INFANTS' LEARNING OF NOVEL SEGMENTS IS MODULATED BY PROSODY

Kateřina Chládková[a,b], Václav Jonáš Podlipský[c], Natalia Nudga[a,b], Nikola Paillereau[a,b], Kateřina Kynčlová[a,b], Šárka Šimáčková[c]

[a]Faculty of Arts, Charles University, [b]Institute of Psychology, Czech Academy of Sciences, [c]Faculty of Arts, Palacký University Olomouc

katerina.chladkova@ff.cuni.cz

ABSTRACT

Young infants recognize atypical realisations of native-language speech. Later they learn words better from native-accented talkers. However, 6-month-olds preferentially listen to unfamiliar speech. We tested whether the learning of new vowels matches 6-month-olds' listening preferences, being more effective from nonnative-accented speech. We exposed Czech six-month-olds to delexicalised utterances with consonants replaced by [f] and vowels by 405 tokens sampled from a bimodal [ɛ]-[æ] distribution, a contrast absent from Czech, and with either native or atypical rhythm. Discrimination of [ɛ]-[æ] was then tested in an alternating/non-alternating paradigm. Longer first-look duration to non-alternating than to alternating trials – indicating a learning effect – was found in infants familiarised with the novel contrast in atypical rhythm; such effect was not detected after familiarisation with native rhythm. Six-month-olds thus more effectively exploit distributional information about novel vowels from non-native rhythm, which matches their previously reported preferences for listening to novel over familiar accents.

Keywords: distributional learning, selective learning, non-native rhythm, vowels, infancy

1. INTRODUCTION

Infants recognize an atypical accent when their native language is being spoken and have accent-specific listening (and social) preferences. For instance, 5- to 6-month-olds prefer watching a talker who speaks the infants' native language with a native accent over a talker with a foreign accent [1], 10-month-olds prefer engaging with native-accented talkers [1], 17-month-olds learn new words better from native- and familiar-accented speakers [2, 3], and pre-school children preferentially learn the function of new objects form native-accented talkers over foreign-accented ones [4].

To recognize accented speech, infants strongly rely on prosodic cues [5–8], as well as on phonotactic patterning [5] and even on vowel properties [9]. Recently, it has been shown that infants as young as 3.5 to 4.5 months discriminate atypical accents in their native-language speech even when those deviate from the native norm only in minimal changes to durational relations between strong and weak syllables and all other cues are equal [10]. Not only could the infants in that study discriminate native and accented speech solely on the basis of the durational patterning, but they also manifested listening preferences with stimuli differing only in durational rhythm cues: 4-month-olds preferentially listened to the native accent and 6-month-olds preferred the non-native accent.

Interestingly, this behaviour was observed at the developmental stage when infants acquire the categories for their native segments. For vowels, perceptual narrowing occurs between the 5th and 6th month of age [11]. This leads us to ask: how does infants' early sensitivity to, and listening preferences for, durational rhythmic patterns interact with the acquisition of native vowel categories?

A wide-spread view is that prosody bootstraps further language development. Previous studies have demonstrated that prosodic cues facilitate word recognition [12], the acquisition of the lexicon and grammar [13, 14], and that rhythmic priming facilitates language comprehension in typically-developing as well as language-delayed children [15]. Here we ask whether prosody can bootstrap the formation of segmental categories. The specific question is whether infants exploit distributional information about a novel vowel contrast more effectively from speech with native versus atypical rhythm, implemented as atypical patterning of durational relations between strong and weak syllables.

The literature suggests that infants might learn about segments more effectively from the type of input for which they also have listening preferences. Infants preferentially listen to infant-directed speech (IDS) over adult-directed speech (ADS) [16] and it has been proposed that they may also learn new segments selectively better from IDS than ADS [17]. One of the functions of IDS, differing from ADS in suprasegmental as well as segmental cues [18], may be to facilitate language acquisition. The question is whether selective learning would align with preferential listening if that means not only learning from more perceptually attractive and often also more salient input (as in the case of IDS) but even learning from *less native* input. This is what we put to test here by investigating whether 6-month-old infants, who have been shown to preferentially listen to rhythmically atypical speech, also learn novel segments more readily from such rhythmically atypical speech.

