Computational modelling of infants’ word acquisition

Jing Liu

Master Thesis intermediate presentation

Department of Computer Science, KU Leuven

Table of contents

Context overview of the work

Twɪŋkəltwɪŋkəllɪtlstar

twinkle, twinkle, little star

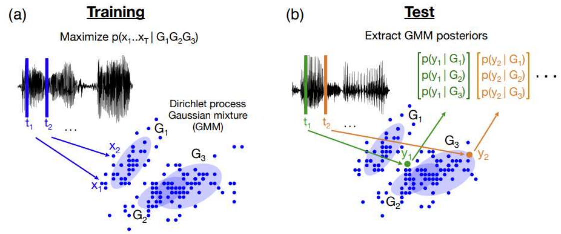
[1] Vallabha et al(2007). *PNAS*. [2] Schatz et al. (2021). *PNAS; [3] Kamper et al. (2015).ICASSP; [4]* Matusevych et al. (2015*).CogSci; [*5] Kamper et al. (2019)*.ICASSP*; [6] Goldwater et al. *(2019).Cognition;* [7] Lavechin et al. (2017). *Psyarxive*

Context overview of the work

Twɪŋkəltwɪŋkəllɪtlstar

twinkle, twinkle, little star

Computational models of language acquisition



- Phonetic learning
- GMM [1][2]
- (Correspondence) autoencoder [3][4]
- (Correspondence) autoencoding recurrent neural network[4][5]

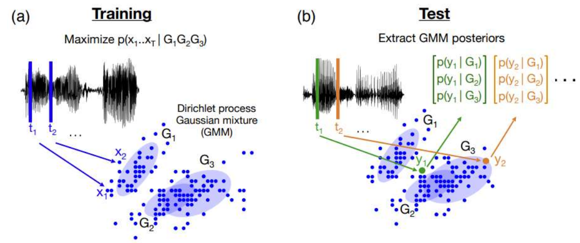
[1] Vallabha et al(2007). *PNAS.* [2] Schatz et al. (2021). *PNAS; [3] Kamper et al. (2015).ICASSP; [4]* Matusevych et al. (2015)*.CogSci;* [5] Kamper et al. (2019)*.ICASSP*; [6] Goldwater et al. *(2019).Cognition;* [7] Lavechin et al. (2017). *Psyarxive*

Context overview of the work

Twɪŋkəltwɪŋkəllɪtlstar

twinkle, twinkle, little star

Computational models of langue acquisition



- Phonetic learning
- GMM [1][2]
- (Correspondence) autoencoder [3][4]
- (Correspondence) autoencoding recurrent neural network[4][5]

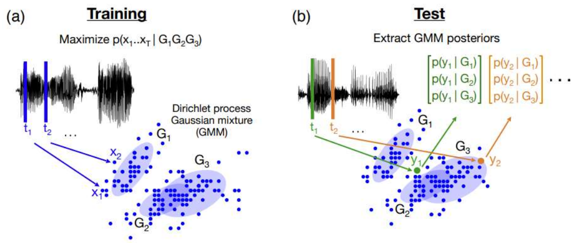
- Word learning

minimum description length(MDL) Bayesian model [6]

[1] Vallabha et al(2007). *PNAS.* [2] Schatz et al. (2021). *PNAS; [3] Kamper et al. (2015).ICASSP; [4]* Matusevych et al. (2015)*.CogSci;* [5] Kamper et al. (2019)*.ICASSP*; [6] Goldwater et al. *(2019).Cognition;* [7] Lavechin et al. (2017). *Psyarxive*

Context overview of the work

Computational models of langue acquisition



- Phonetic learning

- GMM [1][2]
- (Correspondence) autoencoder [3][4]
- (Correspondence) autoencoding recurrent neural network[4][5]

- Word learning

minimum description length(MDL) Bayesian model [6]

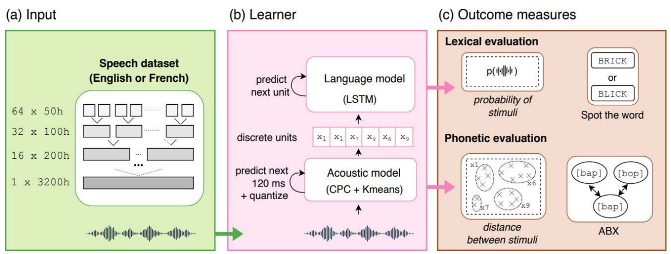
- Joint learning in both phonetic category and lexical knowledge

