Infant Behavior and Development 76 (2024) 101977

Contents lists available at ScienceDirect

Infant Behavior and Development

journal homepage: www.elsevier.com/locate/inbede

The developmental pattern of native and non-native speech perception during the 1st year of life in Japanese infants

Irena Lovčević[a,*], Sho Tsuji[a,b]

[a] International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo Institutes for Advanced Studies, Japan
[b] Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'Études Cognitives, ENS, EHESS, CNRS, PSL University, France

ARTICLE INFO

Keywords:
Perceptual attunement
Non-native speech perception
Native speech perception
Vowel length
Japanese

ABSTRACT

Language development during the 1st year of life is characterized by perceptual attunement: following language-general perception, a decline in the perception of non-native phonemes and a parallel increase in or maintenance of the perception of native phonemes. While this general pattern is well established, there are still many gaps in the literature. First, most evidence documenting these patterns comes from "Minority world countries" with only a limited number of studies from "Majority world countries", limiting the range of languages and contrasts assessed. Second, few studies test both the developmental patterns of native and non-native speech perception in the same group of infants, making it hard to draw conclusions on simultaneous decline in non-native and increase in native speech perception. Such limitations are in part due to the effort that goes into testing developing speech sound perception, where usually only discrimination of one contrast per infant can be tested at a time. The present study thus set out to assess the feasibility of assessing a given infant on their discrimination of two speech sound contrasts during the same lab visit. It leveraged the presence of documented patterns of the improvement of native and the decline of non-native phoneme discrimination abilities in Japanese, therefore assessing native and non-native speech perception in Japanese infants from 6 to 12 months of age. Results demonstrated that 76 % of infants contributed discrimination data for both contrasts. We found a decline in non-native speech perception evident in discrimination of the non-native /ɹ/-/l/ consonant contrast at 9–11, but not at 11–13 months of age. Additionally, a parallel increase in native speech perception was demonstrated evident in an absence of native phonemic vowel length discrimination at 6–7 and 9–11 months and a discrimination of this contrast at 11–13 months of age. These results, based on a simultaneous assessment of native and non-native speech perception in Japanese-learning infants, demonstrate the feasibility of assessing the discrimination of two contrasts in one testing session and corroborate theoretical proposals on two hallmarks of perceptual attunement: a decrease in non-native and a facilitation in native speech perception during the first year of life.

1. Introduction

To acquire a specific language, infants need to learn the phonemes of that language. This process starts with language-universal

* Correspondence to: International Research Center for Neurointelligence (IRCN), The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo 113-0033, Japan.

E-mail address: irena.lovcevic@ircn.jp (I. Lovčević).

https://doi.org/10.1016/j.infbeh.2024.101977
Received 31 January 2024; Received in revised form 27 June 2024; Accepted 4 July 2024
Available online 14 July 2024

speech sound discrimination; however, infants become attuned to the phonemes of their ambient language during the first year of life. In other words, infants start their lives with a remarkable ability to discriminate many of the phoneme contrasts present in world languages. During the first year this ability gradually decreases such that infants' ability to discriminate non-native contrasts declines. This decrease in non-native speech perception is positively related to infants' later language skills (Kuhl et al., 2005; Kuhl & Rivera-Gaxiola, 2008). Indeed, efficient experience-dependent pruning has been demonstrated to be an important part of development during highly plastic periods (Doupe & Kuhl, 1999).

Attunement refers not only to the experience-dependent decline in non-native, but also the experience-dependent improvement in native perception. Hence, research on perceptual attunement should benefit from examining the parallel development of non-native and native speech discrimination.

Extensive research has shown that infants' ability to discriminate non-native phonetic contrasts declines during the first year of life. In a seminal study, Werker & Tees (1984) demonstrated discrimination of the Hindi retroflex-dental contrasts and Nthlakampx glottalized velar versus uvular contrasts in English-learning 6–8-month-old but not in 10–12-month-old infants. Additional studies emerged attesting successful non-native consonant discrimination in 6–8-month-old and a lack thereof in 10–12-month-old infants (Best & McRoberts, 2003; Best et al., 1995; Kuhl et al., 2006; Tsao et al., 2006; Tsushima et al., 1994). These studies suggest that younger infants can discriminate phonetic contrasts absent in their ambient language without specific experience, whereas the same is not the case in older infants.

Multiple theoretical models have been proposed to account for the observed decline in non-native speech perception (Best & McRoberts, 2003; Burnham et al., 1991; Diamond et al., 1994; Lalonde & Werker, 1995; Werker, 1995; Werker & Pegg, 1992). The role of language exposure in the decline in non-native speech perception is unanimously emphasized, although assumed pathways through which this happens differ: The Perceptual Assimilation Model/Articulatory Organ account (PAM/AO, Best, 1994, 1995; Best & McRoberts, 2003) assumes that exposure to the ambient language leads to acquisition of the articulatory patterns underlying the ambient speech. When non-native speech sounds are perceived, they are assimilated to the articulatory patterns of native sounds, which results in a decline in the discrimination of non-native sounds. The Native Language Magnet account (NLM, Kuhl, 1994, 2000) suggests that the exposure to native speech sound categories warps perception, such that speech sound prototypes start acting as perceptual magnets that decrease the distance between such a prototype and surrounding sounds within an ambient language phonetic category and increase the distance in the region of ambient language phonetic boundaries. As a result, the discrimination of non-native boundaries decreases. The Native Language Neural Commitment account (NLNC, Kuhl, 2004; Kuhl et al., 2005) suggests that the strengthening of neural connections to ambient sounds lead to decline in perception of non-native sounds. The Processing Rich Information from Multi-dimensional Interactive Representations framework (PRIMIR, Curtin & Werker, 2007; Werker & Curtin, 2005) assumes that infants use general learning mechanisms together with filters such as perception biases, developmental stage, and specific language learning tasks to track statistical regularities in the input and learn only information that are linguistically relevant. This subsequently leads to decline in non-native sound discrimination as this is linguistically irrelevant. Furthermore, three of these models (NLM, Kuhl, 1994, 2000, NLNC, Kuhl, 2004; Kuhl et al., 2005, and PRIMIR, Curtin & Werker, 2007; Werker & Curtin, 2005) propose that changes in perception are not limited only to non-native speech perception but occur in native speech perception as well. Indeed, these models suggest that the exposure to and acquisition of acoustic and statistical properties of the ambient language facilitates infants' native speech perception.

Regarding the facilitation in infants' native speech perception during the first year of life, the behavioral evidence is limited (Choi et al., 2023; Kuhl et al., 2006; Liu & Kager, 2016). Kuhl et al. (2006) assessed 6–8- and 10–12-month-old American and Japanese infants' discrimination of the /ɹa/-/la/ consonant contrast. This contrast is native for American English- and non-native for Japanese-learning infants. In particular, Japanese infants are exposed to an /ɹ/ sound that is articulatory different from American English /ɹ/ and /l/ with variable pronunciation that might be realized anywhere on the continuum between /ɹ/ and /l/ (Ladefoged & Maddieson, 2001). Hence, Japanese infants learn a unimodal distribution for these sounds perceptually assimilating them to Japanese /ɹ/ (Best & Strange, 1992). Whereas no difference in the discrimination of American English /ɹ/ and /l/ was found between American and Japanese-learning infants at 6–8 months, at 10–12 months there was an increase in discrimination in American infants and a decrease in Japanese infants, suggesting an increase in native and a decrease in non-native speech perception. The majority of other studies assessing non-native versus native speech perception document a pattern of maintenance, rather than improvement, of native discrimination ability. For instance, Tsushima et al. (1994) examined 6–8-month-old and 10–12-month-old Japanese-learning infants' perception of non-native /ɹ-l/ and native /w-y/ contrasts. Their results demonstrated that younger infants could discriminate both contrasts, whereas older infants could discriminate only the native /w-y/ contrast demonstrating the maintenance in the perception of native contrast.

