Using Machine Learning to Identify Articulatory Gestures in Time Course Data

Will Styler[1], Jelena Krivokapić[1,2], Ben Parrell[3], Jiseung Kim[1]

[1]University of Michigan Department of Linguistics - [2]Haskins Laboratories - [3]University of Wisconsin-Madison Department of Communication Sciences and Disorders

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

Objectives and Research Questions

Articulations during pauses have been studied (c.f. Gick et al. 2005, among others) and Katsika 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

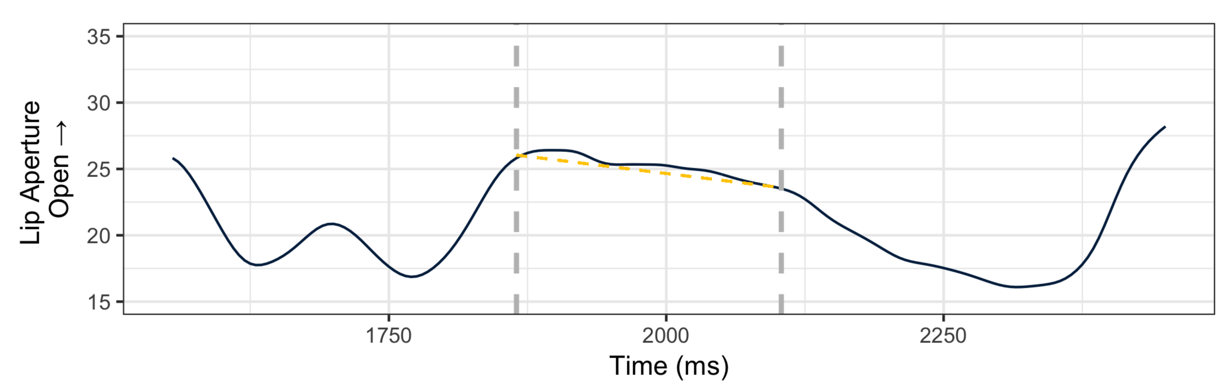
Data were collected using Electromagnetic Articulography (EMA), tracking positions of articulators in time and space.

1. Recording 12 sentences with present or absent prosodic boundaries (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
4. Annotator also provided gradient judgements ("Yes", "Maybe", "Unlikely", "No") for 3/7 speakers

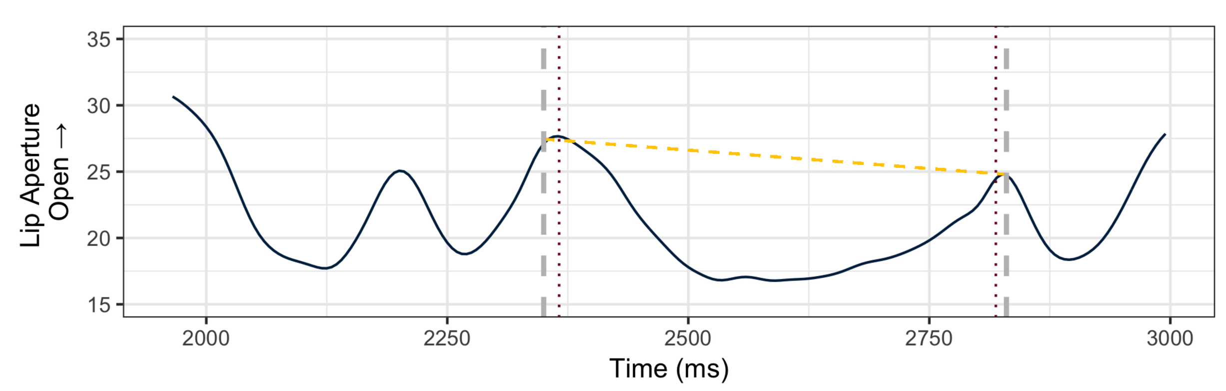
Total Data: **1891 trajectories total across seven speakers**, with pause postures marked as present in 30% of trajectories.

