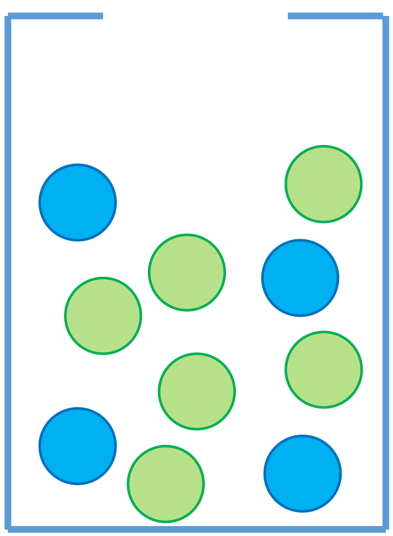
Note: $P(X=i,\ Y=j) = P(Y=j,\ X=i)$

Conditional Probability

- Probability of an event given that another event has occurred

Example

Randomly picks up two balls sequentially from a box containing 4 blue and 6 green balls



X : Color of the first ball

Y : Color of the second ball

$$P(Y = blue \mid X = geen) = \frac{4}{9}$$

$$P(Y = blue \mid X = blue) = \frac{3}{9}$$

Two Principal Rules

- Sum rule
  - Summing joint probability $P(X,Y)$ for all possible values of Y gives probability of $P(X)$
  - $P(X)$ is called the marginal probability

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$$

- Product rule
  - Product of probability $P(X)$ and joint probability $P(Y|X)$ is equal to joint probability $P(X,Y)$

$$P(X = x_i, Y = y_j) = P(Y = y_j \mid X = x_i) P(X = x_i)$$

Bayes' Theorem

- From the product rule, we obtain:

$$P(Y = y_j \mid X = x_i) = \frac{P(X = x_i \mid Y = y_j) P(Y = y_j)}{P(X = x_i)} \quad for\ \forall x_i, y_i$$

If we simplify the notation, we have:

$$\boxed{P(Y \mid X) = \frac{P(X \mid Y) P(Y)}{P(X)}}$$

(Bayes' theorem)

Using the sum rule, $P(Y|X)$ is obtained from joint probability as:

$$P(Y \mid X) = \frac{P(X \mid Y) P(Y)}{P(X)} = \frac{P(X, Y)}{\sum_Y P(X, Y)}$$

Independence

- If the joint distribution of two variables X and Y factorizes into the product of the marginals, then X and Y are said to be "independent"

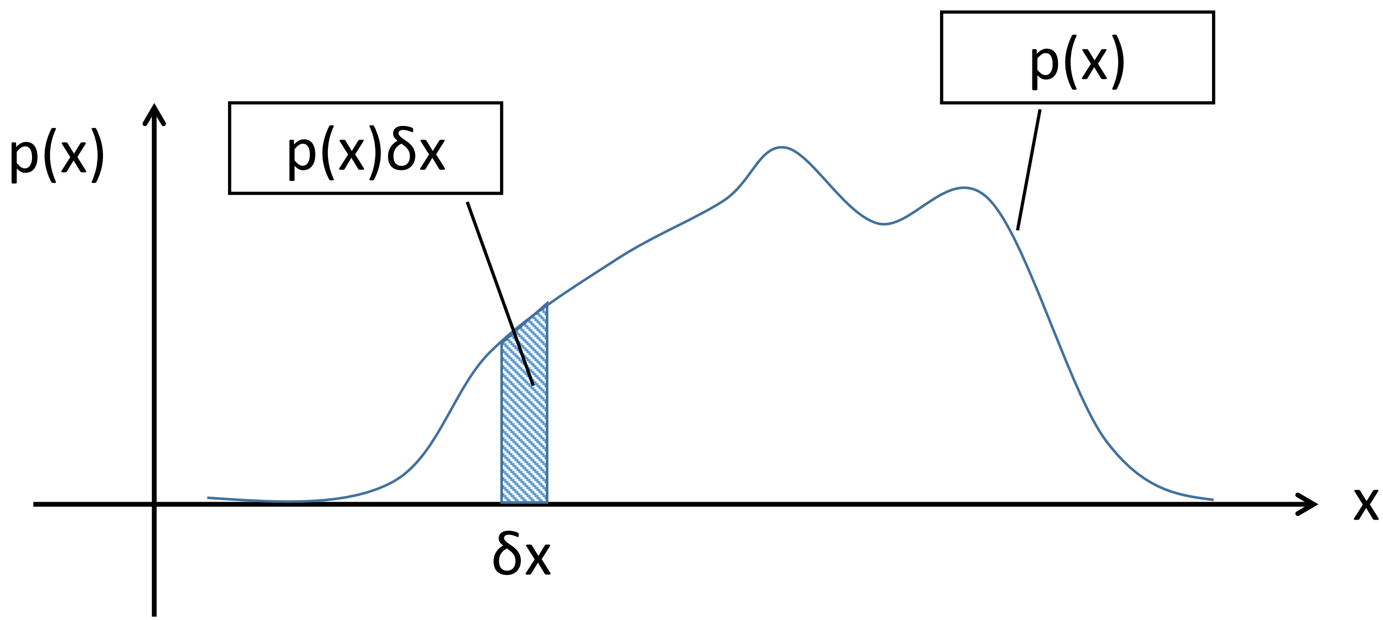
$$P(X,Y) = P(X \mid Y)P(Y) = P(X)P(Y)$$ ⇨ X and Y are independent

