Learning vowels before and after birth: A simulation of how prenatal input may affect perceptual attunement throughout infancy*

Kateřina Chládková
Charles University and the Czech Academy of Sciences, Prague

How can discrete representations of speech sounds emerge from continuous phonetic experience? In an earlier study, we asked whether category emergence for speech sounds can begin before birth. Here I ask what happens right after birth, when the learner encounters a much richer speech signal than before, and whether it matters that the learner has experienced prenatal speech at all. I extend our earlier two-layer neural-network to simulate virtual infants exposed to prenatal and postnatal Czech and Spanish vowels with those that begin learning only after birth. Comparing similarities at the learners' "phonology" layers suggest that prenatal learning may affect subsequent perceptual narrowing, with the prenatally trained network generally developing more differentiated vowel representations compared to the network in which learning does not start until birth. These effects seem to be at least to some extent contrast-specific, and some developmental trajectories suggest U-shaped patterns: some vowels sometimes become similar before differentiating again. I then consider how these simulation outcomes relate to what we know about the development of speech perception in real infants and how the virtual learners could generate predictions about vowel perception in real babies with different prenatal auditory experiences.

* One of the many things I learned from Paul Boersma, my dear PhD promotor, is that questions about language acquisition can be approached by trying to make a computer learn and inspecting what and how it has learnt. Rather than assuming that linguistic representations are available to the learner from the outset, or become given to her at any point in development by the parent (or any other kind of supervisor), one can ask whether, and under what conditions, they emerge from experience. This approach has inspired much of my own research, starting with chapter 5 of my PhD dissertation, and continuing through our subsequent (though not as frequent as I would wish for) collaborations and my (frequent but for the most part not executed) thinking about virtual babies and language learning simulations. At a *voortgangsgesprek* ('progress interview') sometime halfway through my PhD, when asked where I saw myself in five years, I said I would have a babylab in Prague and study language acquisition in newborns. It took a little longer than five years, but it did happen. Paul played an important role in getting me there. In this contribution, I bring together language learning in the youngest possible human participants and in virtual learners. I have learned from Paul not only how to build and interpret computational models (well, a very small part of it), but also how to use them to question assumptions that might otherwise remain unexamined. Much of this learning happened rather implicitly, and perhaps even in a relatively unsupervised fashion. And what I learned has proven remarkably persistent.

Linguistics in Amsterdam, 2026: 151–166
© Kateřina Chládková

1 Introduction

A central question in Paul Boersma's work has been how language-specific speech sound categories emerge from continuous phonetic input, what these categories are, how they change with further learning, and how such changes may accumulate across generations and transform between languages (Boersma 1998; Boersma & Hamann 2009; Boersma 2012; Boersma et al. 2020; Boersma et al. 2022). Human learners are not provided with an inventory of phonological categories, nor are the boundaries between categories marked in the speech signal. And yet, during development, children come to perceive the continuous acoustic variation in their linguistic environment in increasingly language-specific ways (Werker & Tees 1984).

But when does this process begin? The acquisition of language-specific speech sound categories is typically investigated in infants during their first year or two after birth. Most studies have focused on when infants reach speech perception capabilities comparable to those of adults (if such a comparison is at all possible), rather than on how this process begins. For instance, the widely cited developmental timeline of Kuhl (2004) starts at birth, and even Roy's ambitious recordings of virtually the entire early language environment of one child, started only when the baby came home from the hospital (Roy et al. 2015). Earlier theoretical accounts considered different possible developmental trajectories of speech perception (Aslin & Pisoni 1980), but the initial state of the learner has remained somewhat obscure and difficult to establish.

While we know quite a bit about when language-specific perception develops, we know considerably less about its beginnings. Even newborns may already be to some extent experienced listeners, and the initial stages of language-specific perceptual learning may have to be sought before birth. During late gestation, the auditory system is sufficiently developed to support learning from speech (Urbanec et al. 2022), in line with which newborns show remarkable sensitivity to properties of their native language (Dvořáková et al. 2026). Studying the very beginnings of language development in actual fetuses is, by all means, challenging (though certainly not impossible). So, before we jump into fetal neuroimaging, why not first see what virtual fetuses can learn?

This is what we did in Chládková, Nudga, and Boersma (2020). We asked whether phonological category formation could begin before birth, and simulated two virtual fetuses, one exposed to Czech and the other to Spanish vowels. The acoustic input approximated the speech signal available in the intrauterine environment, where higher-frequency information is substantially attenuated (Richards et al. 1992), and the learners' auditory resolution was relatively immature, as approximated by broader spreading of auditory activation in the network. Both learners developed broad vowel proto-categories, primarily on the

basis of first-formant information. Importantly, although neither learner acquired five adult-like vowel categories, the resulting representations differed between the two languages. Thus, even acoustically impoverished prenatal speech appeared sufficient for the emergence of some language-specific phonological structure. The categories were not adult-like, but neither were they arbitrary: they reflected properties of the language to which the learner had been exposed.

The learners in our 2020 study, however, stopped at birth. This leaves an interesting question unanswered. What happens to the initially formed vowel proto-categories when the learner subsequently encounters the acoustically richer speech signal available after birth? And, relatedly, does prenatal learning have any effect on the subsequent development of adult-like vowel representations?