- STELA (Statistical Learning of Early Language Acquisition)[7]

Acoustic model: Contrastive Predictive Coding

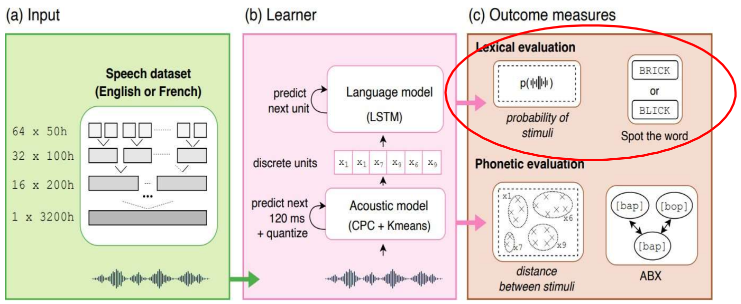
Quantizer: K-means (to simulate phonemes)

Language model: 3-layer LSTM



[1] Vallabha et al(2007). *PNAS.* [2] Schatz et al. (2021). *PNAS; [3] Kamper et al. (2015).ICASSP; [4]* Matusevych et al. (2015).*CogSci;* [5] Kamper et al. (2019).*ICASSP*; [6] Goldwater et al. *(2019).Cognition;* [7] Lavechin et al. (2017). *Psyarxive*

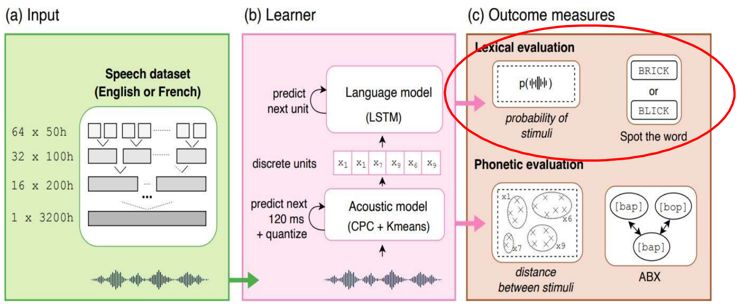
Model evaluation



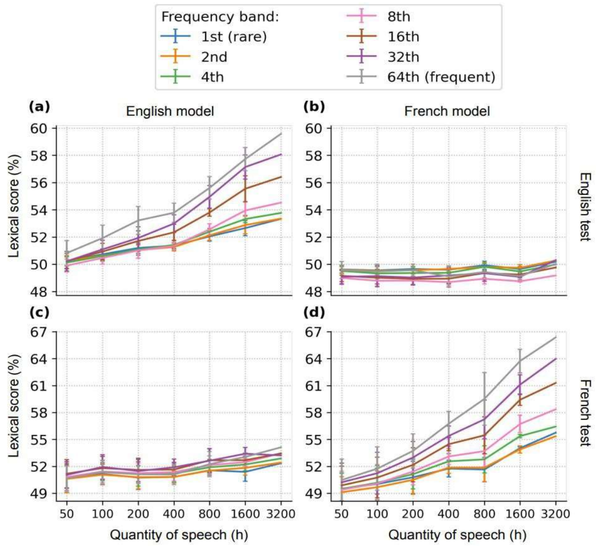
- **Lexical score**

- Spot the word task: present the network with a minimal pair of word and non-word (e.g., ’brick’ versus ’blick’)
- The accuracy score was averaged across all of the pairs in the test set
- Non-words are generated by the Wuggy toolbox

Model evaluation



- **Lexical score**
- Spot the word task: present the network with a minimal pair of word and non- word (e.g., 'brick' versus 'blick')
- The accuracy score was averaged across all of the pairs in the test set
- Non-words are generated by the Wuggy toolbox



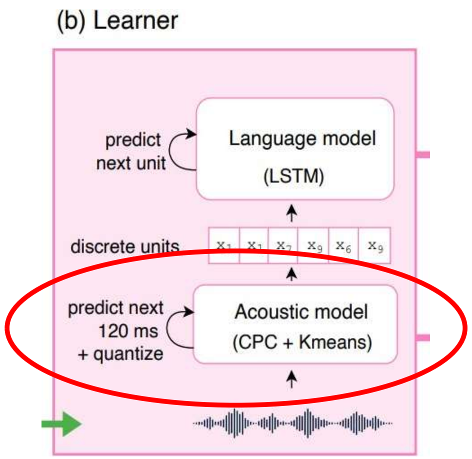
- **Training efficiency**

-> slower increase

Q: How to increase the training efficiency to better simulate infants' acquisition of lexical knowledge?

