

Reassessing the labial-coronal asymmetry in infants' discrimination of medial consonants

Sharon Peperkamp

CNRS & Ecole normale supérieure – Paris Sciences et Lettres, Paris, France

Megha Sundara

UCLA, Los Angeles, USA

Abstract: Asymmetric speech sound discrimination has been observed in both adults and infants. Here we focus on the labial-coronal asymmetry documented in Dutch- and Japanese-learning infants, who successfully discriminate between /ɔmpa/ and /ɔnta/ only after habituation to /ɔmpa/ (Tsuji et al., 2015). We hypothesized that this asymmetry is due to the inherent greater intensity of coronal compared to labial stop consonants. However, testing 64 French- and English-learning five-month-olds with Tsuji et al.'s stimuli, we observed successful discrimination regardless of language and of habituation condition, with only a numeric asymmetry in the predicted direction, i.e. a numerically larger looking time difference between switch and same trials after habituation to /ɔmpa/ than to /ɔnta/. So, we propose alternative ways to test our hypothesis in future research.

Keywords: infant speech perception; consonant discrimination; labial-coronal asymmetry; English; French.

Corresponding author: Sharon Peperkamp, Laboratoire de Sciences Cognitives et Psycholinguistique (ENS-PSL, EHESS, CNRS), 29 Rue d'Ulm, 75005, Paris, France. Email: sharon.peperkamp@ens.psl.eu.

ORCID IDs: Sharon Peperkamp: <https://orcid.org/0000-0001-5985-8878>; Megha Sundara: <https://orcid.org/0000-0003-0028-750X>

Citation: Peperkamp, S., & Sundara, M. (2026). Reassessing the labial-coronal asymmetry in infants' discrimination of medial consonants. *Language Development Research*, 6(1), 1–20. <http://doi.org/10.34842/lvr2025-971>

Introduction

Asymmetries in speech sound discrimination have been reported in both adults and infants. For vowel contrasts, many studies have reported better discrimination from less to more focal vowels than in the opposite direction (for a review of infant studies, see Polka & Bohn, 2011). It has been argued that such asymmetries can be attributed to the phonetic properties privileging one of the two sounds. For instance, some (focal) vowels show formant frequency convergence, which produces salient prominences in the vowel spectra; and such vowels are more common across language inventories (Schwartz et al., 2005). Following Schwartz et al., Polka and Bohn (2011) argue that asymmetries in vowel discrimination are attributable to the fact that focal vowels act as perceptual anchors and therefore discrimination is harder when listeners' attention is drawn away from focal vowels. Meta-analyses of infant (Tsuji & Cristia, 2014) and adult (Polka et al., 2019) data provide evidence supporting this account.

Asymmetries in infants' consonant discrimination have been reported as well (e.g., Kuhl et al., 2006; Tsuji et al., 2015; Nam & Polka, 2016). In this paper, our goal was to evaluate whether one such asymmetry in the discrimination of medial consonant place of articulation can also be attributed to the phonetic properties, this time of coronal stop consonants. To do so, we focused on the labial-coronal asymmetry documented by Tsuji et al. (2015). Tsuji et al. (2015) tested Japanese- and Dutch-learning 4-to-6-month-olds. Infants in both language groups successfully discriminated after habituation to labial /ɔmpa/ but not after habituation to coronal /ɔnta/. In most languages, including Dutch, coronals are more frequent than labials, but in Japanese labials are more frequent than coronals. Thus, these results rule out that the asymmetry is due to a difference in input frequency.

Instead, Tsuji et al. (2015) raise two other possibilities. One concerns differences in token distributions of labials compared to coronals. In particular, coronal consonants typically show more variability in their exact place of articulation (with acoustic consequences) than labials. The overlap of the two categories therefore contains a larger percentage of coronal than of labial tokens. In other words, a coronal token has a higher chance of being misidentified as a labial one than vice versa, and as a result, it should be more difficult to distinguish labials from coronals than vice versa (see also Bien & Zwitserlood, 2013). The other possibility concerns a difference in acoustic salience, with coronals considered to be more salient than labials. Asymmetric patterns in infants' consonant discrimination, between /la/ and /ra/ (Kuhl et al., 2006) and between /vas/ and /bas/ (Nam & Polka, 2016) have also been attributed to the greater salience of one of the categories (/r/ and /b/, respectively).