Pause Postures in the Data

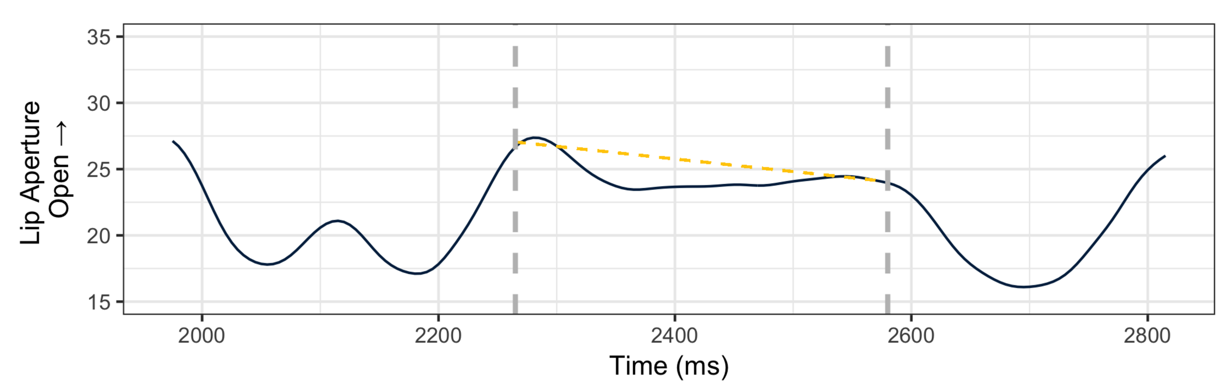
Direct Interpolation ('No' PP)



Pause Posture ('Yes' PP)



Possible Pause Posture ('Maybe' PP)



Characterizing the Curvature

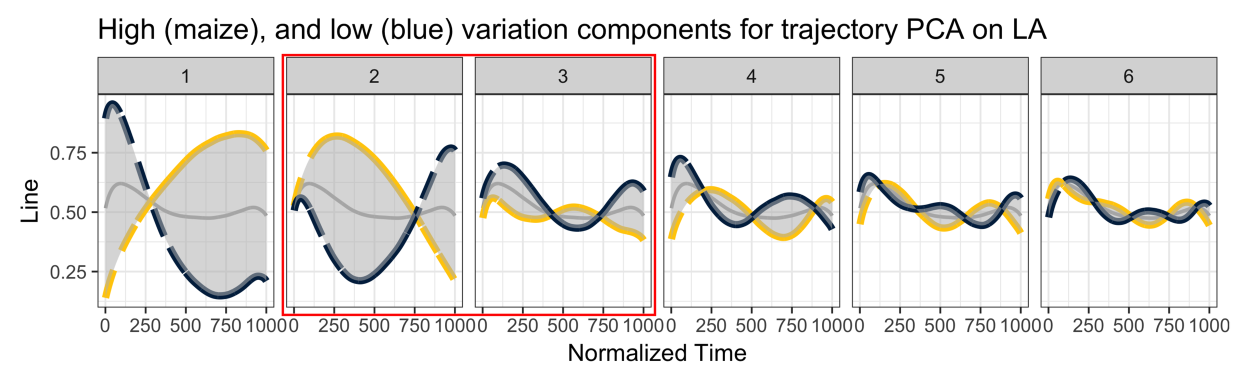
To analyze the data, we must find a non-aprioristic way to describe the curves and turn them into 'features' which can be used for machine learning (after Shaw and Kawahara 2017). This was done using **Functional Principal Component Analysis (fPCA)**.

- We extracted all pause trajectories as curves over time, and used them as input to an fPCA model
- This model extracts the dominant, orthogonal patterns ('components') of variation among curves (6 PCs per analysis).
- This fPCA was run on the data in two different formats, to capture two different aspects of pause posture trajectory

Raw Trajectory PCA

fPCA using as input the raw LA trajectory during pauses, to optimally capture changes in lip aperture.

