

# Speech Production and Perception

# Fundamentals of Speech Processing

- What is speech?
- What is speech processing?
- Speech processing for telecommunication, human-computer interaction and biomedical tasks
- Why speech processing is needed?
- Elements of speech communication
- Steps in human speech production and perception
- Speech production and perception mechanisms
- Language constraints on speech processing
- Acoustic phonetic characterization of speech
- Knowledge sources in speech
- Technological goals in speech processing
- History of speech processing

# Elements of Speech Communication

## Human:



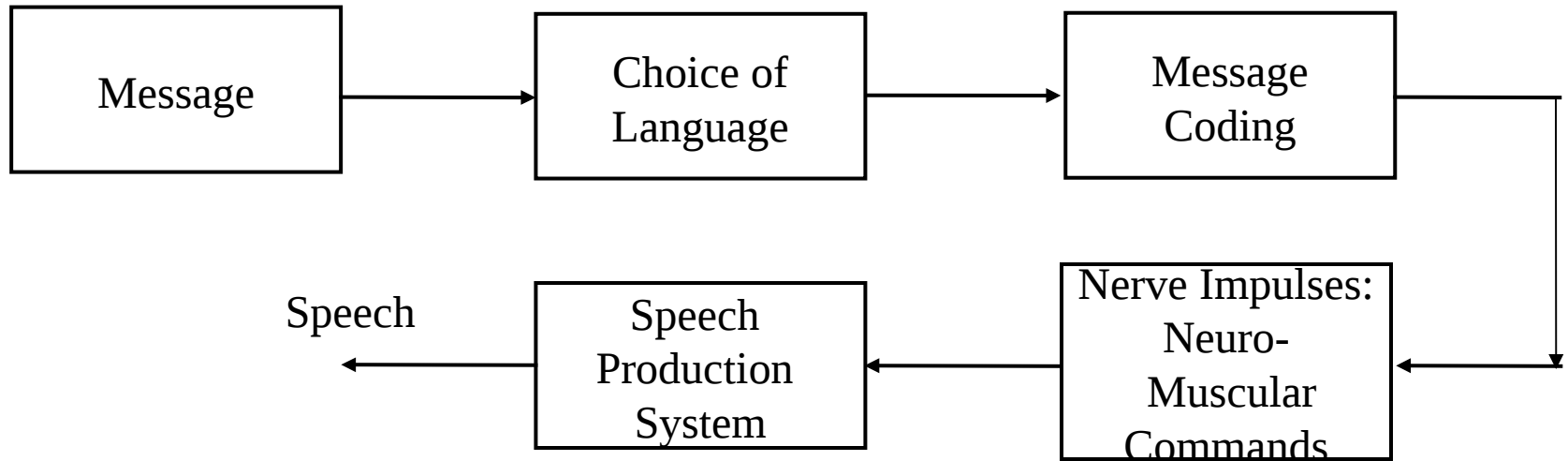
## Electrical:



- **Talker:** Message formulation and conveying via speech mode.
- **Listener:** Reception of speech and message comprehension.
- **Medium:** Physical medium which carries speech from talker to listener.

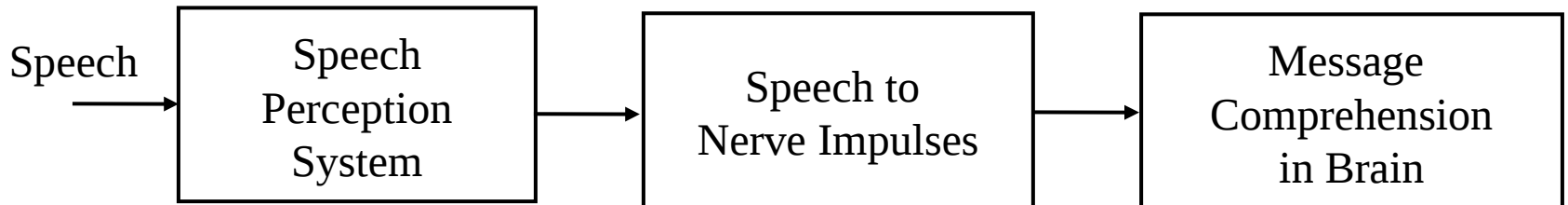
# Speech Chain: Steps in Human Speech Communication

