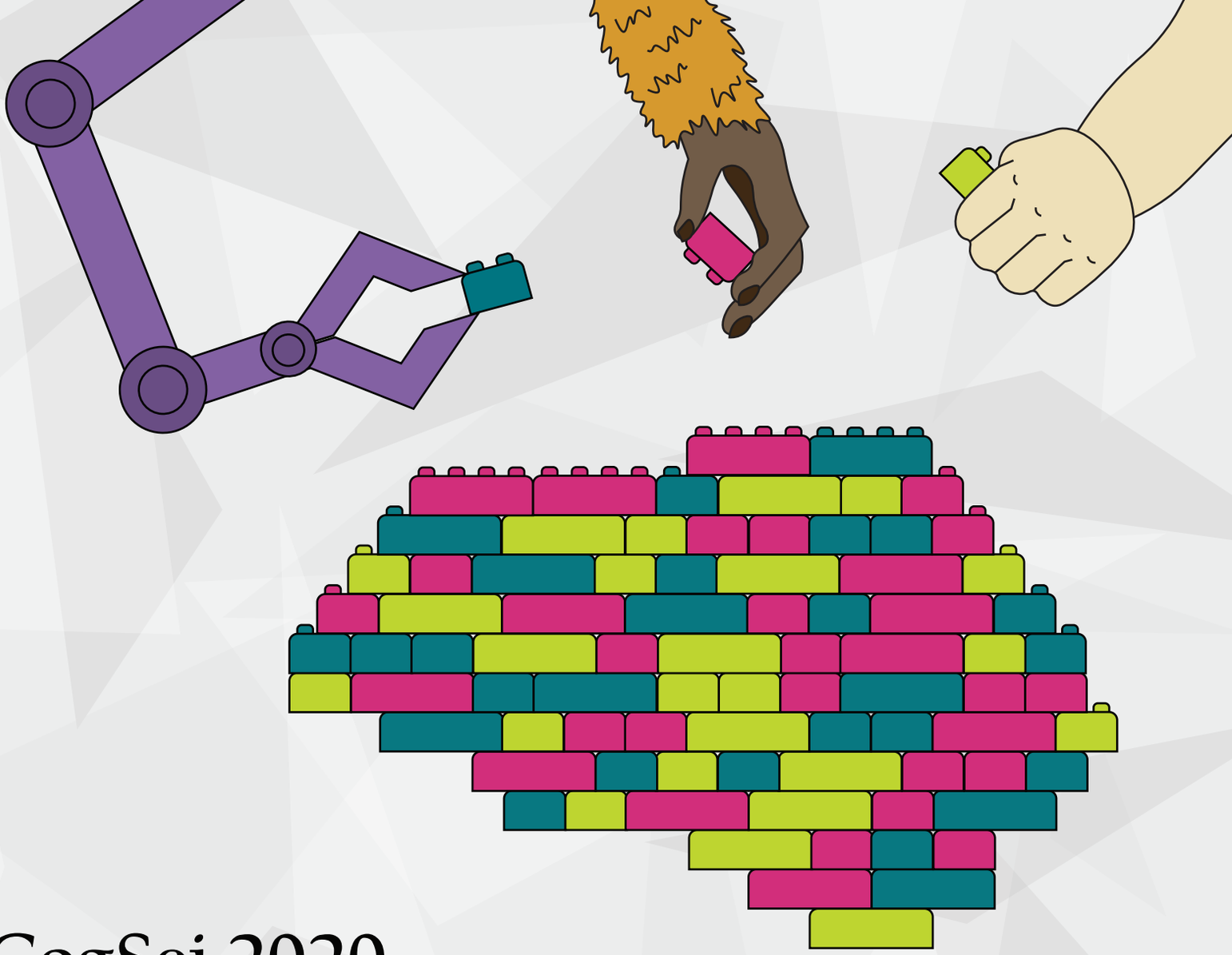
A model of prenatal acquisition of vowels

Chládková, K.; Nudga, N.; Boersma, P.