Here, we set out to test whether there is a phonetic basis for the salience-based account for the labial-coronal asymmetry: We hypothesized that this asymmetry is due to the inherent greater intensity of coronal compared to labial stop consonants. Although

labials and coronals are typically distinguished by spectral cues, specifically the shape of the burst spectrum, the amplitude of the release burst is also used by listeners to distinguish them (Blumstein & Stevens, 1979; Chodroff & Wilson, 2014; Edwards, 1981; Ohde & Stevens, 1983; Repp & Lin, 1989; Winitz et al., 1972). Bilabials have the lowest amplitude with respect to the following vowel resulting from the differences in intra-oral pressure due to the volume of the cavity and rate of airflow after the release of the closure (Stevens et al., 1999; Edwards, 1981).

We preregistered two experiments with English- and French-learning 5-month-olds, one with the original stimuli of Tsuji et al. (2015) to replicate and extend the finding that infants are able to discriminate /ɔmpa/ from /ɔnta/ but only after habituation with the labial /ɔmpa/, and one – to be run only if in the first one we indeed found the predicted discrimination asymmetry – with the same stimuli but in which the intensity profiles of the stop consonants in /ɔmpa/ and /ɔnta/ tokens have been swapped. Thus, while acoustic measurements can show that the labial and coronal consonants in the stimuli differ in intensity, in accordance with a salience-based account, our aim was to show that the asymmetry is because of this difference in intensity. Therefore, we predicted that with the swapped intensity profiles the asymmetry would switch direction, with successful discrimination after habituation to /ɔnta/ but not after habituation to /ɔmpa/. To preview the results, in the first experiment we observed successful discrimination regardless of language and of habituation condition, with only a numeric asymmetry in the predicted direction. As preregistered, we therefore did not proceed to the second experiment. In the discussion, we will comment on the magnitude of perceptual asymmetries that are rooted in phonetic substance and discuss the results in the context of what we know about infants' discrimination of medial consonant place distinctions.

Method

Participants

The final sample consisted of 64 5-month-olds (English: N = 32, 13 girls, mean age: 4;27, range: 4;10 – 5;16; French: N = 32, 13 girls, mean age: 4;29, range: 4;18 – 5;24). This sample size is based on Tsuji et al. (2015), who tested 64 infants divided into four groups of equal size (two languages: Dutch- and Japanese learning, and two age groups: 4 and 6 months). Thus, we tested the same number of infants, also in two language groups but at a single, intermediate, age.

Except for one English infant who was born at 36 weeks (tested at 4;27), all infants were full term. Like the infants tested by Tsuji et al. (2015), all English-learning infants were from monolingual families; according to a parental questionnaire they had at least 90% exposure to English. The French-learning infants had a more mixed language exposure: their exposure to French ranged from 20% to 100% (mean: 83%); languages with

more than 10% additional exposure were English (N=6), Italian (N=3), Spanish (N=2), Swiss German (N=1), Albanian (N=1), and Russian (N=1). Two additional French-learning infants were excluded and replaced because they were outliers: the absolute difference in their looking times to switch and same trials was more than 2 standard deviations above the group mean (both looked longer at the switch than the same trials). Twenty-two further infants (10 English) were tested but excluded due to experimenter error (1), technical problem (2), parental interference (4), fussiness/crying (4), failure to look at the screen (5), not habituating (4), born preterm (1), too old (1).

Stimuli

For the pre- and post-test, we used multiple tokens of /pɔk/, produced by a native French speaker in an infant-directed register. For the habituation and test phase, we used the stimuli from Tsuji et al. (2015). They included eight unique tokens of /ɔmpa/ and eight of /ɔnta/, produced by a native Dutch speaker in an infant-directed register and matched on duration and vowel formant values. Five unique tokens of each item were used for habituation, and five – including two also used for habituation – for testing. (While the habituation-test procedure typically uses only different tokens for the test phase, Tsuji et al. (2015) argued that the presence of a few habituated tokens helps exclude the possibility that dishabituation is due to the presence of only novel tokens.)

