Pronunciation Tuning for Voice Search

Bradford Starkie

http://www.inferencecommunications.com
What is Pronunciation Tuning?
  – Why it’s Hard
Existing Methods for Generating Pronunciations for Unknown Words
Developments at Inference Communications
  – Automated Tools for Pronunciation Tuning
What is Pronunciation Tuning?

• Speech Recognizers need to know how to pronounce words to recognize them.

• Pronunciation tuning is “the process of creating pronunciations for new words or removing unused pronunciations to improve speech recognition accuracy.”
Why Is Pronunciation Tuning Important to Voice-Search?

• The more obscure a word is
  – The more useful it becomes as a key word for searching
  – The more likely it is that a custom pronunciation is required for it

• Examples
  – Surnames and place names in directory applications.
English Pronunciation Is Difficult

• Spelling of words based upon linguistic origin rather than a simple set of rules
  – thought, enough, through, though, cough, bough, byte, bite, bight, gnaw, know.

• Foreign words spelt like foreign words.

• Pronunciation depends upon the context of the word
  – Read Vs Read, Sewer Vs Sewer

• Accents, Abbreviations, Regional differences
Generating Pronunciations for Unknown Words

• Commercial recognisers shipped with letter to sound rule pronunciation systems for unknown words.
• Best systems 75% accurate (for a single utterance)
• Speech Recognition Systems can have multiple pronunciations
  – Solution : Generate more pronunciations than are used
  – For small vocabularies works well but for large vocabularies can lead to confusion between phrases
Speaker Enrolment

- Many speech recognition systems ship with user enrolment capability.
- Speaker can add a word to a dynamic grammar by speaking.
  - The user speaks the phrase several times
  - phoneme string stored in database.
- Best suited to small vocab systems as per phoneme accuracy low for unconstrained phoneme sequences (enrolment).
An Alternative Method (Hybrid Rule/Enrolment Method)

• Collect Speech Samples of the new words
  – More than one utterance per new word.
• Generate a set of alternative pronunciations using letter to sound rules
• Use the speech recogniser to select the best matching one
• Remove pronunciations that cause errors (where possible)
Benefits of Hybrid System

• Faster, Less effort than manual process.
• Don’t need skilled labour to generate speech samples
  – need native speakers.
• Measures speech recognition accuracy as you develop.
• Pronunciation can be tuned using live samples from platform.
Research into Hybrid Systems at Inference Communications

- Inference has been experimenting with Hybrid Pronunciation Systems
- Vendor Independent
- Using VXML standards
  - Pronunciation Lexicon Specification (PLS)
  - Speech Assessment Methods Phonetic Alphabet (SAMPA)
  - Speech Recognition Grammar Specification (SRGS)
Automated Pronunciation Pruning System

• Analyses confusion matrices to prune out pronunciations that cause recognition errors.
• Used on commercially deployed applications
• Performance is comparable to manual tuning of pronunciation lexicons.
• Inference Communications have been developing automated pronunciation tuning tools.
  – Faster & cheaper than manual process.
  – Measures speech recognition accuracy as you develop.
  – Pronunciation can be tuned using live samples from platform.

• Inference has developed an automated pronunciation pruning tool which in commercial use.