CogSci 2020 Proceedings

Developing a Mind:
Learning in Humans, Animals, and Machines

Invited Speakers
Cecilia Heyes
Geoffrey Hinton
Janet Werker

Invited Panels
Deep Integration of Development and Cognitive Science
Social, Cultural, and Linguistic Constraints on Development
Statistical Learning and Development

Co-Chairs: Stephanie Denison | Michael Mack | Yang Xu | Blair C. Armstrong

Proceedings for the 42nd Annual Meeting of the Cognitive Science Society

29 July – 1 August 2020

Developing a Mind: Learning in Humans, Animals, and Machines

Program Chairs: Stephanie Denison, Michael Mack, Yang Xu, Blair C. Armstrong

http://www.cognitivesciencesociety.org/cogsci-2020

Sponsors

The Robert J. Glushko and Pamela Samuelson Foundation
Duolingo
MIT-IBM Watson AI Lab
The Cognitive Science Society

Thank you again for your support!

How to Cite Your Paper

APA formatted citation for a paper:

Author, A. & Author, B. (2020). This is the title of the paper. In S. Denison., M. Mack, Y. Xu, & B.C. Armstrong (Eds.), Proceedings of the 42nd Annual Conference of the Cognitive Science Society (pp. PAGES). Cognitive Science Society.

APA formatted citation for a published abstract:

Author, A. & Author, B. (2020). This is the title of the abstract. In S. Denison., M. Mack, Y. Xu, & B.C. Armstrong (Eds.), Proceedings of the 42nd Annual Conference of the Cognitive Science Society (p. NUMBER). Cognitive Science Society.

APA formatted citation for a talk/poster presentation:

Author, A. & Author, B. (2020, July). This is the title of the talk or poster. Paper (or Poster) presented at the 42nd Annual Conference of the Cognitive Science Society (p. NUMBER).

A Model of Prenatal Acquisition of Vowels

Kateřina Chládková (katerina.chladkova@ff.cuni.cz)
Institute of Phonetics, Charles University, Nám. Jana Palacha 2, 116 38, Praha, Czechia
Institute of Psychology, Czech Academy of Sciences, Hybernská 8, 110 00, Praha, Czechia

Natalia Nudga (nat.nudga@gmail.com)
Institute of Phonetics, Charles University, Nám. Jana Palacha 2, 116 38, Praha, Czechia

Paul Boersma (paul.boersma@uva.nl)
Amsterdam Center for Language and Communication, University of Amsterdam
Spuistraat 134, 1012VB Amsterdam, The Netherlands

Abstract

Humans learn much about their language while still in the womb. Prenatal exposure has been repeatedly shown to affect newborn infants' processing of the prosodic characteristics of native language speech. Little is known about whether and how prenatal exposure affects infants' perception of speech sound segments. Here we simulated prenatal learning of vowels in two virtual fetuses whose mothers spoke (slightly) different languages. The learners were two-layer neural networks and were each exposed to vowel tokens sampled from an existent five-vowel language (Spanish and Czech, respectively). The input acoustic properties approximated the speech signal that could possibly be heard in the intrauterine environment, and the learners' auditory system was relatively immature. Without supervision, the virtual fetuses came to warp the continuous acoustic signal into "proto-categories" that were specific to their linguistic environment. Both learners came to create two categorization patterns and did so in language-specific ways, primarily on the basis of the vowels' first-formant characteristics. Such prenatally formed proto-categories were not adult-like in that they entirely collapsed some of the native-language contrasts. At the same time, the categories reflected features of the adult language in that they were language-specific. These results can inspire future work on speech and language acquisition in real young humans.

Keywords: prenatal learning; speech sound acquisition; vowels; models of language development; neural network

Introduction

Humans start to learn about their native language well before they are even born. At birth infants (or near-term fetuses) recognize the voice of their mother over a female stranger, their native language from an unfamiliar one, and recognize a rhyme they had heard during the last weeks of intrauterine development (Mehler et al., 1988; Moon, Cooper, & Fifer, 1993; DeCasper & Spence, 1986; Kisilevsky et al., 2009). The global prosodic, or suprasegmental, patterns of native-language speech are thus learnable, and begin to be learnt, even before an individual is born (Abboub, Nazzi, & Gervain, 2016). Besides their prosodic patterns, languages are distinguished by how they implement the individual speech sound segments: the number and acoustic properties of vowels and consonants differ vastly across languages and language varieties. Contrary to the acquisition of prosody, however, the acquisition of native-language segmental phonology was typically assumed to start only after a child is born (see e.g. the timeline in Kuhl, 2004). Here we challenge that assumption and employ a computational model to test whether language-specific perceptual categorization of segmental vowel categories could begin before birth.