The acoustic characteristics of the stimuli in terms of duration, pitch and vowel formant frequencies are presented in Table 1 in Tsuji et al. (2015). Most importantly, the tokens of /ɔmpa/ and /ɔnta/ were similar in mean duration (/ɔmpa/: 773 ms, $SD = 46$; /ɔnta/: 740 ms, $SD = 16$; $p = .07$) and mean pitch (/ɔmpa/: 6.56 Erb, $SD = 0.54$; /ɔnta/: 6.16 Erb, $SD = 0.39$; $p > 1$), but differed substantially in the F2 of the second vowel (/ɔmpa/: 11.58 Bark, $SD = 0.07$; /ɔnta/: 11.81 Bark, $SD = 0.04$; $p < .0001$), a cue that typically distinguishes stops and nasals differing in place of articulation. Additional analyses concerning the amplitude of the stop burst (measured by absolute and relative intensity with respect to the following vowel) as measured in PRAAT (Boersma & Weenink, 2009) are shown in Table 1. Importantly, the relative burst intensity of /t/ in /ɔnta/ tokens was larger than that of /p/ in /ɔmpa/ tokens, $t(30) = 10.49$, $p < .0001$.

Table 1. Intensity profile of /ɔmpa/ and /ɔnta/ tokens

	/ɔmpa/ mean (SD)	/ɔnta/ mean (SD)
Mean stop burst intensity (dB)	58.6 (1.7)	66.3 (2.5)
Mean vowel /a/ intensity (dB)	80.0 (1.7)	79.7 (1.4)
Mean relative burst intensity (dB)	-21.4 (2.7)	-13.4 (1.5)

Note that Tsuji et al. (2015, supplementary material) report no differences between the

spectral or duration measures or even absolute intensity of the burst between the labials and coronals. We suspect that the critical difference is that we normalized the burst amplitude with respect to the intensity of the following vowel, providing a fine-grained measure of the burst amplitude. Normalization of burst intensity, typically measured as rms amplitude, is recommended in acoustic phonetics (Edwards, 1981; Zue, 1976), likely because measurement of absolute burst intensity can be easily affected even with small changes in microphone placement.

Procedure

Infants were tested using the central fixation procedure (Werker et al., 1998), the English-learning ones in Los Angeles and the French-learning ones in Paris. Stimulus presentation was controlled by Habit2 (Oakes et al., 2019). The infant sat on a caregiver's lap, facing a screen, in a sound-attenuated booth. Loudspeakers were located on each side of the screen, and a video camera filmed the infant's behavior. An experimenter observed the infant from an adjacent control room on a monitor connected to the camera and coded infants' gaze towards or away from the screen online. The audio stimuli were played at ~60 dBspl. In Los Angeles, both the caregiver and the experimenter wore sound-attenuating headphones with masking music; in Paris, masking for the caregiver was the same, while the experimenter wore noise-cancelling headphones.

All trials were completely infant-controlled, with the presentation of the audio stimulus being contingent on the infant's looks to a colorful, static, checkerboard. Trials started with an attention getter, consisting of colored appearing and disappearing geometrical figures against a black background. Once the infants looked at it, it was replaced by the checkerboard and the audio started. Trials ended when the infant looked away from the screen for more than one second or when the maximum duration, i.e. 19 sec., was reached. Trials with listening times less than two seconds were repeated. Time looked away and time during trials that were repeated was not included in the listening times.

The experiment consisted of four phases. In the pre-test phase, infants heard one trial of repetitions of /pɔk/, for a maximum of 19 seconds. During the habituation and test phase, trials lasted a maximum of 14 seconds, with a 1 sec inter-stimulus interval. During habituation, half the infants listened to /ɔmpa/-stimuli and the other half to /ɔnta/-stimuli. Once infants' looking time declined to 60% of the looking time to the four longest consecutive trials, they entered the test phase, consisting of two same and two switch trials. Same and switch trials alternated, but half of the infants first heard a same trial and the other half first heard a switch trial. Finally, in the last, post-test phase, infants heard the same /pɔk/-stimuli that were presented during pre-test.

Note that Tsuji et al. (2015) calculated habituation criteria with respect to the mean of

the first four trials. This had escaped our attention when analyzing their procedure and we thus failed to mention this difference in our preregistration. Comparing habituation to the four longest trials provides a more conservative estimate of infant habituation in contrast to one where comparisons are made to the first four trials, because sometimes infants have very short looks at the beginning of the experiment while they are learning the contingency between looking and listening. Thus, the habituation results reported here are more stringent than the ones in their study.

