

Modelling Robustness in Infant Word Recognition

The Effect of Noise and Speaker Familiarity

Christina Bergmann^{1, 2}
 Louis ten Bosch¹
 Paula Fikkert¹

¹ Centre for Language Studies
² International Max Planck Research School
 for Language Sciences

Radboud University Nijmegen



Research Questions

- Can we model infants' behaviour assuming only episodic representations and without higher-level linguistic knowledge?
- How robust is word recognition based on such internal representations to different sources of variability in the input?

Background

Infant language acquisition

- Discovery of word-like units in continuous speech starting around 7.5 months
- Episodic storage (e.g. Newman, 2008)
- 7.5 month olds have problems when confronted with a *new speaker* (Houston & Jusczyk, 2000/2003)
- Particularly when speaker gender changes
- 7.5 month olds can deal with *moderate background noise* (Barker & Newman, 2004)

Experiments

Learning an Episodic Lexicon

- Adjust memory to optimally accommodate new information
 → Sentence by sentence
- Short sentences containing 1 *keyword* (out of 20) with associated semantic object label
- One female Speaker (Mother) or one male Speaker (Father)
- Learn from 1000 Utterances (50 per *keyword*)

Testing Word Recognition

- No learning during testing
- Unknown sentences, no semantic object label
- Mother/Father (familiar speaker) or female/male stranger (unknown speakers)
- Added noise (multitalker babble) with decreasing SNR (Signal-to-Noise Ratio)

The Model

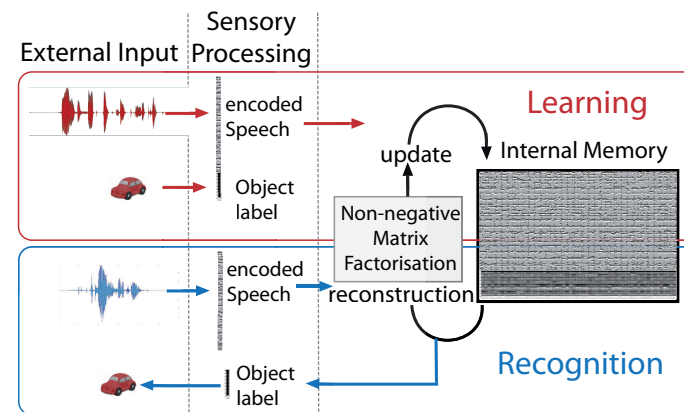


Fig 1: Schematic outline of the computational model, with a division into learning and recognition (testing)

- Multi-modal input (real speech and semantic object labels)
- Incremental learning from one sentence at a time
 Idea: Recognise new input by interpreting it in terms of fragments of previous experiences
- Episodic, language-general representations
 Idea: Store information on short stretches of the speech
 → No symbols, no phonetic categories, no word boundaries

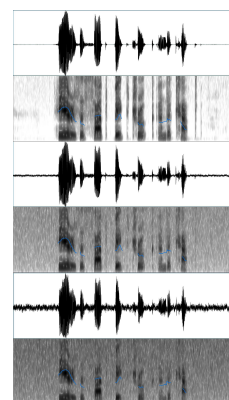


Fig 2: Clean speech (top) and added babble noise at SNR 20 dB and 10 dB

What is Noise doing to the Signal?

- With decreasing Signal-to-Noise Ratios (SNR), noise gets louder in comparison to the speech signal
- Acoustic information is gradually destroyed

→ Adults can use linguistic and lexical knowledge for top-down recognition

Results

- Same speaker during training and testing
 → perfect recognition
- *Stranger* always below *Mother/Father*
- Gradual decrease of accuracy with noise
- Stranger more sensitive to noise
- Differences according to gender of familiar speaker only

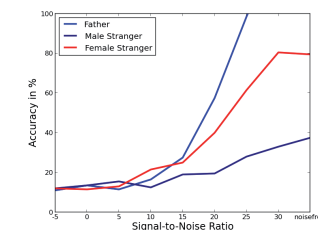
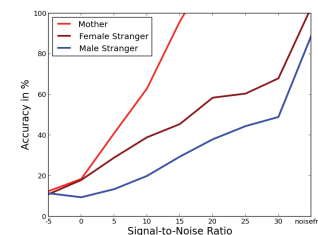


Fig 3: Word recognition accuracy for each familiar speaker and unknown female/male Strangers. Performance is displayed as word recognition accuracy across SNRs in dB at the end of learning.

Discussion

Replications

- No strong gender effect (contrary to Houston & Jusczyk, 2000/2003)
- **But:** No gender effect found in more natural learning conditions (van Heugten & Johnson, 2012)
- Yes, slow decrease with more noise and strong advantage for known speaker (Barker & Newman, 2004)

Representations

- Fairly robust to noise and speaker change
- Necessary to use episodic information that can deal with partially destroyed speech signal
 → Not possible with discrete string of phones or words

Conclusion

- A model using representations with minimal assumptions and no language-specific knowledge can model infant behaviour
- Noise and speaker change only gradually affect word-recognition
 → These effects are additive, there is no interaction
- Using a model we could replicate, combine, and extend findings on the word representations of 7.5 month olds

References

B. A. Barker & R.S. Newman. (2004). *Cognition*, **94**(2).
 C. Bergmann, M. Gubian, & L. Boves. (2010). *Proceedings of the Eleventh Annual Conference of the International Speech Communication Association (Interspeech)*.
 D. M. Houston & P.W. Jusczyk. (2000). *Journal of Experimental Psychology: Human Perception and Performance*, **26**(5).
 D. M. Houston & P.W. Jusczyk. (2003). *Journal of Experimental Psychology: Human Perception and Performance*, **29**(6).
 R.S. Newman. (2008). *Current Directions in Psychological Science*, **17**(3).
 M. van Heugten & E.K. Johnson. (2012). *Journal of Speech, Language, and Hearing Research*, **55**(2).