Recent work suggests that (some) segmental categories of native-language speech could, comparably to the suprasegmentals, begin to be acquired already during intrauterine development. Moon, Lagercrantz, and Kuhl (2013) tested American English and Swedish newborns in a high-amplitude sucking paradigm on discrimination of within-category variants of an American English /i/ and a Swedish /y/. Overall, the infants suckled more to the variants of the non-native vowel than to the variants of the native vowel. This result was interpreted as evidence for a stronger within-category discrimination – and thus lesser extent of perceptual warping, or categorization – for the non-native vowel phoneme. One might question to what extent the Swedish /y/ and the American English /i/ differed perceptually from an American English /u/ and Swedish /i/, respectively. Nevertheless, the language-specific performance of the newborn infants suggests that prior experience with the ambient language (i.e. prenatal and/or short post-natal) is what affected the infants' behavior at the time of the experiment.

Although not testing the effects of naturalistic language exposure, several other studies demonstrate that before and at birth, humans can learn about the speech sounds in their environment. Cheour et al. (2002) showed that several hours of overnight auditory training with frequent [i]'s and rare [ɨ]'s and [y]'s helped Swedish newborns to perceptually distinguish amongst those vowels at post-test. Along similar lines but this time with auditory training done over several weeks *before* birth, Partanen et al. (2013) found facilitating effects of prenatal exposure on newborns' discrimination of vowel length and vowel pitch. It thus seems that fetuses learn from exposure to the ambient speech sounds.

The ambient speech signal for a developing fetus is, however, different from the speech signal that infants learn from after they are born. A relatively large body of research addressed the intrauterine (speech) sound properties. Despite

some variation in the reported cut-off frequency values and attenuation degrees, the findings indicate that, *in utero*, the externally generated acoustic signal below 500 Hz or even 1000 Hz is well preserved and that higher frequencies get progressively attenuated (e.g. Gerhard & Abrams, 1996; Richards et al., 1992). It is thus primarily temporal and low-frequency information that can guide the prenatal learning of speech (Granier-Deferre et al., 2011). This explains why near-term fetuses and newborns demonstrate language-specific perception of prosody, i.e. temporal and intonational patterns that are inherently realized within the low-frequency range.

At the segmental level, adults' recognition of phonemes from intrauterine recordings of speech seems to be mainly cued by the sounds' first formant (Querleu et al., 1988), which is the lowest of the vocal tract resonating frequencies and in adults typically ranges from 200 Hz to 1200 Hz. The ideal candidates for prenatal learning are vowels because, compared to most consonants, they are loud and are distinguishable in the low- to mid-frequency range (mostly by their first two, and sometimes the first three, formant frequencies). Arguably, the main cue on the basis of which fetuses learn to categorize vowels in the ambient speech signal (if they do perceptual categorization at all) will be the lowest, i.e. the first, formant (F1). The prominent role of F1 in prenatal learning is quite intuitive but it remains unclear whether and to what extent individual vowel categories could be *in utero* distinguished solely on the basis of their F1 properties, and whether and to what extent these would interact, or collide, with information from the second formant (F2) or higher ones.

Computational modelling can help examine the potential fetal and perinatal learning processes without the need to recruit the rather sensitive population and measurement techniques. Seebach et al. (1994) presented neural-network simulations of prenatal acquisition of plosive place of articulation. Seebach et al.'s neural network learned to categorize three plosive places of articulation [pa], [ta], [ka], and did so on the basis of acoustic information in lower as well as higher-frequency range.

To get a picture of how prenatal acquisition could unfold for vowels we simulated unsupervised intrauterine learning of five-vowel inventories. The goals were to find out (1) whether and on which vowel dimensions, fetuses could perform perceptual warping, or categorization, and (2) whether they could do so in language-specific ways. We employed a biologically plausible model, a two-layer symmetric neural network, and had it listen – inside the womb – to vowels' first three formants. Two learners were simulated: one listening to Spanish and the other listening to Czech (both these languages contrast 5 short vowel phonemes and slightly differ in some of the vowels' acoustic realizations). The resulting perceptual behavior was evaluated to see whether categorical-like structures emerged and whether they were specific to the learners' ambient language.

Experiment

Network

Prenatal learning of vowels is simulated here with a bidirectional neural network that has previously been used to successfully demonstrate several phonological phenomena, such as category creation or auditory dispersion (Boersma, Benders, & Seinhorst, 2020). When exposed to the sounds of a 5-vowel language at some point *after* birth, this network comes to successfully create the 5 adult-like phoneme categories (Boersma, Chládková, & Benders, in prep.).

