Learning Phone Embeddings
for Word Segmentation of Child-Directed Speech

Jianqiang Ma   Çağrı Çöltekin   Erhard Hinrichs

{jma,ccoltekin,eh}@sfs.uni-tuebingen.de

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

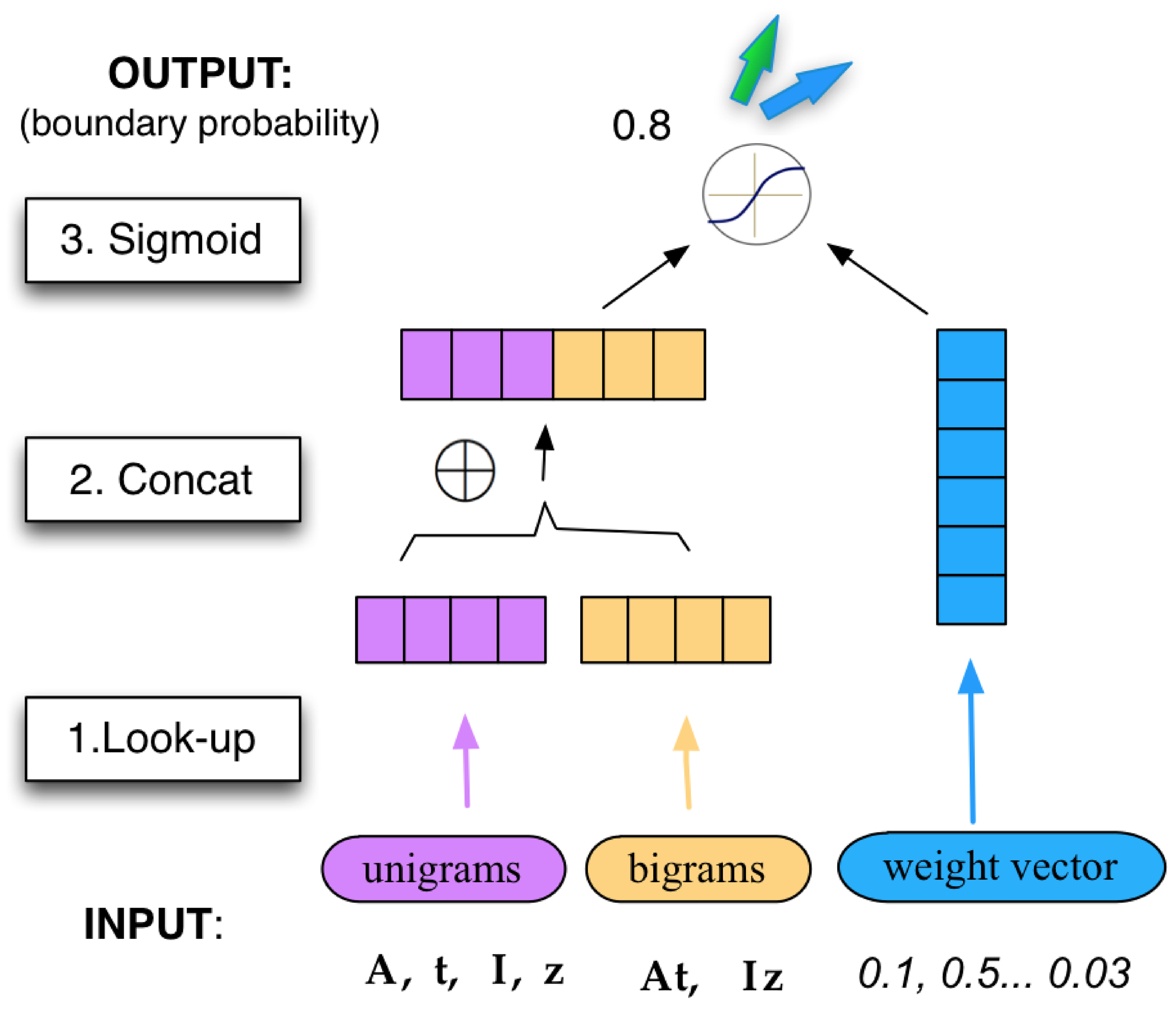
Introduction

- Segmentation of words from continuous speech is one of the first tasks infants solve
- Embeddings allow learner to exploit similarities between the phones
- The model learns from unlabeled data, exploiting information from utterance boundaries