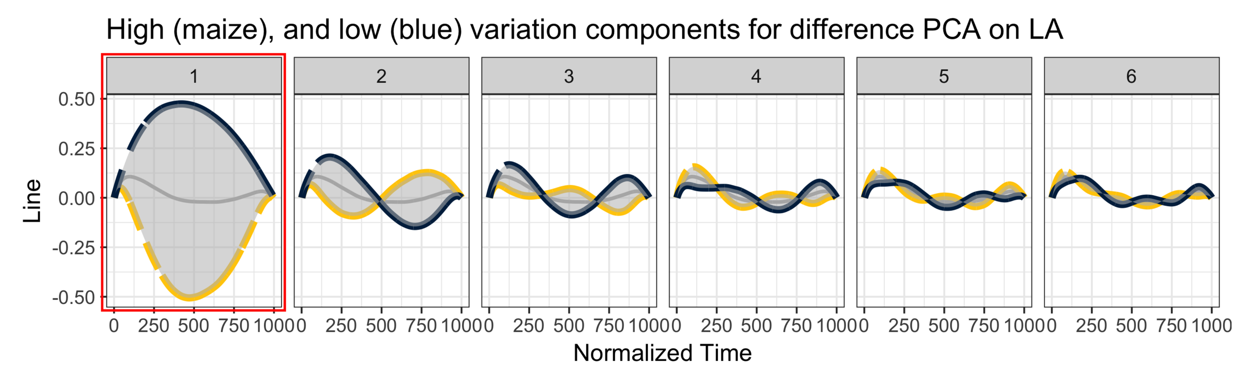
- Describes the changes in position over time



Trajectory Difference PCA

fPCA using the difference between direct interpolation and the LA trajectory during pauses.

- Captures movement towards a separate target



Machine Learning with Support Vector Machines

To evaluate the measurability and gradience in these data, we trained a supervised machine learning algorithm using the "Yes" or "No" Pause Posture judgements to define two 'classes' from the 12 PC scores for each curve.

About Support Vector Machines (SVMs)

- A very common, very accurate machine learning algorithm
- Examines the data in a multi-dimensional space (one dimension per feature)
- Finds a hyperplane which optimally separates the classes
- Classification consists of finding where each new token falls relative to this line
- Radial Kernel SVM used here to allow for non-linear classification

Training the SVMs

- Randomly split the data into 80% for training, 20% for testing
- SVM is trained using the 12 fPCA features along with 'Yes' vs. 'No' judgements
- 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'.

	'No'	'Yes'
Pred.'No'	268	107
Pred. 'Yes'	0	0
Pred. Accuracy	71.4%	
Cohen's Kappa	0	

Model B: Raw Trajectory PCs Only
Only includes the six 'Raw Trajectory' PCA Features.

	'No'	'Yes'
Pred.'No'	254	8
Pred. 'Yes'	14	99
Pred. Accuracy	94.1%	
Cohen's Kappa	0.85	

Model C: All Twelve PC Features
Uses all twelve PCA features from Raw Trajectory and Trajectory Difference.

	'No'	'Yes'
Pred.'No'	258	7
Pred. 'Yes'	10	100
Pred. Accuracy	95.4%	
Cohen's Kappa	0.89	

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
- fPCA analysis characterizes the curves with sufficient detail
- This mirroring of annotator judgements implies PP have a replicable pattern
- **Machine Learning can simulate annotator judgements!**