## Talker:



Message->Language Coding->Nerve Impulses-> Mechanical Motion-> Acoustic Pressure Variations

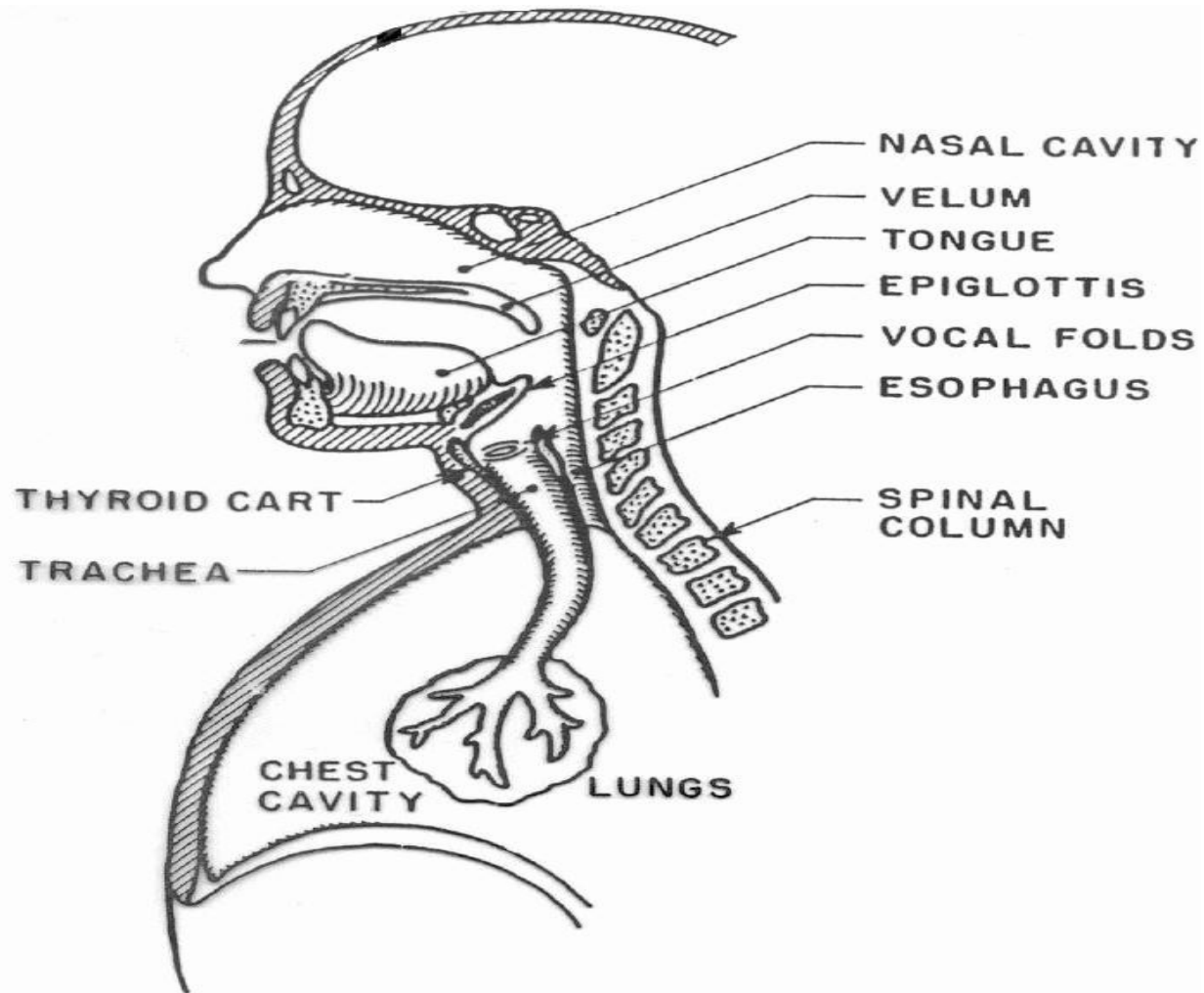
## Listener:



Acoustic Pressure Variations->Mechanical Motion->Nerve Impulses->Language Decoding->Message

