

Speech Signal Characteristics

Arto Teräs (ajt@iki.fi)

February 15, 2002

Based on chapter 2 of the book
Handbook for Neural Networks for Speech Processing
(edited by Shigeru Katagiri, Artech House 2000)

Contents

- Speech signal information contents
- Physical processes in speech production
- Mathematical models
- Synthetic speech production and recognition, recent trends

Types of Information in Speech

- Information embedded in speech can be divided in three categories: linguistic, paralinguistic and nonlinguistic information
1. Linguistic Information
 - “Symbolic information that is represented by a set of discrete symbols and rules for their combination”
 - The meaning of words, dividing words to syllables and syllables to phonemes, composing sentences
 - Word accent: “‘permit’ vs. ‘per’mit”
 - Controlled by the speaker

Types of Information in Speech (continued)

2. Paralinguistic Information
 - “Information that is not inferable from a written counterpart but is deliberately added by the speaker to modify or supplement linguistic information”
 - Changing a declarative sentence to a question
 - Emphasizing words
 - Mostly discrete — information local to one sentence, but also continuous elements (like adding a tone of being suspicious)
 - Controlled by the speaker

Types of Information in Speech (continued)

3. Nonlinguistic Information

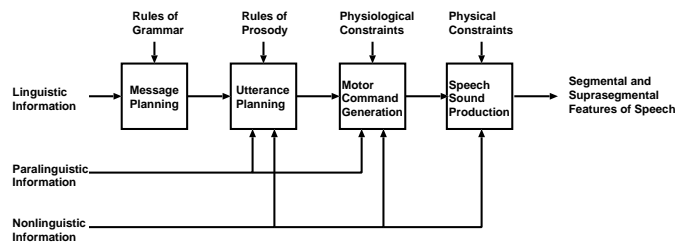
- Speaker properties: age, gender, individual morphological characteristics, condition of health
 - Identifies the speaker
 - Emotional factors: anger, sadness, delight
 - Not controlled by the speaker
- ⇒ Data must be collected outside experimental environments!

Segmental and suprasegmental features

- Segmental features are characteristics local to a discrete segment of speech: selecting the word “cake”, the pronunciation of the vowel “a” in this word
- Suprasegmental, or prosodic, features are acoustic characteristics distributed over several segments: the base frequency of the speaker voice, influence of emotional state
- Unit length matters:
 - Composing word “permit” of two syllables (segmental, base unit words)
 - Changing accent between the syllables (suprasegmental, base unit syllables)

Speech Production Process

- Hierarchical model (in reality the stages overlap in time):



Physical Production of Sounds

- Speech organs convert motor commands to sounds
- *Source*: Expiratory airflow makes vocal folds vibrate and produces a fundamental frequency F0
- *Filter*: Moving the jaws and tongue resizes the hollow spaces and modifies the sound
- Consonants produced by complex movements and blocking the airflow at different places

Mapping language to the physical process

Example sentence: “Jukka teki taian”

- Movements in base cases can be photographed
- Contact and air pressure towards edges when producing consonants can be measured

Problems:

- Adjacent phones affect each other’s pronunciation
- Articulation speed and many other situation-dependent factors matter

Mathematical models

- The base frequency produced by the airflow and vocal folds can be modeled as a filter system
- Recently, three-dimensional dynamic models have been created to model the articulators (tongue, lips, jaws)
- The tongue plays the most important role (highest number of parameters)
- Vertical movements of the larynx are crucial in distinguishing voiced and voiceless consonants and also in controlling the base frequency
- Studying the central nervous system may help understand the whole process of controlling articulatory movements

Speech synthesis and recognition

- Traditionally, efforts have concentrated around finding efficient parameters for modeling speech waveforms and correlating acoustical features and linguistic information
- Paralinguistic and nonlinguistic features have not been considered adequately
- Proper processing of suprasegmental features like emotional conditions could help in extracting the linguistic information in speech recognition

Home Exercise

Say aloud the sentence “Kissa käveli katolla”. Which articulatory postures do you find (see figures 2.12 and 2.14)? Consider especially the pronunciation of the “l” consonant. What differences in posture and sound do you find while pronouncing the “l” in the words “käveli” and “katolla”?

Compare these with other words which contain single or double “l” consonants adjacent to various consonants and vowels. Describe briefly your observations.

If you don’t speak Finnish, you can alternatively compare the “l” sounds in the phrase “Please give me the pillow” and other English words.