Syllable Position Prominence in Unsupervised Neural Network Segment Categorization

Fengyue(Lisa) Zhao, Sam Tilsen
Department of Linguistics
Cornell University, NY, USA

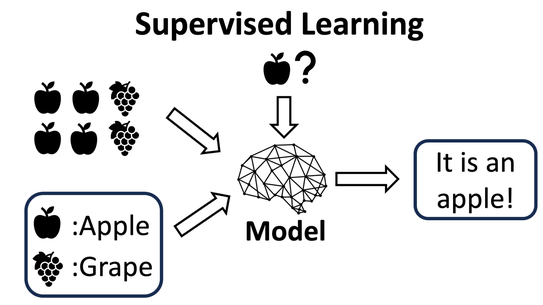
Cornell Phonetics Laboratory

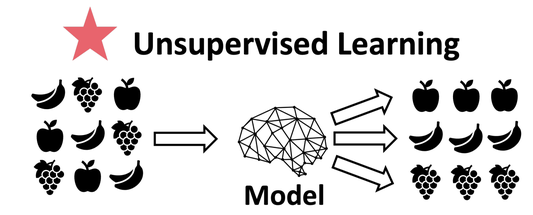
COGNITIVE SCIENCE @ CORNELL

1 Motivation:

- Many arguments for the cognitive reality of phoneme presuppose their existence. [1]
- Unsupervised learning may provide evidence for categories that avoids this problem. [2][3]
- English obstruents exhibit diverse phonetic realizations across syllable positions (e.g. /t/ and /p/ in *top* and *pot*). [4]
- Linguistically we assume that phone identity—(e.g. /p/ vs. /t/) is a strong predictor of representational similarity, while syllable position—e.g. onset vs. coda—is perhaps a secondary factor. But is this always the case?

This study: Unsupervised learning of English obstruents /t/ and /p/ in different syllable positions

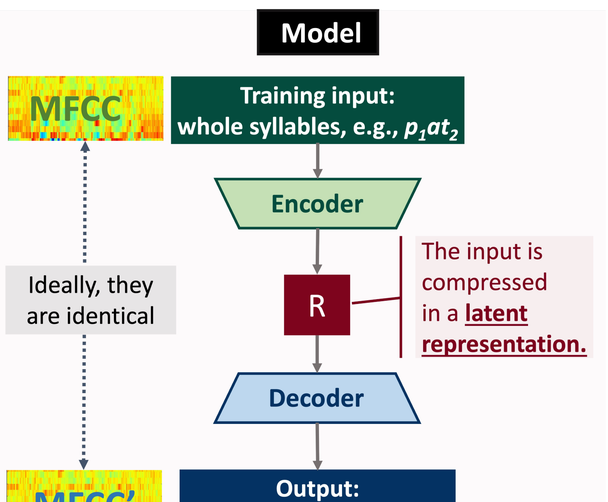


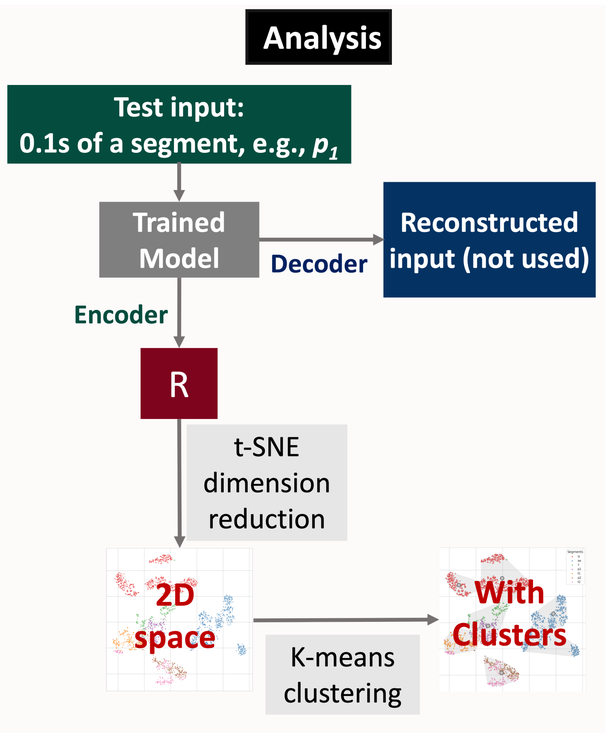


2 Methods

Experimental data:

