Modeling age of exposure in L2 learning of vowel categories

Meghan Clayards¹, Joe Toscano²
¹Centre for Research on Language Mind and Brain, McGill University, ²Department of Psychology, University of Iowa

Background
- Age of exposure is a predictor of second language proficiency
- Native-like phonological proficiency is attained only by learners exposed at the earliest ages (simultaneous bilinguals)

Does early commitment to phonological patterns account for age of exposure effects?
- Two computational models were trained using unsupervised learning (Toscano & McMurray, 2008; McMurray et al., 2009)
  - 2D Gaussian Mixture Model
  - Hebbian Normalized Recurrence Network

Exposure to L2 decreases with later exposure to L2

Language data

Spanish monolingual

Quichua monolingual

Early bilingual

Mid bilingual

Late bilingual

Results

Models

2D Gaussian Mixture Model (GMM)
- Categories defined by 2D (F1 x F2) Gaussian distributions
- Each Gaussian defined by six parameters:
  - \( \phi \): likelihood of category
  - \( \mu_1 \) and \( \mu_2 \): mean along each cue dimension
  - \( \sigma_1 \) and \( \sigma_2 \): standard deviation along each dimension
  - \( \rho \): correlation between F1 and F2

- Each model is initialized with 200 Gaussians with equal \( \phi \), \( \sigma \), and \( \rho \) and random \( \mu \)-values

- During training, model is presented with F1 and F2 values
- Parameters updated via maximum likelihood estimation and winner-take-all competition

- Trained on 80,000 training trials per run (30 runs total)

Hebbian Normalized Recurrence Network (HNRN)
- Two-layer neural network
- 200 topographically organized input units with Gaussian tuning curves for each dimension (F1 and F2)
- 20 output units corresponding to possible vowel categories
- On each training trial:
  - Pair of F1 and F2 values are presented at input layer
  - Activation feeds forward to output units
  - Output units complete
  - Activation feeds back to input units
  - Weights are updated via Hebbian learning
- Process repeats until output unit activation settles
- Trained on 40,000 training trials per run (30 runs total)

Simulation groups
- Monolingual Quichua and Spanish groups
- Simultaneous bilingual: trained on tokens from both languages
- Early bilingual: started training with Quichua and switched to input from both languages after 25% of total training time
- Mid bilingual: switched after 50% of total training
- Late bilingual: switched after 75% of total training

Hebbian Normalized Recurrence Network (HNRN)
- Weighted connections between neuronal populations
- Activation feeds forward to output units
- Output units compete
- Activation feeds back to input units
- Weights are updated via Hebbian learning

3210

Analyses

- Total number of categories
- Mean F1 and F2 values of model categories
- Number of front vs. back vowel discriminations

- GMM: components with \( \phi > \) mean
- HNRN: output units with weights > mean (averaged across F1 and F2 weights)

- Category means
- GMM: \( \mu_1 \) and \( \mu_2 \)
- HNRN: input unit with greatest weight

Conclusions
- Both models learn the vowel systems of Quichua and Spanish when trained on monolingual data
- When exposed to bilingual data, both models distinguish different numbers of categories depending on when L2 (Spanish) is introduced, paralleling data from bilingual speakers with different ages of acquisition
- Age of exposure effect in the models arises as a consequence of learning without any changes in model plasticity (learning rates)
- Language users may be restricted in learning a second language not because of a critical period, but by the commitments that the system has already made to the first language; a separate mechanism is not necessary

Future directions
- GMM categories are sometimes unstable and with too much training data, the model may end with fewer categories than in the language (unlike human learners); further exploring the parameter space of the model may lead to a more stable set of parameters
- HNRN can be given additional types of input, such as context information or labels for different lexical items; could allow the network to maintain distinctions between more categories
- Look at whether it is more or less difficult to learn an L2 with more categories or to learn one with fewer categories (Spanish->Quichua)

References