We measure infants' ability to exploit distributional information about a novel vowel contrast. Infants have been reported to be able to employ a distributional learning mechanism to uncover category-like structures in their speech sound environment [19, 20], although the nature of the (to-be-)learnt speech categories has been questioned [21]. The present experiment tests whether infants discover the input's distributional structure of segmental categories with different success depending on whether it is embedded in native versus nonnative rhythm, implemented, respectively, as native-like versus atypical patterning of durational relations between strong and weak syllables.

2. METHOD

2.1 Participants

Forty-one 5.5- to 6.5-month-old infants acquiring Czech participated (5 other infants were excluded for fussiness). All were born full-term, had normal hearing and no familial risk of dyslexia. They were randomly assigned to the native (n = 21) or the non-native rhythm (n = 20) condition.

2.2 Stimuli

The training stimuli were adapted from our prior study on rhythm preferences [10]. Three women, native speakers of Czech, recorded utterances in Czech with native rhythm and with atypical rhythm, in which the durational ratio between word-initial and non-initial syllables was increased (to achieve consistent rendition, the speakers imitated a resynthesized model recording, see [10]).

In the present "[fɛfæfɛ] design," the utterances were delexicalized by replacing all Cs with the same token of [f] and all Vs with one of 405 unique tokens sampled from a continuous bimodal distribution [22] with peaks at [ɛ] and [æ], whose midpoint F1 and F2 values were 730 Hz and 2125 Hz, and 1015 Hz and 1885 Hz, respectively. Original segment durations were always preserved. Different [fɛfæfɛ] utterances from the 3 speakers were combined to create 8- to 9-sentence trials with native or non-native rhythm. The post-exposure test stimuli were isolated [ɛ] and [æ] with F1 and F2 of the training-distribution peaks and 125 ms long.

2.3 Procedure

In the training phase, the infants were exposed to the bimodal [ɛ]-[æ] distribution embedded in delexicalized utterances either in native or non-native rhythm. There were 5 training trials, with the average duration of 24.5 s, amounting to ~2 minutes of training. The subsequent test phase comprised of two blocks, each containing an alternating [ɛ]-[æ] and a non-alternating [ɛ]-[ɛ] trial (order counterbalanced). Each test trial contained 16 vowels (8 pairs) and lasted 8.3 s.

The infants were tested in a central fixation paradigm implemented in PyHab [23] while seated on the lap of their parent, in a sound-treated booth. A camera mounted on top of the central screen recorded the infants' looking behaviour. Stimulus presentation was not contingent on infant looking. The data were coded offline by a coder blind to the stimuli and trial types.

2.4 Looking time measures

We analysed first-look duration, following [24, 10]. In a paradigm such as ours, with stimulus presentation *not* contingent on infant looking – i.e. a paradigm in which trials do not end as soon as the infant stops looking – first look duration may better capture infants' true discrimination abilities or listening preferences than does total looking time. Arguably, the latency of the first looking away reflects the infant's genuine attention to the presented auditory stimulus, while any repeated looks towards the stimulus throughout the trial that goes on irrespective of the infant's attention, may as well be just random. Besides first look duration we also analysed total looking time because it is a more commonly used measure in the literature.

Both measures were analysed separately for the Training and for the Test. The models for the Training phase assessed preferential listening, and those for the Test phase assessed distributional learning.

2.5 Statistical analyses

First look duration and total looking time were log-transformed and each submitted to a linear mixed-effects model, *lme4, lmerTest* in R [25–27].

The analyses of Training modelled the fixed effects of Familiarization type (sum-coded: -native, +non-native) and Trial number (mean-centred), and random per-participant intercepts and slopes for Trial number.

In the analyses of the Test phase, the fixed factors were Familiarization (-native, +non-native), Test-type (-alternating, +nonalt.), Test-block (-first, +second), and their interactions. The random effects were per-participant intercepts and slopes for Test-type and Test-block.

