Spoken Language Structure

Notes on Spoken Language Processing

Chia-Ping Chen

Department of Computer Science and Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan ROC

Hierarchy

- sounds (speech production and perception)
- phones (phonetics and phonology)
- syllables
- words (morphology)
- phrases
- sentences (syntax, semantics)
- discourse

Sounds

- Physically, a sound is a longitudinal pressure wave.
- The amplitude of a sound is measured on a logarithmic scale called decibels (dB).
- The *sound pressure level (SPL)* measures the absolute sound pressure P in dB,

$$SPL = 20 \log_{10} \frac{P}{P_0} \text{ dB},$$

where P_0 corresponds to the threshold of hearing for a 1kHz tone. Note that the threshold of hearing is a function of frequency.

Articulators

- lungs: source of air
- vocal cords: control of voiced/unvoiced
- velum (soft palate): valve to nasal cavity
- hard palate: roof inside mouth
- tongue
- teeth
- lips

Voicing

- The vocal cords vibrate during the production of voiced sounds. They are open and relaxed for voiceless sounds.
- The vibrating frequency of vocal cords during a voiced sound is called the *fundamental frequency*.

Formants and Spectrograms

- The vibration of vocal cords excites the vocal tract, resulting in resonances.
- The resonant frequencies corresponds to *formants*.
- The formants can be read from a *spectrogram*, which shows short-time distribution of energy in different frequency bins.
- The dark strips in spectrograms show the formants, which are indicative of the sounds produced.

Speech Perception

- physical vs. perceptual attributes
  - intensity vs. loudness
  - fundamental frequency vs. pitch
  - spectral shape vs. timbre
  - phase difference vs. location
- The perception may well depend on more than one physical factor.

Frequency Analysis

- The cochlea (a part of human ear) acts like a set of overlapped filters.
- The bandwidth of these filters are the critical bands determined by psycho-acoustic experiments.
- It is found that the perceptual resolution is finer in the lower frequencies.
- The *Bark-scale* and *mel-scale* frequencies all incorporate such characteristics.

Phon*

- *Phonetics* refers to the study of speech sounds and their production, classification and transcription.
- *Phonology* is the study of the distributions and patterning of speech sounds and the rules governing pronunciation.
- A *phoneme* is an abstract label for word pronunciation while a *phone* is an acoustic realization.

Allophones

- The realizations of a phoneme (phone) differs from context to context. They are called allophones.
- Allophones are categorical, limited to a small number of cases.
- For example,
  - aspirated vs. unaspirated /p/
  - light vs. dark /l/
  - longer /ae/ before voiced consonant

Coarticulation

- There are some non-categorical variations in speech in addition to allophones.
- A common source of variation is coarticulation, meaning a phone changes due to neighbor phones.

Did you hit it to Tom?

  - palatalization (*did you*)
  - reduction of unstressed sound (*you*)
  - flapping (*hit it*)
  - reduction of schwa (*to*)
  - reduction of geminate (double consonant) (*it to*)

Syllables

- Syllables are cohesive units formed by phonemes.
- A syllable is generally centered around a vowel. It can be decomposed into *onset*, *nucleus*, and *coda*. Nucleus and code combine to be *rime*.
- The word *strengths* has a very lengthy syllable.
  - onset: /str/
  - nucleus: /eh/
  - coda: /nx th s/

Words

- A larger unit than syllable is *word*.
- Every word in a sentence has a *part of speech* (POS) tag, such as noun, verb, adjective, etc.
- *Morphology* is about word formation, including inflection, derivation, conpounds, and so on.
  - In English, prefixes and suffixes are often used to deal with morphology.
- Word class is not the same as part of speech tag. It can be automatically derived from data and is often used in class-based language model.

Syntax

- A sentence is a string of words. However, not all strings of words are sentences. A legitimate (grammatical) sentence needs to comply to certain rules.
- Syntactics is the study of rules that regulate the relationship between words in a sentence.
- A grammatical sentence can be represented by a *parse tree*.

Semantics

- Semantics is the *meaning* of sentences.
- Semantics is deeper than syntax.
  - syntactic roles, such as subject, verb, object specify the direction of action.
  - Semantic roles, such as agent, theme, instrument, result, specify *who did what to whom*.
- Meaning can be represented by FOPC (a logical form).

Lexical Semantics

- Words, by themselves, cannot be judged to be true or false, literal or metaphorical.
- However, it is obvious that the meaning of a sentence is dependent on the *senses* of component words.
- Lexical semantics is the study of word senses. In particular, certain relations among word senses, such as contrasts, similars, class inclusion, part-whole, case relations, and so on.