The network has two layers, which a phonologist might call levels of representation. As illustrated in Figure 1, the bottom, auditory layer consists of 33 nodes. These correspond to the auditory frequency range from 4 to 28 ERB and thus represents a part of the basilar membrane (the distance between two adjacent nodes is 0.75 ERB). The second layer represents a higher, abstract level of processing, and we may call it the phonological form. The phonology layer consists of 15 nodes. We are using a 5-vowel system for network training and expecting category creation; in the ideal scenario that a learner acquires 5 categories, reserving 3 nodes for each category seems reasonable (note that we do not force the network to create 5 categories exactly).

There are connections across all nodes between the two layers, as well as within layers. In the network's initial state all connections between the layers have random low weights (ranging from 0 to 0.1). The connections within layers are inhibitory and held constant (set at -0.1 and -0.25, within the auditory and the phonological layer, respectively).

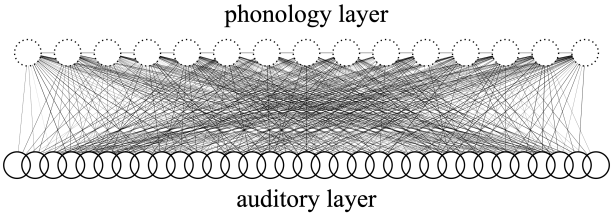


Figure 1: The neural network before learning.

Input

The virtual learner is to be exposed to Spanish or Czech vowels (see Figures 2 and 3, and Table 1; data from Chládková et al., 2011, 2019). The network is trained with auditory data transformed to approximate the intrauterine speech sound acoustics.

Since the values of the reported acoustic modulations vary across studies, we base our model on the transformations reported in Richards et al. (1992), who found that: "Low-frequency sounds (125 Hz) generated outside the mother were enhanced by an average of 3.7 dB. There was a gradual increase in attenuation for increasing frequencies, with a maximum attenuation of 10.0 dB at 4 kHz." (p.186)

According to the definition of the unit decibel, enhancement by 3.7 dB means an increase in sound intensity

by the factor of $10^{0.37}$ which is approx. 2.34, whereas $-10\ dB$ would mean a decrease by the factor of 0.1.

$$sound\ level\ [dB] = 10 \times \log_{10} \frac{I}{I_0}$$

The increase in attenuation seems to be gradual. We could apply a linear approximation in order to predict the factor of increase or decrease for other frequencies. The equation of a line passing through the two given points can be expressed as follows:

$$p = \max(0.1, -0.00058 \times f + 2.42)$$

where f is frequency measured in Hz and the obtained value of p will be the intensifying factor for the given frequency. The value of p will always be at least 0.1, since that is the factor corresponding to the reported maximal attenuation at 4kHz (Richards et al., 1992).

Table 1: The formant values of Spanish and Czech vowels (in ERB). Means and between-speaker standard deviations from 10 Castilian Spanish and 17 standard Czech female speakers.

	Spanish vowels				
	/i/	/e/	/a/	/o/	/u/
F1 mean	8.88	10.67	13.63	11.12	9.35
(sd)	(0.54)	(0.47)	(0.31)	(0.41)	(0.37)
F2 mean	22.97	21.57	19.46	16.36	14.67
(sd)	(0.50)	(0.50)	(0.51)	(0.49)	(0.55)
F3 mean	24.41	23.99	23.54	23.17	23.91
(sd)	(0.45)	(0.49)	(0.57)	(1.74)	(0.63)
	Czech (short) vowels				
	/i/	/e/	/a/	/o/	/u/
F1 mean	8.93	11.34	12.99	9.47	8.15
(sd)	(0.80)	(0.87)	(1.06)	(1.25)	(0.49)
F2 mean	21.99	20.58	18.00	15.86	15.14
(sd)	(0.93)	(0.42)	(1.18)	(0.98)	(1.31)
F3 mean	24.09	23.74	23.12	23.94	23.94
(sd)	(0.60)	(0.46)	(1.43)	(0.71)	(0.71)