One likely reason for the paucity of behavioral studies on native improvement is the very fact that infants can discriminate many contrasts at birth; thus, infants' initial sensitivity to certain native contrasts might explain the lack of further improvement in their perception. We want to remark that meta-analytic evidence nevertheless finds an increase in effect sizes for native perception over the first 14 months of life, but this pattern might be hard to find in a single study (Tsuji & Cristia, 2014). Therefore, in order to assess the relationship between non-native decline and native improvement, it would be informative to use a native contrast that is not well discriminated early in life, and that can hence improve over exposure. Only a limited number of studies demonstrates that some native contrasts are hard for infants to discriminate at a younger age, but the perception of these improves over time. For example, Korean infants do not discriminate the native stop contrast at 4–6 and 7–9 months of age, with discrimination appearing only around 12 months of age (Choi et al., 2023). Similarly, Dutch infants showed an improvement in discrimination of native /I/-/ɪ/ vowels from 5 to 15 months of age (Liu & Kager, 2016). Furthermore, Japanese infants cannot discriminate the long-short vowel contrasts at 4 months, with discrimination emerging sometime between 7.5 and 9 months of age (Mugitani et al., 2009; Sato et al., 2010).

Another gap in the literature is the lack of studies that assess both native and non-native speech perception in infants from “Majority World Countries” (Alam, 2008) with almost 90 % of all speech perception studies coming from “Minority World Countries” and 45 % being conducted with English-learning infants (Singh et al., 2022). Hence, the present study assessed the developmental patterns of native and non-native speech perception in Japanese-learning infants. Speech perception in Japanese-learning infants constitutes only 6 % of total speech perception infancy research (Singh et al., 2022). Furthermore, previous studies on speech perception in Japanese-learning infants assessed solely native (Mugitani et al., 2009; Sato et al., 2010) or solely non-native speech perception (Kuhl et al., 2006) with only one study (Tsushima et al., 1994) assessing both native and non-native speech perception in the same group of Japanese-learning infants. Having only one study that tested both native and non-native speech perception in same group of Japanese-learning infants warrants replication, especially given that this study suffers from a small sample size (10 to 13 participants per group in final analysis).

The goal of the present study was to assess two hallmarks of perceptual attunement, the decline in non-native speech sound discrimination and the improvement in native speech sound discrimination, in infants from the same Japanese language background. In order to do so, it is necessary to choose a non-native contrast that shows declining discrimination, as well as a native contrast that shows improving discrimination over the same time span. In Japanese, the previously mentioned studies on non-native /ɹ/-/l/ discrimination (Kuhl et al., 2006; Tsushima et al., 1994) document a decline in the discrimination of this contrasts. In addition, Japanese also offers a documented case of native improvement in speech sound discrimination, namely Japanese-acquiring infants’ discrimination of the native vowel length contrast. In the Japanese language the long-short vowel distinction is phonemic, used to distinguish word meanings. For example, the words /obasan/ and /oba:san/ differ only in the length of the vowel /a/, but they have different meanings (aunt and grandmother, respectively). Hence, Japanese infants eventually need to acquire this distinction in order to distinguish word meanings. Previous studies suggest that whereas 4-month-old infants cannot discriminate the long-short vowel contrasts, this ability develops sometime between 7.5 and 9 months of age (Mugitani et al., 2009; Sato et al., 2010). One proposed reason why this contrast is hard to distinguish without ample language experience is its distributional properties. Although a canonical short vowel in Japanese is around 100 ms long and a canonical long vowel is around 200 ms long, in reality these two vowel types show two properties that could impede their discrimination. First, short vowels are much more frequent than long vowels, making up the majority of vowel sounds (94 %). Second, these short vowels vary in actual pronunciation length between 30 and 1000 ms (Bion et al., 2013), with lengthening happening due to a multitude of reasons, among them phrase-final lengthening or emphasis. The distribution of long vowels falls completely within that of short vowels, meaning that infants need to extract the vowel length distinction from a rather unimodal distribution. As such, Japanese offers a native (vowel length) and non-native (/ɹ/-/l/) speech sound contrast with documented improvement and decline, respectively, and which can be well grounded in the particular distributional characteristics of these contrasts in the context of the native language.

Infants’ native speech perception was assessed via the long-short /a/-/a:/ vowel contrast. Regarding non-native speech perception, this was examined via the North American English /ɹ/ - /l/ consonant contrast, reviewed further above. Based on previous literature, we hypothesised: i) an increase in native speech perception evident in no discrimination at 6 months, and discrimination at 10 months (Mugitani et al., 2009; Sato et al., 2010); ii) a decrease in non-native speech perception evident in discrimination at 10 months and an absence or a decrease in discrimination at 12 months (Kuhl et al., 2006; Tsushima et al., 1994).

In addition, we examined the feasibility of assessing discrimination of both types of speech sound contrasts in one testing session per infant, by sequentially testing each infant on both native and non-native discrimination. Our study design allowed us to assess the within-participant relationship between native and non-native perception, for those age groups where we assessed both types of contrasts at the same time. This design choice offers both a methodological contribution on the feasibility to run two habituation experiments in one session and allowed us to assess iii) whether native and non-native perception were correlated.

1.1. Experiment 1

In Experiment 1, we assessed the ability of 6–7-month-old Japanese-learning infants to discriminate a native long-short vowel duration contrast that is phonemic in Japanese. Based on previous studies, we expect no discrimination of this contrast in 6-month-old infants (Sato et al., 2010).

2. Method

2.1. Participants

Thirty (18 females) full-term born monolingual Japanese-learning 6–7-month-old infants participated in Experiment 1 (*Mean Age* = 6.79, *SD* = 0.53, *Age range* = 6.03 – 8.00). An additional 9 infants were tested but excluded from the final sample due to crying (7) and experimental error (2). According to G*Power 3 (Faul et al., 2007) calculations, the minimum sample size for an alpha of 0.05 and effect size of 0.58 (Sato et al., 2010) to achieve a 90 % power for our planned analysis is 27 participants indicating that our sample size is sufficient to detect an effect if there is a one. The University of Tokyo ethical committee approved the study. Participants were recruited via a university laboratory database of families who have expressed interest to take part in infancy research. Before the experiment, caregivers signed informed consent form on their infant’s behalf. All infants received a 1000-yen book voucher as a gift for their participation in the study.

2.2. Stimuli

To test infants' discrimination of the long-short vowel contrasts, we used the two-syllable non-words /mana/ versus /ma:na/ differing only in the duration of the first vowel /a/, with the mean duration of /ma/ being significantly shorter ($M = 133.12$ ms, $SD = 40.36$) than /ma:/ ($M = 239.20$ ms, $SD = 18.56$, 95 % *CIs* of difference [86.16, 125.99], $t(17) = 11.24$, $p < .000$, $d = 3.38$). The words /mana/ and /ma:na/ did not differ in mean pitch. These stimuli were previously used in Sato et al.'s (2010) study on vowel perception in Japanese infants. The /ma:na/ stimuli were recorded by a female native speaker of Japanese, whereas the /mana/ stimuli were created by shortening the steady part of the first vowel in /ma:na/ to ensure that stimuli did not differ in acoustic features other than the duration of the vowel. From these instances both /ma:na/ and /mana/ were chosen randomly and four stimulus lists were created, two with 10 repetitions of the word /ma:na/ and two with 10 repetitions of the word /mana/ with around 1.5-seconds interval between the repetitions and with total duration of the list around 15 s. They are described in detail in Sato et al. (2010). As for pretest and posttest trials, a female native speaker of Japanese was recorded repeating the word /panta/.

2.3. Procedure

The experiment was conducted in a dimly lit soundproof laboratory room and was controlled via the Habit X 2.0 software (Habit X 2.0; Oakes et al., 2019) on a Windows computer. Visual stimuli were presented on a 23-inch screen and audio stimuli were delivered through two forward-facing loudspeakers positioned below the screen. Infants sat on their caregiver's lap in front of the screen. Caregivers listened to the music over noise-cancelling headphones to conceal the auditory stimuli presented to an infant. The experimenter observed the infant from an observation room through a video camera and recorded infants' looks to the screen by pressing a key on the experimental computer keyboard. The experimenter was holding the key when infants were looking to the screen and was releasing the key when infants were not looking to the screen. These key presses were logged via experimental software and were used in analyses. Infants' looks were recorded on a video camera as well.

