Speech Perception and Phonetics

What is a sound?

Sound is a longitudinal pressure wave formed by compression and rarefaction of air molecules.

What is an audio signal

Is a representation of a sound in audio frequency range.

$$\forall t \in \mathbb{R}, x(t) = \dots$$

The phases of speech production

1. Communication intent
2. Message formulation
3. Language system
4. neuromuscular maping
5. vocal tract
6. speech signal

Which are the systems involved in speech production

1. Respiratory system: provides the energy for speech
2. Phonatory system: produces the sound
3. Articulatory system: shapes the sound

7 types of phonation

1. Modal: normal voice
2. Vocal fry: creaky voice
3. Creaky voice: low pitch
4. Breathy voice: air escapes
5. Harsh voice: rough voice
6. Whispery voice: no vocal fold vibration
7. Falsetto: high pitch

What is vowel sound

A sound produced with an open vocal tract. Vibration of the vocal cords with no obstruction of the vocal tract.

Single tube model of the vocal tract

The vocal tract is modeled as a single tube with a sound source at one end and the lips at the other end.

Muliple tube model of the vocal tract

The vocal tract is modeled as a series of tubes with a sound source at one end and the lips at the other end. The tubes are connected by a series of constriction points.

What is vowel triangle

A triangle that represents the position of the tongue in the mouth when producing a vowel sound. The corners of the triangle represent the three main vowel sounds: /i/, /a/, /u/.

Consonant sounds

A sound produced with a constriction in the vocal tract. There are two types of consonants: voiced and voiceless.

The speech perception process

1. Speech signal
2. speech analysis
3. feature extraction
4. words, prosody, and phonemes
5. semantics
6. perception

What is a decibel

A unit of measurement for the intensity of a sound. It is a logarithmic scale that measures the ratio of a sound to a reference sound.

What is the maximum frequency of human hearing

The maximum frequency of human hearing is 20,000 Hz.

What is the minimum frequency of human hearing

The minimum frequency of human hearing is 20 Hz.

What is the mel scale

The mel scale is a perceptual scale of pitches that is based on the human ear's perception of frequency.

What is phonetics

Phonetics is the study of the sounds of human speech.

Difference between phone and phoneme

A phone is a speech sound that can be distinguished from other sounds in a language, while a phoneme is a speech sound that can change the meaning of a word.

What is IPA

The International Phonetic Alphabet (IPA) is a system of phonetic notation that represents the sounds of spoken language.

Translate "hello" in IPA

/həˈloʊ/

What is the difference between a vowel and a consonant

A vowel is a speech sound that is produced with an open vocal tract, while a consonant is a speech sound that is produced with a constriction in the vocal tract.

What is the difference between a voiced and voiceless sound

A voiced sound is produced with vibration of the vocal cords, while a voiceless sound is produced without vibration of the vocal cords.

[What are some IPA subsets]

- ARPAbet
- SAMPA
- X-SAMPA

Classification of Speech Sounds

- Voicing:
  - Voiced
  - Unvoiced
- Manner of Articulation:
  - Stops
  - Fricatives
  - Affricates
  - Nasals
  - Liquids
  - Glides
- Place of Articulation:
  - Bilabial
  - Labiodental
  - Dental

    - Alveolar
    - Palatal
    - Velar
    - Glottal

What is prosody

Prosody is the rhythm, stress, and intonation of speech.

List those

- Average f0 male voice: 130 Hz
- Average f0 female voice: 220 Hz
- Normal speech average range: 1 octave
- Singing average range: 2 octaves
- Voiced unvoiced transition: 100ms-200ms

What is pitch countour and where is used?

Pitch contour is the pattern of pitch changes in speech. It is used in speech analysis to identify prosodic features. Like, in question intonation, the pitch rises at the end of the sentence.

Prosodic boundaries

Separate units of speech, like words, phrases, and sentences. They are marked by changes in pitch, loudness, and duration.