3. RESULTS

3.1 Training phase

The model for first look duration in the Training phase yielded a significant intercept and an effect of Trial number, indicating that the infants' attention decreased throughout the familiarisation phase (estimated slope [log scale] = -0.192, *SE* = 0.094, *df* = 40.134, *t* = -2.034, *p* = 0.049). The effect of Familiarization type was not significant: the mean first look duration to nonnative rhythm was numerically (but not significantly) longer than the mean first look duration to native rhythm; see left panel in Figure 1.

The model for total looking time detected a significant intercept and an effect of Trial number, again confirming that the infants' looking times decreased throughout familiarisation (estimated slope = -0.084, *SE* = 0.031, *df* = 39.881, *t* = -2.750, *p* = 0.009). The effect of Familiarization type with the *t* value of 1.961 (estimated at 0.062, *SE* = 0.032, *df* = 38.502, *p* = 0.057) indicated that infants in the non-native rhythm condition tended to look longer during familiarisation than infants in the native rhythm condition (native: mean = 17.6 s, 95% CI = 16.0–19.3 s; non-native: mean = 19.9 s, 95% CI = 18.1–21.9 s); see left panel in Figure 2.

3.2 Test phase

The model summary for first look duration in the Test phase is shown in Table 1.

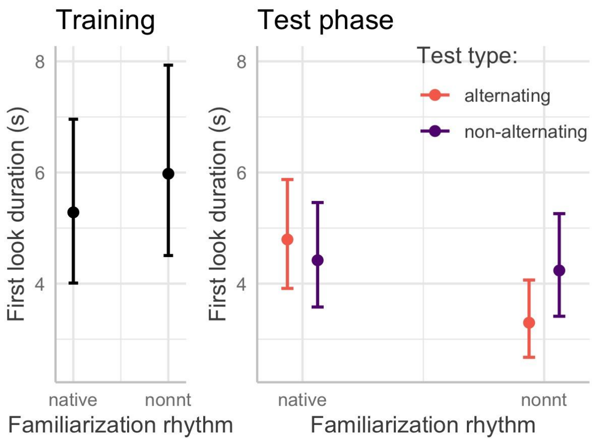

Figure 1. *First look duration* in Training and Test. Estimated means and 95% confidence intervals.

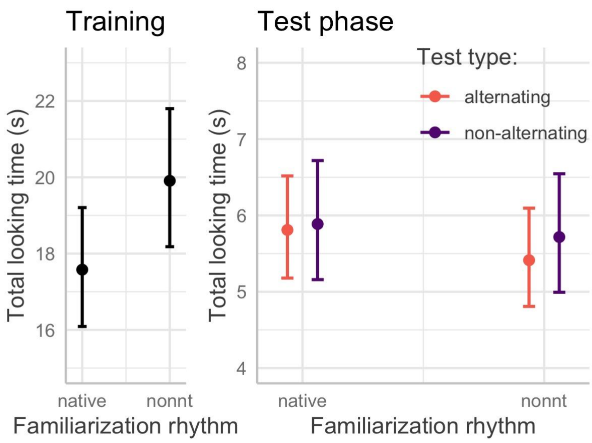

Figure 2. *Total looking time* in Training and Test. Estimated means and 95% confidence intervals.

There was a significant intercept, meaning that the infants' first looks were real, i.e. they were reliably longer than 0. For two other interaction parameters, the absolute *t*-value approached 2, just missing out the commonly used alpha threshold of 0.05. Unpacking the interaction of Familiarization and Test-type (see right panel in Figure 1) revealed that infants trained with non-native rhythm looked longer to non-alternating than to alternating trials at test, by nearly 1 second (mean difference = 0.93 s, *t* = -2.065, *p* = 0.046); no such effect was detected for native-rhythm exposure (mean difference = -0.37, *t* = 0.690, *p* = 0.494).

