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

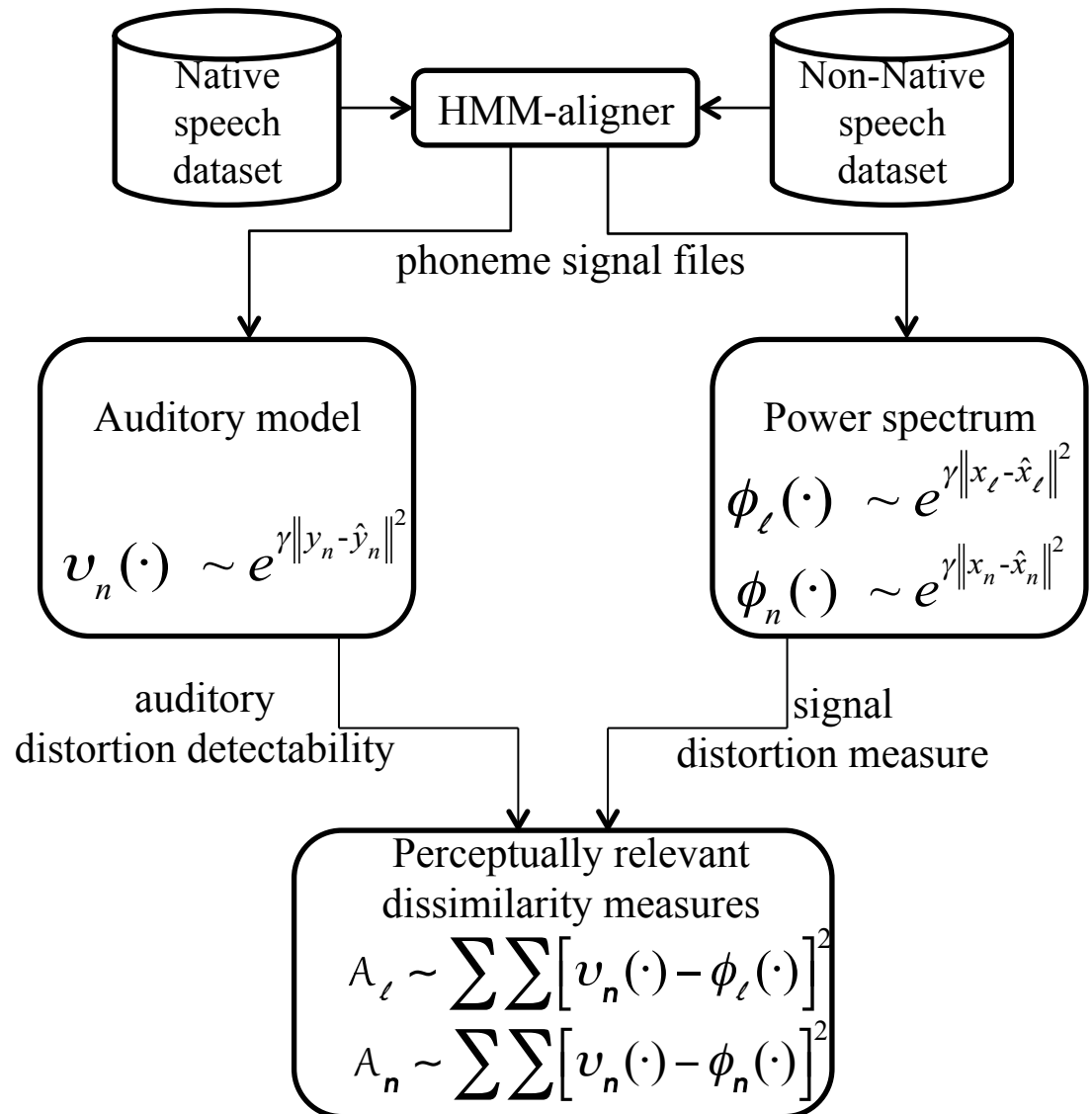Christos Koniaris

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 → HMM-aligner ← Non-Native speech dataset

phoneme signal files

Auditory model

$$\upsilon_n(\cdot) \sim e^{\gamma\|y_n\text{-}\hat{y}_n\|^2}$$

Power spectrum

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

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

auditory distortion detectability

signal distortion measure

Perceptually relevant dissimilarity measures

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

$$A_n \sim \sum\sum\left[\upsilon_n(\cdot) - \phi_n(\cdot)\right]^2$$

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	nPAD Problematic vowels [*ordered*]
German	Θ_ℓ^{eucl}	æː, ɛ, yː, uː, ʊ, ɛː, ø, œː, øː, iː, ə, ɑː
	Θ_ℓ^{rbf250}	æː, eː, Y, øː, ɑː, ʉː, ɛː, i, yː, a, ə, iː
	Θ_ℓ^{rbf500}	æː, eː, Y, øː, ɑː, ɛː, ʉː, a, i, yː, ə, iː
	$\Theta_\ell^{rbf1000}$	æː, øː, Y, eː, ɑː, ɛː, a, ʉː, i, œː, yː, ə
Chinese	Θ_ℓ^{eucl}	ɵ, æː, ɛ, yː, ʉː, ɛː, ø, øː, iː, ɑː, eː, e, ɔ, ə, a, ʉː, i, oː
	Θ_ℓ^{rbf250}	ɵ, ɑː, ʉː, eː, i, ɛː, øː, æː, iː, ɔ, a, yː, oː, ə, ʉː, e, ʊ, œː
	Θ_ℓ^{rbf500}	ɵ, ɑː, æː, ʉː, ɛː, eː, øː, i, a, ɔ, iː, yː, oː, œː, e, ə, ʉː, ɛ
	$\Theta_\ell^{rbf1000}$	ɑː, æː, ɛː, øː, ʉː, eː, i, a, œː, ɔ, e, iː, ɛ, oː, yː, ø, ə, ʉː

Examples of the results (consonants)

L1 bkgr.	Type	nPAD Problematic consonants [*ordered*]
German	$\Theta\ell^{eucl}$	ɦ, ŋ, ʋ, n, m, b, r, d, l, k, ʂ, t, p, h, f, ɕ, s
	$\Theta\ell^{rbf250}$	ʂ, ɕ, s, r, l, ɦ, g, ŋ, d, k, t, b, h, f, ʋ, n, p
	$\Theta\ell^{rbf500}$	ʂ, ɕ, s, r, l, ɦ, g, ŋ, d, k, t, b, h, f, ʋ, n, p
	$\Theta\ell^{rbf1000}$	ɕ, ʂ, s, r, l, g, ŋ, k, d, t, h, b, f, ʋ, n, ɦ, j
Chinese	$\Theta\ell^{eucl}$	ɦ, ŋ, ʋ, m, n, b, r, l, d, k, t, f, g, ʈ, p, j, h, s
	$\Theta\ell^{rbf250}$	ɦ, l, r, ŋ, j, g, f, k, b, ʋ, m, n, t, ʈ, p, d, h, s
	$\Theta\ell^{rbf500}$	l, r, ɦ, ŋ, j, g, k, f, b, ʋ, m, n, t, ʈ, p, d, h, s
	$\Theta\ell^{rbf1000}$	l, r, ŋ, j, g, k, ɦ, b, f, m, ʋ, n, t, ʈ, 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