The current study employed the modified visual habituation paradigm (Stager & Werker, 1997). In this type of paradigm, infants are habituated to one type of stimuli in the habituation phase, and then are presented during the test phase with another kind of stimuli (switch trials) and the same kind of stimuli (same trials). The paradigm consists of 4 phases: pre-test, habituation, test, and post-test. Before each trial, an attention getter was presented, and as soon as the infant looked to the screen, the experimenter pressed the key on the experimental computer keyboard to start the next trial. The pre-test and post-test phases consisted of one trial each. During the trial, infants saw visual animation of rotating wheel, and listened to a repetition of the audio stimulus "panta". A comparison of looking behavior during pre- and post-test can indicate whether infants become fatigued or uninterested over the course of the experiment. During the habituation phase, one of the four lists of either vowel length was presented randomly together with a visual stimulus (red and black static checkerboard). The habituation phase lasted until infants reached the habituation criterion or the maximum of 28 habituation trials. The habituation criterion was reached when average looking time in one habituation block/window (four trials) was less than 65 % of infants' average looking time to the first/basis window (first four habituation trials). Next, infants proceeded to the test phase, which consisted of four test trials, with the first two test trials presenting stimuli of different vowel length (switch trials) followed by two trials presenting stimuli of the same vowel length (same trials) as in the habituation phase. Thus, infants who were habituated with /ma:na/, in the test phase heard two trials of /mana/ (switch trials) and then two trials of /ma:na/ (same trials). On the other hand, infants who were habituated with /mana/, in the test phase were presented with two trials of /ma:na/ (switch trials) and then with two trials of /mana/ (same trials).

2.4. Data analysis

To compare infants' looking time between the same and switch trials, a linear mixed effects model (LME) was fitted using the *lmer* function of the *lme4* package (Bates et al., 2015) in R (R Core Team R, 2020). The LME model was fitted with Trial (same vs. switch), Habituation Stimulus (short vs. long vowel), and their interaction as the independent variables, a random intercept[1] for Participants, and infants' looking time[2] as the dependent variable. Habituation Stimulus was included in the model given previous findings on asymmetrical perception of phonemes in infants, that is a better discrimination of sound contrast presented in one direction compared to discrimination of the same sound contrast presented in the opposite direction (Mugitani et al., 2009; Polka & Bohn, 2011). All factor variables were dummy coded. The significant effects of predictors and interactions were assessed using ANOVAs with Satterthwaite's method using the *anova* function of the *lmerTest* package (Kuznetsova et al., 2017). As a measure of effect size Cohen's d was calculated from the *F*-statistic using the function *F_to_d* from the package *effectsize* (Ben-Shachar et al., 2021). Whenever models showed significant effects, post hoc analyses were conducted using Tukey tests from the package *emmeans* (Lenth, 2019). We additionally assessed infants' sustained interest in the task by comparing their looking times during the pre-test and post-test trial with a paired t-test, as well as the potential influence of the habituation phase on results by comparing habituation times of infants habituated with the long versus short vowel using a t-test for independent samples.

[1] Based on Akaike information criterion (AIC), the model with random intercept for participant (AIC: 2228.341) provided a better fit than model with Trial by-participant slope (AIC: 2230.619).

[2] An additional LME model with log-transformed looking times was fitted, and the results were comparable to those of this model (see Supplementary Material S1 for model output).

3. Results

The LME results demonstrated no significant main effect of Trial, $F(1, 88) = 4.09$, $p = 0.05$, *Cohen's d* = .43 (95 % *CIs* [.01,.85]) (see Table 1 for full model). Infants' looking time to the switch trials ($M = 6433$ ms, $SD = 3008$) did not differ significantly from their looking times to the same trials ($M = 5375$ ms, $SD = 2710$) (Fig. 1). Furthermore, there was no significant main effect of Habituation Stimulus on infants' looking time ($F(1, 28) = .25$, $p = .62$, *Cohen's d* = .19 (95 % *CIs* [−.55,.93])), and no significant interaction between Trial and Habituation Stimulus ($F(1, 88) = .17$, $p = .68$, *Cohen's d* = .09 (95 % *CIs* [−.33,.51])).

Regarding habituation times, a t-test for independent samples demonstrated no difference in infants' mean looking times during the habituation phase between infants habituated with short versus long vowels (short: $n = 16$, $M = 8447.19$ ms, $SD = 2336$; long: n = 14, $M = 8750.64$ ms, $SD = 1916$), $t(27.9) = .39$, $p = .70$, *Cohen's d* = 0.15 (95 % *CIs* [−.60,.89]).

Finally, a paired t-test showed that looking time during pretest trials ($M = 16{,}334$ ms, $SD = 2885$) was significantly longer than looking time during posttest trials ($M = 14{,}806$ ms, $SD = 3846$), $t(29) = -2.37$, $p < .05$, *Cohen's d* = 0.44 (95 % *CIs* [0.06, 0.82]), suggesting that infants might have lost interest or were fatigued during the test phase.

Hence, the results suggest that 6–7-month-old Japanese infants failed to discriminate the vowel length contrast between /mana/ and /ma:na/. Though fatigue or loss of interest during the test phase might have contributed to these results, not discriminating the vowel length contrast at this age is consistent with previous findings (Mugitani et al., 2009; Sato et al., 2010). There was no evidence for asymmetrical perception of vowel length contrast.

3.1. Experiment 2

The results of Experiment 1 demonstrated that 6–7-month-old Japanese-learning infants could not perceive the difference between the short and long vowel contrast. In Experiment 2, we examined the discrimination of the vowel length contrast in 9–11-month-old infants. This age was chosen based on previous studies that demonstrated discrimination of the vowel length contrast around this age (Mugitani et al. 2009; Sato et al., 2010). Additionally, at this age we also tested the same infants' discrimination of the non-native /ɹ/ versus /l/ consonant contrast in order to examine native and non-native speech perception in parallel. Since discrimination of this contrast is well attested at the younger age (Kuhl et al., 2006; Tsushima et al., 1994) and in order to balance the resources with study goals, this contrast was not tested in Experiment 1.

3.2. Participants

Forty-three full-term born monolingual Japanese-learning 9–11-month-old infants participated in Experiment 2. Of these, 9 participants were excluded from native vowel condition due to fussiness (4), not completing the task due to the caregiver judging the infant to be too upset to continue (3), not habituating (1), and crying (1). Hence, the final sample for the analysis of native vowel discrimination consisted of 34 infants (19 females, *Mean Age* = 10.43, *SD* =.84, *Age range* = 9.08 – 11.41). Regarding the non-native experiment, 10 participants were excluded due to fussiness (2), not completing the task due to the caregiver judging the infant to be too upset to continue (5), not habituating (2), and crying (1). One of the excluded infants overlapped between the native and non-native experiment. Thus, the final sample for analysis of non-native consonant discrimination included 33 infants (20 females, *Mean Age* = 10.44, *SD* =.87, *Age range* = 9.05 – 11.41). Based on G*Power 3 calculations (Faul et al., 2007), the minimum sample size for native vowel discrimination analysis was 27 participants (see Experiment 1), and for non-native consonant discrimination analysis the minimum sample size for an alpha of 0.05 and effect size of 0.60 (Tsushima et al., 1994) to achieve a 90 % power is 26 participants. Hence, our sample size for both native and non-native analysis is sufficient to detect an effect if there is an effect.[3]

3.3. Stimuli

For native speech perception, stimuli were the same as in Experiment 1. Regarding non-native speech perception, we chose the words "right" /ɹaɪt/ and "light" /laɪt/ that differ only in the first sound. Note that although these are real words in English, they are not known words to Japanese infants. To create the audio stimuli, a female native speaker of Canadian English was recorded repeating these words. The eight best instances of each word were chosen, and two stimulus lists were created, one with 10 repetitions of the word "right" /ɹaɪt/ and one with 10 repetitions of the word "light" /laɪt/ with around 1.5-seconds interval between the repetitions (two words in each of the lists occurred twice) and with total duration of the list around 15 s. The stimuli did not differ significantly in duration $t(7) = -1.82$, $p = .11$, *Cohen's d* = 0.69 (95 % CIs [−0.15, 1.49]) ("right" /ɹaɪt/: $M = 422.63$, $SD = 44.4$; "light" /laɪt/: $M = 387.63$, $SD = 42.2$), nor in mean fundamental frequency $t(7) = -0.05$, $p = .96$, *Cohen's d* = 0.02 (95 % CIs [−0.72, 0.76]) ("right" /ɹaɪt/ $M = 174.66$, $SD = 9.04$, "light" /laɪt/: $M = 174.28$, $SD = 23.8$). For posttest/pretest trials, the same audio stimuli were used as in Experiment 1.