Results

Infants' mean listening times during pre-test, post-test and habituation trials, as well as the mean number of trials to habituation are presented in Table 2. A comparison with the looking times in Tsuji et al. (2015) is provided in Table S1 of the Supplementary material.

Table 2. Mean listening times during pre-test, post-test and habituation trials in seconds, as well as the mean number of trials to habituation. Standard errors are shown in parentheses.

	French <i>mean (SD)</i>	English <i>mean (SD)</i>
Pre-test	12.6 (1.09)	13.4 (1.01)
Habituation time		
labial	98 (11.1)	117 (16.7)
coronal	100 (10.9)	95 (8.87)
Trials to habituate		
labial	11.9 (1.38)	12.5 (1.38)
coronal	11.8 (1.00)	11.3 (0.90)
Post-test	6.00 (0.63)	6.50 (0.86)

Mean looking times to the post-test were shorter than those to the pretest ($t = 8.28$, $p < .0001$), and a majority of infants ($N=36$) failed to dishabituate to the post-test trial, i.e. their listening time for this trial was not longer than the mean of their listening times for the last four habituation trials. Due to their large number, we diverged from the preregistration and did not exclude data from these infants.

The times to habituate were submitted to a linear mixed model using the *lme4* package (Bates et al., 2015) in the R environment (R Core Team, 2021), with contrast-coded fixed factors Language (French vs English), Condition (labial vs coronal habituation), and their interaction, as well as a random intercept for Participant. Neither of the main effects nor the interaction was significant (all $|t| < 1$). The numbers of habituation trials were submitted to a linear model with the same factors. Again, neither of the main

effects nor the interaction was significant (all $|t| < 1$). Thus, across languages and across habituation conditions, infants did not differ either in their total time to habituate or in their number of habituation trials. Listening times to same and switch trials during the test phase are presented in Figure 1.

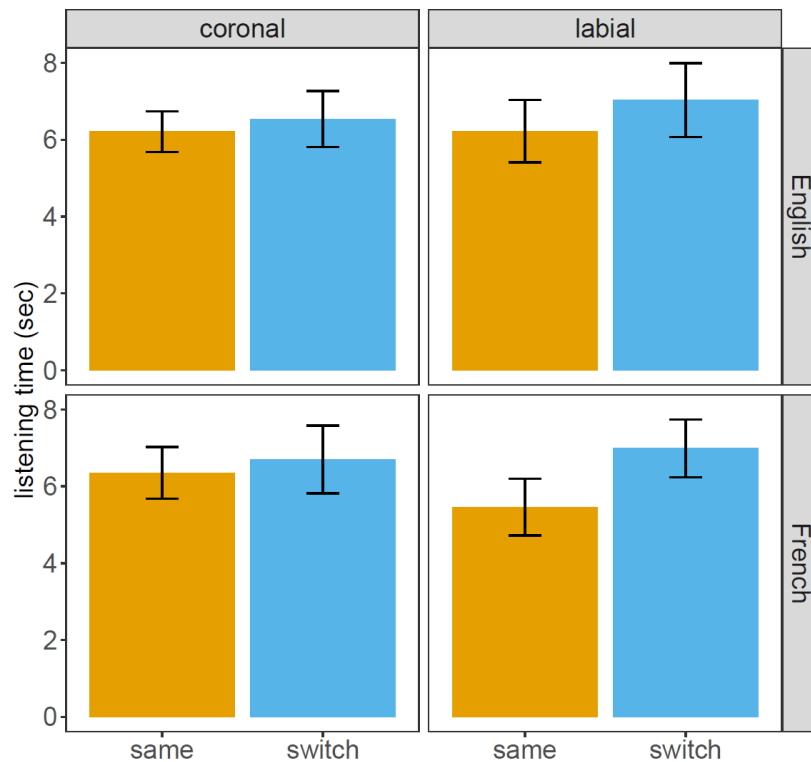


Figure 1. Mean listening times to same and switch trials as a function of habituation condition and language group.