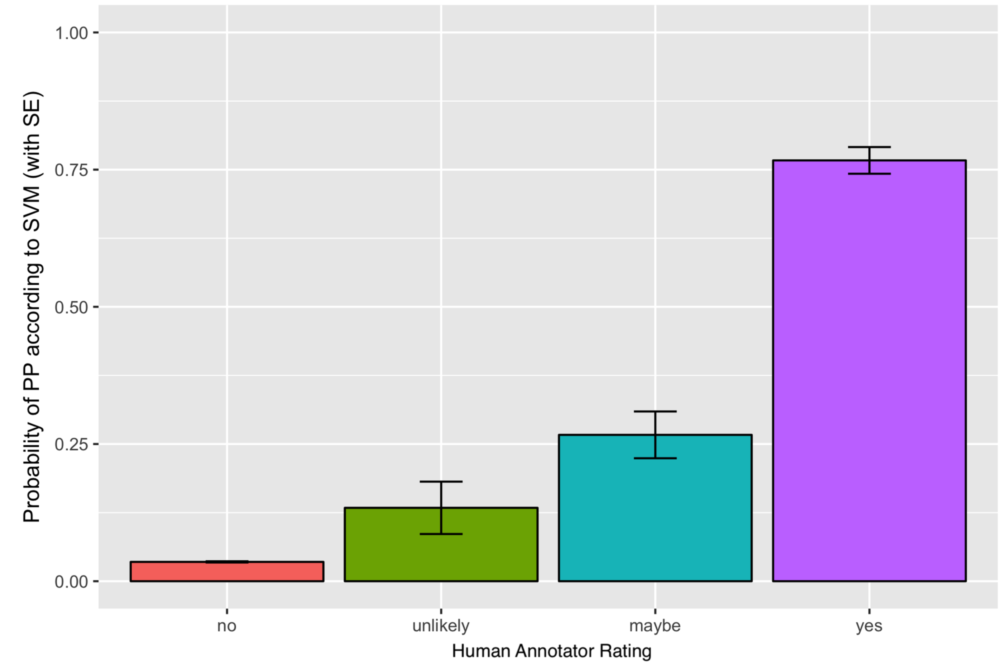
Evaluating individual features

- 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 = 0.832
- **Specific types and timings of curvature characterize PPs!**

Can we model gradiency in judgement?

Per-token probability judgements extracted from SVM (e.g. "How sure is the model that this is a PP?") then compared to annotator judgements of "Yes", "Maybe", "Unlikely", "No" pause posture likelihood for three speakers.

- If annotator judgements match SVM judgements, SVM can be used to model gradient likelihood of Pause Postures



Is automatic annotation good enough?

- Our best-performing model (C) finds PPs in novel data with 95.4% accuracy
- Out of 375 unknown items, it misclassified 17 tokens
  - It's slightly more prone to false positives
- Human-to-machine agreement of 0.891 is widely considered to be excellent
- Some evidence that variation between speakers in posture introduces additional error

Conclusions

1. **Pause Postures are empirically findable in the data**
   - We can automatically identify and characterize them on the basis of PCs on trajectories, implying that PPs have a unique and replicable pattern
2. **We can capture the gradient nature of pause postures using probabilities derived from SVMs**
   - Tokens can be classed as 'stronger' or 'weaker' postures mathematically
3. **This method is effective for all types of curvilinear data**
   - Classification based on fPCA-derived features from curves is widely applicable to many problems in speech (e.g. formant tracking, f0, ultrasound contours)

Acknowledgements and References

Work supported by NSF Grant BCS-1348150 to Beddor and Coetzee, NSF Grant 551513 to Krivokapić, NIH Grant DC003172 to Byrd, and DC002717 to Whalen. Thanks also to Anna Stone.

Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., and Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. Journal of Phonetics, 44:62-82.

Shaw, J., Kawahara, S. (2017) Modelling articulatory dynamics in the frequency domain. Poster at Workshop on Dynamic Modeling in Phonetics and Phonology 2017, Hosted by the Chicago Linguistic Society.

Gick, B., Wilson, I., Koch, K., and Cook, C. (2005). Language-specific articulatory settings: Evidence from inter-utterance rest position. Phonetica, 61(4):220-233.