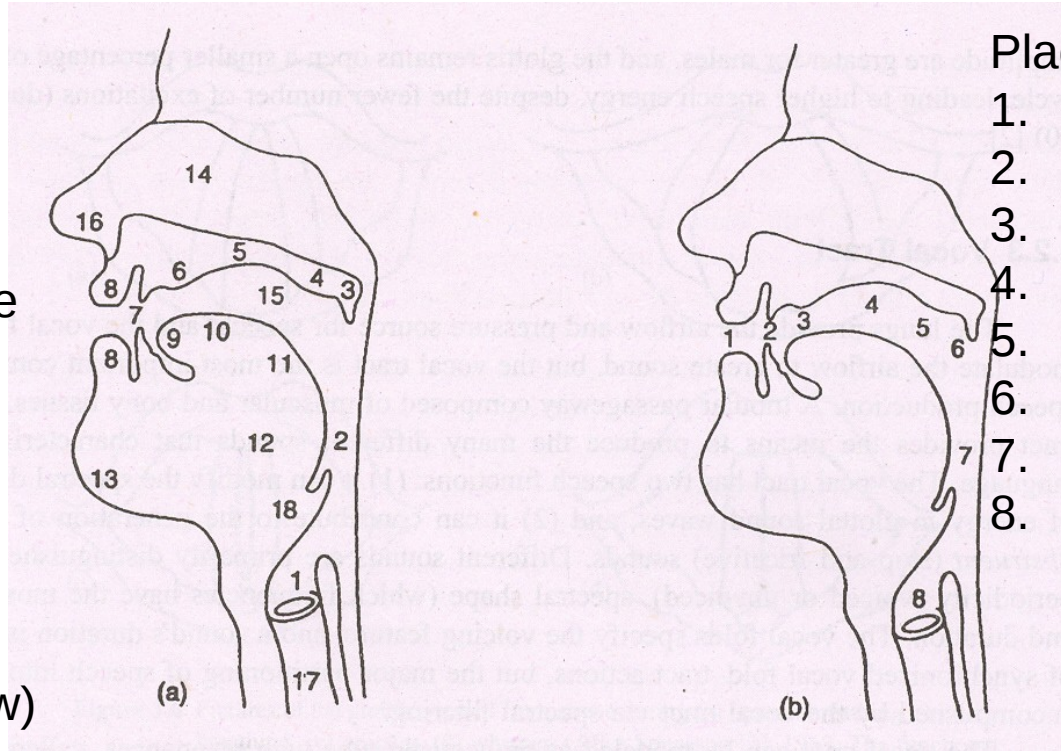
- **Dual role of a person as talker/listener**

# Speech Production System



# Vocal tract

1. Vocal folds
2. Pharynx
3. Velum
4. Soft palate
5. Hard palate
6. Alveolar ridge
7. Teeth
8. Lips
9. Tongue tip
10. Blade
11. Dorsum
12. Root
13. Mandible (jaw)
14. Nasal cavity
15. Oral cavity
16. Nostrils
17. Trachea
18. epiglottis



## Places of articulation

1. Labial
2. Dental
3. Alveolar
4. Palatal
5. Velar
6. Uvular
7. Pharyngeal
8. glottal

# Speech Production Mechanism

- Speech is produced during exhalation of air
- Lungs & associated structure provides required energy
- Vocal-folds inside larynx is the main excitation source and constriction (total or partial) along vocal tract is an additional source
- Supra-glottal system which includes pharynx, oral cavity and nasal cavity behave as time-varying resonator