We start with the preregistered analysis. The listening times were analyzed in a linear mixed-effects model with contrast-coded fixed factors Trial Type (same vs switch), Condition (labial vs coronal habituation), Language (French vs English), and the 2-way and 3-way interactions, as well as a random intercept for Participant (but no slope for Trial Type as it yielded a singular fit). Statistical significance was established by way of model comparison, using likelihood ratio tests. As Language did not interact with Trial Type in any way (all $|t| < 1$), we removed this factor and its interactions in a new model. This model revealed a marginal effect of Trial Type ($\beta = -0.38$, $SE = 0.21$, $t = -1.82$, $\chi^2 = 3.29$, $p = .070$), with longer listening times to switch than to same trials. The effect size was small, Cohen's $d = 0.29$. Neither the effect of Condition nor the Trial Type \times Condition interaction were significant (both $t < 1$). Thus, while the marginal effect of Trial Type indicates a trend towards discrimination, the absence of a Trial Type \times Condition

interaction shows that there is no evidence in support of discrimination being asymmetric, although it is evident from the figure that there is a numerical trend in the direction reported by Tsuji et al. (2015).

Infant experiments with a habituation paradigm typically present two rather than four test trials. Given the large number of infants who did not dishabituate to the post-test, the test phase may have been too long for infants to remain on task. We therefore additionally ran a non-preregistered analysis of infants' listening times during the first two test trials. Listening times to the first same and switch trial are presented in Figure 2.

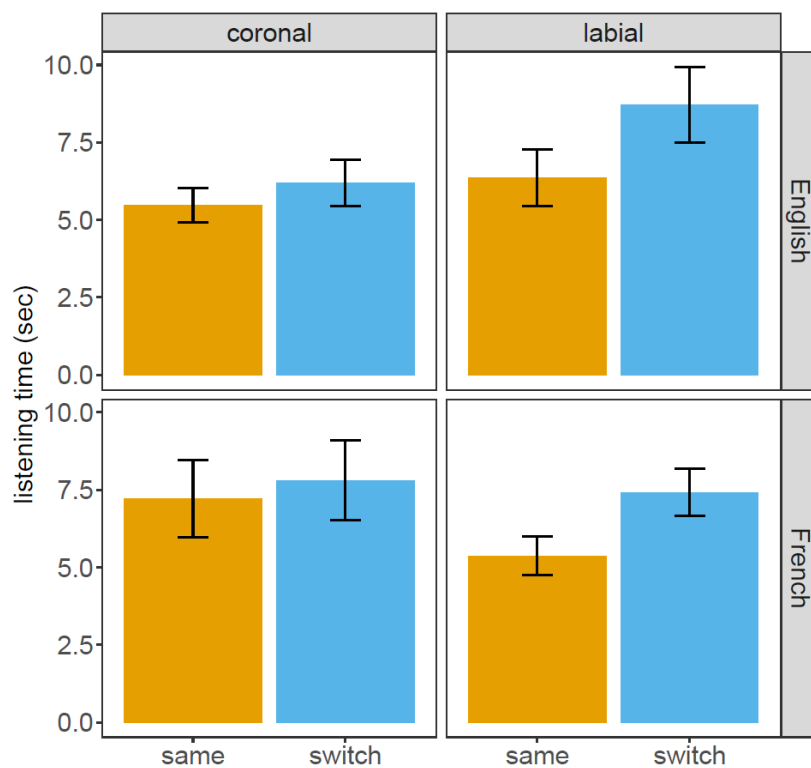


Figure 2. Mean listening times to the first same and switch trial as a function of habituation condition and language group

A model with the same structure as before revealed a significant effect of Trial Type ($\beta = -0.71$, $SE = 0.26$, $t = -2.78$, $\chi^2 = 7.30$, $p < .007$). The size of the effect was medium, Cohen's $d = 0.34$. There was also a marginal interaction between Language and Condition ($\beta = -0.71$, $SE = 0.39$, $t = -1.82$, $\chi^2 = 3.24$, $p = .072$). Post-hoc analyses using the *emmeans* package (Lenth, 2025) revealed that French infants habituated to /ɔnta/ had numerically longer listening times in the test phase than those habituated to /ɔmpa/ ($t < 1$), while for English infants the pattern was reversed ($\beta = -1.71$, $SE = 1.14$, $t = -1.51$, p

> .1). No other main effect or interaction reached significance (all $|t| < 1$, except for the Trial Type \times Condition interaction: $\beta = 0.39$, $SE = 0.26$, $t = 1.51$, $\chi^2 = 2.23$, $p > .1$). Thus, these analyses of just the first two trials show robust evidence of discrimination for medial consonant place distinctions and again no sign of discrimination being asymmetric, despite there being a greater numerical difference in the right direction as seen in Figure 2.¹