One possibility is that prenatal learning provides a head start. If a learner already has some organization of the vowel space, subsequent learning might proceed more efficiently than in a learner who starts without prior experience. Alternatively, the broad proto-categories established prenatally might hinder the differentiation of vowels that initially fell within a single proto-category. A third possibility is that postnatal input eventually overrides the effects of prenatal experience, so that learners with different initial states arrive at similar representations.

To get a first impression of the possible effects of prenatal learning, I extend our earlier two-layer neural-network model to simulate the acquisition of Czech and Spanish vowels during postnatal development. For each language, I compare a virtual learner who has already experienced prenatal speech with one whose exposure to speech begins only at birth. I am particularly interested in what happens to vowels that were grouped into the same proto-category during prenatal learning: these were /a e/ versus /i o u/ in Czech, and /a/ versus /i e o u/ in Spanish. I therefore inspect the developmental trajectories of all vowel contrasts over 18 simulated postnatal months, looking not only for lasting effects of prenatal experience, but also for clues about which learning mechanisms might be needed for the eventual emergence of adult-like phonological representations.

2 Simulations

2.1 Network architecture

The network architecture, learning algorithm, and general parameter settings are retained from the prenatal model (Chládková et al. 2020). The neural network has two layers: a bottom, auditory layer with 33 nodes representing frequencies between 4 and 28 ERB in steps of 0.75 ERB, and an upper, phonology layer containing 15 nodes; as shown in Figure 1. I call the upper layer ‘Phonology’ to

indicate that it represents a layer with *discretized* representations[1], in contrast to the 'Auditory' layer that represents continuous psychoacoustic dimensions of resonating frequencies. The connection weights between the auditory and the phonology layer are free to update during learning using the *inoutstar* learning algorithm (Boersma et al. 2020). Learning is unsupervised: the network receives auditory input without information about the identity or number of vowel categories.

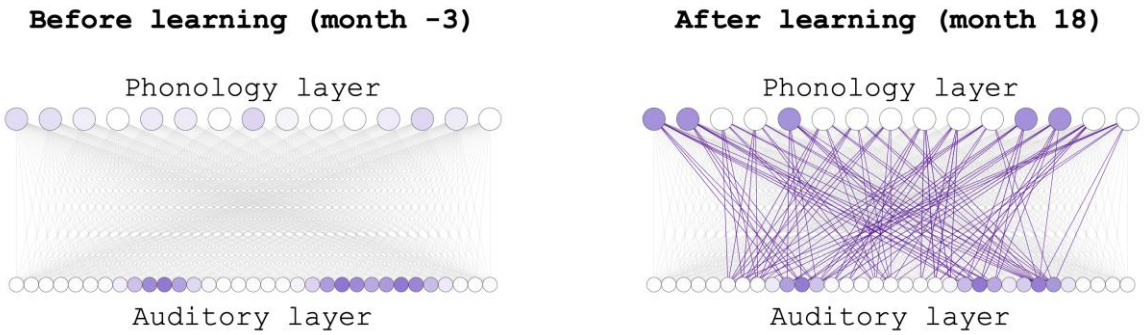


Figure 1: The two-layer network. Before learning (left), all connections between the auditory and the phonology layer have random low weights, shown by thin grey lines. After learning (right), some connections have large positive weights, shown by purple lines, such that a specific auditory input (in this case, the three formants of Czech /e/, specifically, the unfiltered postnatal version of this vowel) activates a distinct set of nodes at the phonology layer. Notice the wider prenatal vs. the narrower postnatal spreading at the auditory layer.

The present simulations extend the duration of learning beyond birth, which means that there are more learning steps in total and that, at birth, the prenatal filter is removed, changing the acoustic characteristics of the input compared to the prenatal period.

The network employed here is relatively simple – if not for the prenatal learner, then definitely for the developing postnatal learner, for whom we know that lexical knowledge begins to emerge by around 6 months of age, and probably earlier (Bergelson & Swingley 2012). Rather than introducing a more elaborate architecture, I retain the architecture of the original model for reasons of parsimony. I believe this is a necessary first step if we want to isolate the consequences of prenatal learning history: in this way, we can ask whether differences in developmental experience alone are sufficient to produce different phonological representations during subsequent learning.

[1] Calling the upper layer 'Phonology' is not meant to indicate any specific level of phonological abstraction, and it could as well be called a lexicon of phonetic categories.

2.2 Input and learning conditions

I simulate the acquisition of the five monophthongal short vowels /i e a o u/ in Czech and Spanish (in Czech, the first two vowels are in fact /ɪ/ and /ɛ/, for simplicity, the same symbols are used here for both languages). The vowels represent the five short monophthongs of each of the two languages, and their distributions were defined by the means and between-speaker standard deviations of F1, F2, and F3 measured in female speakers of Castilian Spanish and standard Czech (Chládková et al. 2011, 2019); see Table 1.

Table 1: The first three formants used to define the input distributions for the simulations. Values are in ERB, table is showing means and standard deviations (in parentheses).