$$P(X,Y) = P(X \mid Y)P(Y) \neq P(X)P(Y)$$ ⇨ X and Y are not independent

Probability Densities

- If the probability of a real-valued variable x falling in the interval $(x, x+\delta x)$ is given by $p(x)\delta x$ when $\delta x \rightarrow 0$, $p(x)$ is called the probability density of x

p(x)δx is probability ⇒ $p(x) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} p(x)dx = 1$



The Sum and The Product Rules For Continuous Variable

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$$

$$\Rightarrow \quad p(x) = \int p(x, y) dy$$

$$P(X = x_i, Y = y_j) = P(Y = y_j \mid X = x_i) P(X = x_i)$$

$$\Rightarrow \quad p(x, y) = p(y \mid x) p(x)$$

Expectation

- Expectation of a function $f(x)$ under a probability distribution $p(x)$ is denoted by $E[f]$

$$E[f] = \sum_x p(x) f(x)$$

(x is discrete)

$$E[f] = \int p(x) f(x) dx$$

(x is continuous)

Mean and Variance

- Mean
  - Synonym of the expectation $E[f(x)]$
- Variance
  - A measure of how much variability there is in $f(x)$ around its mean value $E[f(x)]$

$$\operatorname{var}[f] \equiv E\left[\left(f(x) - E[f(x)]\right)^2\right] = E\left[f(x)^2\right] - E[f(x)]^2$$

  - In particular, the variance of the variable x itself is:

$$\operatorname{var}[x] = E\left[\left(x - E[x]\right)^2\right]$$

Covariance

- Covariance
  - The extent to which x and y vary together

$$\mathrm{cov}[x,y] \equiv E_{x,y}\left[\left(x - E[x]\right)\left(y - E[y]\right)\right]$$
$$= E_{x,y}[xy] - E[x]E[y]$$

Expectation with respect to joint probability of x and y

Entropy

- Amount of randomness in the random variable

$$H[x] = E[-\log(p(x))] = -\sum_x p(x) \log p(x)$$

Example

x	0	1
p(x)	0.5	0.5

$$H[x]$$
$$= -0.5\log(0.5) - 0.5\log(0.5)$$
$$= 0.693$$

x	0	1
p(x)	0.1	0.9

$$H[x]$$
$$= -0.1\log(0.1) - 0.9\log(0.9)$$
$$= 0.325$$

Relative Entropy

- A measure of dissimilarity of two distributions p and q
  - Also called as kullback-Leibler (KL) divergence

$$KL(p \| q) = E_p\left[\log\left(\frac{p(x)}{q(x)}\right)\right]$$

$$= -\int p(x)\log\left(\frac{q(x)}{p(x)}\right)dx$$

- KL(p||q) is nonnegative.
  KL(p||q)=0 if and only if p(x)=q(x)

Note: KL(p||q) ≠ KL(q||p)

Approximating Expectation with Sampling

When $x_1, \mathrm{x}_2, \cdots, x_N$ are samples independently drawn from a distribution $p(x)$

$$E[f] = \sum_x p(x) f(x) \approx \frac{1}{N} \sum_{n=1}^{N} f(x_n)$$ (x is discrete)

$$E[f] = \int p(x)\, f(x) dx \approx \frac{1}{N} \sum_{n=1}^{N} f(x_n)$$ (x is continuous)

Exercise 1.1

Joint probability of random variables A and B are given in the following table. According to the table, for example,
$P(A = 0, B = 0) = 0.2$

$P(A, B)$	B=0	B=1	B=2
A=0	0.2	0.1	0.1
A=1	0.3	0.2	0.1

1. Obtain $P(A = 0)$
2. Obtain $P(A = 0|B = 0)$