Using Machine Learning to Identify Articulatory Gestures in Time Course Data
Will Styler 1, Jelena Krivokapić 2, Ben Parrell 3, Jiseung Kim 1
1University of Michigan Department of Linguistics - 2Haskins Laboratories - 3University of Wisconsin-Madison Department of Communication Sciences and Disorders

will@savethevowels.org - http://savethevowels.org/tbl

Objectives and Research Questions

Articulations during pauses have been studied (e.g. Gick et al. 2005, among others) and Katiska et al. 2014 noted ‘Pause Postures’, specific configurations of the articulators at strong prosodic boundaries. To study this, we first must know:

1. Are there measurable, reproducible patterns which identify pause postures in these data?
2. Can we empirically capture the gradience and uncertainty of these pause postures using machine learning?

Can these sorts of curvilinear patterns be (gradiently?) classified and studied using machine learning?

About the EMA Data

Data were collected using Electromagnetic Articulography (EMA), tracking positions of articulators in time and space.

1. Recording 12 sentences with present or absent prosodic boundary (e.g. “I don’t know about Mima. # Mini does, though.” or “There’s a lovely story I know about biBU. # Mini doesn’t like it though. ‘) with AG500 EMA System.
2. Two sensors (above and below the lips) used to generate lip aperture (LA) trajectory, reported here.
3. For all speakers, annotator marked span of each pause indicating presence (‘Yes’) or absence (‘No’) of PP

Total Data: 1891 trajectories total across seven speakers, with pause postures marked as present in 30% of trajectories.

About Support Vector Machines (SVMs)

A very common, very accurate machine learning algorithm

Model created from these three features

Randomly split the data into 80% for training, 20% for testing

Class weights were adjusted to compensate for PP rarity

Returns classifications, accuracy, and probability of PP for each token

Cohen’s Kappa calculated to measure human/computer agreement (controlling for agreement due to chance)

Results: Modeling the Data

Model A: “Ignore the Data”

Model guesses that all tokens are ‘not a pause posture’.

Model B: Raw Trajectory PCs Only

Only includes the six ‘Raw Trajectory’ PCA features.

Model C: All Twelve PC Features

Uses all twelve PCA features from Raw Trajectory and Trajectory Difference.

Are there measurable patterns associated with Pause Postures?

Model Results

• SVM reliably finds the same patterns as human annotator
• Accurate annotation is possible using non-aprioristic curve measurements
• EMA PCA analysis characterizes the curves with sufficient detail
• This mirroring of annotator judgements implies PP have a replicable pattern

Evaluating individual features

• Post-Hoc analysis to estimate feature ‘importance’ using RandomForest algorithm
• From Raw Trajectory PCA, PC2 and PC3 were best features
• From Trajectory Difference PCA, PC1 was best feature by far

Model created from these three features alone yields 92% accuracy with Kappa = 0.852

Specific types and timings of curvature characterize PPs!

Conclusions

1. Pause Postures are empirically findable in the data
2. We can automatically identify and characterize them on the basis of PCs on trajectories, implying that PPs have a unique and replicable pattern
3. We can capture the gradient nature of pause postures using probabilities derived from SVMs

Tokens can be classed as ‘stronger’ or ‘weaker’ postures mathematically

This method is effective for all types of curvilinear data

Classification based on PCA-derived features from curves is widely applicable to many problems in speech (e.g. formant tracking, ultrasound contours)

Acknowledgements and References

Supported by NSF Grant 1241695 to B.R. and NSF Grant 1313806 to J.K. Also thanks to Sunita Raman, Katsika et al., Krivokapić et al.