Language	Formant	/i/	/e/	/a/	/o/	/u/
Spanish	F1	8.88 (0.54)	10.67 (0.47)	13.63 (0.31)	11.12 (0.41)	9.35 (0.37)
	F2	22.97 (0.52)	21.57 (0.50)	19.46 (0.51)	16.36 (0.49)	14.67 (0.55)
	F3	24.41 (0.45)	23.99 (0.49)	23.54 (0.57)	23.17 (1.74)	23.91 (0.63)
Czech	F1	8.93 (0.80)	11.34 (0.87)	12.99 (1.06)	9.47 (1.25)	8.15 (0.49)
	F2	21.99 (0.93)	20.58 (0.42)	18.00 (1.18)	15.86 (0.98)	15.14 (1.31)
	F3	24.09 (0.60)	23.74 (0.46)	23.12 (1.43)	23.94 (0.71)	23.94 (0.71)

For each language, two learning conditions are implemented. In the *prenatal-plus-postnatal* condition, the network first receives three simulated months of prenatal input. Prenatal speech is acoustically filtered to approximate the signal available in the intrauterine environment, using the following frequency-dependent attenuation function, which is based on the *in-utero* acoustic measurements by Richards et al. (1992):

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

where *f* is frequency in Hz and *p(f)* is the frequency-dependent amplitude factor. At birth, prenatal attenuation is removed, and the network continues learning from unfiltered vowel input, which means that the postnatal *p(*f) = 1. In the *postnatal-only* condition, the network begins learning at birth, without previous exposure to prenatally filtered speech.

Prenatally, a parameter called auditory spreading is set to 1.36, simulating immature auditory frequency resolution at the basilar membrane (this parameter is 1.5 times larger here than in Boersma et al. 2020). The auditory spreading gradually narrows during postnatal development, following a linear function:

$$s(m) = s_0 + \frac{m}{18}(s_{18} - s_0)$$

where m denotes postnatal age in simulated months, and $s(m)$ is the auditory spreading parameter; here, $s_0 = 1.36$ and $s_{18} = 0.907$. During each month's learning, spreading is set to the value at the midpoint of that month.

Each month contains 20,000 learning steps. In each step, one vowel is randomly selected from the five vowels of the Czech-based or Spanish-based exposure language, its F1, F2, and F3 values are sampled from the corresponding distributions, and input at the auditory layer. The resulting auditory activation is calculated as the sum of three Gaussian activation patterns, centered on the sampled formant frequencies:

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

where p_{F1}, p_{F2}, and p_{F3} are the intrauterine attenuation or enhancement factors for the three formants defined above, s is auditory spreading, k is the auditory-node index (1–33), and $F1$, $F2$, *and* $F3$ are the formant frequencies expressed on the auditory-node scale. After the auditory input is presented, activation spreads through the network over 500 iterations, following the formula:

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

where where η_a is the spreading rate (here 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. The connection weights are updated using the *inoutstar* learning algorithm (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 η_w is the learning rate, here $\eta_w = 0.001$, a_i and a_j are the activities of the connected nodes, and the *instar* and *outstar* contributions are equal (*instar* = *outstar* = *weightLeak* = 0.5 each). This means that the connection is strengthened when both nodes are active, weakened when only one of the two nodes is active, and slightly weakened when neither of the two connected nodes is active. After each learning step, the connection weights between the auditory and phonology layers are normalized:

$$w_{ij\ norm} = 0.1 \times n \frac{w_{ij}}{\sum_{j=1}^{33} w_{ij}}$$

where w_{ij} is the connection weight from auditory node i to phonological node j, and n is the total number of auditory nodes.

2.3 Evaluation

Both conditions are evaluated before learning and at the end of each month over the simulated period, that is, from prenatal month –3 through postnatal month +18 for the prenatal-plus-postnatal condition[2], and from month 0 through postnatal month +18 for the postnatal-only condition. The evaluation simulates how the learner perceives five auditory F1–F2–F3 tokens corresponding, respectively, to mean acoustic representations of the five selected vowels in each language. The same unfiltered auditory vowel tokens are used in both conditions and at all developmental stages. The activation patterns in the phonological layer are taken as a proxy of phonological representation status and are plotted in Figure 2.