Pairwise comparisons for the interaction of Familiarization and Test-block showed that infants exposed to native rhythm during familiarisation looked longer to the first block of test trials than to the second block, by about 1.1 second (mean difference = 1.13 s, *t* = 2.350, *p* = 0.024); no such effect was detected for infants exposed to non-native familiarisation (mean difference = -0.2 s, *t* = -0.507, *p* = 0.615).

The model for total looking time in the Test phase yielded a significant intercept, confirming that, overall, total looking time during test trials was greater than 0. No other effects were significant, only one had a potentially meaningful *t*-value, namely, the effect of Test-block (estimated mean effect = -0.040, *SE* = 0.023, *df* = 35.907, *t* = -1.752, *p* = 0.088), indicating that the total looking time tended to be longer in the first than in the second block.

Parameter	Estim.	*SE*	*df*	*t*	*p*
Intercept	1.422	0.064	39.100	22.251	<.001
Familiarization (-nat +nonnt)	-0.110	0.064	39.100	-1.722	.093
Test type (-alt+nonalt)	0.042	0.042	38.844	1.002	.323
Test block (-1st+2nd)	-0.047	0.037	37.786	-1.272	.211
Familiarisation * Test type	0.083	0.042	38.844	1.963	.057
Familiarisation * Test block	0.074	0.037	37.786	2.000	.053
Test type * Test block	-0.022	0.032	39.173	-0.680	.501
Familiarisation *T.type*T.block	-0.007	0.032	39.173	-0.229	.820

Table 1: Fixed-effects model summary for first look duration (log transformed) in the Test phase.

4. DISCUSSION

This experiment tested whether infants' learning of novel segments is modulated by the rhythm of the speech in which the to-be-learnt segments are embedded. Given that 6-month-old Czech-exposed infants have demonstrated listening preferences for non-native over native rhythm [10], we tested whether this would be reflected in selective learning from non-native rhythm.

We exposed 6-month-old monolingual Czech infants to a bimodal distribution representing a novel vowel contrast, an English-like [ɛ]-[æ] distinction. Infants passively listened to 405 vowel tokens sampled from the [ɛ]-[æ] distribution embedded in delexicalized [fɛfæfɛ] utterances with either native Czech rhythm or with atypical, non-native rhythm. Immediately after the ~2-minute training phase, the infants were tested on their discrimination of [ɛ]-[æ] in an alternating/ non-alternating, central fixation, paradigm.

The present results for the training data corroborate previous findings [10] that Czech 6-month-olds preferentially listen to non-native rather than native rhythm. Note however that this effect was found here for total looking time (and not for first look duration) while in our prior study, with the same population and similar stimuli, we found this effect for first look duration. This discrepancy in outcomes for the different measures might be due to the nativeness being a within-subjects factor in the prior work (and thus more robust) and only a between-subjects factor here. It could also be attributed to stimulus properties: the previous study used low-frequency words that had rich segmental structure while the present study used completely delexicalized stimuli with reduced segmental complexity.

The present results for the test phase indicate that infants learn novel segments better from input to which they preferentially attend, even if it is less native-like. Based on the distributional training literature [19], learning effects should be reflected in longer looking to non-alternating than to alternating trials. This is because the preference for non-alternating test trials translates as infants' preference for *novelty at test* as opposed to the – at test already *familiar – bimodal alternation* between the two peaks, [ɛ] and [æ], that they were exposed to during training. Our analyses indicate longer first look duration to non-alternating than alternating test trials in infants trained with non-native rhythm. This suggests that infants were able to discover the underlying bimodal distribution better when it was presented in non-native than in native rhythm.

In sum, the present experiment suggests that infants' distributional learning of novel segmental contrasts is modulated by prosody, specifically, the rhythm of the training speech input. When both measures (i.e. first look duration and total looking time) are considered, the selectivity of the learning mechanism matches the infants' preferential listening behaviour. Future research is needed to further test the trends observed here, as well as to tackle the question of which perceptual and learning mechanisms are tapped at by first look versus total looking times, and compare these across paradigms contingent versus non-contingent on infant looking behaviour.