Finally, given the numerical listening time difference in the direction of the labial-coronal asymmetry predicted by Tsuji et al. (2015), meta-analysis can determine whether or not there is an overall asymmetry in the experiments carried out so far. Proposed and carried out by Sho Tsuji using the *metafor* package (Viechtbauer, 2010), this meta-analysis combines the data from her four groups, consisting of 4- and 6-month-old Japanese and Dutch-learning infants, with those from our two groups, consisting of 5-month-old English and French learning infants. All listening times were scaled, and the mean listening times to same and switch trials were entered. First, standardized coefficients, with Standard Errors, were obtained for the interaction term, Trial Type \times Condition, from linear mixed-effects effects models applied to the data for each experiment; the model included main effects of Trial Type and Condition and its interaction, with a random intercept for Participant. Then, the standardized coefficients were combined using a random-effects model to get an estimate of effect size, weighted by the inverse of the squared Standard Error. The weighting is meant to reflect the precision of the estimate. A forest plot for the estimate for each group is shown in Figure 3. Note that the Standard errors for the two groups reported here are smaller than those for the groups tested by Tsuji et al., likely because we have twice the number of infants in each group.

¹ These results align with two other non-preregistered analyses shown in the Supplementary material. (Both were carried out after the one in the main text.)

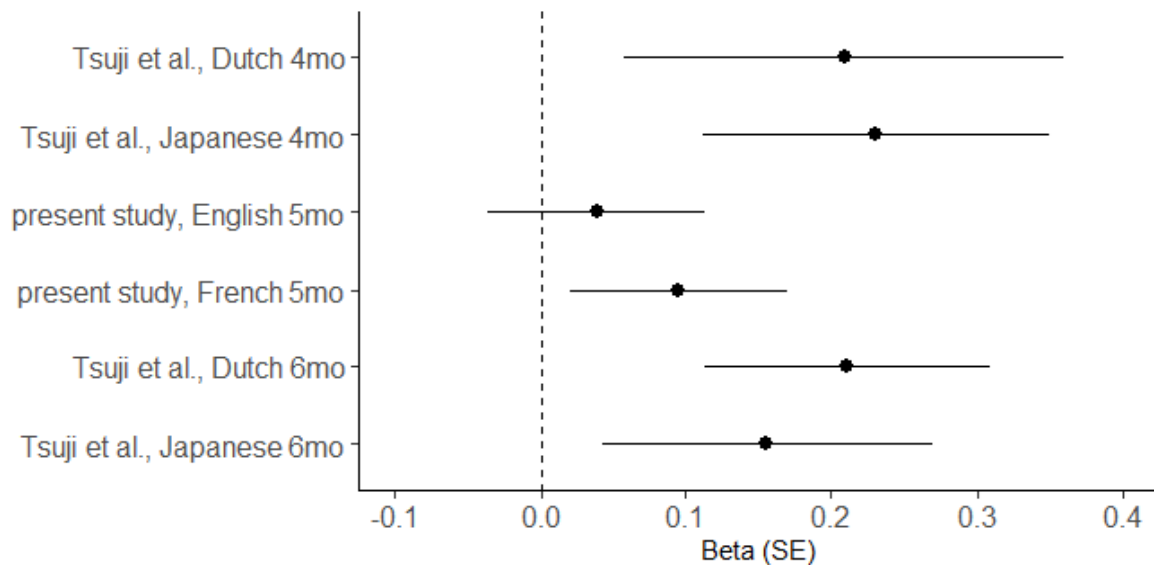


Figure 3. Forest plot. In each row, the center of the dot indicates the effect size (i.e., the estimate of the Trial Type \times Condition interaction term) and the whiskers delimit the standard error.

These estimates are interpreted following Cohen's guidelines for effect sizes. As shown in the figure, the effect size estimates for the asymmetry were different from 0 in the experiment with French but not English infants. Overall, there was a small but significant asymmetry across all six groups ($\beta = 0.13$, $SE = 0.04$, 95% CI: [0.05, 0.20], $z = 3.26$, $p < .002$).