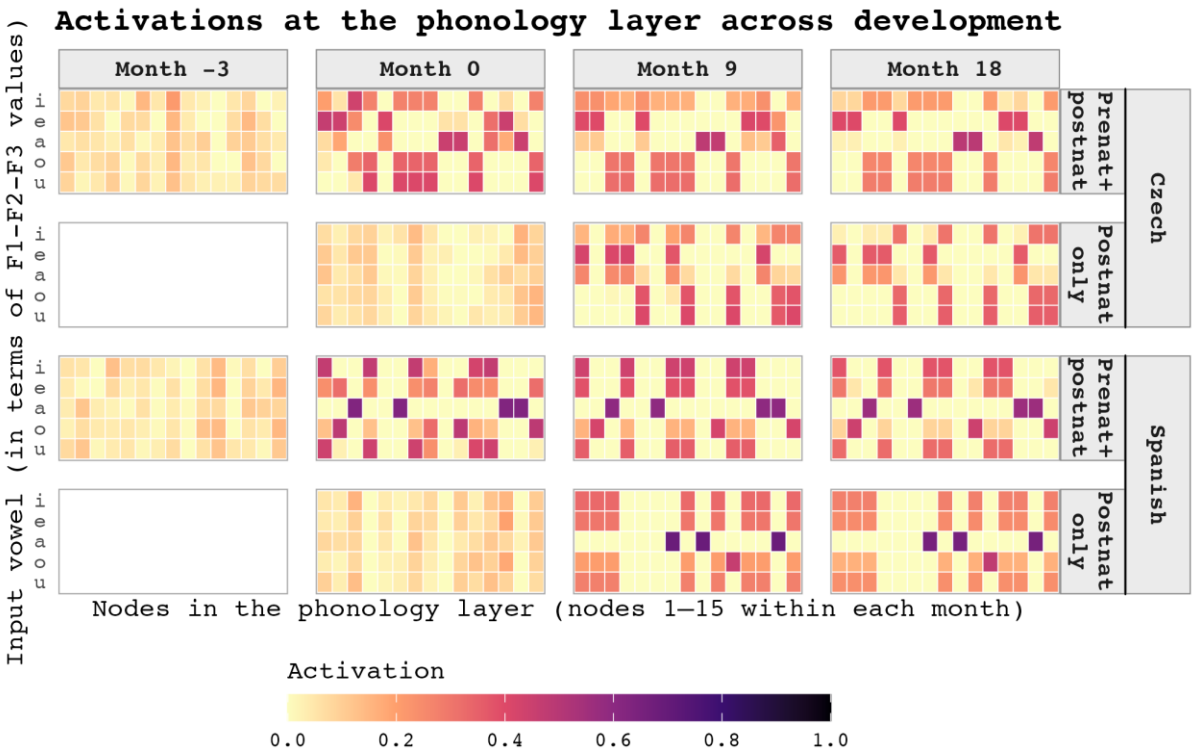


Figure 2: Tile plots of simulated perception of prototypical F1-F2-F3 tokens per vowel in terms of activation patterns across the 15 nodes in the phonology layer. Fill indicates activation strength; similar activation patterns between vowels suggest less differentiated phonological representations.

[2] Evaluating both networks at month +18 means that the prenatal-plus-postnatal babies had three more months of learning than the postnatal-only babies. Crucially, all of the differences discussed below show the same patterns if one considers the prenatal-plus-postnatal simulations at month +15.

As in Boersma et al. (2022), similarity between the vowels' phonological representations is calculated as the cosine similarity between the activation vectors elicited by each pair of vowels. A cosine similarity of 1 indicates identical activation-vector directions, whereas lower values indicate more differentiated activation patterns. I calculate pairwise similarities for all ten possible vowel pairs and plot them in Figure 3 to inspect how they change during development. The overall similarity at the phonology layer, i.e., the mean similarity across the ten vowel pairs, is shown in Figure 4. Changes in similarity are interpreted here as changes in the organization of phonological representations. The assumption here is that high similarity in the vowels' phonological representation is associated with their poor perceptual discrimination and, likewise, low similarity is associated with good perceptual discrimination of the vowel contrast at hand.

3 Inspecting and discussing the learning outcomes

3.1 Does prenatal learning affect subsequent phonological development?

My question was whether virtual learners with prenatal experience would develop differently from learners that received postnatal input alone. Inspecting the per-vowel-pair and the mean curves in Figures 3 and 4, respectively, suggests that the two learning conditions may lead to different developmental trajectories in both languages.

At birth, the prenatal-plus-postnatal networks already exhibit relatively differentiated phonological activation patterns. The mean cosine similarity across the ten vowel pairs is 0.374 for Czech and 0.416 for Spanish. As a sanity check – and unsurprisingly – the corresponding values in the postnatal-only learners, who had not yet received any learning input, are rather high in both languages (0.904 and 0.891, respectively, for Czech and Spanish).

A perhaps more intriguing question is whether the difference persists as both networks per language learn after birth. The overall activation patterns gradually become more differentiated: at month 18, the mean phonological similarity is 0.330 in the Czech prenatal-plus-postnatal network and 0.480 in the Czech postnatal-only network. The corresponding values for Spanish are 0.390 versus 0.508. In both languages, prenatal experience thus seems to be associated with lower overall phonological similarity even after prolonged exposure to the same postnatal input, as far as data from our four virtual babies show. The representations established before birth appear to leave a trace in the subsequent organization of the phonological layer.[3]

[3] Whether this is an overestimation produced by the lack of a lexicon or other layers remains to be tested in future work, as I discuss in more detail in Section 3.4.