[3] Additional analysis was conducted with only infants who completed both native and non-native tasks (N = 25). These results are reported in Supplementary Material S2.

Table 1
Summary of LME model fitted with Trial (same vs. switch), Habituation Stimulus (short vs. long vowel), and their interaction as the independent variables, a random intercept for Participants, and infants' looking time as the dependent variable.

Predictors	*β*	*SE*	*t*	*CI*	*p*
(Intercept)	5250.29	750.47	7.00	3763.61–6736.96	**< 0.001**
Trial [Switch]	829.82	753.61	1.10	-663.08–2322.72	0.27
Habituation stimulus [short]	233.81	1027.62	0.23	-1801.90–2269.52	0.82
Trial [Switch] * Habit vowel [short]	427.93	1031.93	0.42	-1616.31–2472.17	0.68
Marginal R^2 / Conditional R^2				0.03 / 0.35	

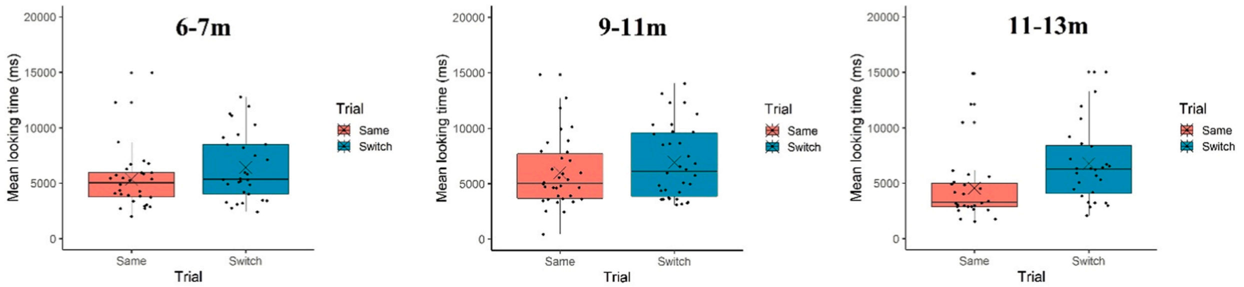

Fig. 1. Mean looking times for the same (red) and switch trials (blue) in response to long-short vowel contrast in 6–7- (left), 9–11- (middle), and 11–13-month-old (right) infants. *Note.* The circles represent the individual data points, "x" symbols represent the means, the horizontal lines within the boxes represent the median, the ends of the vertical lines represent the minimum and maximum values, and the boxes encompass the inter-quartile ranges.

3.4. Procedure

The same procedure as in Experiment 1 was used for both native and non-native stimuli. Regarding non-native speech perception, during the habituation phase infants heard repetitions of either "right" /ɹaɪt/ or "light" /laɪt/. Once they habituated, they proceed to the test phase, consisting of four trials, two switch and two same trials. Infants who were habituated with "right" /ɹaɪt/, in the test phase heard two trials of "light" /laɪt/ (switch trials) and then two trials of "right" /ɹaɪt/ (same trials). On the other hand, infants who were habituated with "light" /laɪt/, in the test phase were presented with two trials of "right" /ɹaɪt/ (switch trials) and then with two trials of "light" /laɪt/ (same trials). Each infant participated in both experiments with a short break between the sessions. The break lasted on average between five and 10 min during which caregivers and infants played with toys and books in the child friendly laboratory main room. The order of experiments was randomized.

3.5. Data analysis

The main data analysis was conducted in the same way as in Experiment 1, separately for native and non-native contrasts, except that an additional factor was included in the LME model. Thus, the LME model was fitted with Trial (same vs. switch), Habituation Stimulus (short vs. long vowel), and Task Order (native vs. non-native) and their interaction as the independent variables, a random intercept for Participants,[4] and infants' looking time[5] as the dependent variable. Inclusion of Task Order in the model allowed an assessment of whether there was a difference in looking time to switch and same trials between infants who conducted native experiment first and infants who did non-native experiment first.

4. Results

4.1. Native speech perception

The LME results demonstrated no significant main effect of Trial, $F(1, 84) = .61, p = .44$, *Cohen's d* $= .17$ (95 % *CIs* [−.26,.60]) (see Table 2 for full model). There was no significant difference in infants' looking time to the switch ($M = 6929$ ms, $SD = 4068$) compared to their looking times to the same trials ($M = 6044$ ms, $SD = 3526$) (Fig. 1). Hence, contrary to our expectations, the results suggest that 9–11-month-old Japanese infants failed to discriminate the vowel length contrast between /mana/ and /ma:na/. Furthermore, there

[4] This model was better fit than a model with Trial by-participant slope for both native (AIC: 2134.46, 2138.09) and non-native speech perception (AIC: 2345.34, 2345.47).
[5] LME models with log-transformed looking times yielded comparable results (see Supplementary Material S1 for model outputs).

Table 2
Summary of LME model fitted with Trial (same vs. switch), Habituation Stimulus (short vs. long vowel), and Task Order (native vs. non-native) and their interaction as the independent variables, a random intercept for Participants, and infants' looking time as the dependent variable.

Predictors	*β*	*SE*	*t*	*CI*	*p*
(Intercept)	4830.58	1237.74	3.90	2377.17–7284.00	**< 0.001**
Trial [Switch]	-753.11	1150.30	-0.66	-3033.19–1526.98	0.51
Habituation stimulus [short]	1396.80	1260.22	1.11	-1101.17–3894.77	0.27
Task order [native first]	904.58	1279.17	0.71	-1630.95–3440.12	0.48
Trial [Switch] * Habituation stimulus [short]	293.52	1171.19	0.25	-2027.97–2615.01	0.80
Trial [Switch] * Task order [native first]	2082.60	1188.80	1.75	-273.80–4439.00	0.08
Marginal R^2 / Conditional R^2				0.09 / 0.45	

was no significant main effect of Habituation Stimulus on infants' looking time ($F(1, 26) = 1.91$, $p = .18$, *Cohen's d* = .54 (95 % *CIs* [−.25, 1.32])), and no significant interaction between Trial and Habituation Stimulus ($F(1, 84) = .06$, $p = .80$, *Cohen's d* = .05 (95 % *CIs* [−.37,.48])) suggesting no asymmetry in vowel length perception in 9–11-month-old Japanese infants. Also, the fitted model showed no significant main effect of Task Order ($F(1, 26) = 2.95$, $p = .10$, *Cohen's d* = .67 (95 % *CIs* [−.12, 1.46])) and no significant interaction between Task Order and Trial ($F(1, 84) = 3.07$, $p = .08$, *Cohen's d* = .38 (95 % *CIs* [−.05,.81])) suggesting no difference in looking time to switch versus same trials between infants who completed the native compared to infants who completed the non-native task first.

Next, regarding habituation times, a t-test for independent samples demonstrated no difference in infants' mean looking times during the habituation phase between infants habituated with short versus long vowels (short: $n = 14$, $M = 8170.07$ ms, $SD = 2480$; long: $n = 20$, $M = 7847.06$ ms, $SD = 2373$), $t(27.33) = −.38$, $p = .71$, *Cohen's d* = −.15 (95 % *CIs* [−.90,.61]).

A paired t-test showed that there was no significant difference in looking times during pretest trials ($M = 16{,}151$ ms, $SD = 3547$) compared to looking times during the posttest trials ($M = 14{,}377$ ms, $SD = 4456$), $t(33) = −1.81$, $p = .08$, *Cohen's d* = −.32 (95 % *CIs* [−.66,.04]), suggesting that infants did not lose interest in the task or become fatigued during the test phase.