CoML KU LEUVEN

Literature review and hypotheses



- Hypothesis I: Speech segmentation algorithms

1) matching-first models [1-4]

Aim to find high quality pairs of identical segments and cluster them based on similarity

2) segmentation-first models[5-7]

Bayesian Segmental GMM and Embedded Segmental K-Means

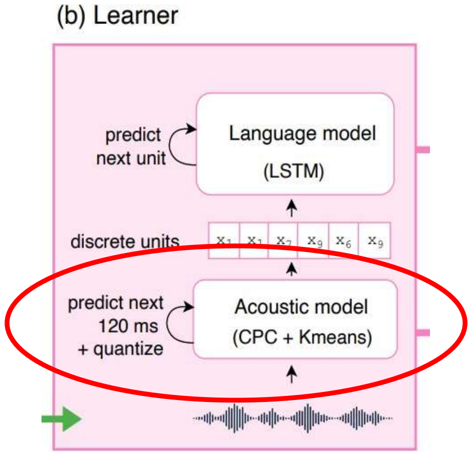
(ES-KMeans)

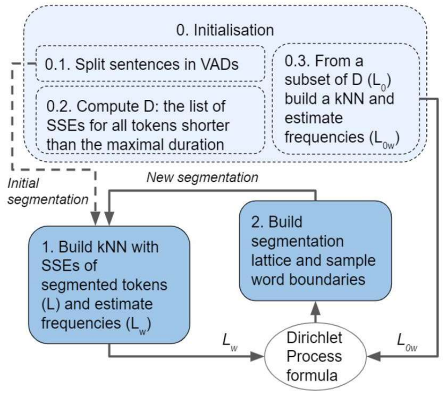
3) segmentation-only models [8-9]

discover directly the likely word boundaries

[1] Park & Glass. (2007). *Trans on Audio, Speech, and Language Processing;* [2] Jansen & Van Durme. 2011. *IEEE Workshop on ASR Understanding;* [3] Bhati et al. (2020). *arXiv preprint*
[4] Rasanen et al.(2015). *Interspeech; [5]Lee et al.(2015) Trans ACL; [6]Kamper et al.(2017). CS and Language; [7] Kamper & van Niekerk(2021)Interspeech ;* [8] Bhati et al. (2021). *arXiv preprint*
[9] Cuervo et al.(2021).ICASSP

Literature review and hypotheses





- Hypothesis I: Speech segmentation algorithms

1) matching-first models [1-4]

Aim to find high quality pairs of identical segments and cluster them based on similarity

2) segmentation-first models[5-7]

Bayesian Segmental GMM and Embedded Segmental K-Means

(ES-KMeans)

3) segmentation-only models [8-9]

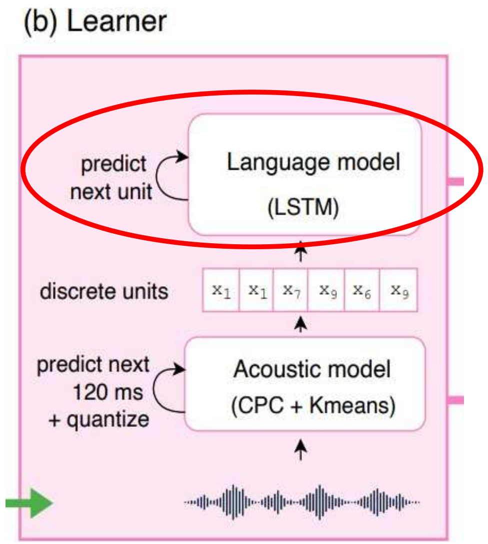
e.g. segmental contrastive predictive coding (SCPC) framework

4) combination of 2) and 3)

DP-Parse[10]

[1] Park & Glass. (2007). *Trans on Audio, Speech, and Language Processing;* [2] Jansen & Van Durme. 2011. *IEEE Workshop on ASR Understanding;* [3] Bhati et al. (2020). *arXiv preprint*
[4] Rasanen et al.(2015). *Interspeech; [5]Lee et al.(2015) Trans ACL; [6]Kamper et al.(2017). CS and Language; [7] Kamper & van Niekerk(2021)Interspeech ;* [8] Bhati et al. (2021). *arXiv preprint*
[9] Cuervo et al.(2021).ICASSP [10] Algayres et al. (2022). Trans ACL