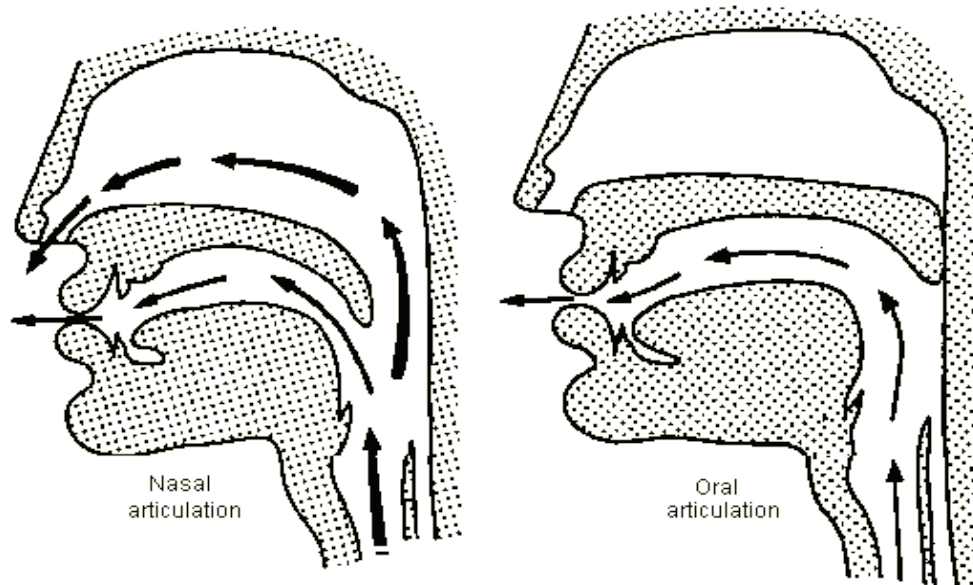
# Excitation Sources

- **Voiced Excitation**
  - Vibration of vocal folds
  - Voiced speech
- **Unvoiced Excitation**
  - Total constriction along the vocal tract
  - Partial constriction along the vocal tract
  - Unvoiced speech
- **Mixed Excitation**
  - Combination of above
  - Mixed speech

