December 4, 2003

Pronunciation dictionaries & TTS

Chapter 4.6–4.8

Overview

- TTS overview
- Pronunciation dictionaries
- FST-based pronunciation lexicon
- Prosody
- Presentations
- Next time: Review, evaluations

Text-To-Speech

- Map orthography to phonetic transcription
- Add in prosody
- Map phonetic transcription + prosody to acoustic signal

Pronunciation dictionaries

- List words and their pronunciations
- No morphological or phonological rules
- PRONLEX: 90,694 wordforms
- CMUdict: 100,000 wordforms
- CELEX: 160,595 wordforms
- Designed for ASR, but can be adapted for speech synthesis
- In what way do the requirements on dictionaries differ between these two applications?
- What problems might arise for this approach?

Problems for simple listing

- Highly variable pronunciations (*and*, *I*, *the*, *of* etc.)
- Names:
  - 21% of 33 million words of AP newswire were names (Liberman & Church 1992).
  - Includes not only people's names but also company names and product names.
  - ... named entity recognition
- Morphological productivity
- Number names, with different possible pronunciations:
  - Serial, combined, paired, hundreds, trailing unit, (trailing unit with a decimal)

FST-based approach

- Components:
  - large morpheme pronunciation dictionary, encoded as an FST
  - FSAs for morphology
  - FSTs for morphphonology (like spelling change rules)
  - heuristics and LTS rules/transducers for names and acronyms
  - default LTS rules/transducers for other unknown words
  - (Named-entity recognizer)

Architecture

- Lexical, intermediate and surface levels all contain two tapes, one for pronunciation and one for orthography.
- Lexicon-FST: composed of two-level lexicon plus FSAs/FSTs for morphology (+PL| ϵ:s|z) [4.21–23]
- FST_1 ... FST_n: orthographic and phonological rules, run in parallel

Architecture

Lexical:	f o:aa x:ks +N +PL
	LEXICON-FST
Intermediate:	f o:aa x:ks ^ s:z
	FST_1 ... FST_n
Surface:	f o:aa x:ks e:ix s:z

- Map from lexical entry (plus inflection, etc) to a pronunciation
- Map from surface orthography to surface pronunciation via lex entry

Names

- Donnelly marketing organization: 1.5 million name "tokens" (for 72 million US households)
- Liberman & Church (1992) attempt to handle most frequent 250,000 (1/6) of these
  - Dictionary of 50,000 names covers 59%
  - Stress-neutral suffixes (*-s*, *-son*, *-ville*): 84%
  - Name-name compounds and rhyming heuristics: 89%
  - Prefixes, stress-changing suffixes and suffix-exchanges: ??
  - LTS rules for the remainder.

Prosody

- Prominence: stress (lexical and sentential)
- Structure: intonational phrases/units, intermediate phrases
- Tune: F0 pattern, component parts include pitch accent

English pitch accents (Pierrehumbert 1980)

- **H***: high (on a stressed syllable)
- **L***: low (on a stressed syllable)
- **L*+H**: rise, starting on a stressed syllable
- **L+H***: rise, ending on a stressed syllable
- **H+L***: fall, ending on a stressed syllable
- (**H*+L**: apparently not needed)

Other components of the English system

- Phrase accents:
  - **L-**
  - **H-**
- Boundary tones:
  - **L%**
  - **H%**

Text-To-Speech

- Map orthography to phonetic transcription
- Add in prosody
- Map phonetic transcription + prosody to acoustic signal

Overview

- TTS overview
- Pronunciation dictionaries
- FST-based pronunciation lexicon
- Prosody
- Presentations
- Next time: Review, evaluations