

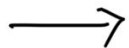
Text to Speech

M.Tech. Artificial Intelligence, Second Year, NMIMS

By,

Bilal Hungund, Data Scientist, Halliburton

Raw or
Tagged
Text



Text
Analysis

- Document Structure
- Text Normalization
- Linguistic Analysis



Phonetic
Analysis

- Homograph disambiguation
- Grapheme-Phoneme Conversion



Prosodic
Analysis

- Pitch
- Duration



Speech
Analysis

- Voice Rendering

SYNTHESIS

→ Text Analysis:

→ Text normalization: Converting tokens to a uniform sequence.
Such as upper case → lower case, languages, tokenization

→ Document Structure: This involves deciding order and grouping of sentences.

→ Linguistic Analysis: Dealing with different languages with its grammar.

Homograph Disambiguation

→ Represent statistical procedure for lexical ambiguity resolution in TTS synthesis.

→ Types of homographs, different part of speech, same part of speech, proper names, roman numerals, fractions/dates, years/quantifiers, abbreviations

→ It is a classification problem in which target is pronunciation of different words

→ The goal of the problem is to identify patterns to correctly classify ambiguous words

Pronunciation	Context
(1) led	.. it monitors the <i>lead</i> levels in drinking
(1) led	... median blood <i>lead</i> concentration was
(1) led	.. found layers of <i>lead</i> telluride inside ..
(1) led	... conference on <i>lead</i> poisoning in ...
(1) led	.. strontium and <i>lead</i> isotope zonation ..
(2) lid	maintained their <i>lead</i> Thursday over ...
(2) lid	.. to Boston and <i>lead</i> singer for Purple
(2) lid	Bush a 17-point <i>lead</i> in Texas, only 3
(2) lid	his double-digit <i>lead</i> nationwide. The
(2) lid	the fairly short <i>lead</i> time allowed on ..

Pronunciation Modelling

- It is a baseline linear sequence of basic units (e.g. phones) describing a pronunciation of the lexical keys
- The pronunciation lexicons contains
 - orthographic word (keyword)
 - optional grammatical information
 - pronunciation information
- Widely used for speech synthesis

→ Phonetic Analysis

→ Phonetic is the study of the physical properties of speech sounds in three aspects, production, transmission and reception.

→ Three major branches of phonetics,

Articulatory

Articulation involving different organs

Sounds of p, t, n, e...
pit, pan, pin

Acoustic

Amplitude, frequency and formants

Auditory

Way of sounds perceived

Pronunciation Modelling includes,

→ Prosodic Analysis

- It is a study of the tune and rhythm of speech
- Sequence representation of the phones
- Characterised by, vocal pitch (fundamental frequency), acoustic intensity and rhythm (phoneme and syllable duration)

Features

Prosodic: pitch, formants, energy, jitter, shimmer

Spectral characteristics: spectral flux, slope, centroid, entropy, roll-off, flatness

Voice Quality: zero crossing, harmonic-to-noise ratio, noise-to-harmonic ratio

Phonation: vocal pitch, pitch period entropy

→ Phonemic Analysis

→ The way we understand those sounds

→ Phonemic analysis is a crucial part of phonetic analysis which aims at investigating the entire sound system of a language

→ Identifying the contrastive units of speech sounds and their systematic variants in a language

Challenges in Phonemic Analysis

→ Consists discrete units

→ Languages have unique phonemic systems

→ Sounds are affected by the environments in which they occur

→ Speech sounds have two levels of representations

Words
↳ Phonemes

⇒ Tokenization of all possible phonemes

Generating sequence from tokenization

Embeddings of sequences

Padding of sequences

Sequence Modelling

Unsupervised Learning on embeddings

Words to Phoneme Prediction

References

- http://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/linguistics/02.introduction_to_phonetics_and_phonology/02._main_areas_of_phonetics_/et/7652_et_et_02.pdf
- T. Svendsen, "Pronunciation modeling for speech technology," 2004 International Conference on Signal Processing and Communications, 2004. SPCOM '04., 2004, pp. 11-16, doi: 10.1109/SPCOM.2004.1458347.
- https://repository.upenn.edu/cgi/viewcontent.cgi?article=1056&context=ircs_reports
- <https://shahabks.github.io/myprosody/>
- [Homograph Disambiguation in Text-to-speech Synthesis](#)
- http://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/linguistics/02.introduction_to_phonetics_and_phonology/17._phonemic_analysis/et/7660_et_et_17.pdf