

# Multilingual TTS

SyRG Discussion 2/26/04

Brian Langner

# Issues with Multilingual TTS

- Quality relative to monolingual TTS
- Processing text in multiple languages
- Handling phonetic/phonemic differences between languages
- Obtaining sufficient coverage

# “Multilingual TTS”

- What is meant by it?
  - Synthesizer that can read text from multiple languages
    - paragraph in German, then paragraph in English, etc.
  - Synthesizer that can read mixed-language text
    - German sentence containing English (and French, and ...) words/phrases

# Synthesis Quality

- Unit selection voices sound “pretty good”
  - Sets bar pretty high, better quality voices are expected now
- Difficulties with other issues tend to degrade quality of multilingual voices

# Processing Mixed-Language Text

- Lots of text that we might want to synthesize contains foreign words
  - Either “loan” words or absorbed words
- Why do we care?
  - Prosody and pronunciation of foreign words is different than for native words
  - Loan words vs. absorbed words are also different
  - Quality depends on doing this right

# Text Processing

- Can't have a lexicon with just full word forms
  - Arbitrary combinations of stems, affixes, etc. make size intractable
  - Must analyze morphology
- Since foreign words often are very different syntactically from native words, need to analyze syntax as well
- Also have problems with homographs
  - Word is present in multiple languages, which one is “right” and should be pronounced?

# Method

- Build lexicon + grammar for each language
- Build inclusion grammar
  - Combines the individual grammars
  - Provides rules describing how to map foreign constituents into the native equivalent
    - Since features between languages do not always overlap, also provide a feature mapping so that unification is possible
  - Apply penalty for using foreign rules
    - Solves problem of homographs (with proper penalty weights)

# How to handle phonetic/phonemic differences?

- Full database for each language
  - Benefit: native database
  - Drawback: lots of recording, voice talent that can speak all languages required
  - If the voice talent is not fluent in all the languages, the resulting voice will be accented
    - not clear if this is good or bad
- Single database with IPA coverage
  - Benefit: one database, covers “all” sounds
  - Drawback: less useful for non-european languages, phoneme->sound mapping not necessarily consistent across languages



# Phonetic/phonemic differences (con't)

- Fake it
  - Use a single language's database, with mappings to foreign phones
  - Benefit: fast and easy, requires only native database
  - Drawback: resulting voice is accented (at best) if languages are similar, “weird” if not
- Differences in dictionaries (both inter- and intra- language) only compound this problem

# Phonetic/phonemic differences (con't)

- Apply voice conversion techniques
  - Still requires at least one bilingual database
  - Not for new languages, but useful for crosslingual voice modifications