



HAL
open science

From Sound Primitives Learning to Structure Extraction: Brain-inspired Model of Infant Early Vocal Learning

Xiaodan Chen, Alexandre Pitti, Mathias Quoy

► **To cite this version:**

Xiaodan Chen, Alexandre Pitti, Mathias Quoy. From Sound Primitives Learning to Structure Extraction: Brain-inspired Model of Infant Early Vocal Learning. IMOL, Stéphane Doncieux, Sep 2023, Paris Sorbonne Université, France. hal-04372439

HAL Id: hal-04372439

<https://hal.science/hal-04372439v1>

Submitted on 4 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

From Sound Primitives Learning to Structure Extraction: Brain-inspired Model of Infant Early Vocal Learning

Xiaodan Chen, Alexandre Pitti, Mathias Quoy
ETIS, CY Cergy Paris Université - ENSEA – CNRS, UMR 8051
firstname.lastname@cyu.fr

The underlying mechanism of the biological neural machinery of the human brain receives an upsurge of attention from machine learning scientists, as developmental learning gives insight into open-ended unsupervised learning. In particular, language acquisition remains an exploitable field in not only machine learning but psychology and biology as well.

Psychologically speaking, infants' perception of speech sounds in early infancy contributes to later language acquisition, making it one of the main paths of exploring how infants' brains manage to process early information. It is noticeable that adults' perception of speech sounds becomes constrained compared to that of newborns, as influenced by the phonetic patterns of their native language, which appears by the end of the first year after birth. Correspondingly, in machine learning's perspective, it is doubtful if unsupervised learning in adulthood might happen without supervised learning [4]. Thus developmental learning until infants' first anniversary is believed to be the most similar mechanism to unsupervised machine learning.

As a result, our research focuses on modeling computationally the learning mechanism of speech sound in early infancy, more precisely before 12 months of age, and we propose a bio-inspired early-infant-based neural network for the learning of speech sounds, which lays a foundation for the later grammatical structure extraction neural network for language acquisition.

Our neural network is mainly based on three hypotheses. The first one is distributional learning. [2] argues that infants' perception of phonetic categories is affected by the sound in the environment. For example, as English speakers hear -r and -l as distinct while Japanese speakers hear them as identical, American infants' discrimination of English -r and -l increases and that of Japanese infants decreases between 8 to 10 months of age. The second one is that imitation is critical for early vocal learning. As is shown in Fig. 1, formant frequency values of produced vowels form corresponding clusterings in relation to one another as phonetic categories develop. Therefore, [1] highlights that hearing the vocalization of others (perception) and hearing oneself produce sounds (production) are both essential for speech sound perception. The third one is that interaction reinforces infants' perception during language exposure. Experimental results in [2] show that infants that are exposed to foreign language material via audio-type and television only showed no learning. Their performance equaled that of infants in the control group without being exposed to foreign language material, while infants interacting with teachers showed better performance. This proves that the presence of human

interacting with infants is essential for speech sound perception.

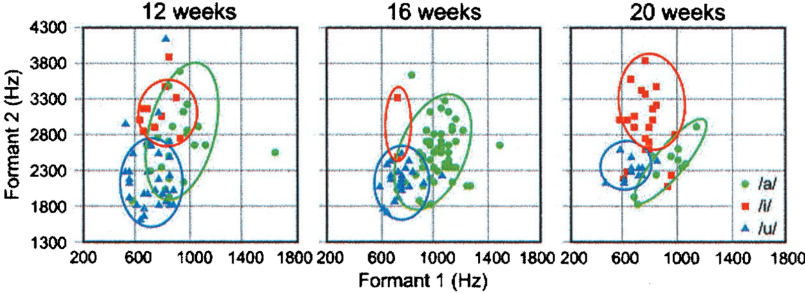


Figure 1: Infants’ vowels /a/, /i/, or /u/ recorded show developmental change between 12 and 20 weeks of age during imitations [1].

Consequently, we propose a neural network consisting of unsupervised clustering algorithms with a reinforcement loop for the reproduction of sounds, and we analyze biologically neuron activities and characteristics to see if they show something in common with the neurons in natural-born intelligent system. As mentioned before, native language’s influence on infants’ sensitivity towards phonetic distribution takes place by the end of 12 months of age. This can be explained by the infants’ selective attention to only those acoustic patterns that are highly native-language-related [3]. Therefore, undoubtedly, such a speech sound learning neural network may play a role in the language acquisition system in our future work.

References

- [1] Patricia K. Kuhl. A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, 2000.
- [2] Patricia K. Kuhl. Brain mechanisms in early language acquisition. *Neuron*, 67(5):713–727, 2010.
- [3] Jessica Maye, Janet F. Werker, and Lou Ann Gerken. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 2002.
- [4] Larijn Zaadnoordijk, Tarek R. Besold, and Rhodri Cusack. Lessons from infant learning for unsupervised machine learning. *Nature Machine Intelligence*, 2022.