Word Pronunciation

Julia Hirschberg

CS 4706
Today

• Motivation
• Challenges for automatic word pronunciation
• Standard methods
• Innovative solutions
• TTS demos:
  – ScanSoft/Nuance
  – AT&T
  – IBM
  – Cepstral
• SNL Robot Repair
Motivation

• Intelligibility
• Naturalness
• Applications to language learning
  – Unlimited vocabulary
  – Type a word or phrase and hear it spoken in your target language
    • To imitate
    • To learn to recognize
• Speech therapy
Word Pronunciation

• What determines how a word is pronounced?
  – History/Language Origin/Dictionaries:
    • shoe (ME shoo), phoenix (Gr)
    • mole, attaches, resume
  – Part-of-speech:
    • use, close, dove, multiply, coax
  – Morphology:
    • ferryboat, ferryboats
    • Popemobile (pope+mobile)
Letter-to-Sound Rules

• Define correspondences between orthography and phonemic representation, e.g.
  – i _{C}e$ → /ai/  rise
  – Else i → /ih/  rip

• Deals with any input
Problems

• Must be built by hand
• Many exceptions, e.g.
  • i $_{C}e$ → /ai/ matches ripen/risen/riser/river/ripper
  • Proper names: Nice, Ramirez, Ribeiro, Rise, Infiniti
• Symbols and abbreviations: &c, evalu8, cu, tsp
• Assigning lexical stress
• Solutions
  – More complex rules
  – Exceptions dictionary
    • Consulted first
    • But how do we handle morphological variation? E.g.
      – Rise’s hat
Dictionary-based Approaches

- Rely on very large dictionary with orthography and pronunciation for each word
- Typically created by hand or by expansion of online pronouncing dictionary
Problems

• Redundancy of representation
  – Cat, cats, cat’s, cats’
• Out-of-vocabulary (OOV) items
  – Proper names: covering all U.K. surnames would require >5,000,000 entries
  – New words: …
    • Technical terms: liposuction, anova, bernaise
    • Foreign borrowings: frappe, ciao, louche
• Solutions
  – Larger dictionary
  – Morphological preprocessing before dictionary look-up
  – Fall back to L2Sound rules if no dictionary ‘hit’
Major Challenges for TTS

• Disambiguating homographs
  – bass/bass

• Pronouncing new words
  – New names in the news:
  – New words: iPad, Kindle

• Expanding abbreviations and acronyms correctly
Homograph Disambiguation by Decision List Classifiers (Yarowsky ‘97)

- E.g., bass/bass, nice/Nice, live/live, desert/desert, lead/lead

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

- Rank by

$$\text{Abs}(\text{Log} \left( \frac{P(\text{Sense} | f_i=v_j)}{P(\text{Sense}_2 | f_i=v_j)} \right))$$
Pronouncing OOV Words

• Techniques for handling OOVs
  – Inferring country of origin:
    • Takashita, Leroy, Kirov, Lima, Infiniti
  – Pronunciation by analogy
    • Analog/dialog
    • Risible/visible
    • Proper names: Alifano/Califano
Bootstrapping Phonetic Lexicons (Maskey et al ’04)

• For some languages, online pronouncing lexicons exist – but for others….e.g. Nepali
  – How to minimize effort in creating lexicons?

• Approach
  – Given a native speaker and a large amount of online text in the language…
    • Native speaker builds small lexicon by hand for seed set of N most common words in text, e.g.
      – is: /izh/
      – the: /dhax/
• Derive L2S rules from lexicon automatically, e.g.
  – is \(\rightarrow\) ih{zh}
  – the \(\rightarrow\) {dh}ax …

• Loop: Choose the next N most common set of words from the text and use the lexicon + L2S rules to predict pronunciations, e.g.
  – telephone \(\rightarrow\) /telaxfown/
  – He \(\rightarrow\) /hax/?
  – Rise \(\rightarrow\) /rihzhax/?

• Assign a confidence score to each prediction by comparing each word to all words in lexicon
  – If is \(\rightarrow\) /ihzh} in lexicon and no other orthographically similar words are pronounced differently, new rule his \(\rightarrow\) /hihzh/ scores high
• For low confidence pronunciations, Active Learning step:
  – Inspect and calculate error rate
  – Hand correct errors and add all to lexicon
  – Iterate from Loop until performance stabilizes
    • Build a new set of L2S rules from augmented lexicon

• Results
  – English:
    • 94% success on test set after 23 iterations, 16K entry lexicon
    • Performance comparable to CMUDict and 1/7 the size
  – German:
    • 90% accuracy after 13 iterations, 28K lexicon
  – Nepali
    • 94.6% accuracy after 16 iterations, 5K lexicon
Improving Pronunciation Dictionary Coverage (Fackrell and Skut ’04)

• Idea: Many proper names have more than one spelling (e.g. More/Moore; Smith/Smythe)
  – Homophones
  – Find a ‘fuzzy’ mapping between OOV (Out of Vocabulary) words and words already in the lexicon
  – Identify spelling alternations that are ‘pronunciation-neutral’ in an existing lexicon to produce rewrite rules for OOVs
• Pros?
• Cons?
Deriving Pronunciations from the Web (Ghoshal et al ’09)

• Extract candidate orthography/pronunciation pairs (ad-hoc and IPA)
  – E.g. bruschetta (pronounced broo-SKET-uh)
• Validate the candidates: how likely are these pairs to represent a word and its pronunciation
• Normalize ad-hoc and IPA pronunciations
• Pros?
• Cons?
Pronunciation Evaluation

• How would you evaluate the pronunciation module of a TTS system?
Next Class

• Readings
• Download the ToBI cardinal examples (see http://www1.cs.columbia.edu/~agus/tobi/)
  – You will first need to download WaveSurfer
    • http://www.speech.kth.se/wavesurfer/
  – Then download the cardinal examples
• Listen to each of the cardinal examples
  – Try to imitate each one and to decide what it ‘means’