5. ACKNOWLEDGMENT

This work was funded by a Czech Science Foundation grant 21-09797S.

6. REFERENCES

[1] Kinzler, K.D., Dupoux, E. Spelke, E.S., 2007. The native language of social cognition. *Proceedings of the National Academy of Sciences* 104(30), 12577-12580.

[2] Mattock, K., Polka, L., Rvachew, S. Krehm, M., 2010. The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science* 13, 229-243.

[3] Byers-Heinlein, K. Fennell, C.T., 2014. Perceptual narrowing in the context of increased variation: Insights from bilingual infants. *Developmental Psychobiology* 56, 274-291.

[4] Kinzler, K.D., Corriveau, K.H. Harris, P.L., 2011. Children's selective trust in native-accented speakers. *Developmental Science* 14, 106-111.

[5] Ramus, F., 2002. Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition* 2, 85-115.

[6] Gasparini, L., Langus, A., Tsuji, S. Boll-Avetisyan, N., 2021. Quantifying the role of rhythm in infants' language discrimination abilities: A meta-analysis. *Cognition* 213, 104757.

[7] White, L., Floccia, C., Goslin, J. Butler, J., 2014. Utterance-Final Lengthening Is Predictive of Infants' Discrimination of English Accents. *Language Learning* 64, 27-44.

[8] Martinez-Alvarez, A., Benavides-Varela, S., Lapillonne, A. Gervain, J., 2022. Newborns discriminate utterance-level prosodic contours. *Developmental Science*, e13304.

[9] Zacharaki, K. Sebastian-Galles, N., 2021. The ontogeny of early language discrimination: Beyond rhythm. *Cognition* 213, 104628.

[10] Paillereau, N., Podlipský, V.J., Smolík, F., Šimáčková, Š., & Chládková, K. 2021. The development of infants' sensitivity to native versus non-native rhythm. *Infancy*, 26, 423-441.

[11] Tsuji, S. Cristia, A., 2014. Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology* 56, 179-191.

[12] Johnson, E.K. Jusczyk, P.W., 2001. Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language* 44, 548-567.

[13] Mandel, D.R., Jusczyk, P.W. Nelson, D.G.K., 1994. Does sentential prosody help infants organize and remember speech information? *Cognition* 53, 155-180.

[14] Gleitman, L., Wanner, E. 1982. The state of the state of the art. In Wanner E., Gleitman, L. (Eds.), *Language acquisition: the state of the art* (pp. 3-48). Cambridge, UK: CUP.

[15] Ladányi, E., Lukács, Á. Gervain, J., 2021. Does rhythmic priming improve grammatical processing in Hungarian-speaking children with and without developmental language disorder?. *Developmental Science* 24, e13112.

[16] ManyBabies Consortium, 2020. Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science* 3, 24-52.

[17] Adriaans, F. Swingley, D., 2017. Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America* 141, 3070-3078.

[18] Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., Fusaroli, R., 2022. A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour*, 1-20.

[19] Maye, J., Werker, J.F., Gerken, L., 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101-B111.

[20] Wanrooij, K., Boersma, P., Van Zuijen, T.L., 2014. Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study. *Frontiers in Psychology* 5, 77.

[21] Schatz, T., Feldman, N.H., Goldwater, S., Cao, X.N. Dupoux, E. 2021. Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *Proceedings of the National Academy of Sciences* 118, e2001844118.

[22] Wanrooij, K., Boersma, P., 2013. Distributional training of speech sounds can be done with continuous distributions. *The Journal of the Acoustical Society of America* 133, EL398-EL404.

[23] Kominsky, J.F. 2019. PyHab: Open-source real time infant gaze coding and stimulus presentation software. *Infant Behavior and Development* 54, 114–119.

[24] Lewis, M., Kagan, J., Kalafat, J. 1966. Patterns of fixation in the young infant. *Child Development* 37, 331– 341.

[25] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1–48.

[26] Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82, 1–26.

[27] R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.