Summary

- We model joint learning of segmentation and phone embeddings
- The embeddings model outperforms the symbolic model (using one-hot vectors)
- The embedding model learns linguistically meaningful classes of phones

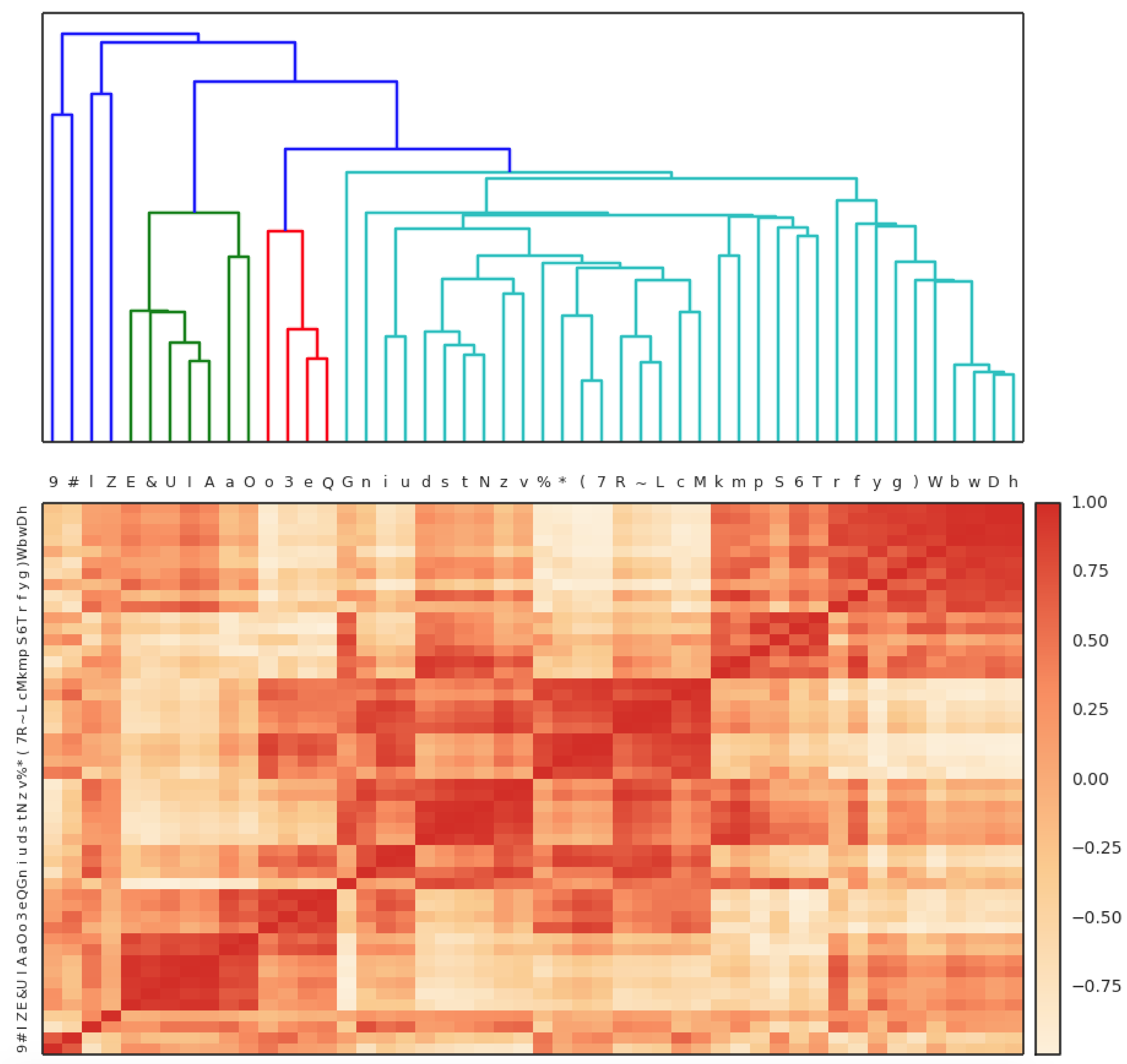
The model



The position between t and I in "WAtIzIt" is being predicted.