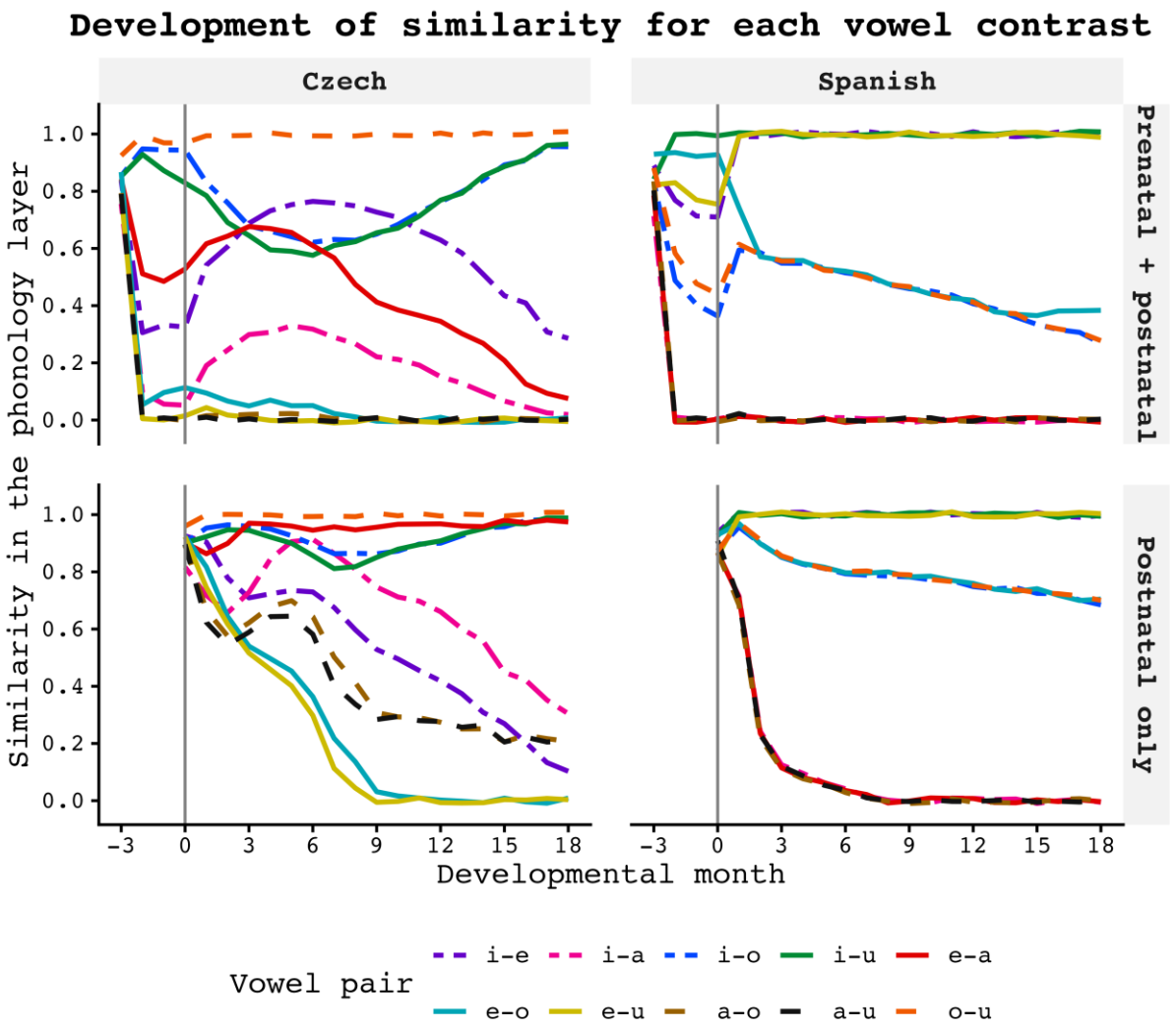


Figure 3: Development of vowel similarity at the phonology layer. Curves show cosine similarity in phonological activation for all ten vowel pairs across months. Higher similarity indicates more overlap within a vowel contrast, whereas lower similarity indicates greater differentiation. Slight vertical jitter (= 0.01) was added to the individual curves to improve visibility of overlapping trajectories.

Could such an effect also be observed in real infants? One interesting population is infants born preterm, whose prenatal auditory experience is interrupted and replaced by an acoustically different, one could say richer, extrauterine environment. When matched to full-term infants on corrected age, i.e., age calculated from the expected rather than the actual birth date, infants born prematurely do not seem to benefit from their longer extrauterine listening experience with the acoustically rich speech signal (Jansson-Verkasalo et al. 2010; Peña et al. 2012). This might partly be because prenatal auditory history matters for subsequent perceptual attunement, although prematurity involves many developmental differences besides the duration of prenatal speech exposure. Another interesting population would be full-term infants born to deaf parents, particularly deaf mothers who communicate primarily through sign language.

These infants may have had less systematic exposure to maternal speech before birth, despite having typical auditory development and access to speech from other people in their environment. The present simulations offer a way to isolate prenatal exposure as a factor and generate predictions for future experiments comparing infants with different prenatal auditory histories.