Thus, the results suggest that 9–11-month-old infants could not discriminate the native vowel length contrast and did not show asymmetries in their speech perception.

4.2. Non-native speech perception

The LME results demonstrated a significant main effect of Trial, $F(1, 96) = 13.38$, $p < 0.001$, *Cohen's d* = .75 (95 % *CIs* [.33, 1.16]) with infants' looking times to the switch trials ($M = 4769$ ms, $SD = 3014$) significantly longer compared to infants' looking time the same trials ($M = 3475$ ms, $SD = 1985$) (see Table 3 for full model and see Fig. 2). Thus, the results suggest that 9–11-month-old Japanese infants discriminated non-native /ɹ/-/l/ consonant contrast. Additionally, results showed no significant main effect of Habituation Stimulus on infants' looking time ($F(1, 30) = 2.93$, $p = .10$, *Cohen's d* = .62 (95 % *CIs* [−.11, 1.35])), and no significant interaction between Trial and Habituation Stimulus ($F(1, 96) = 0.96$, $p = .33$, *Cohen's d* = .20 (95 % *CIs* [−.20,.60])) suggesting no asymmetry in vowel length perception at 9–11 months of age. Furthermore, there was no significant main effect of Task Order ($F(1, 30) = 1.47$, $p = .24$, *Cohen's d* = .44 (95 % *CIs* [−.29, 1.16])) and no significant interaction between Task Order and Trial ($F(1, 96) = .004$, $p = .95$, *Cohen's d* = .01 (95 % *CIs* [−.39,.41])) suggesting no difference in looking time to switch versus same trials between infants who completed the native compared to infants who completed the non-native task first.

Regarding habituation times, a t-test for independent samples demonstrated no difference in infants' mean looking times during the habituation phase between infants habituated with /ɹ/ versus /l/ sound (/ɹ/: $n = 17$, $M = 5176.01$ ms, $SD = 2043$; /l/: $n = 16$, $M = 6104.84$ ms, $SD = 2257$), $t(30.21) = 1.24$, $p = .23$, *Cohen's d* = .45 (95 % *CIs* [−.28, 1.17]).

A paired t-test showed no significant difference in infants' total looking time during pretest ($M = 13{,}424$ ms, $SD = 4797$) compared to the posttest phase ($M = 13{,}471$ ms, $SD = 4684$), $t(32) = .05$, $p = .96$, *Cohen's d* = .01 (95 % *CIs* −.34,.36]).

The results of this experiment demonstrated that 9–11-month-old Japanese infants are able to discriminate the non-native consonant contrast with no asymmetry in their discrimination.

4.3. Correlation between native and non-native speech perception

Finally, Pearson correlation coefficients were calculated to examine the relationship between performance on native versus performance on non-native speech perception. This was only done for the 25 infants who completed both native and non-native experiments by calculating the correlation between difference in looking time to switch versus same trials. Results showed a weak negative correlation between the variables that did not reach statistical significance, $r(23) = −0.14$, $p = .51$.

4.4. Experiment 3

The results of Experiment 2 demonstrated discrimination of the non-native speech contrast by 9–11-month-old Japanese-learning infants and no evidence of discrimination of the native vowel length distinction. The latter was unexpected based on previous studies,

Table 3
Summary of LME model fitted with Trial (same vs. switch), Habituation Stimulus ("right" /ɹaɪt / vs. "light" /laɪt/), and Task Order (native vs. non-native) and their interaction as the independent variables, a random intercept for Participants, and infants' looking time as the dependent variable.

Predictors	*β*	*SE*	*t*	*CI*	*p*
(Intercept)	4273.15	553.94	7.71	3176.75–5369.54	**< 0.001**
Trial [Switch]	1660.53	566.38	2.93	539.50–2781.55	**< 0.01**
Habituation stimulus ["right" /ɹaɪt/]	-734.36	758.68	-0.97	-2236.01–767.28	0.34
Task [native]	-814.56	758.68	-1.07	-2316.21–687.08	0.29
Trial [Switch] * Habituation stimulus ["right" /ɹaɪt/]	-761.88	775.73	-0.98	-2297.26–773.50	0.33
Trial [Switch] * Task [native]	50.42	775.73	0.07	-1484.96–1585.80	0.95
Marginal R^2 / Conditional R^2				0.15 / 0.42	

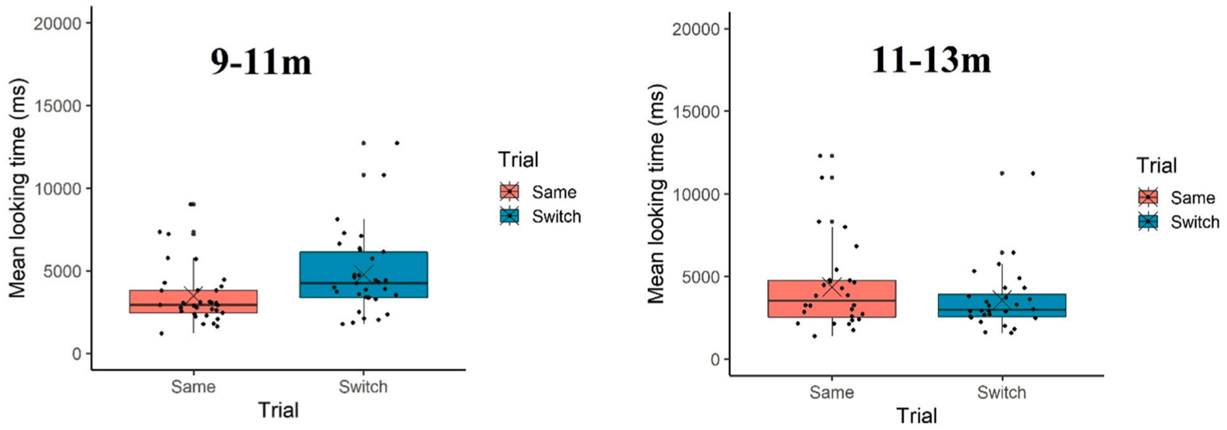

Fig. 2. Mean looking times for the same (red) and switch trials (blue) in response to the non-native /ɹ/-/l/ consonant contrast in 9–11- (left) and 11–13-month-old (right) infants. *Note.* The circles represent the individual data points, "x" symbols represent the means, the horizontal lines within the boxes represent the median, the end of the vertical lines represent the minimum and maximum values, and the boxes encompass the inter-quartile ranges.

and we therefore decided to assess the native contrast again in 11–13-month-old infants. In Experiment 3, we therefore assessed the discrimination of both vowel length and non-native consonant contrasts in 11–13-month-old Japanese-learning infants. We hypothesised that 11–13-month-old infants will be able to discriminate the native vowel contrast (Mugitani et al., 2009; Sato et al., 2010), while at the same time failing to discriminate the non-native consonant contrast (Kuhl et al., 2006; Tsushima et al., 1994).

4.5. Participants

Forty-one full-term born monolingual Japanese-learning 11–13-month-old infants participated in Experiment 3 (see Experiment 1 and Experiment 2 for the minimum sample size calculations). Of these, 12 participants were excluded from native vowel analysis due to not completing the task due to the caregiver judging the infant to be too upset to continue (8), not habituating (3), and crying (1). Hence, the final sample for vowel analysis consisted of 28 infants (16 females, *Mean Age* = 12.22, *SD* = 0.68, *Age range* = 11.51 – 13.77). Regarding the non-native experiment, 13 participants were excluded due to not completing the task due to the caregiver judging the infant to be too upset to continue (8), not habituating (3), and crying (2). Thus, the final sample for consonant analysis included 28 infants (14 females, *Mean Age* = 12.23, *SD* = 0.67, *Age range* = 11.51 – 13.70). Seven of the excluded infants overlapped between the native and non-native experiment.[6]

4.6. Stimuli and procedure

The stimuli and procedure were identical to those in Experiments 1 and 2.

[6] Additional analysis was conducted with only infants who completed both native and non-native tasks (N = 21). These results are reported in Supplementary Material S2.

4.7. Data analysis

The data analysis was conducted in the same way as in Experiment 2.

