

# INVESTIGATING WORD LEARNING PROCESSES IN AN ARTIFICIAL AGENT



Michele Gubian, Christina Bergmann, Lou Boves  
Radboud University and Max Planck Institute  
Nijmegen, The Netherlands



## OUTLINE

**CONTEXT:** Simulation of language acquisition in infants.

**PROBLEM:** Computational models depend on many parameters and implementation choices. It can be hard to interpret results in cognitive terms.

**GOAL:** To analyse **simulations** with a state-of-the-art model (**ACORNS**) and tell apart results that can be generalised to infant learners from what is due to the implementation.

## THE SIMULATIONS

**GOAL:** To investigate how infants cope with different speakers while learning to discover 'words' and associate them with concepts.

### DETAILS

- **Model** trained incrementally, **70 columns** MEMORY MATRIX
- Training set: **480 English sentences**
- Each sentence contains one out of **10 keywords** (*car, dog, ...*)
- **4 speakers** (2 male, 2 female)
- Modalities: **speaker-mixed** vs. **speaker-blocked**
- **Model** does not know the number of keywords/speakers nor where the keywords occur in the audio input

## TOOLS FOR THE ANALYSIS



Devise a **dissimilarity measure** between 110,000 long audio columns.

We chose a normalised symmetric KL-divergence.



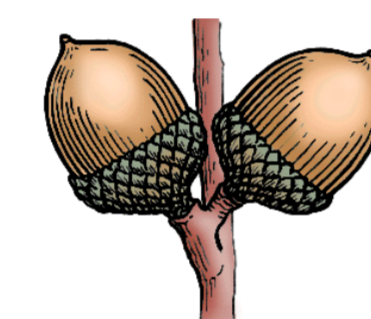
Perform a **cluster analysis** of the column space during training to locate isolated columns associated to keywords.

We used hierarchical clustering and silhouette coefficients.

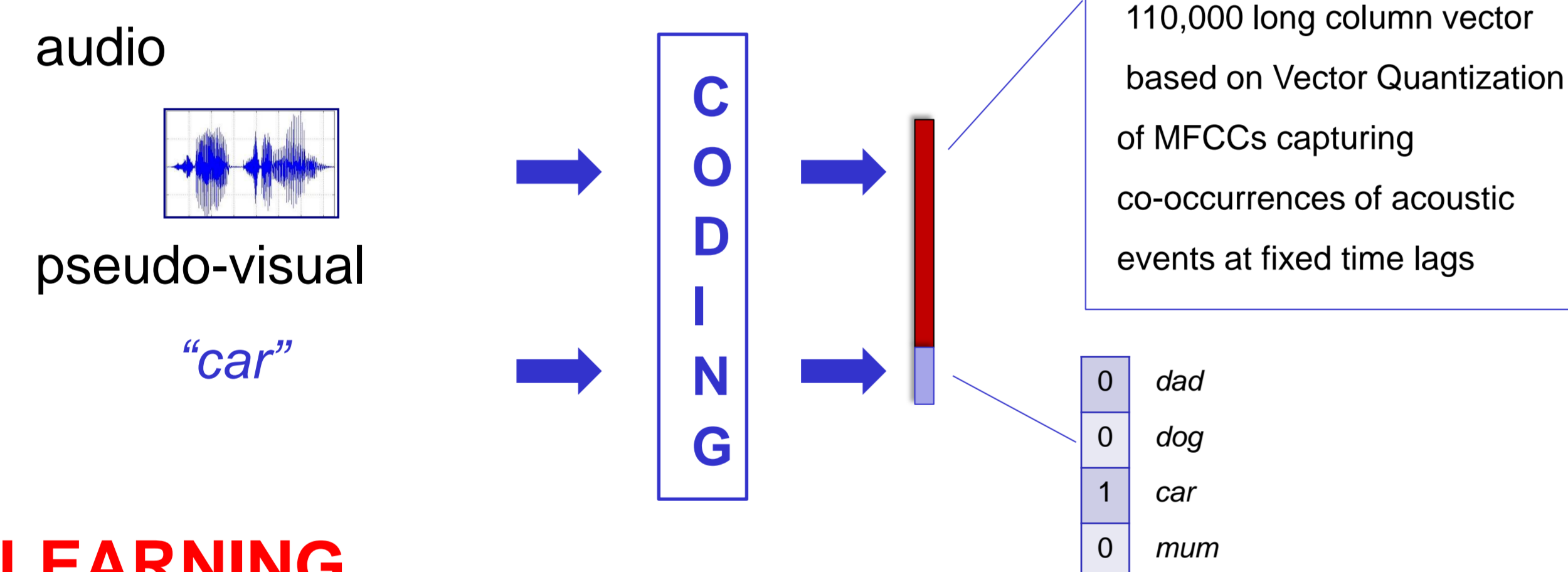


Deploy a battery of **linear models** in the (failed) attempt to understand the role of columns not associated to keywords.