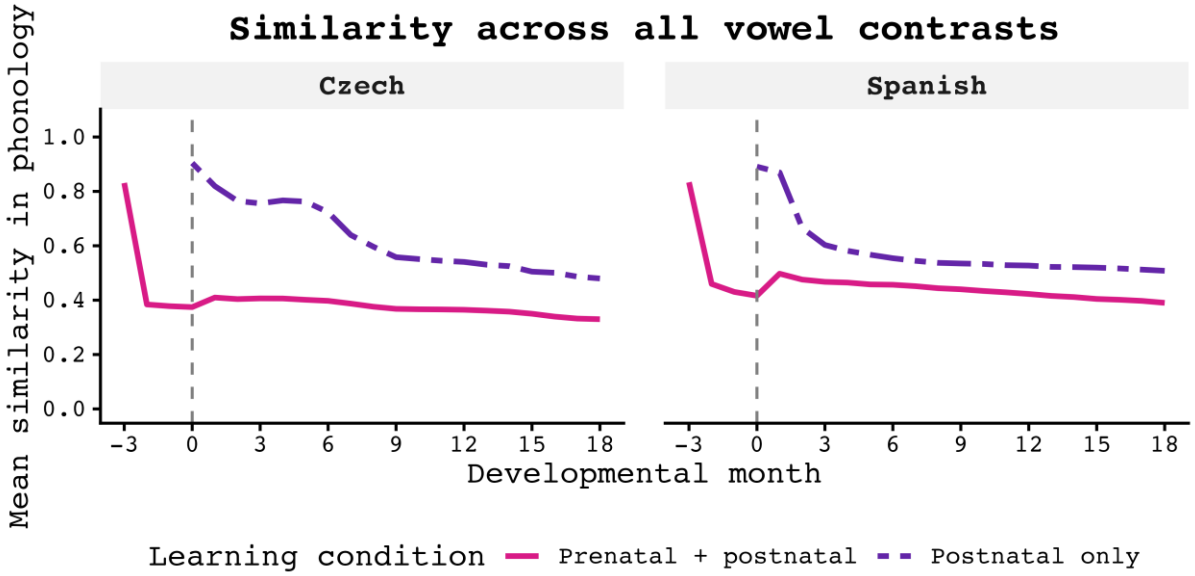


Figure 4: Mean similarity in the phonology layer across development. Curves represent mean cosine similarity across all ten vowel pairs at each developmental month, per language and condition. Lower similarity indicates greater mean differentiation at the phonology layer.

3.2 What happens to vowels that share a prenatal proto-category?

Averaging across vowel pairs may conceal important contrast-specific differences. In Czech, for example, prenatal learning substantially affected the differentiation of /a/ and /e/, even though these vowels formed a single prenatal proto-category, both here and in our 2020 simulations. At birth, the phonological representations of these vowels have a cosine similarity of 0.525 in the prenatal-plus-postnatal network (and, as seen in Figure 2, the two vowels activate identical nodes, which is why we called them a single proto-category in the 2020 paper), compared with 0.904 in the postnatal-only network. After nine months of postnatal learning, the similarity of /a/ and /e/ decreases to 0.410 in the prenatally trained network but *increases* to 0.957 in the postnatal-only network. At month 18, the two vowels' representations are substantially more differentiated in the prenatal-plus-postnatal network (0.082) than in the postnatal-only network (0.978).

It might seem counterintuitive that having grouped tokens of [a] and [e] into a single proto-category before birth helps the learner distinguish /a/ from /e/ later on. One might have expected precisely the opposite, namely, that merging two vowels early in development would make their subsequent differentiation more

difficult. It should be noted, however, that not all Czech vowels show this pattern. For example, /o/ and /u/ which were also collapsed into a single proto-category before birth (along with /i/), remain virtually identical at the phonology layer in both conditions up to month 18, with cosine similarity approaching 1.

In Spanish, prenatal experience also resulted in stronger differentiation of certain contrasts. At month 18, phonological similarity between Spanish /e/ and Spanish /o/ is 0.376 in the prenatal-plus-postnatal network, compared with 0.696 in the postnatal-only network (see Figure 2). Similarly, /i/ and /o/ have a similarity of 0.269 in the prenatally trained network and 0.692 in the postnatal-only one. Again, note also here that [e]s, [o]s, and [i]s (along with [u]s) all belonged to a single prenatal proto-category, yet some of these vowels become less similar precisely in the prenatally exposed learner. In contrast, the representations of Spanish /i/ and /u/ remain virtually identical in both learning conditions.

The simulated difficulty with /i/–/u/ has a possible parallel in infant word learning in the real world. Curtin et al. (2009) found that 15-month-old Canadian English-learning infants succeeded in learning words distinguished by /i/ versus /ɪ/, but not words distinguished by /i/ versus /u/ or /ɪ/ versus /u/ (which are both relatively large contrasts in F2, although /u/ is somewhat more front in English than it is in Spanish or Czech). This does not necessarily mean that the infants could not discriminate the vowels in a non-lexical task, but it shows that some vowel differences are more difficult to use in word learning than others (in this specific case with the F1 cue perhaps being relied on more strongly than the F2 cue).

Why would prenatal learning facilitate the differentiation of some vowels that initially belonged to the same proto-category (but perhaps not so much the learning of other vowels)? One possibility concerns the acoustic information that survives prenatal filtering. In Chládková et al. (2020), the prenatal proto-categories emerged largely on the basis of F1 which is not as attenuated by the *in-utero* filter as higher formants. Vowels within a proto-category were therefore relatively similar in F1, though not necessarily identical. Perhaps creating a broad category before birth is beneficial for subsequent learning when its members already differ slightly in the acoustic dimension that helped establish that category (like /a/ and /e/ did in our Czech-inspired exposure language).

This bears some resemblance to findings from adult L2 learning, where the effects of distributional exposure depend on prior language experience. Learning a new vowel contrast may initially involve becoming more sensitive to variation along the relevant acoustic dimension, without yet establishing separate categories. This increased acuity may then provide the basis for subsequent category formation (Chládková & Šimáčková 2021). Similarly, prenatal learning

might help[4] organize the acoustic space in a way that facilitates the later differentiation of vowels, even if those vowels initially belong to the same proto-category.