5. Results

5.1. Native speech perception

The LME results[7] demonstrated a significant main effect of Trial, $F(1, 81) = 18.21$, $p < 0.0001$, *Cohen's d* = .95 (95 % *CIs* [.49, 1.40]) with infants' looking times to the switch trials ($M = 6773$ ms, $SD = 4048$) significantly longer than looking time to same trials ($M = 4552$ ms, $SD = 3393$) (see Table 4 for full model and see Fig. 1). Hence, the results suggest that 11–13-month-old Japanese infants discriminated the vowel length contrast between /mana/ and /ma:na/. Additionally, results showed no significant main effect of Habituation Stimulus on infants' looking time ($F(1, 25) = .001$, $p = .97$, *Cohen's d* = .01 (95 % *CIs* [−.77,.80])), however, there was a significant interaction between Trial and Habituation Stimulus ($F(1, 81) = 5.47$, $p < .05$, *Cohen's d* = .52 (95 % *CIs* [.08,.96])) suggesting an asymmetry in vowel length perception in 11–13-month-old infants. A posthoc Tukey test demonstrated significantly longer looking times to the to the switch ($M = 7389$ ms, $SD = 4364$) compared to the same trials ($M = 4764$ ms, $SD = 3226$) in infants who were habituated with /ma:na/ ($p = .02$), but no such effect in infants habituated with /mana/. Furthermore, the fitted LME model demonstrated no significant main effect of Task Order ($F(1, 25) = .32$, $p = .58$, *Cohen's d* = .23 (95 % *CIs* [−.56, 1.01])), but a significant interaction between Trial and Task Order ($F(1, 81) = 5.16$, $p < .05$, *Cohen's d* = .50 (95 % *CIs* [.06,.95])). The posthoc Tukey test showed that infants who first participated in the non-native experiment exhibited significantly longer looking times to the switch trials ($M = 6299$ ms, $SD = 3860$) compared to the same trials ($M = 3721$ ms, $SD = 2916$), $p = .003$). However, it should be noted that there is an overlap between habituation stimulus that infants heard and the order of experiment. Specifically, 15 infants participated in the native experiment first and were habituated with the long vowel, 11 infants did the non-native experiment first and were habituated with the short vowel, and 2 infants participated in native experiment first and heard the short vowel in habituation phase. Thus, it is hard to disentangle whether the observed effects were due to the habituation stimulus or the order of experiments.

With regards to habituation times, a t-test for independent samples demonstrated a significant difference in infants' mean looking times during habituation phase, $t(25.59) = 2.32$, $p < 0.05$, *Cohen's d* = .92 (95 % *CIs* [.10, 1.72]). Infants who were habituated with long vowels exhibited a longer mean looking time during habituation phase ($n = 15$, $M = 8699.75$ ms, $SD = 2290$) compared to infants habituated with short vowels (short: $n = 13$, $M = 6707.15$ ms, $SD = 2239$). This might suggest that infants were slower in processing the long vowels compared to short vowels.

A paired t-test showed no significant difference in looking times during pretest trials ($M = 15{,}534$ ms, $SD = 3709$) compared to the looking times during the posttest trials ($M = 15{,}981$ ms, $SD = 4150$), $t(27) = .42$, $p = .68$, *Cohen's d* = .08 (95 % *CIs* [−.30,.46]), suggesting that infants did not lose interest in the task or become fatigued during the test phase.

Thus, the results of this experiment demonstrated that 11–13-month-old Japanese infants are able to discriminate the native vowel length contrast. In addition, they exhibited asymmetrical perception with infants perceiving the difference from long-to-short, but not the short-to-long vowel change. Asymmetries were also found in habituation times, where infants took longer to habituate to long than to short vowels.

5.2. Non-native speech perception

The results[8] demonstrated no significant main effects for Trial ($F(1, 81) = 3.97$, $p = .05$, *Cohen's d* = .44 (95 % *CIs* [.00,.88])) suggesting no significant difference in infants' looking time to the switch trials ($M = 3581$ ms, $SD = 2283$) compared to their looking times to the same trials ($M = 4368$ ms, $SD = 3204$) (see Table 5 for full model and see Fig. 2). Also, there was no significant main effect of Habituation Stimulus ($F(1, 25) = 1.04$, $p = .32$, *Cohen's d* = .41 (95 % *CIs* [−.39, 1.20])), and no significant interaction between Trial and Habituation Stimulus ($F(1, 81) = .33$, $p = .57$, *Cohen's d* = .13 (95 % *CIs* [−.31,.56])). Additionally, the results showed no significant main effect of Task Order ($F(1, 25) = .54$, $p = .47$, *Cohen's d* = .30 (95 % *CIs* [−.50, 1.08])) and no significant interaction between Trial and Task Order ($F(1, 82) = .33$, $p = .57$, *Cohen's d* = .13 (95 % *CIs* [−.31,.56])).

Next, regarding the habituation times, a t-test for independent samples demonstrated no difference in infants' mean looking times during habituation phase between infants habituated with /ɹ/ versus /l/ sound (/ɹ/: $n = 13$, $M = 5733.6$ ms, $SD = 1655$; /l/: $n = 15$, $M = 5287.6$ ms, $SD = 1696$), $t(25.60) = -.70$, $p = .49$, *Cohen's d* = −.28 (95 % *CIs* [−1.05,.50]).

Furthermore, a paired t-test demonstrated no significant difference in infants' looking time during pretest ($M = 14{,}040$ ms, $SD = 4765$) compared to the posttest phase ($M = 13{,}572$ ms, $SD = 5365$), $t(27) = -.51$, $p = .61$, *Cohen's d* = −.10 (95 % *CIs* −.48,.28]).

5.3. Correlation between native and non-native speech perception

There was no significant correlation between performance on native versus performance on non-native speech perception, $r(19) =$

[7] AIC for model with random intercept for participant was 2061.60 and AIC for model with Trial by-participant slope was 2050.98.

[8] Based on AIC model with random intercept for participant (AIC: 1995.60) was better fit than model with Trial by-participant slope (AIC: 1995.70).

Table 4
Summary of LME model fitted only with infant with Trial (same vs. switch), Habituation Stimulus (short vs. long vowel), and Task Order (native vs. non-native) and their interaction as the independent variables, a random intercept for Participants, and infants' looking time as the dependent variable.

Predictors	*β*	*SE*	*t*	*CI*	*p*
(Intercept)	945.49	2461.33	0.38	-3935.42–5826.39	0.70
Trial [Switch]	7988.25	2490.99	3.21	3048.52–12,927.98	**< 0.01**
Habituation stimulus [short]	2775.47	2284.86	1.22	-1755.51–7306.44	0.23
Task order [native first]	3818.55	2333.23	1.64	-808.33–8445.42	0.11
Trial [Switch] * Habituation stimulus [short]	-5409.75	2312.40	-2.34	-9995.33–-824.17	**< 0.05**
Trial [Switch] * Task order [native first]	-5363.25	2361.35	-2.27	- 10,045.89–-680.61	**< 0.05**
Marginal R^2 / Conditional R^2				0.13 / 0.41	

Table 5
Summary of LME model fitted with Trial (same vs. switch), Habituation Stimulus ("right" /ɹaɪt / vs. "light" /laɪt/), and Task Order (native vs. non-native) and their interaction as the independent variables, a random intercept for Participants, and infants' looking time as the dependent variable.

Predictors	*β*	*SE*	*t*	*CI*	*p*
(Intercept)	4276.92	657.4	6.51	2973.28–5580.56	**< 0.001**
Trial [Switch]	-854.58	606.34	-1.41	-2056.96–347.81	0.16
Habituation stimulus ["right" /ɹaɪt/]	2106.04	1800.35	1.17	-1464.13–5676.21	0.25
Task order [native first]	-1655.92	1800.35	-0.92	-5226.09–1914.25	0.36
Trial [Switch] * Habituation stimulus ["right" /ɹaɪt/]	-953.69	1660.52	-0.57	-4246.57–2339.18	0.57
Trial [Switch] * Task order [native first]	953.58	1660.52	0.57	-2339.30–4246.45	0.57
Marginal R^2 / Conditional R^2				0.04 / 0.43	

0.27, $p = .23$.