## THE ACORNS MODEL



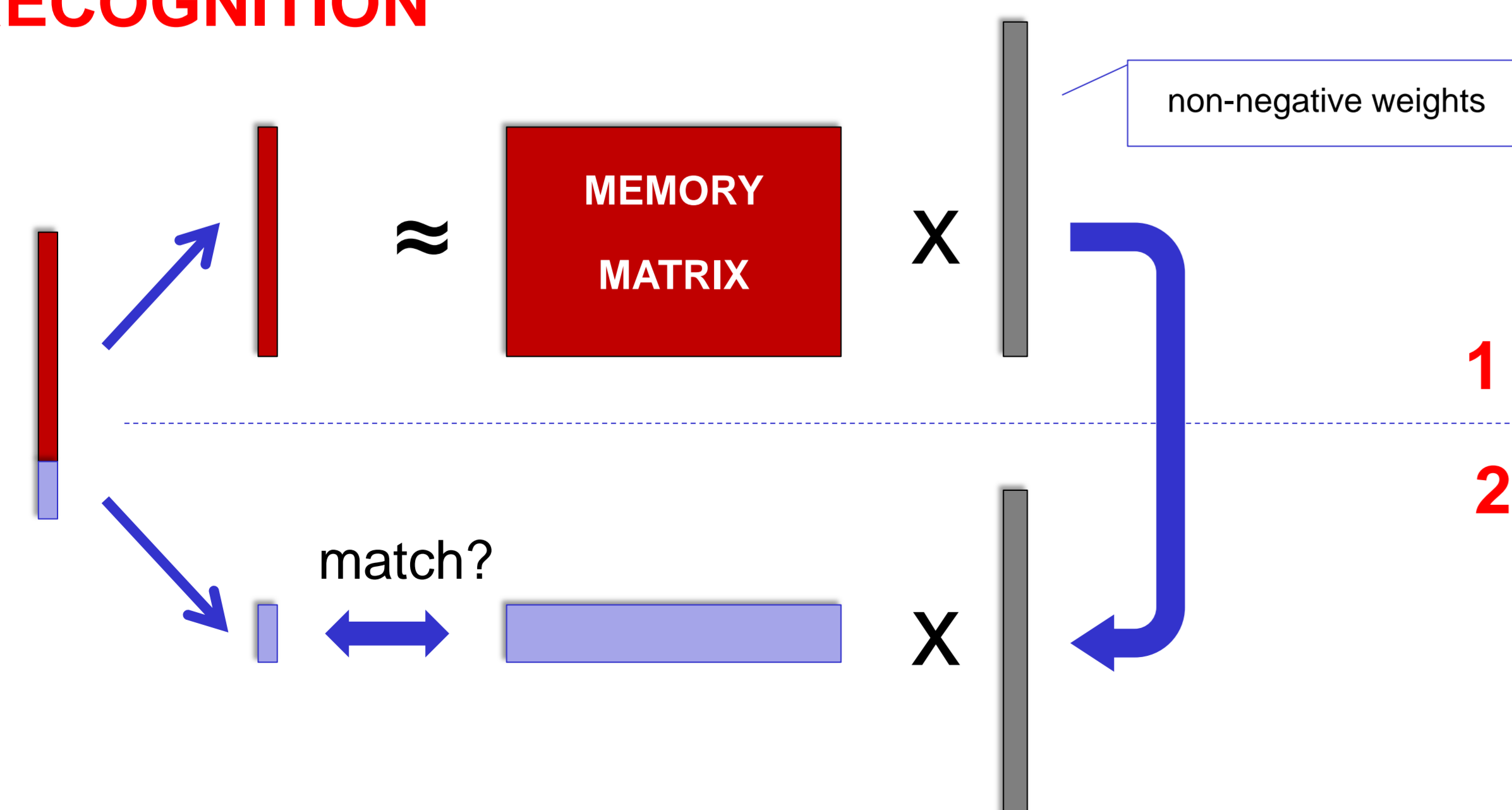
### INPUT and CODING



### LEARNING

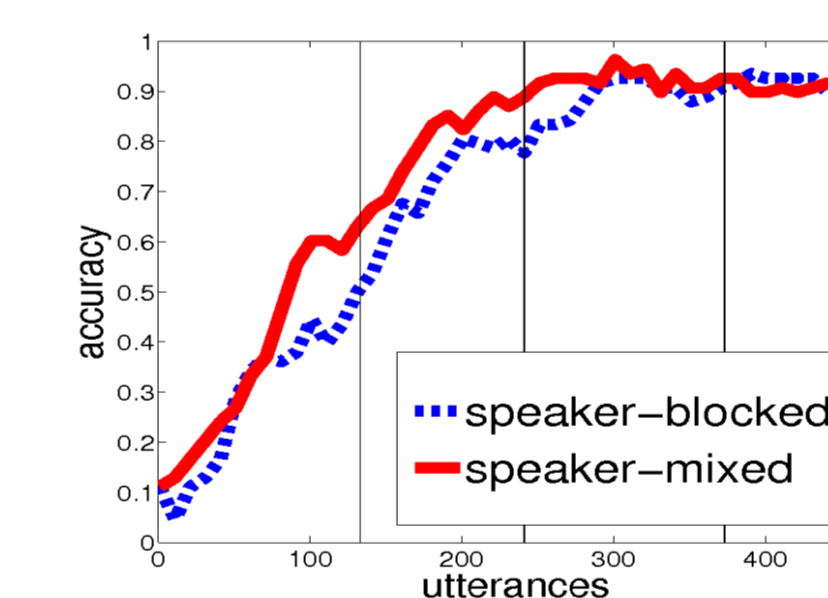


### RECOGNITION

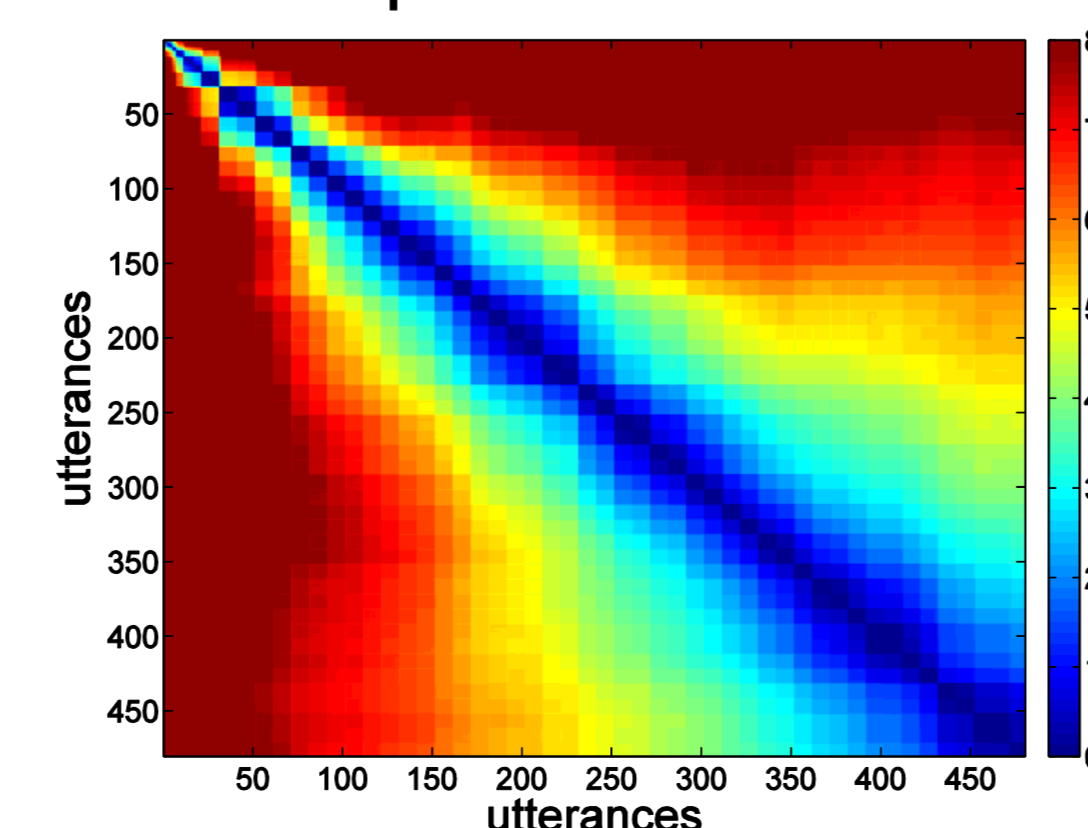


## ANALYSIS OF THE SIMULATIONS

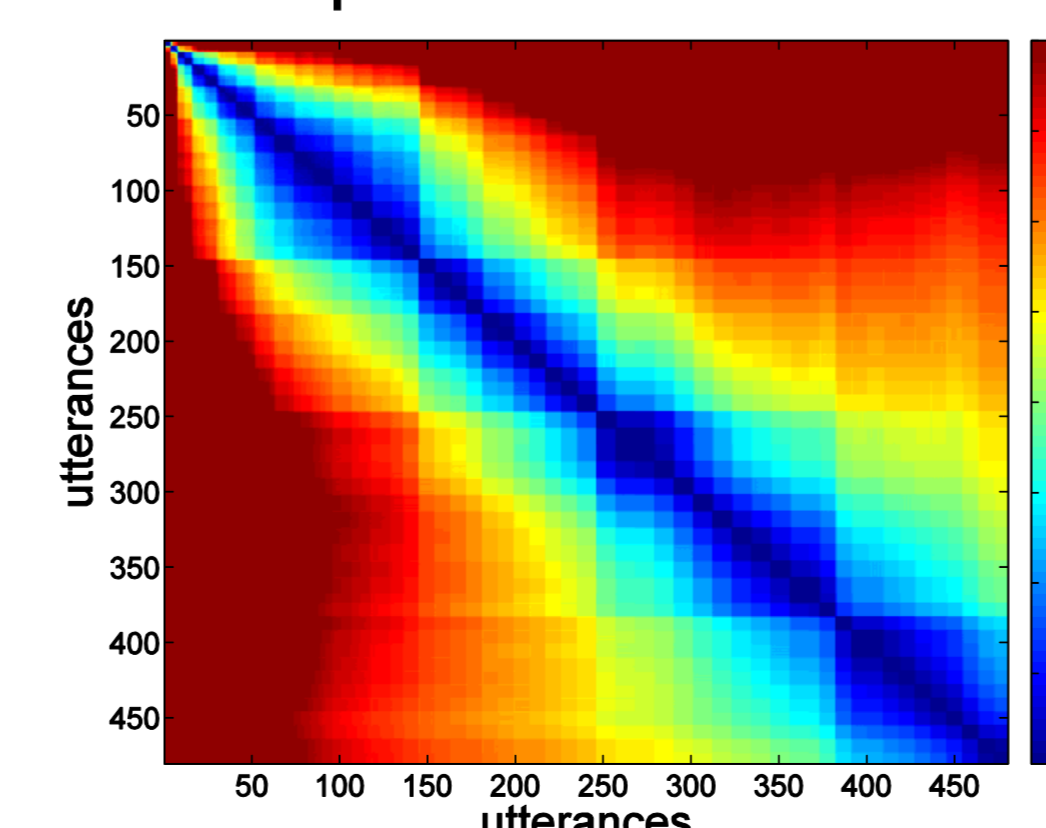
- 1) Sound-keyword association accuracy of 90% after 200 training utterances for speaker-mixed and speaker-blocked.
- 2) After 3-4 presentations of each sound-keyword pair some columns in the MEMORY MATRIX encode only one pairs.
- 3) The other columns are used in audio reconstruction but we were not able to determine what exactly they encode.
- 4) All columns keep evolving during training. Sound-keyword columns change only when the keyword is present. All react to new speakers in the speaker-blocked modality, but no speaker dependent structures appear.



speaker-mixed



speaker-blocked



## FROM SIMULATIONS TO INFANTS

- ACORNS simulates supervised word learning, with the amount of speech heard in 2-3 days by an infant.
- The fast one-to-one sound-keyword binding in ACORNS mirrors data on fast recognition of familiar words in infants as in [Newman, 2008].
- Simulations assume undivided attention of the learner to the object referred to in the spoken utterance.
- Sensitivity to speaker change in ACORNS was only marginal.
- Our analysis of internal representations suggests that both the sound-keyword binding and the moderate speaker-dependent dependency in ACORNS are due to **the orthogonality of the keyword coding**. Major re-organisations are not necessary.

## TAKE-HOME MESSAGE

- We emphasise the need for a detailed inspection of internal representations in computational models.
- We encourage the use of transparent models and to provide tools to inspect their internal behavior.