An approach to measure pronunciation similarity in second language learning using radial basis function kernel

Christos Koniaris
University of Gothenburg, Sweden
Problem Formulation

• Perceptual diagnostic evaluation of non-native speech vs. spectral-based similarity measure

• Quantitatively measure the degree of difference in pronunciation of phonemes by a group of non-native speakers as compared to a group of native speakers

• Examine the radial basis function kernel or RBF kernel as an alternative similarity measure to Euclidean distance
Approach

• Block diagram of the method

• Compute native perceptual assessment degree (nPAD)

\[ \Theta_\ell = \frac{A_\ell}{A_n} \]

Native speech dataset \rightarrow HMM-aligner \rightarrow Non-Native speech dataset

Auditory model

\[ u_n(\cdot) \sim e^{\gamma \| y_n - \hat{y}_n \|^2} \]

Power spectrum

\[ \phi_\ell(\cdot) \sim e^{\gamma \| x_\ell - \hat{x}_\ell \|^2} \]

\[ \phi_n(\cdot) \sim e^{\gamma \| x_n - \hat{x}_n \|^2} \]

Auditory distortion detectability

Signal distortion measure

Perceptually relevant dissimilarity measures

\[ A_\ell \sim \sum \sum \left[ u_n(\cdot) - \phi_\ell(\cdot) \right]^2 \]

\[ A_n \sim \sum \sum \left[ u_n(\cdot) - \phi_n(\cdot) \right]^2 \]
Application

- **Spectral model**: frequency domain psychoacoustic model

- **Linguistic study**: a survey on identifying common problems for speakers of a certain L1 background

- **Data**: repeating text after a natively speaking virtual language tutor (two sessions). Recordings from:
  - 37 non-native speakers, 11 L1 backgrounds
  - 11 native speakers (Swedish)
## Examples of the results (vowels)

| L1 bkgr. | Type | nPADA Problematic vowels [ordered] |
|----------|------|-----------------------------------|
| German   | $\Theta_{eucl}$ | æː, eː, yː, uː, øː, æː, øː, iː, ø, ïː |
|          | $\Theta_{bf250}$ | æː, eː, y, ø, æ, u, æː, i, y, a, ø, ï |
|          | $\Theta_{bf500}$ | æː, eː, y, ø, æ, u, æː, a, i, y, ø, ï |
|          | $\Theta_{bf1000}$ | æː, ø, y, e, æ, æː, a, u, i, æː, y, ø |
| Chinese  | $\Theta_{eucl}$ | ò, æː, e, y, u, æː, ø, øː, iː, æː, e, ø, ø, ï |
|          | $\Theta_{bf250}$ | ò, æː, ï, e, i, æː, æː, iː, æ, æ, ï, ø, ï, æ, ï, e, ø, ø, æ |
|          | $\Theta_{bf500}$ | ò, æː, ï, e, i, æː, ø, æ, ï, i, ï, ï, æ, ï, æ, æ, ï, æ, æ |
|          | $\Theta_{bf1000}$ | ï, æ, æ, e, ï, ï, æ, e, ï, æ, æ, ï, æ, æ, ï, æ, æ |
Examples of the results (consonants)

| L1 bkgr. | Type | nPAD Problematic consonants [ordered] |
|----------|------|--------------------------------------|
| German   | $\Theta_{\text{eucl}}$ | $\tilde{f}, \eta, \nu, \eta, m, b, r, d, l, k, \tilde{s}, t, p, h, f, e, s$ |
|          | $\Theta_{\text{bf}250}$ | $\eta, e, s, r, l, \tilde{f}, g, \eta, d, k, t, b, h, f, v, n, p$ |
|          | $\Theta_{\text{bf}500}$ | $\eta, e, s, r, l, \tilde{f}, g, \eta, d, k, t, b, h, f, v, n, p$ |
|          | $\Theta_{\text{bf}1000}$ | $e, s, s, r, l, g, \eta, k, d, t, h, b, f, v, n, \tilde{f}, j$ |
| Chinese  | $\Theta_{\text{eucl}}$ | $\tilde{f}, \eta, \nu, m, n, b, r, l, d, k, t, f, g, t, p, j, h, s$ |
|          | $\Theta_{\text{bf}250}$ | $\tilde{f}, l, r, \eta, j, g, f, k, b, y, m, n, t, t, p, d, h, s$ |
|          | $\Theta_{\text{bf}500}$ | $l, r, \tilde{f}, \eta, j, g, k, f, b, y, m, n, t, t, p, d, h, s$ |
|          | $\Theta_{\text{bf}1000}$ | $l, r, \eta, j, g, k, \tilde{f}, b, f, m, v, n, t, t, p, d, h, s$ |
Conclusions

• Method to automatically, quantitatively evaluate non-native speakers’ pronunciation

• Compare similarities between power spectrum domain and auditory perception domain

• nPAD to quantify non-native similarities in comparison to native variations

• Compare RBF kernel with Euclidean distance