Thus, the results suggest that 11–13-month-old Japanese infants could not discriminate the non-native phoneme contrast /ɹ/ - /l/ and did not exhibit asymmetry in their non-native consonants discrimination.

6. Discussion

The current study assessed the developmental patterns of native and non-native speech perception in monolingual Japanese-learning infants from 6 to 12 months of age. Native speech perception was assessed via the long-short vowel contrast in 6–7- (Experiment 1), 9–11- (Experiment 2), and 11–13-month-old infants (Experiment 3). Results confirmed previous findings of an improvement in native discrimination, though at a later time-point than previously reported (Mugitani et al., 2009; Sato et al., 2010). Non-native speech perception was assessed via the North American English /ɹ/ - /l/ consonant contrast at 9–11 (Experiment 2) and 11–13 months of age (Experiment 3). Results demonstrated discrimination at 9–11, but not at 11–13 months of age corroborating the previous findings (Kuhl et al., 2006; Tsushima et al., 1994). In the following paragraphs, these results will be discussed in detail.

Regarding native speech perception, infants at 6–7 and 9–11 months of age failed to discriminate the difference between short and long vowels, whereas they were able to discriminate the same contrast at 11–13 months. These results accord previous findings demonstrating a facilitation in Japanese infants' discrimination of the native vowel length contrast during the first year of life (Sato et al., 2010).

Compared to earlier findings (Mugitani et al., 2009; Sato et al., 2010), infants in the present study acquired the long-short vowel distinction at a slightly later age (between 11–13 as opposed to between 7.5–9 months of age). A possible explanation for this might be cohort effects (Schaie, 1965), a proposal that social, cultural, or historical events of any given era systematically affect infants' development. Social factors relevant to the present study are: being born during the covid19 pandemic (Charney et al., 2021), excessive exposure to various electronic devices (Madigan et al., 2020), use of media devices by caregivers during caregiver-infant interaction (Nomkin & Gordon, 2021), and a lower number of breastfeeding mothers (Lovcevic, 2023). These factors have been found to cause delays in various aspects of children's language development during the first years of life, and might have affected the development of speech perception, one of the first developmental milestones in language development. Although we did not collect data on these factors, these societal changes should be considered when comparing the results of current developmental studies with the studies conducted a more than a decade ago. Another possible explanation for the absence of vowel length discrimination at 9-11 months might be the low power of infant studies, especially in studies with younger infants (Bergmann et al., 2018), that might lead to variation in the results of statistical evaluations between samples. Both possible explanations illustrate the importance of replicating infant results in order to reach generalizable conclusions.

Regarding non-native speech perception, our results demonstrated a decrease in infants' ability to discriminate the non-native /ɹ/ - /l/ consonant contrast at around 11 months of age. Hence, these findings are consistent with a large body of evidence indicating discrimination of non-native consonant contrasts in younger infants without previous exposure that declines around 12 months of age (Best & McRoberts, 2003; Best et al., 1995; Kuhl et al., 2006; Tsao et al., 2006; Tsushima et al., 1994; Werker & Tees, 1984).

To our knowledge, this is one of the rare studies that tested both native and non-native speech perception in one group of Japanese-learning infants, allowing a parallel assessment of native and non-native speech perception. Whereas our results suggest an increase in native and a decrease in non-native speech perception, we did not find a significant correlation between infants' performance on the native and the non-native speech perception tasks. This does not necessarily mean that there is no association between native and non-native speech perception, but it might be due to the cross-sectional nature of the study design. Indeed, due to the cross-sectional design we could not assess a correlation in decline versus improvement in speech perception which might provide more insights on the association between native and non-native speech perception over development.

Previous evidence demonstrated the existence of perceptual asymmetries in both vowel and consonant perception evident in easier discrimination of vowel or consonant quality contrasts presented in one direction compared to discrimination of the same contrast presented in the opposite direction (Kuhl et al., 2006; Polka & Bohn, 2011; Tsuji et al., 2015). Our results demonstrated asymmetries in native speech perception in 11–13-month-old Japanese-learning infants evident in easier discrimination of the change from long-to-short vowel compared to change from short-to-long vowel. Different from qualitative contrasts, which have often been found early on, this asymmetry in vowel quantity perception has previously been suggested to stem from experience-dependent development from more acoustically based towards more phonologically based perception. Mugitani et al. (2009) demonstrated symmetrical perception of the vowel length contrast in 10-month-old Japanese-learning infants, and asymmetrical discrimination in 18-month-old infants. Their interpretation of developing phonological discrimination of the vowel length contrast is consistent with near infrared spectroscopy evidence indicating bilateral hemodynamic activation during the acoustic processing of the vowel length contrast in younger infants, with phonemic discrimination appearing around 13 months of age evident in left dominant hemodynamic activation (Minagawa-Kawai et al., 2007).

The asymmetry within this framework could be interpreted such that the short vowel acts as a reference point in the discrimination of the vowel length contrast, making it easier to detect the change towards it than vice versa. It is possible that the exposure to a much higher quantity of short compared to long vowels in Japanese infants' input (Bion et al., 2013) leads to this status as referent points.

Our additional results of longer habituation times to long compared to short vowels in 11–13-month-olds are consistent with this interpretation, such that these longer habituation times might be related to slower processing speed (Poli et al., 2024) with infants needing more time to process the less encountered and more variable long compared to the frequently encountered short vowel stimuli.

The methodological contribution of the current study consists in the completion of two habituation sessions during one lab visit. This allowed an assessment of both native and non-native speech perception in one group of infants all acquiring the same native language. This, in turn, enabled us to examine a parallel developmental pattern in infants' native and non-native speech perception. Our results demonstrated that the order of task presentation (native first versus non-native first) did not influence infants' performance in the test phase. This suggests that it is not only possible to conduct the two habituation sessions in one visit, but that infants' performance on a subsequent task is not influenced by their participation in the first task.

A final contribution of the present study was to replicate the effects found in only two previous studies each, which, given the low power and resulting low replicability of infant perception studies (Bergmann et al., 2018) is a worthwhile endeavour. At the same time, we contribute data from a relatively underrepresented population, given that most research has been limited to "Minority World Countries" (Alam, 2008) with only 10 % of speech perception research coming from "Majority World Countries" including Asian countries although they represent 60 % of world population (Singh et al., 2022). Furthermore, our study contributes to perception of vowel length research given that only 2 % of vowel perception studies assess vowel length with majority of studies focusing on vowel quality perception (Singh et al., 2022).

One of the limitations of the current study is testing two different types of contrasts, a non-native qualitative consonant contrast, and a native quantitative vowel contrast. Despite these differences, we chose these two contrasts since there is a documented decline in sensitivity for the former, and a documented improvement in sensitivity for the latter in the previous literature on Japanese infants, providing a good test case for assessing perceptual attunement to native and non-native contrasts. Moreover, although perceptual attunement generally is documented to occur earlier for vowel quality contrasts (around 6 months, but see Götz et al., 2024) compared to consonant quality contrasts (between 10 and 12 months) (Best & McRoberts, 2003; Polka & Bohn, 1996), previous studies in Japanese have already documented that this timeline is delayed with regard to the Japanese phonemic vowel length contrast (Mugitani et al., 2009; Sato et al., 2010), aligning the timeline of attunement in the two chosen contrasts.

7. Conclusion

The present study was designed to assess the two defining features of perceptual attunement, the decline in non-native and the improvement in native speech perception during the first year of life. Unlike most previous studies that examined these in infants from distinct language backgrounds, we assessed both non-native and native sound discrimination in infants from the same language background (Japanese). The results demonstrated a decrease in non-native consonant contrast discrimination at around 11 months of age, and a parallel increase in native vowel length discrimination. Hence, our results support the theoretical proposals of perceptual attunement during the first year of life marked by a decrease in non-native and a parallel increase in native speech perception. Furthermore, the present results indicate the feasibility of testing both non-native and native speech perception in one session providing the pathway for future studies on perceptual attunement and its relations with infants' language development.

CRediT authorship contribution statement

Sho Tsuji: Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Irena Lovčević:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare no competing interests.