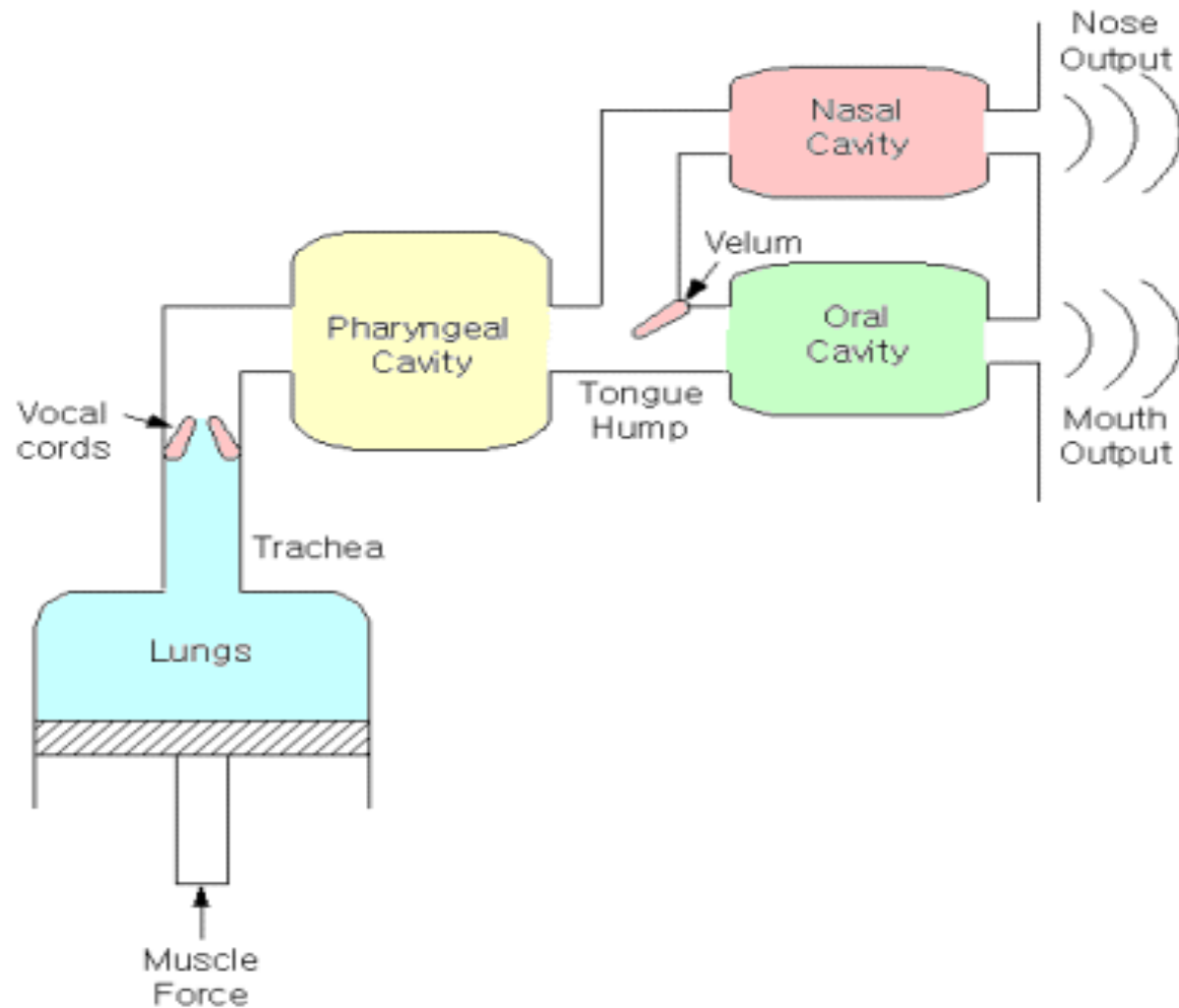


# Production of Speech Sounds

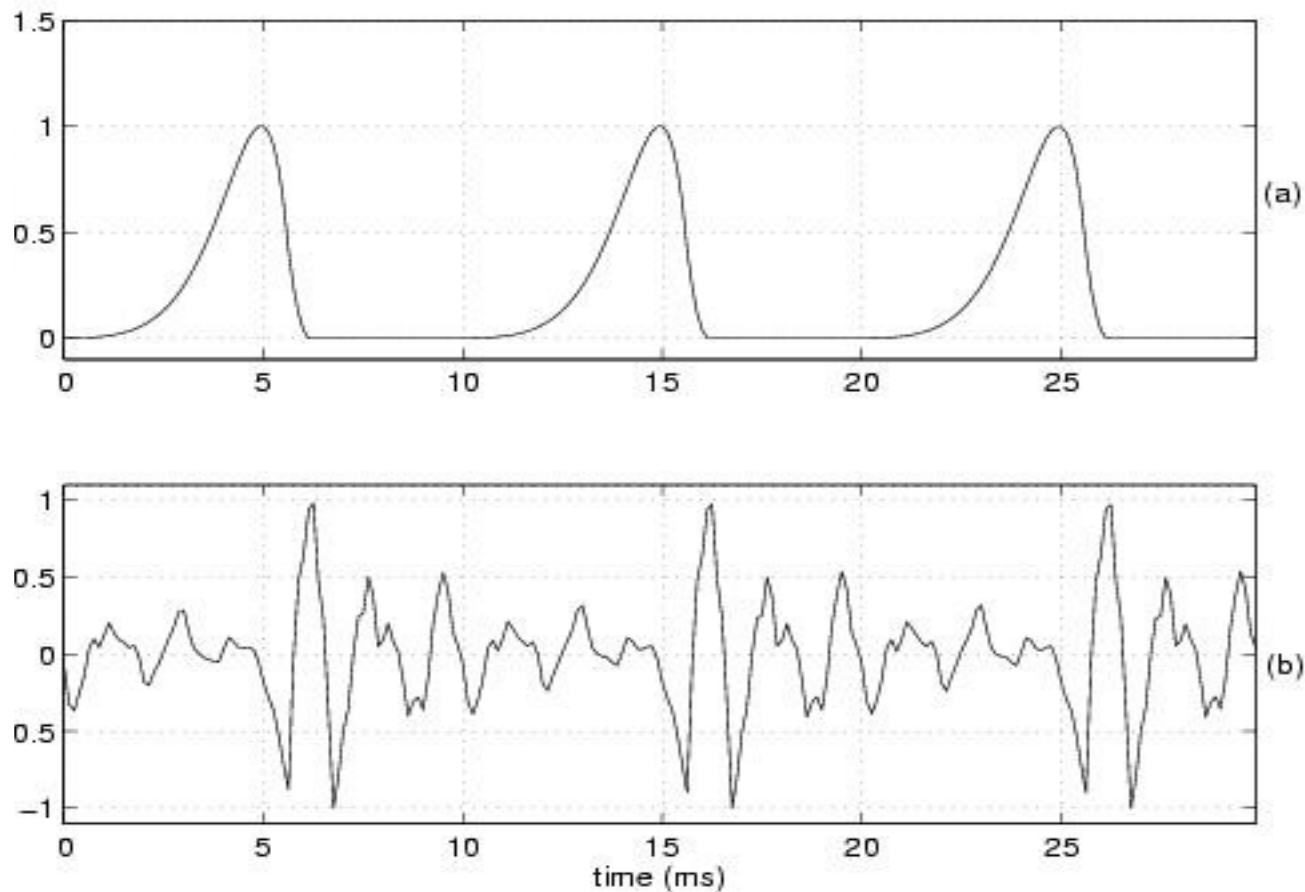
- Vowels - Oral cavity is wide opened, tongue hump, glottal vibration
- Unvoiced Consonants - constriction (Complete closure)
- Voiced Consonants - constriction & glottal vibration
- Nasal Sounds - nasal cavity is coupled
- Fricatives – Partial closure, Narrow opening