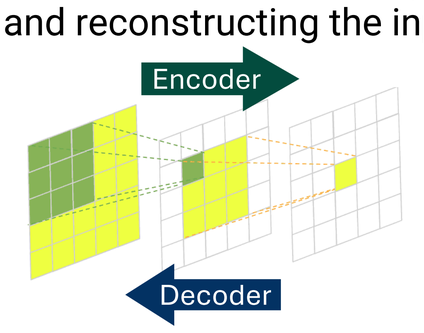
- Nine syllable combinations: $\{p_1, t_1, ʔ\}_{onset}$ /a/ $\{p_2, t_2, ∅\}_{coda.}$ e.g. $[p_1at_2]$.
- /p/ and /t/ in onset position (p_1 and t_1) and coda position (p_2 and t_2).
- N_{subj}=6, N_{item}=3456
- The syllables were articulated following an initial prolonged [i:] (iy).
- Training (60%), validation (20%), and test sets (20%)





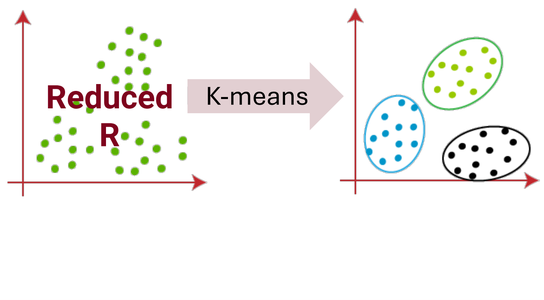
Autoencoder neural network (NN):

- **Encoder**: compressing input to latent representation (**R**).
- **R:** most compressed representation of the input
- **Decoder**: decompressing **R** and reconstructing the input.