Training

The network is trained in a bottom-up direction, in 40.000 steps. At each iteration, a random vowel category is selected and its F1-F2-F3 values are drawn from the Gaussian distributions defining this vowel's first three formants in ERB (see Table 1 and Figures 2 and 3; the formula for converting Hz to ERB is: $y_{ERB} = 11.17 \ln((x_{Hz} + 312) / (x_{Hz} + 14680)) + 43$). The auditory nodes are clamped (meaning that their activities cannot change) and the phonological nodes are unclamped (meaning that their activities will adapt). The auditory node activities are set with the formula:

$$\Delta e = 0.5 \times \left(p_{F1} \times e^{\frac{(k-F1)^2}{s^2}} + p_{F2} \times e^{\frac{(k-F2)^2}{s^2}} + p_{F3} \times e^{\frac{(k-F3)^2}{s^2}} \right)$$

where p is the factor of in-utero increase or decrease of energy, s – auditory spreading, k – index of the auditory node (from 1 to 33), F1-F3 are the frequencies of the first, second and third formant respectively measured on our arbitrary scale from 1 to 33. Auditory spreading here is 1.5 times larger than in Boersma et al. (2020) to represent the fact that the basilar membrane (and the topography in the A1 as well) is a bit less developed before than after birth.

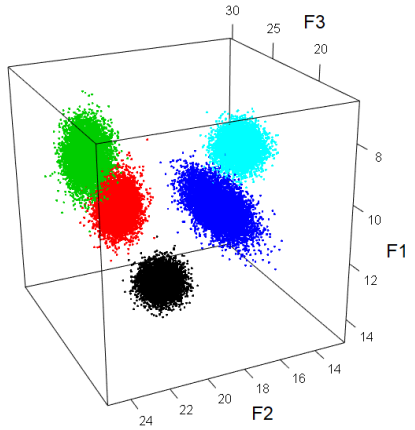


Figure 2: A plot of 40,000 F1-F2-F3 Spanish input vowels. Color-coding of 5 adult categories, unknown to the learner.

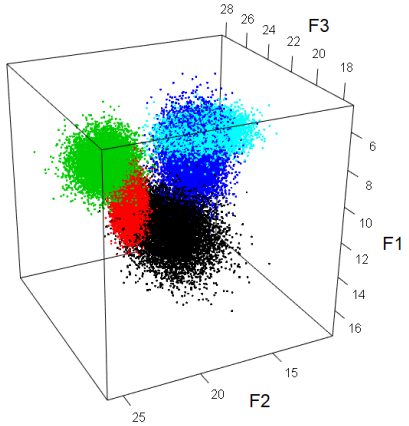


Figure 3: A plot of 40,000 F1-F2-F3 Czech input vowels. Color-coding of 5 adult categories, unknown to the learner.

Once the auditory node activities are set (i.e. fetus encounters an auditory stimulus), the activity is allowed to spread through the network. The activities of the unclamped nodes at the phonological layer are initially set to zero and during the spreading of activity they are being updated in 500 small steps according to the following formula:

$$\Delta e = \eta_a \left(\sum_{connected\ nodes\ i} w_{ij} a_i - e_j \right)$$

where η_a is the spreading rate (in our simulation kept constant at 0.01), w_{ij} is the weight of connection between

nodes i and j, a_i is the activity of node i, and e_j stands for the current excitation of node j.

After activity spreading, the weights of between-layer connections are updated. The weight of a connection between an input node i in the bottom auditory layer and an output node j in the top phonological layer is changed according to the *inoutstar* learning rule (Boersma et al., 2020):

$$\Delta w_{ij} = \eta_w(a_i a_j - instar\ a_i w_{ij} - outstar\ a_j w_{ij} - weightLeak\ w_{ij})$$

where $instar = outstar = weightLeak = 0.5$, and η_w is the learning rate (in our simulation equal to 0.001). The formula means that the weight of a connection is strengthened when both nodes that it connects are on, weakened when only one of the nodes is on, and slightly weakened if none of the nodes is on.

Once the weights are updated, they undergo normalization so that the total weight of a node's connections is kept at a fixed value (see Rumelhart & Zipser, 1985), by applying:

$$w_{ij\ \text{normalized}} = w_{norm} \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}}$$

where w_{ij} is the current weight of the connection between node i from the top layer and node j from the bottom layer, and w_{norm} is the value at which the sum of connection weights incoming to node i is maintained. In the present simulation, $w_{norm} = 0.1 \times n$, where n is the number of auditory nodes.