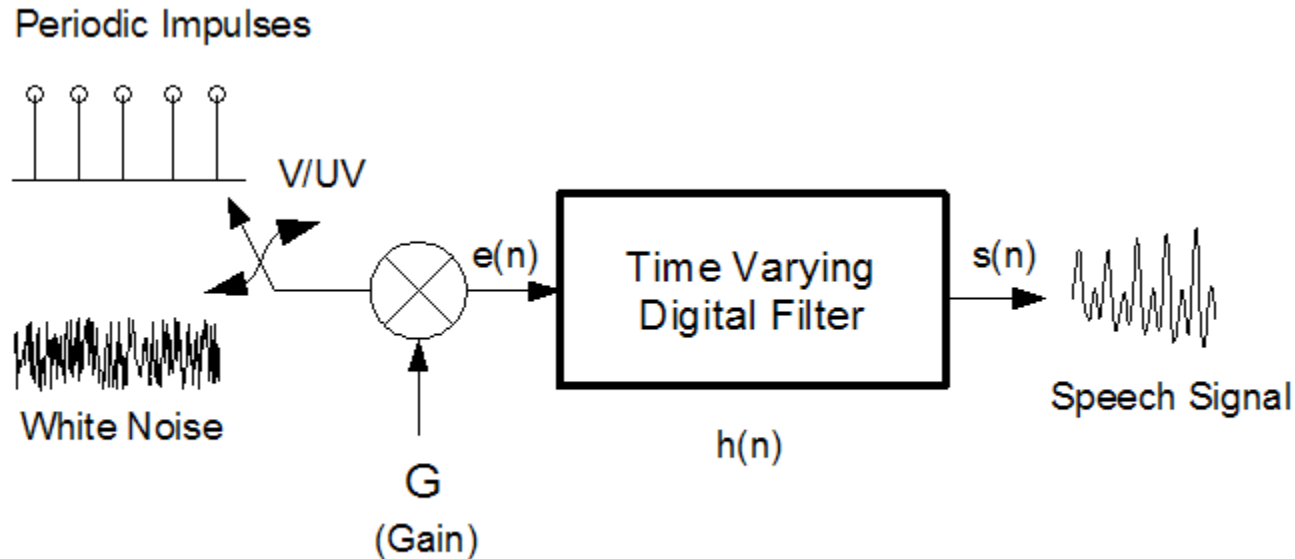
# Schematic Representation of Speech Production Process