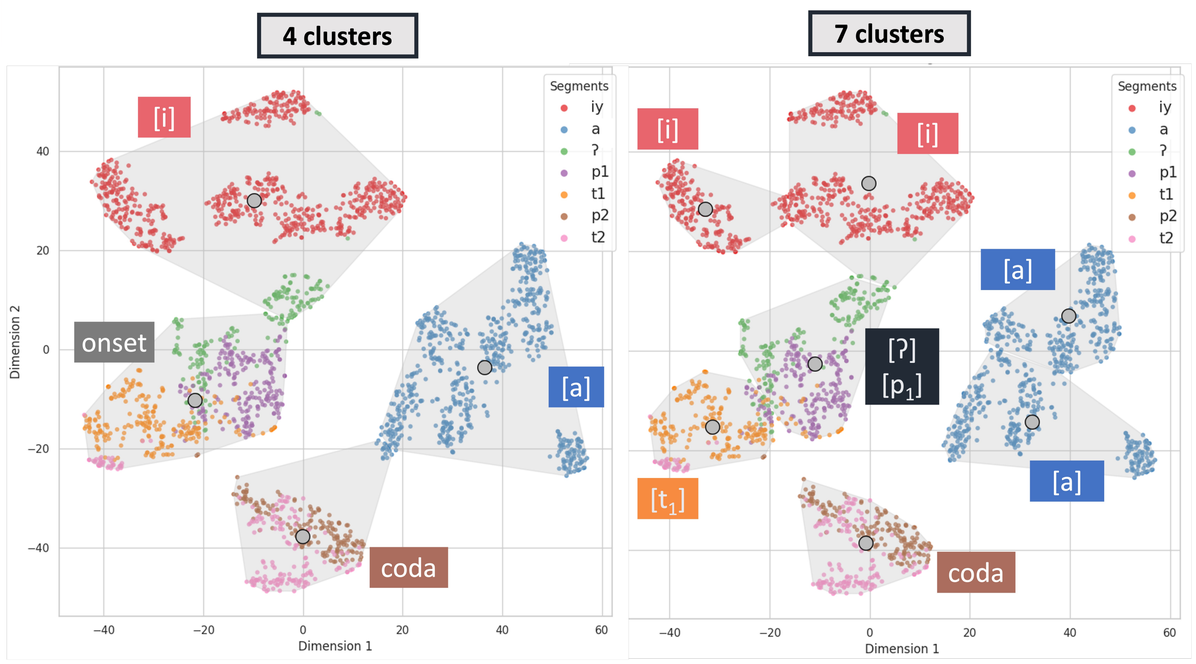


Clustering analysis:

- **Get R**: feed the trained model with test data.
- **Reduce dimensions of R:** t-distributed Stochastic Neighbor Embedding (t-SNE)
- **Access similarities**: K-means clustering within the **reduced R** space.

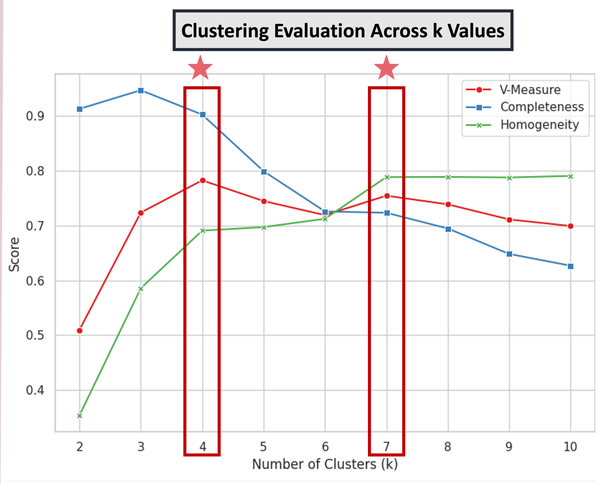


3 Results: Syllable position emerged as a stronger predictor of representational similarity than segment identity.



- Consonants with the same syllable position (e.g. onset p_1, t_1, ʔ) were closer to each other compared to the same identity (e.g. onset p_1 and coda p_2).
- **4 clusters:** onsets, codas, [i]'s, [a]'s
- **7 clusters:** Increasing *k* does not lead to clusters for segment.
- (Sub-clusters for [i] and [a] were from individual speakers).

How to choose the number of clusters k?



- Three evaluation scores:
  - Homogeneity (H)
  - Completeness(C)
  - V-measure(V): a harmonic mean between H and C.
- Best performance: k=**4** and k =**7.**

k	3	**4**	5	6	**7**	8
H	0.59	0.69	0.70	0.71	0.79	0.79
C	0.95	0.90	0.80	0.72	0.72	0.69
V	0.72	**0.78**	0.74	0.71	**0.76**	0.74

4 Summary

- Developed an unsupervised learning method for segment categorization.
- Applied it to English obstruents in different syllable positions.
- Found that syllable position is more prominent than segment identity in **R** learned by the unsupervised NN, suggesting that the role of syllable position in human representations may be underappreciated.

5 Discussion

- Unsupervised learning allows for theoretical constructions (like phonemes and syllable positions) **to be discovered**, rather than presupposed.
- Future follow-ups:
  - Encoding **articulatory data** to input to compare acoustic / motor features.
  - **Larger and noisier dataset**: ensure that the model is not simply learning distributional information of the training data and expand to sounds beyond obstruents.

6 References

[1] Port, R. F. (2010). Rich memory and distributed phonology. [2] Turk, A. (1994). Phonological Structure and Phonetic Form: Articulatory phonetic clues to syllable affiliation: gestural characteristics of bilabial stops. [3] Shain, C., & Elsner, M. (2019). Measuring the perceptual availability of phonological features during language acquisition using unsupervised binary stochastic autoencoders. [4] Shain, C., & Elsner, M. (2020). Acquiring language from speech by learning to remember and predict.

Take a picture to download the poster

fz227@cornell.edu