Evaluation

When the network reaches an equilibrium, i.e. when the connection weights no longer change with training, we evaluate how the learner perceived an F1-F2-F3 combination typical for each of the five adult vowel categories. During evaluation, auditory nodes are clamped and the respective F1-F2-F3 auditory nodes (and their neighbors) are activated, and activity is allowed to spread through the network causing a particular activity pattern at the unclamped phonological level. Figures 4 and 5 show, for each language, how the learner perceived a typical rendition of each of the five adult phonemes. The activation patterns at the phonological layer reveal whether the learner warps the auditory space into category-like structures.

Let us first examine the outcome of Spanish-exposed network. As can be seen in the top left corner of Figure 4, an incoming sound [i], i.e. the F1-F2-F3 of the adult phoneme /i/, activates the auditory nodes corresponding to [i]-like low F1 and high F2 and F3 . Hearing this particular sound [i] also activates some nodes in the phonology layer, namely nodes 4, 6, 7, 8, 9, 12, and 13 (counting at the top layer from left to right). The same phonological nodes are activated upon hearing three other sounds, namely [e], [o], and [u]. This means that the four vowels [i], [e], [o], and [u], all of which differ in their acoustic F1, F2, and/or F3 values, share a discretized specification at the level of phonology. The sound of [a] does not activate any of the phonological nodes that are relevant for [i], [e], [o], and [u], but elicits a unique distinct pattern at the layer of phonology, namely, nodes 1, 2, 3, 5, and 14. This means that the Spanish virtual learner perceives [a] as a one category and she perceives [i]-[e]-[o]-[u] as another category. The vowel /a/ is thus perfectly distinguishable from any of the other four Spanish vowels. At the same time, the adult /i/, /e/, /o/, and /u/, are for the (near-term) Spanish-exposed fetus overlapping and hardly discriminable from each other.

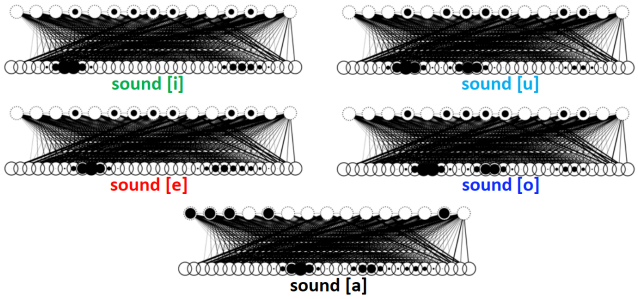


Figure 4: The Spanish-exposed network after learning: perceiving typical realizations of each of the 5 native vowels. Font color-coding of sound as in the scatter plot in Fig. 2.

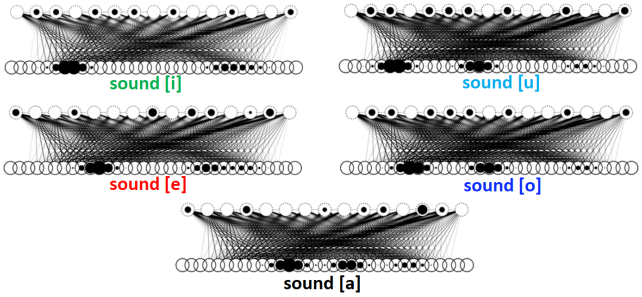


Figure 5: The Czech-exposed network after learning: perceiving typical realizations of each of the 5 native vowels. Font color-coding of sound as in the scatter plot in Fig. 3.

The fetus exposed to Czech comes to warp the auditory signal into slightly different categories. Figure 5 shows that there is a single pattern of phonological activity for the sounds [a] and [e], namely, nodes 1, 4, 8, 10, 11, 13, and 14 (perhaps with slightly varying category goodness ratings as indicated by e.g. node 13 blackening up more for [a] and less for [e] and vice versa for e.g. node 14). Besides that, there is a completely different pattern of phonological activity for [i], [u], and [o], namely, nodes 2, 3, 5, 6, 7, 9, 12, and 15.

Both the Spanish-exposed and the Czech-exposed fetus formed two category-like structures in their phonology. The emerged categories were qualitatively different across the two languages. Three out of the four emerged phonological structures (namely, the Spanish /i-e-o-u/ and the Czech /a-e/ and /i-o-u/) only partially corresponded to how adult speakers

of the language categorize their native speech sounds. One could thus as well call those "proto-categories".

Evaluating the outcomes across the two languages, it seems that learners based their categorization exclusively on the F1 dimension. The Czech fetus created one proto-category for [a]'s and [e]'s because in Czech /a/ and /e/ are relatively close on the F1 dimension. Along the same lines, Czech short /i/, /u/, and /o/ have relatively similar F1 values, which – despite the fact that /i/ is rather far from /u/ and /o/ on the F2 dimension – lead the Czech-exposed fetus to group them into a single proto-category.