# Glottal Volume Velocity and its Speech Signal



# Digital Model for Speech Production

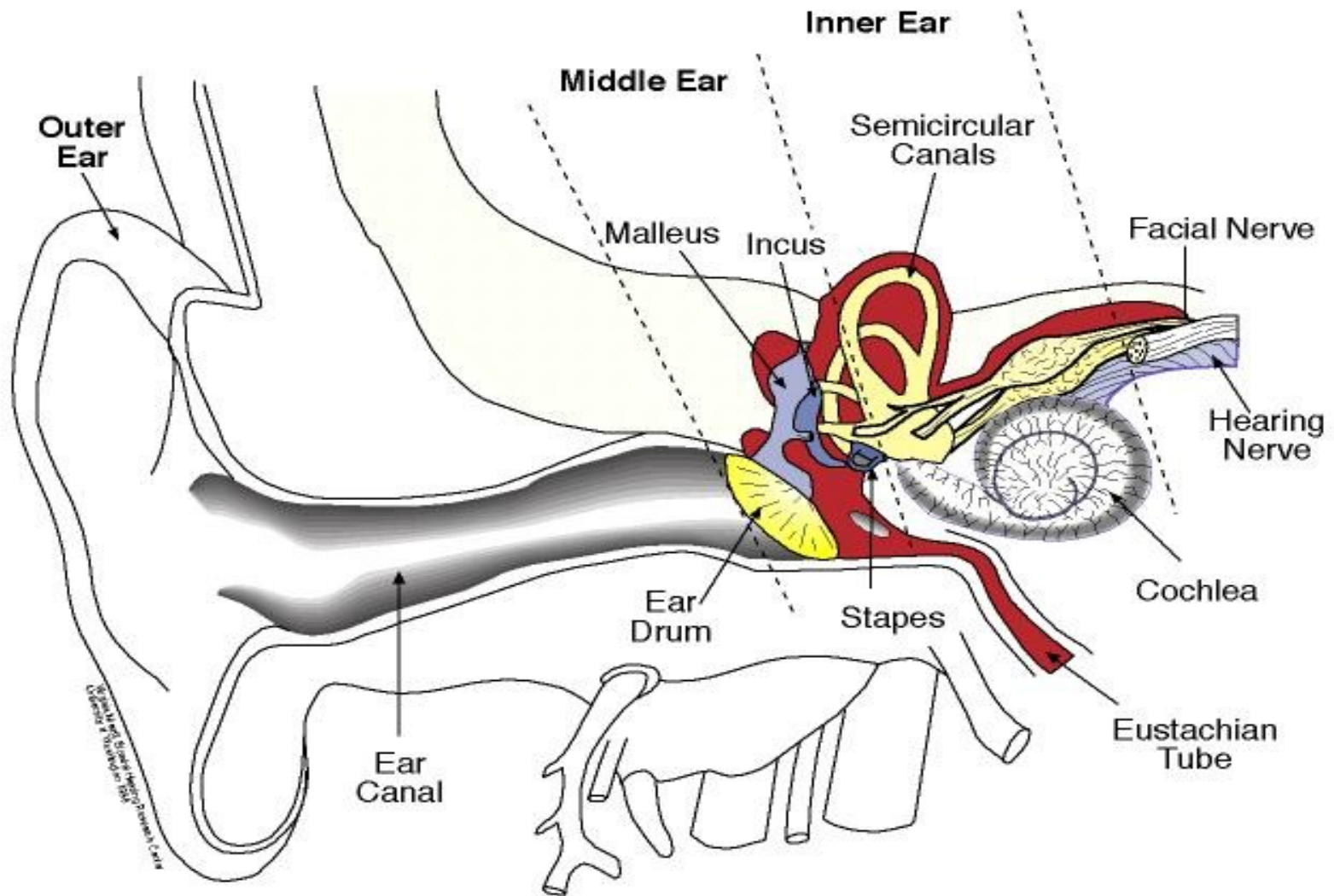


$$e(n) \blacklozenge E(e^{j\omega})$$

$$h(n) \blacklozenge H(e^{j\omega})$$

$$s(n) \blacklozenge S(e^{j\omega})$$

# Speech Perception System



# Speech Perception Mechanism

- Mainly three regions - outer ear, middle ear & inner ear
- Outer ear - directs speech pressure variations towards the middle ear
- Middle ear - transforms pressure variations into mechanical motion
- Inner ear - converts mechanical vibrations into electrical firings in the auditory neurons, which leads to brain
- Language decoding and message understanding at the higher centers of learning which is less understood

# Steps in Speech Reception and Message Comprehension

- Acoustic pressure variations funnelled into middle ear by outer ear.
- Eardrum converts acoustic pressure variations to mechanical vibrations.
- Mechanical vibrations are transferred to inner ear by middle ear bones.
- Standing wave patterns are generated in inner ear liquid.
- Standing waves are converted into neural firings on auditory nerve.
- Neural firings are decoded and message comprehension is done in brain.

# Digital Model for Speech Perception

