PUE 4113 Speech Processing.

Prof. Ciira Maina
ciira.maina@dkut.ac.ke

29th May, 2023

Course Content

1. Speech production and perception
2. Speech signal analysis
3. Feature extraction
4. Modeling speech
5. Speech coding
6. Speech systems: Speech recognition, speaker recognition
7. Machine learning for speech processing
    - Gaussian Mixture Models
    - Hidden Markov Models
    - Neural Networks - Deep neural networks, generative models, sequence-to-sequence models

Course website:
`www.ciirawamaina.com/speech-processing.html`

Today's Lecture

1. Introduction to speech processing
2. Speech production

Speech Systems

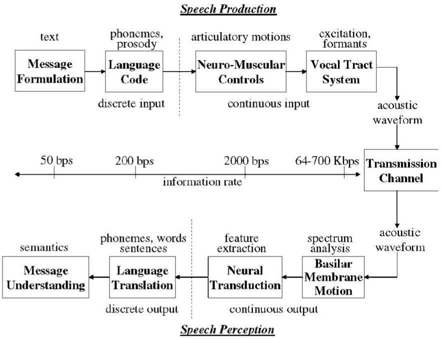
- Speech technology is now ubiquitous
- Human machine interaction using voice is becoming common place

Speech Systems

- Speech is the primary communication medium for human beings
- The speech signal conveys a lot of information
  - What was said: speech recognition
  - Who said it: speaker recognition
  - Speaker's emotional state: Emotion recognition
  - Speaker's age, gender, ...
- Other applications include
  - Text to speech systems
  - Speech coding

The speech chain

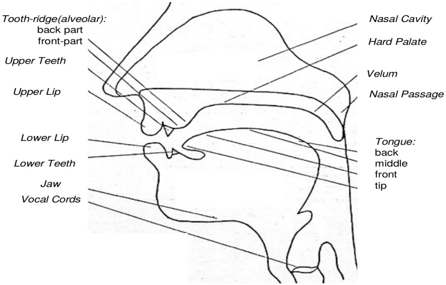
- Speech begins as a thought in the speakers mind
- A corresponding speech signal is generated
- The speech is perceived and interpreted by the listener



Source: Rabiner, L. R., & Schafer, R. W. (2007). Introduction to digital speech processing. Foundations and Trends® in Signal Processing, 1(1–2), 1-194.

Speech production

- The speech production apparatus



Source: Huang, X., Acero, A., Hon, H. W., & Reddy, R. (2001). Spoken language processing: A guide to theory, algorithm, and system development (Vol. 1). Upper Saddle River: Prentice hall PTR.

Speech production

- Speech consists of sound waves emanating from the mouth and nostrils of a speaker
- Sound waves are longitudinal pressure waves consisting of compressions and rarefractions of air molecules.
- Two major sound classes exist
    - Consonants - produced in presence of constrictions in the throat or obstructions in the mouth
    - Vowels - Produced without major constrictions or obstructions
- Major parts involved in speech production: Lungs, vocal cords, soft palate (velum), hard palate, tongue, teeth, lips

Voiced and unvoiced sounds

- Voiced sounds are created when the vocal folds vibrate
- Otherwise the sound is unvoiced
- Vocal cords vibrate at frequencies ranging from about 60Hz to 300Hz
- The rate of opening and closing of the vocal folds in the larynx during production of voiced sounds is the fundamental frequency (F0)
- The fundamental frequency contributes to the perception of pitch

Formants

- The vocal tract can be modeled as a tube that is closed at the vocal cords and open at the lips
- Resonances within this tube occur at a given set of frequencies corresponding to nodes at the open end and antinodes at the closed end
- The tube is excited by the periodic glottal wave produced by the vibration of the vocal cords.

Formants

- Harmonics of this wave that occur at the tube resonance frequencies are emphasized.
- This will be explored further when we consider the source-filter model of speech
- When the shape of the vocal tract changes, the resonances change
- The resonances of the oral cavities for a particular articulator configuration are called formants

Readings

- HAH - Chapter 1-2
- RS - Chapter 1-3