Discussion

We simulated prenatal learning of vowel inventories from two languages. Our virtual learners were modeled as two-layer symmetric neural networks where the bottom layer corresponded to the basilar membrane and the upper layer to a more abstract, "phonological", level of representation. These two-layer neural nets had been previously shown to exhibit phonological learning behaviors such as (post-natal) category creation or auditory dispersion (Boersma et al. 2020). To simulate learning in individuals with a relatively immature auditory system, our learners' basilar membrane had a less accurately developed topography than that of previously reported virtual language learners. To simulate learning in utero, before reaching the basilar membrane the input underwent modulations that are typical of the sounds' passing from the external to the intrauterine environment.

Prenatal learning of vowels was assessed for two five-vowel languages that are similar to one another in the number of vowel phonemes that they contrast but slightly differ in how they implement some of the phonological categories acoustically. The languages were Castilian Spanish and (the short phonemes of) standard Czech. A Spanish and a Czech fetus were each trained with 40,000 vowel tokens drawn from their respective (mother's) language, hearing each token's (modulated) first, second, and third formant values. Training was unsupervised, meaning that the learner did not know which vowel category was intended, which is the only learning mechanism plausible *in utero* as the fetus does not have access to any other cues (e.g. visuals on objects that is being talked about) that could inform them on the intended category membership, and neither did they how many categories they should eventually acquire.

Both the Spanish and the Czech learner came to perceptually categorize the continuous auditory world. Their categorization of vowels was not adult like: they did not warp the auditory space into the five adult categories but each arrived at two category-like structures. Interestingly, even the few categories that the learners created were specific to the language that they had been exposed to. The Spanish-exposed fetus created one precategory for the [i-e-o-u] sounds and one for [a]'s. The Czech-exposed fetus created one proto-category for [i-o-u] sounds and one for [a-e] sounds. Given the vowel acoustics in each language, it is apparent that the learners' perceptual categorizations are based on the vowels' first formant (in line with the observations in e.g. Querleu et al., 1988, and McCarthy, Skoruppa, & Iverson, 2019).

The argument that humans may perceptually categorize vowels in language-specific ways (perhaps entirely) on the basis of the vowels' F1 does not align with the results of Moon et al. (2013) who reported language-specific warping effects for vowel differences cued primarily by F2 (and higher formants). An explanation might be that in the tested infants, the warping of F2 occurred within the few days or hours that the newborns had before the experiment was administered. Possibly, the individuals might have first developed a coarse perceptual categorization of F1 *in utero* (as suggested by the present simulations). Once born, they could have promptly started warping also the suddenly well audible and structured F2 (and higher formant) space and exhibit those newly-acquired F2 categorization effects in the experiment. Learning to categorize within several hours or days of exposure is not unlikely since infants can reportedly learn novel speech sound contrasts even after brief, several-minute, exposure (Maye et al., 2008; Wanrooij et al., 2014).

Conclusions and Future Research

We extend on previous modeling work that found (language-specific) bottom-up perceptual categorization of speech sounds through a distributional learning mechanism (such as Guether & Gjaja, 1996; Vallabha et al., 2007). Our simulations suggest that language-specific categorization behavior acquired through unsupervised learning could develop already *in utero*.

The learners in our simulations developed rough rather than precise representations of the ambient language environment. Nevertheless, those coarse perceptual categorizations make predictions for both language general as well as language-specific effects in vowel discrimination (that humans might exhibit at around the time of birth perhaps).

Both virtual individuals came to be able to distinguish /a/ from /i/ and from /u/, which means that /a/-/i/ and /a/-/u/ should be reliably discriminated (and perceived as a between-category difference) by both a Spanish and a Czech near-term fetus or newborn infant. Human fetuses and newborns, as well as non-human animals, indeed discriminate these vowel contrasts (Kujala et al., 2004; Shahidullah & Hepper, 1994; Baru, 1975). Since even animals who are not exposed to human language do so, these discrimination capabilities can be driven by the large acoustic F1 distance between /a/ and the other two corner vowels and not necessarily by perceptual warping.

