Outline of Report to the Berkeley Zoom Phonologists Group of August 30, 2021,

describing a pure telephony pronunciation intelligibility remediation system

by James Salsman, jim@phoneclearly.com

Background describing the previous mobile app/web system: arxiv.org/abs/1709.01713
Learner experience: call (+1) 351-253-2759 **[note: this number has degraded performance as we are waiting to upgrade from narrowband to wideband audio input from learners]**; get asked to say prompts; try to say them; maybe hear a score; if score is low, then hear a native speaker exemplar pronunciation remixed to amplify and lengthen the portion of the prompt which would improve the score most if improved, and try again, up to three times; repeat.

Application level logs describe the server-side activity.

Diphone learner analytics versus phonemes or words: numeric and neurophysiological motivations.

Vocal tract articulation derived from n-best speech recognition results of each phoneme in prompts, not words. Eleven features per phoneme: duration, acoustic score from speech recognition system, substitutions, insertions, deletions, place, closedness, roundedness, voicing, nasality, and the proportion of physiologically adjacent phonemes that appeared to be more likely than the expected phoneme.

Data collection: transcriptions, accuracy, source bias, coverage, cross-validation issues. Transcription collection can be done with SMS/MMS but Amazon Mechanical Turk is easier to handle transcriptionist remuneration.

> From O'Brien, *et al.* (2019) "Directions for the future of technology in pronunciation research and teaching," *Journal of Second Language Pronunciation* **4**(2):182-207: jbe-platform.com/content/journals/10.1075/jslp.17001.obr
> page 186: "pronunciation researchers are primarily interested in improving L2 learners' intelligibility and comprehensibility, but they have not yet collected sufficient amounts of representative and reliable data (speech recordings with corresponding annotations and judgments) indicating which errors affect these speech dimensions and which do not."
> Please see also page 192: "Collecting data through crowdsourcing...."

Coverage includes most Indo-European languages capable of fitting into the CMUBET/ARPABET phoneme set, although adding a tonal channel for Mandarin, Cantonese, Vietnamese, etc., would involve weeks to months of additional work.

Diphthongs in CMUBET/ARPABET:

AW -> AA UH

AY -> AA IY
ER -> UH R (substitutes /ʊ/ for /ɜ/)
EY -> EH IY (substitutes /ɛ/ for /e/)
OW -> AO UH (substitutes /ɔ/ for /o/)
OY -> AO IY

Twilio bandwidth issues: telephone number location and calling telephone network can both cause bandwidth limitation and concordant voice quality degradation issues. Twilio doesn't support high bandwidth telephone connection requests yet, but their support has been somewhat responsive concerning a recording timeout bug.

Internationalization: beyond pronouncing prompts in the learners' first languages, involving recording ideally or speech synthesis, transcriptions should be collected for the target learner population. Otherwise, accuracy can suffer substantially.

Homographs: speech synthesis issues (e.g. "live") are addressed by first language internationalization.

Code size: PocketSphinx is hundreds of thousands of lines of C, the feature extraction code is 800, and the Python code for initializing the database and running the telephony app is about 1,700 lines in total. The Python/SqlAlchemy Dataset package is used to access a Postgres database, currently about 6 MB in text dump format.

I still need to tune many parameters: some, like automatic gain control adjustments, don't require collecting large amounts of additional data; others do.

Goals:

Self-sufficiency: e.g. learners may someday send sliding-scale regionalized payment amounts via links sent to them over SMS to pay for hosting and data collection fees.

Integration: Open source? Partnership with academia, in particular phonologists who might benefit? Partnership with private equity? Crowdfunding? Partnerships with existing pronunciation assessment software companies, such as Duolingo, Pearson, Rosetta Stone, Babbel; at least a dozen others? With companies prospecting the opportunity such as Achievable.me.

Interested? Call me at (+1) 650-427-9625, sign up at bit.ly/slig and/or email jim@phoneclearly.com

This work has been generously supported by Justin Pincar and Tyler York of Achievable.me and Kevin Lenzo of Duolingo.

This document is at bit.ly/zoomphon