Data availability

The study datasets and analysis script can be accessed via the following link: https://osf.io/jkech/?view_only=198c20a1ec68465581d4c0678c5fdb44..

Acknowledgements

This research was funded by MEXT World Premier International Research Center Initiative startup grant to IL, and JSPS KAKENHI Grant Numbers 20H05617 and 20H05919 to ST. We would like to thank Professor Reiko Mazuka and Laboratory for Language Development at RIKEN Brain Science Institute for providing us with Japanese stimuli, Dr Monika Barbir for recording of the English stimuli, and Tomoko Takashima for recording of the pretest/posttest stimuli. Furthermore, we thank lab managers who contributed to data collection, especially Rie Ono and Tomoko Takashima, and the participating families for their valuable time. We have no conflicts of interest to disclose.

References

Alam, S. (2008). Majority world: Challenging the West's rhetoric of democracy. *Amerasia Journal, 34*(1), 88–98. https://doi.org/10.17953/amer.34.1.l3176027k4q614v5

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 48. https://doi.org/10.18637/jss.v067.i01

Ben-Shachar, M. S., Makowski, D., Lüdecke, D., Kelley, K., and Stanley, D. (2021). Package 'Effectsize'. Available at: https://cran.r-project.org/web/packages/effectsize/effectsize.pdf.

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development, 89*(6), 1996–2009. https://doi.org/10.1111/cdev.13079

Best, C. T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In J. Goodman, & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Timonium, MD: York: Press.

Best, C. C., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech, 46*(2–3), 183–216. https://doi.org/10.1177/00238309030460020701

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant behavior and development, 18*(3), 339–350. https://doi.org/10.1016/0163-6383(95)90022-5

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*(3), 305–330. https://doi.org/10.1016/S0095-4470(19)30637-0

Bion, R. A. H., Miyazawa, K., Kikuchi, H., & Mazuka, R. (2013). Learning phonemic vowel length from naturalistic recordings of Japanese infant-directed speech. *PLoS ONE, 8*(2), Article e51594. https://doi.org/10.1371/journal.pone.0051594

Burnham, D. K., Earnshaw, L. J., & Clark, J. E. (1991). Development of categorical identification of native and non-native bilabial stops: infants, children and adults. *Journal of Child Language, 18*(2), 231–260. https://doi.org/10.1017/S0305000900011041

Charney, S. A., Camarata, S. M., & Chern, A. (2021). Potential impact of the COVID-19 pandemic on communication and language skills in children. *Otolaryngology–Head and Neck Surgery, 165*(1), 1–2. https://doi.org/10.1177/0194599820978247

Choi, Y., Nam, M., Yamane, N., & Mazuka, R. (2023). Lack of early sensitivity and gradual emergence of native phoneme categories: A pattern from underrepresented language learners. *Developmental Science.* , Article e13422. https://doi.org/10.1111/desc.13422

Curtin, S., & Werker, J. F. (2007). Perceptual foundations of phonological development. In M. Gareth Gaskell, G. T. M. Altmann, P. Bloom, A. Caramazza, & P. Levelt (Eds.), *Oxford handbook of psycholinguistics*. New York: Oxford University Press.

Diamond, A., Werker, J. F., & Lalonde, C. (1994). Toward understanding commonalities in the development of object search, detour navigation, categorization, and speech perception. *Human Behavior and the Developing Brain* (pp. 380–426). New York: Guilford Press.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22*(1), 567–631. https://doi.org/10.1146/annurev.neuro.22.1.567

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Götz, A., Krasotkina, A., Schwarzer, G., & Höhle, B. (2024). Asymmetries in Infants' Vowel Perception: Changes in Vowel Discrimination in German Learning 6- and 9-Month-Old Infants. *Language and Speech, 0*(0). https://doi.org/10.1177/00238309241228237

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology, 4*(6), 812–822. https://doi.org/10.1016/0959-4388(94)90128-7

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences U.S.A., 97*, 11850–11857.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience, 5*(11), 831–843.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period". *Language Learning and Development, 1*(3–4), 237–264. https://doi.org/10.1080/15475441.2005.9671948

Kuhl, P., & Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annual Review of Neuroscience, 31*(1), 511–534. https://doi.org/10.1146/annurev.neuro.30.051606.094321

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Ladefoged, P., & Maddieson, I. (2001). Vowels of the world's languages. *Phonology: Critical Concepts in Linguistics*, 5–39.

Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development, 18*(4), 459–475. https://doi.org/10.1016/0163-6383(95)90035-7

Lenth, R. (2019). Emmeans: Estimated marginal means, aka least-squares means. Available at https://cran.r-project.org/web/packages/emmeans/index.html.

Liu, L., & Kager, R. (2016). Perception of a native vowel contrast by Dutch monolingual and bilingual infants: A bilingual perceptual lead. *International Journal of Bilingualism, 20*(3), 335–345. https://doi.org/10.1177/1367006914566082

Lovcevic, I. (2023). Associations of breastfeeding duration and cognitive development from childhood to middle adolescence. *Acta Paediatrica, 112*(8), 1696–1705. https://doi.org/10.1111/apa.16837

Madigan, S., McArthur, B. A., Anhorn, C., Eirich, R., & Christakis, D. A. (2020). Associations between screen use and child language skills: A systematic review and meta-analysis. *JAMA Pediatrics, 174*(7), 665. https://doi.org/10.1001/jamapediatrics.2020.0327

Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *Journal of Neuroscience, 27*(2), 315–321.

Mugitani, R., Pons, F., Fais, L., Dietrich, C., Werker, J. F., & Amano, S. (2009). Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology, 45*(1), 236–247. https://doi.org/10.1037/a0014043

Nomkin, L. G., & Gordon, I. (2021). The relationship between maternal smartphone use, physiological responses, and gaze patterns during breastfeeding and face-to-face interactions with infant. *PLOS ONE, 16*(10), Article e0257956. https://doi.org/10.1371/journal.pone.0257956

Oakes, L. M., Sperka, D., DeBolt, M. C., & Cantrell, L. M. (2019). Habit2: A stand-alone software solution for presenting stimuli and recording infant looking times in order to study infant development. *Behavior Research Methods, 51*, 1943–1952.

Poli, F., Ghilardi, T., Beijers, R., de Weerth, C., Hinne, M., Mars, R. B., & Hunnius, S. (2024). Individual differences in processing speed and curiosity explain infant habituation and dishabituation performance. *Developmental Science, 27*(3), Article e13460. https://doi.org/10.1111/desc.13460

Polka, L., & Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America, 100*(1), 577–592. https://doi.org/10.1121/1.415884

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics, 39*(4), 467–478. https://doi.org/10.1016/j.wocn.2010.08.007

R Core Team R. (2020). *A language and environment for statistical computing*. R Foundation for Statistical Computing.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology, 46*(1), 106–119. https://doi.org/10.1037/a0016718

Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64*(2), 92–107. https://doi.org/10.1037/h0022371

Singh, L., Rajendra, S. J., & Mazuka, R. (2022). Diversity and representation in studies of infant perceptual narrowing. *Child Development Perspectives, 16*(4), 191–199. https://doi.org/10.1111/cdep.12468

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*(6640), 381–382.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America, 120*(4), 2285–2294. https://doi.org/10.1121/1.2338290

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology, 56*(2), 179–191. https://doi.org/10.1002/dev.21179

Tsuji, S., Mazuka, R., Cristia, A., & Fikkert, P. (2015). Even at 4 months, a labial is a good enough coronal, but not vice versa. *Cognition, 134*, 252–256. https://doi.org/10.1016/j.cognition.2014.10.009

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: Language-specific developmental changes in speech perception abilities. *3rd International Conference on Spoken Language Processing (ICSLP 1994)*, 1695–1698. https://doi.org/10.21437/ICSLP.1994-438

Werker, J. F. (1995). Exploring developmental changes in cross-language speech perception. In L. R. Gleitman, & M. Liberman (Eds.), *Language* (pp. 87–106). MIT Press.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197–234. https://doi.org/10.1080/15475441.2005.9684216

Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, and implications* (pp. 285–311). York Timonium, MD: York: Press.

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