

# Teaching computers to teach people to read and speak

updates: <http://tinyurl.com/osl08>

(Stanford Open Source Lab '08)

see also: <http://talknicer.com/d>

(online demo)

James Salsman

[jim@talknicer.com](mailto:jim@talknicer.com)

# speech recognition for pronunciation evaluation can help most learners acquire language faster

- typically three to five times more useful per time spent practicing than self study with recordings
- details: Jack Mostow's Project LISTEN at CMU
- commercial example: Rosetta Stone's English study packs retail for ~\$300 up from \$30
- billions of people want to learn more language

# Julius open source speech recognition

- from Cambridge Hidden Markov Model Toolkit
- free as in speech and beer
- running on XO
- C, flat files, a few sh scripts
- several megabyte memory footprint for triphones
- expect under 3 MB footprint for diphones (to do!)
- feasible on low-end cell phone equipment

# microphone upload

- Adobe Flash 10 using open Speex vocodec is the best solution for two years now
- W3C rejected Device Upload as “device dependent” in 1999
- Mozilla and Google Chrome have made promises several months ago, but nothing yet

# phoneme alignment and pronunciation scoring

- acoustic scores: fit to models from 5000 speakers
- durations: cadence
- pitch: important for tonal languages, but not English except for punctuation-like information
- amplitude: less important for stress and punctuation, very important for weighting parts of speech when converting word to phrase scores
- can adapt to accent and dialect by comparing phoneme scores to set of exemplar pronunciation to derive word and phrase scores

# agreement with human pronunciation judges

- 65-70% is really easy: about 5-10 recorded exemplars of each phrase from diverse speakers speaking with ordinary pronunciation
- 80% takes 20+ exemplar pronunciations
- 85%+ is impossible even for humans

# patent encumbrance

- “Speech Training Aid” by R. Series *et al* (1991) at U.K. Defence Research Agency, sold to private QnetiQ, then 20/20 Speech, then Aurix, then NXT plc., maker of high-fidelity stereo equipment
- doesn’t cover reading tutoring which is in many cases exactly the same task, algorithms, and completely indistinguishable in all other details
- can be licensed, but it has been very difficult
- patent holders more interested in suing abundant infringers than licensing

# crowdsourced accuracy review systems

- voxforge.org and librivox.org collect exemplars
- vetting exemplar pronunciations can be done with
  - volunteers, including learners and anonymous
  - paid workers, including mostly poor and non-native speakers from e.g. Mechanical Turk or Craigslist
- Wikimedia Strategic Proposal (accuracy review)



# Questions and Answers

Thank you!

[jim@talknicer.com](mailto:jim@talknicer.com)

<http://talknicer.com> [ /d ]

these slides:

<http://talknicer.com/olpcsf.ppt>