3.3 U-shaped development

Some of the trajectories simulated here suggest that learning may not necessarily be monotonic, especially in the Czech prenatal-plus-postnatal condition (see Figure 2). For the contrasts /i/–/a/ and /i/–/e/, phonological similarity appears to increase during the early postnatal months, peaking around months 5 and 6, respectively, before decreasing again. This means that vowels that initially elicited relatively distinct phonological activation patterns temporarily become more similar before differentiating again.

Why might this happen? One possibility is that the introduction of acoustically richer postnatal input reorganizes the representations established during prenatal learning. The network may temporarily develop more overlapping activation patterns as its connection weights adapt to the new input distributions. Interestingly, this temporary *increase* in the similarity of Czech /i/–/a/ and /i/–/e/ is accompanied by a *decrease* in the similarity of /i/–/o/ and /i/–/u/. The latter two contrasts subsequently become more similar again. It seems that the network is reorganizing relationships among vowels rather than progressively separating all vowel representations from one another.

Whether these non-monotonic changes in phonological similarity correspond to changes in perceptual discrimination in the real world would need to be tested directly. The literature to date suggests that developmental changes in infant vowel perception are not always straightforward, and in some populations of infants or speech sounds they may indeed follow U-shaped trajectories (Byers-Heinlein & Fennel 2014; Götz et al. 2018).

3.4 Learning phonology will require the lexicon, too

One important caveat must be flagged. The present model learns exclusively from auditory information, but this is not how actual human babies learn in real life. During the first year after birth, infants increasingly acquire knowledge about words – their forms and their meanings. This additional information may change the way speech sound contrasts are represented. Boersma et al. (2003) distinguished an initial stage of auditory-driven category learning from a subsequent stage in which lexical representations contribute to the refinement of

[4] A reviewer suggests that this implies that F1 will be a better learnable acoustic dimension than F2 for two reasons: not only because (1) humans have better frequency resolution in lower frequencies, but also because (2) the lower frequencies, i.e., F1, represented speech input and intake already prenatally whereas F2 probably did not.

phonological categories. Could the lexicon help learners establish contrasts that remain poorly differentiated even after long months of auditory learning?

Consider Czech /o/–/u/ and Spanish /i/–/u/, whose phonological activation patterns remain almost identical in both learning conditions. One important source of additional information after birth would be visual speech cues, particularly articulatory information from the speaker's face (see Seinhorst 2026), which are absent from the present auditory-only model. Another potential source is lexical information. If the learner acquired words in which these vowels distinguish meanings (e.g., *kop* 'kick-N' vs. *kup* 'buy.IMP.SG' and *piso* 'floor/apartment' vs. *puso* 'put.PST.3SG', in Czech and Spanish, respectively), the emerging lexical representations might provide an additional source of information for separating them. This would turn the perhaps more intuitive learning trajectory partly around: rather than relying only on previously established vowel categories to learn words, the learner could use knowledge of words to help her discover vowel contrasts that auditory learning alone has not established. Studies with real-life language users provide evidence that lexical knowledge can shape phonetic categorization (e.g., Norris et al. 2003), although whether it can support the initial establishment of such contrasts remains an open question.

This idea also connects directly to our earlier joint work, reported in Boersma et al. (2022). Simulations of a five-vowel toy language with a multi-layer neural network showed how interactions between bottom-up phonetic information and top-down morphological information could lead to the emergence of abstract phonological structure. The present model does not include lexical or morphological representations, but the resulting developmental trajectories call for such an additional learning opportunity (most loudly in the case of Czech /o/–/u/ and Spanish /i/–/u/), the modelling of which needs to be done in our follow-up work.

It is possible that, even without modelling top-down effects of an emerging lexicon, simply adding intermediate layers would allow the network to encode phonological abstraction differently. In Boersma (2019), for example, discrete categorical behaviour could emerge through unsupervised distributional learning in a three-layer deep Boltzmann machine. That model differed from the present network not only in its architecture but also in its learning mechanism, employing Hebbian learning and anti-Hebbian unlearning. Future work on the effects of prenatal exposure will need to explore these possibilities, too.

4 From virtual to real babies

One of the advantages of computational models is that they allow us to manipulate aspects of developmental experience that cannot easily be separated when studying human infants. In the present simulations, the two learning conditions

differ in whether they receive prenatal input, while their postnatal experience is held constant.

Beyond the obvious language-specific predictions for infants learning different languages, the present simulations lead to a testable hypothesis that infants acquiring the same language will develop differently depending on the amount and quality of speech experience available before birth. Think, for instance, of a hearing infant born to a deaf mother. Depending on the mother's communication practices and the surrounding environment, this baby may receive different amounts and types of prenatal speech input than a baby whose mother communicates through spoken language. The model can also inform predictions for scenarios when an infant is deprived of auditory input altogether due to congenital hearing loss and their hearing is going to be restored sometime in infancy (via cochlear implantation or another method) such that they can learn to speak (if their families wish them to). Think of international adoptees, where prenatal input is not lacking but it is from a different language than postnatal input. These are all rather small and highly sensitive populations. That does not mean we should not work with them but we need to plan what and when exactly we want to measure. In this respect, computational simulations could provide a useful bridge between theoretical accounts of phonological acquisition and empirical research on early speech and language development, including across a wide variety of populations and languages.