Discussion

Our goal in this paper was to explore the phonetic basis of a perceptual asymmetry reported in the literature in the discrimination of labial versus coronal stop consonants. As a prelude to determine whether the asymmetry can be attributed to the greater relative burst intensity of the coronal stops, we tried to replicate the reported asymmetry in French and English learning 5-month-olds. We used the stimuli from the original study by Tsuji et al. (2015) that documented the asymmetry. Although there was a larger numerical difference in listening time to same and switch trials when infants were habituated to the labial /ɔmpa/, as might be expected based on Tsuji et al.'s findings, we failed to obtain statistical evidence for an asymmetry. Following our preregistration, we thus did not carry out a second experiment with stimuli in which the intensity profiles of /p/ and /t/ are swapped.

Instead, unlike Tsuji et al (2015), we found robust discrimination of /ɔmpa/ and /ɔnta/, at least when we restricted the analyses to the first two test trials. There was also

marginal support for discrimination when the analyses were extended to all four test trials. The effect size for discriminating this medial consonant place distinction was small, smaller even than the effect size for discriminating subtle place contrasts between velar and coronal nasals, or dental-retroflex distinctions in nasals and laterals in initial-position; effect sizes for those subtle contrasts range from 0.53 to 0.61 (Sundara et al., 2018). In our stimuli, both the stop consonants and the preceding nasals differ in place of articulation. The small effect size indicates that infants still found it difficult to distinguish the place difference in the medial consonants. There are two possible reasons for the smaller effect size obtained in the current experiment.

First, we know from research on adults that place distinctions are harder to perceive in clusters, likely because there are fewer acoustic cues available to signal place differences in clusters, and the available ones are not robust (for summaries, see Wright, 2001, 2004). For instance, when stop consonants occur between vowels, they are signaled by vowel formant transitions from the preceding and following vowel, in addition to acoustic cues in the stop burst itself. Because vowels are loud, vowel transition cues preceding and following the stop consonant are robust. In contrast, stop bursts are weak and easily masked by environmental noise. Further, when stop consonants are preceded or followed by another consonant instead of a vowel, listeners do not have access to the robust vowel transition cues from the preceding or following vowel respectively. There is, however, limited data on infants' perception of consonantal contrasts in non-initial position anywhere within the word to confirm the developmental roots of these perceptual limitations. To the best of our knowledge there is only one other study besides Tsuji et al. (2015) that examines word-medial consonant discrimination by infants: Archer et al. (2016) tested the discrimination of word-medial consonants in clusters (/apta/-/akta/ and /abta/-/agta/) in 12- and 20-month-old English learning infants and found that only the older infants successfully discriminated these medial /p/-/k/ and /b/-/g/ contrasts. They argue, and we agree, that infants' difficulty with this contrast may be due to the reduced acoustic salience of the phonetic cues for place distinction when stop consonants are in a cluster (particularly with other stops or nasals).

A second and more plausible reason for the smaller effect size in our experiment is that we habituated infants till their listening time declined by 60%, replicating Tsuji et al.'s design. A more stringent habituation criterion of a decline to 50% is typically recommended as a way to maximize the number of infants who are actually habituated when going into the test phase (Cohen, 2004; Oakes, 2010). In support of this argument, young infants' discrimination of a subtle nasal contrast (/na/ vs. /la/) was reported by Sundara et al. (2018), who used the 50% criterion, while Narayan et al. (2010) observed no discrimination of this contrast with the 60% criterion. Therefore, whether the effect size of discrimination for medial contrasts is really smaller than that for initial contrasts remains to be determined.

So why did we fail to find an asymmetry when we used the same stimuli and an almost identical habituation procedure as Tsuji et al. (2015)? The complete list of differences in the methods between their study and the present one can be found in Table S2 of the Supplementary material. It is unlikely that any of these differences contributes to our diverging results. One aspect that at first sight could raise concern relates to the infants' language background. That is, the stimuli were produced by a native speaker of Dutch. Yet because Tsuji et al. (2015) tested not only Dutch- but also Japanese-learning infants and found no language-related difference, neither in the habituation nor in the test phase, it is unlikely that our results are due to the fact that we tested French- and English-learning infants.