A more interesting situation occurs for the less salient vowel differences. After learning, the two fetuses differed in how they distinguished, for instance, /a/ from /e/ and /e/ from /o/. A Spanish infant may, at birth, not be able to tell apart /e/ from /o/, unlike the Czech infant who will differentiate the two vowels /e/ and /o/ quite reliably. On the contrary, the Spanish near-term fetus or newborn infant will distinguish between /a/ and /e/ in her parents' speech, while a Czech individual of the same age might have troubles doing so.

Our model needs to be developed further to account for speech sound learning on other dimensions (of which duration is a particularly intriguing one) and of other classes of sounds. We have shown that already with 5-vowel inventories, the model provides for informed hypotheses about language-specific listening in speech perception experiments with newborn infants, who had – to date – been mostly considered universal listeners not perceiving speech sounds in language-specific ways.

Acknowledgments

Funded by Charles University (grant PRIMUS/17/HUM/19) and by Czech Science Foundation (grant 18-01799S).

References

Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. *Brain and Language* 162, 46–59.

Baru, A. (1975). Discrimination of synthesized vowels [a] and [i] with varying parameters (fundamental frequency, intensity, duration and number of formants) in dog. *Auditory Analysis and Perception of Speech*, 91–101.

Boersma, P., Benders, T., & Seinhorst, K. (2020). *Neural networks for phonology and phonetics*. Manuscript. University of Amsterdam. http://www.fon.hum.uva.nl/paul/papers/BoeBenSei46.pdf

Boersma, P., Chládková, K, & Benders, T. (in prep.). Phonological features emerge substance-freely from the phonetics and the morphology. Manuscript in preparation.

Cheour, M., Martynova, O., Näätänen, R., Erkkola, R., Sillanpää, M., Kero, P., … Aaltonen, O. (2002). Psychobiology: Speech sounds learned by sleeping newborns. *Nature*, *415*(6872), 599.

Chládková, K., Černá, M., Paillereau, N., Skarnitzl, R., & Oceláková, Z. (2019). Prenatal infant-directed speech: vowels and voice quality. *Proceedings of the 19th ICPhS*, 1525–1529.

Chládková, K., Escudero, P., & Boersma, P. (2011). Context-specific acoustic differences between Peruvian and Iberian Spanish vowels. *Journal of the Acoustical Society of America*, 130, 416-428.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*(2), 133–150.

Gerhard, K.J., & Abrams, R.M. (1996). Fetal hearing: characterization of the stimulus and response. *Seminars in Perinatology*, 20, 11-20.

Granier-Deferre, C., Ribeiro, A., Jacquet, A., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science*, *14*(2), 336–352.

Guenther, F.H., & Gjaja, M.N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, *100*, 1111-1121.

Kisilevsky, B. S., Hains, S. M., Brown, C. A., Lee, C. T., Cowperthwaite, B., Stutzman, S. S., … Huang, H. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development*, *32*(1), 59–71.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kujala, A., Huotilainen, M., Hotakainen, M., Lennes, M., Parkkonen, L., Fellman, V., & Näätänen, R. (2004). Speech-sound discrimination in neonates as measured with MEG. *NeuroReport*, *15*(13).

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111.

McCarthy, K.M., Skoruppa, K. & Iverson, P. (2019). Development of neural perceptual vowel spaces during the first year of life. *Scientific Reports, 9*, 19592.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178.

Moon, C., Cooper, R. P., & Fifer, W.P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development* 16, 495–500.

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica*, *102*(2), 156–160.

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences*, *110*(37), 15145–15150.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crèpin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *28*(3), 191–212.

Richards D.S., Frentzen B., Gerhardt K.J., McCann M.E., & Abrams R.M. (1992). Sound levels in the human uterus. *Obstetrics and Gynecology*, 80(2), 186–190.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science,* 9, 75–112.

Seebach, B. S., Intrator, N., Lieberman, P. & Cooper, L. N. (1994). A model of prenatal acquisition of speech parameters. *Proceedings of the National Academy of Sciences, 104,* 13273–13278.

Shahidullah, S., & Hepper, P. G. (1994). Frequency discrimination by the fetus. *Early Human Development*, *36*(1), 13–26.

Vallabha G.K., McClelland J.L., Pons F., Werker J.F., Amano S. (2007) Unsupervised learning of vowel categories fom infant-directed speech. *Proceedings of the National Academy of Sciences, 104,* 13273–13278.

Wanrooij, K., Boersma, P., & Van Zuijen, T. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology*, *5*, 77.