Literature review and hypotheses



- Hypothesis II: External memory [1-4]

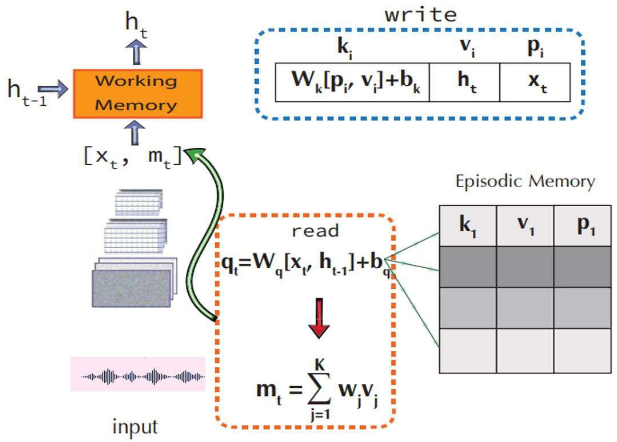
- store representations of examples or experiences

- Memory replay mechanism

-> add previously stored instances either during training[2] or test phases [3]

- Selective mechanism[4]

Similarity-based v.s. Surprise-based

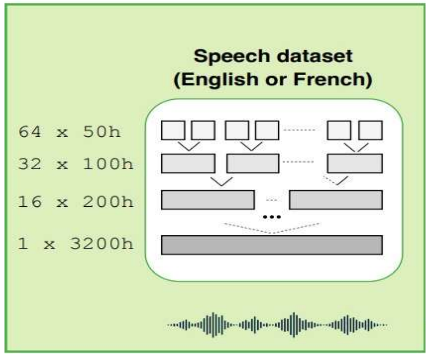


[1] Grave. (2007). *Trans on Audio, Speech, and Language Processing;* [2] Danqi. 2021. *ACL;* [3] Bhati et al. (2020). *arXiv preprint*
[4] Rasanen et al.(2015). *Interspeech*

Experiment Design

- Environment of the infant

- Models trained on varying training lengths



Experiment Design

- Environment of the infant: varying training lengths
- Model Design

Unit level	Boundary type	Model	Abbreviation
Characters	Without	Baseline LSTM	Baseline char without bound
		Memory-augmented LSTM	Memory char without bound
	Gold boundary (word)	Baseline LSTM	Baseline char without bound
		Memory-augmented LSTM	Memory char without bound
	Automatic boundary	Baseline LSTM	Baseline char with auto bound
		Memory-augmented LSTM	Memory char with auto bound
Phonemes	Without	Baseline LSTM	Baseline phon without bound
		Memory-augmented LSTM	Memory phon without bound
	Gold boundary (word)	Baseline LSTM	Baseline phon without bound
		Memory-augmented LSTM	Memory phon without bound
	Automatic boundary	Baseline LSTM	Baseline phon with auto bound
		Memory-augmented LSTM	Memory phon with auto bound
Speech	Automatic boundary	Baseline LSTM	Baseline speech with auto bound
		Memory-augmented LSTM	Memory speech with auto bound

- Input: different input modalities

- Automatic boundary: DP-Parse

- Language models

Baseline: 3-layer LSTM

Memory-augmented: local memory at test time

Progress and future work

		Feb	Mar	Apr	May	Jun	Jul	Aug
Pre-processing	Char with/out bound							
	Phonemize (g2p)							
	DP-Parse							
LSTM construction	Baseline							
	Memory-augmented							
Evaluation	Wuggy test (p2g)							
Model type		Baseline Char with/out bound	Memory char with/out bound	Baseline/Memory char with auto bound	Baseline /Memory Phon	Baseline /Memory Phon	Baseline/Memory speech with auto bound	
Thesis writing								

CoML KU LEUVEN

Progress and future work

		Feb	Mar	Apr	May	Jun	Jul	Aug
Pre-processing	Char with/out bound							
	Phonemize (g2p)							
	DP-Parse							
LSTM construction	Baseline							
	Memory-augmented							
Evaluation	Wuggy test (p2g)							
Model type		Baseline Char with/out bound	Memory char with/out bound	Baseline/Memory char with auto bound	Baseline /Memory Phon	Baseline /Memory Phon	Baseline/Memory speech with auto bound	
Thesis writing								

- If time permits:
- Test different selective mechanisms: surprise-based v.s. similarity-based
- Morphological rules and oov words
- Other aspects of evaluation task: semantic

Thanks for your attention!