This study is a modest extension of our earlier work. It asks whether the speech input a learner experiences before birth – the commonly assumed starting point of language development – matters for subsequent phonological acquisition. The results suggest that it indeed may: in the present simulations, prenatal experience leaves traces in speech sound representations that persist throughout postnatal learning.

5 References

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology, Volume 2: Perception* (pp. 67–96). Academic Press.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences 109*(9): 3253–3258. https://doi.org/10.1073/pnas.1113380109

Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. PhD dissertation, University of Amsterdam. https://dare.uva.nl/record/1/143963

Boersma, P., Escudero, P., & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound

categories. *Proceedings of the 15th International Congress of Phonetic Sciences*. https://www.fon.hum.uva.nl/archive/2003/icphs/BoersmaEscuderoHayesICPhS2003.pdf

Boersma, P., & Hamann, S. (2009). Loanword adaptation as first-language phonological perception. In A. Calabrese & W. L. Wetzels (Eds.), *Loan phonology* (pp. 11–58). John Benjamins. https://doi.org/10.1075/cilt.307.02boe

Boersma, P. (2012). Modelling phonological category learning. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 207–218). Oxford University Press. https://pure.uva.nl/ws/files/1610467/117123_CategoryEmergence11.pdf

Boersma, P. (2019). Simulated distributional learning in deep Boltzmann machines leads to the emergence of discrete categories. *Proceedings of the 19th International Congress of Phonetic Sciences*, 1520–1524. https://www.fon.hum.uva.nl/archive/2019/icphs/2019-icphs-Boersma.pdf

Boersma, P., Benders, T., & Seinhorst, K. (2020). Neural network models for phonology and phonetics. *Journal of Language Modelling 8*(1): 103–177. https://doi.org/10.15398/jlm.v8i1.224

Boersma, P., Chládková, K., & Benders, T. (2022). Phonological features emerge substance-freely from the phonetics and the morphology. *Canadian Journal of Linguistics/Revue canadienne de linguistique 67*(4): 611–669. https://doi.org/10.1017/cnj.2022.39

Byers-Heinlein, K., & Fennell, C. T. (2014). Perceptual narrowing in the context of increased variation: Insights from bilingual infants. *Developmental Psychobiology 56*(2): 274-291. https://doi.org/10.1002/dev.21167

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*(1): 416–428. https://doi.org/10.1121/1.3592242

Chládková, K., Černá, M., Paillereau, N., Skarnitzl, R., & Oceláková, Z. (2019). Prenatal infant-directed speech: Vowels and voice quality. *Proceedings of the 19th International Congress of Phonetic Sciences*, 1525–1529. https://fonetika.ff.cuni.cz/wp-content/uploads/sites/104/2019/08/Chl-etal19-prenatal-IDS.pdf

Chládková, K., Nudga, N., & Boersma, P. (2020). A model of prenatal acquisition of vowels. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*. https://escholarship.org/uc/item/1nw4k415

Chládková, K., & Šimáčková, Š. (2021). Distributional learning of speech sounds: An exploratory study into the effects of prior language experience. *Language Learning 71*(1): 131–161. https://doi.org/10.1111/lang.12432

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science 12*(5): 725–731. https://doi.org/10.1111/j.1467-7687.2009.00814.x

Dvořáková, M., Kikoťová, N., Urbanec, J., Goetz, A., & Chládková, K. (2026). Newborns' brains detect foreign rhythm in native-language speech. *Developmental Science 29*(5): e70254. https://doi.org/10.1111/desc.70254

Götz, A., Yeung, H. H., Krasotkina, A., Schwarzer, G., & Höhle, B. (2018). Perceptual reorganization of lexical tones: Effects of age and experimental procedure. *Frontiers in Psychology 9*: 477. https://doi.org/10.3389/fpsyg.2018.00477

Jansson-Verkasalo, E., Ruusuvirta, T., Huotilainen, M., Alku, P., Kushnerenko, E., Suominen, K., ... & Hallman, M. (2010). Atypical perceptual narrowing in prematurely born infants is associated with compromised language acquisition at 2 years of age. *BMC Neuroscience 11*(1): 88. https://doi.org/10.1186/1471-2202-11-88

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

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology 47*(2): 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Peña, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. *The Journal of Neuroscience 32*(33): 11159–11163. https://doi.org/10.1523/JNEUROSCI.6516-11.2012

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. https://www.ovid.com/jnls/greenjournal/abstract/00006250-199208000-00006~sound-levels-in-the-human-uterus

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences 112*(41): 12663–12668. https://doi.org/10.1073/pnas.1419773112

Seinhorst, K. (2026). The McGurk effect in symmetric neural networks. *Linguistics in Amsterdam 17*(2): XX–XX.

Urbanec, J., Kremláček, J., Chládková, K., & Skálová, S. (2022). The development and neurophysiological assessment of newborn auditory cognition: A review of findings and their application. *Acta Medica 65*(1): 1–7. https://doi.org/10.14712/18059694.2022.9

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development 7*(1): 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Kateřina Chládková

katerina.chladkova@ff.cuni.cz

Charles University & Czech Academy of Sciences, Voršilská 1, 110 00 Praha