Learning acoustic features for English stops with graph-based dimensionality reduction

Patrick F. Reidy, a Mary E. Beckman, b Jan Edwards, c Benjamin Munson, d Allison A. Johnson c

a Caller Center for Communication Disorders, University of Texas at Dallas  b Dept. Linguistics, The Ohio State University  c Dept. Hearing and Speech Sciences, University of Maryland  d Dept. Speech-Language-Hearing Sciences, University of Minnesota

Overview

The computation of spectral features that cue segmental contrasts is a process of dimensionality reduction. Traditional approaches accomplish this reduction by mapping a high-dimensional observation (e.g., a spectrum) to a small number of pre-determined features (e.g., spectral moments; Forrest et al., 1988). Such approaches fail to exploit the distributional structure of the observations in the high-dimensional space and typically ignore superposing relationships among the observations, such as the word in which the segment occurs.

This study adapts the Laplacian Eigenmaps algorithm (Belkin & Niyogi, 2003; Bengio et al., 2003) to learn acoustic features for /t/ versus /k/, consonants that contrast in terms of spectral shape and that differentially exhibit vowel-contextual variation in their spectral shape (see Fig. 1). The algorithm constructs a graph that simultaneously represents the high-dimensional structure of excitation patterns computed from a talker’s productions, and aligns lexical correspondences between talkers. A function that embeds the excitation patterns into a two-dimensional feature-space is learned by computing the eigenvectors of the constructed graph.

Speech Production Data

- 21 adults (10 women, 11 men) completed a picture-prompted word repetition task.
  - Two lists of words were used to elicit a variety of target consonants. Each list contained 32 words in which a target /t/ or /k/ occurred word-initially before a vowel (see footnotes at the bottom for the stop-initial words in the two lists).
  - Participants A50–A65 completed Lists A and B; participants A66–A70, only List A.
  - Training set: List A productions by participants A50–A65 (N = 403).
  - Test sets: List B from A50–A65 (N = 447); List A from A66–A70 (N = 156).
- Multitaper spectra were estimated from 25-ms windows around stop bursts, and then passed through an auditory (gammatone) filter bank, yielding excitation patterns.

Laplacian Eigenmaps Algorithm

1. Let X = {x1, ..., xn} be the training set of 493 excitation patterns (361-dimensional vectors). Each xi is pre-processed to sum to 1, so that it may be treated as a probability mass function.
2. Define a similarity function f on X in terms of Kullback-Leibler divergence DKL on X (see Fig. 2).
3. Define a function f1 on X that induces a weighted graph. Nodes correspond to observations in X (see diagram below, where (x1, x2, x3) versus (x4, x5) represent productions by different talkers). Edges connect nodes corresponding either to productions by the same talker (solid lines) or to productions of the same target word by different talkers (dashed lines). Edge weights encode similarity between excitation patterns. Parameter μ ∈ (0, 1) adjusts the balance between preserving the structure of each talker’s production-space and aligning multiple talkers’ production-spaces.
4. Define a function f1 on X that induces a weighted graph. Nodes correspond to observations in X (see diagram below, where (x1, x2, x3) versus (x4, x5) represent productions by different talkers). Edges connect nodes corresponding either to productions by the same talker (solid lines) or to productions of the same target word by different talkers (dashed lines). Edge weights encode similarity between excitation patterns. Parameter μ ∈ (0, 1) adjusts the balance between preserving the structure of each talker’s production-space and aligning multiple talkers’ production-spaces.
5. Solve the generalized eigenvalue problem γ1 - y1 = Lγ1, y1. The eigenvectors γ1, γ2, γ3, ..., γn that correspond to the two least, non-zero eigenvalues embed X into 2-dimensional space: x = (γ1(x), γ2(x)).
6. Extend eigenvectors γ1, γ2 to eigendensities γ1, γ2. The projection γ1(x) of a test point x ≠ X onto dimension 1 of the embedding is a linear combination of the components of γ1.

Eigenfunction Embedding of Test Set I (A50–A65, List B)

Figure 3. The image of the test data under the Laplacian eigenfunctions, γ1 and γ2. The ellipses denote 95% confidence regions (bivariate γ1). The dashed and dotted ellipses were estimated from the training data. The solid ellipses were estimated from test productions of the 4 words that were not represented in the training data.

Eigenfunction Embedding of Test Set II (A66–A70, List A)

Figure 4. The image of the test data under the Laplacian eigenfunctions, γ1 and γ2. The ellipses denote 95% confidence regions (bivariate γ1), estimated from the training data. Here, the embedding of the test data was guided solely by lexical information; comparison with Fig. 3 underscores the importance of lexical alignment across talkers.

Discussion and Future Directions

- The two-dimensional embedding that is learned by Laplacian Eigenmaps reflects well-established articulatory constriction features. γ1 distinguishes /t/ versus back-vowel /k/, reflecting place of constriction (anterior versus posterior); γ2 distinguishes /t/ versus front-vowel /k/, reflecting tongue shape (apical versus damped).
- We plan to extend this method to develop dynamic spectral features that model the transition from a stop burst to a vowel (see Nossar & Zahnorian, 1991).

Figure 2. The image of the training data under the embedding given by Laplacian eigenvectors, γ1 and γ2, learned with μ = 1/6. The ellipses denote 95% confidence regions, assuming a bivariate t-distribution.

(Note: the front-vowel /t/ data point that overlaps with the /N/ productions was determined post-hoc to be a misclassification.)

Figure 1: Excitation patterns computed from participant A54’s productions of /t/ versus /k/ before the vowels /v/ (left panel) and /o/ (right panel). The dotted lines indicate a subset of the values used to compute the Kullback-Leibler divergence.