- Online learning with SGD, L2 regularization
- Utterance boundaries as positive instances
- Negative instances are sampled randomly from intra-utterance positions

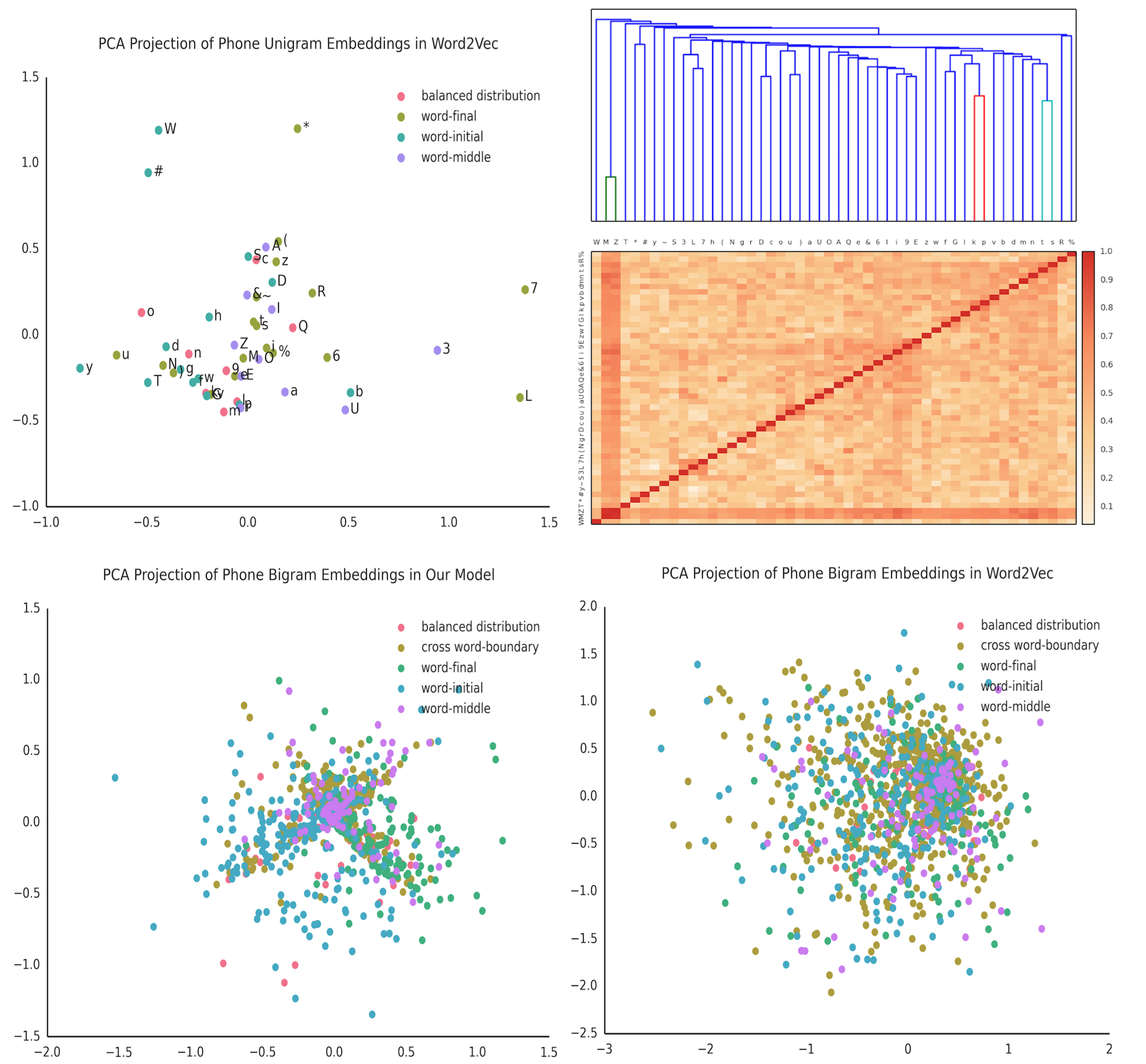
Embeddings capture phonology



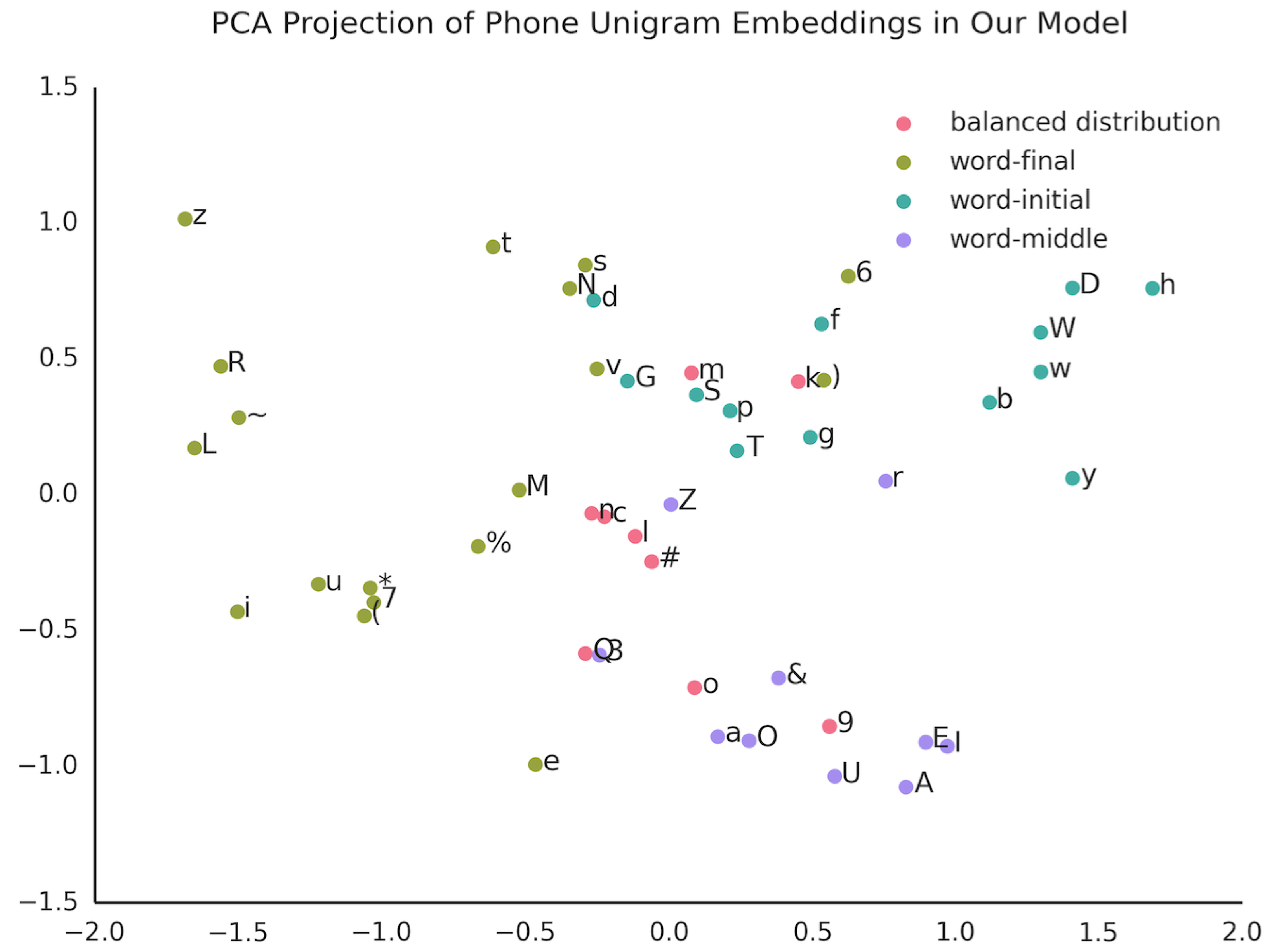
Results on child-directed speech corpus

Model	EO	EU	BF	WF	LF
embedding/all	**6.4**±**0.1**	**17.3**±**0.2**	**82.9**	**68.7**	**42.6**
symbolic/all	8.1±0.1	25.8± 0.2	75.9	60.2	31.6
embedding/unigram	15.8±0.1	**10.6**±**0.3**	**77.4**	**59.1**	**40.7**
symbolic/unigram	**13.2**±**0.1**	21.7±0.2	73.4	54.4	29.4

Comparison with word2vec embeddings



Similarities of embeddings indicate segmentation roles



Jianqiang Ma, Çağrı Çöltekin and Erhard Hinrichs (2016) “Learning Phone Embeddings for Word Segmentation of Child-Directed Speech”
In *Proceedings of ACL 2016 Workshop on Cognitive Aspects of Computational Language Learning* pp. 53–63
http://aclweb.org/anthology/W/W16/W16-1908.pdf