Another difference that is worth discussing concerns the stimuli used for the pre- and post-test. Whereas the pre- and post-test stimulus was an auditory monosyllabic nonce word in both studies (/pɔk/ in our study, /ni:m/ in theirs), in our study it was paired with the same static checkerboard picture that was used during the habituation and test phases whereas in Tsuji et al. (2015) it was paired with a video of a colorful smiley that rolled across the screen, following its contour. With such an engaging video, only infants who definitely do not want to look at the screen anymore will fail to dishabituate to the post-test. Tsuji et al. (2015) did not examine this, but probably only very few infants – if any – failed to dishabituate to the post-test. Thus, this difference in the design no doubt explains why only the infants tested in our experiment had on average longer looking times to the pre- than to the post-test, and why so many of them failed to dishabituate to the post-test. Perhaps the use of a static, but engaging picture, in combination with a shorter test phase consisting of only two trials, would help to exclude infants who are no longer on task.

Finally, our failure to find an asymmetry in discrimination is not due to a sample size issue. Indeed, both the present study and the one by Tsuji et al. report results from 64 infants. Recall that their infants came from two different ages, 4 and 6 months, where we only tested 5-month-olds. Thus, our sample is less heterogeneous, which makes testing the same number of infants as Tsuji et al. more than appropriate. However, a power analysis of their data using G*Power (Faul et al., 2007) revealed that in order to reliably detect their Type \times Condition interaction with a power of .80, 140 infants would need to be tested. This means that the asymmetric discrimination effect is quite small, making its estimation noisy and thus, sensitive to sample size. (This is of course a recurring concern within the field of infant research; for discussion, see Oakes (2017) and Bergmann et al. (2018)). In other words, because of the small effect size of the asymmetry, it may or may not be detected with a sample of 64 infants.

Whatever the reason for our inability to detect the asymmetry, we conclude that asymmetries in the perception of the labial and coronal stop distinction are weaker than those in the perception of vocalic contrasts – the latter being typically reported consistently in individual experiments (for reviews, see Polka & Bohn, 2003, 2011). We

propose two ways to go forward with the endeavor to examine the source of the labial-coronal asymmetry. One is to keep the same, naturally produced, stimuli and set up a larger cross-lab collaboration than the present one. This can not only facilitate larger sample sizes but also provide an opportunity to assess the robustness of small effects. The other is to run a study with newly created, carefully controlled, stimuli. With stimuli that differ only in the intensity of the stop bursts, one could run two experiments as in our planned study, the second one after having swapped the intensity profiles. Alternatively, as a competing hypothesis to account for the asymmetry is based on differences in token distribution (Bien & Zwisserlood, 2013), two sets of stimuli could be created, one differing in intensity and the other in token distribution of the stop consonants. This would hence allow one to pit the two hypotheses against one another.

References

- Archer, S., Zamuner, T., Engel, K., Fais, L., & Curtin, S. (2016). Infants' discrimination of consonants: Interplay between word position and acoustic salience. *Language Learning & Development*, 12(1), 60-78.
<https://doi.org/10.1080/15475441.2014.979490>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
<https://doi.org/10.18637/jss.v067.i01>
- Bergmann, C., Tsuji, S., Piccinini, P., Lewis, M., Braginsky, M., Frank, M., & 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>
- Bien, H., & Zwisserlood, P. (2013). Processing nasals with and without consecutive context phonemes: evidence from explicit categorization and the N100. *Frontiers in Psychology - Section Psychology of Language*, 4.
<https://doi.org/10.3389/fpsyg.2013.00021>
- Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, 66(4), 1001-1017.
<https://doi.org/10.1121/1.383319>
- Boersma, P., & Weenink, D. (2009). Praat: Doing Phonetics by Computer [Computer program].

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28. <https://doi.org/10.18637/jss.v080.i01>

Chodroff, E., & Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. *The Journal of the Acoustical Society of America*, 136(5), 2762–2772. <https://doi.org/10.1121/1.4896470>

Edwards, T. J. (1981). Multiple features analysis of intervocalic English plosives. *The Journal of the Acoustical Society of America*, 69(2), 535–547. <https://doi.org/10.1121/1.385482>

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, 175–191. <https://doi.org/10.3758/BF03193146>

Fox, J., & Weisberg, S., (2019). *An R Companion to Applied Regression*. SAGE.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show facilitation for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2), 13–21. <https://doi.org/10.1111/j.1467-7687.2006.00468.x>

Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models*, 4th Edition. McGraw-Hill Irwin.

Lenth, R. (2025). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.10.7-100003. <https://github.com/rvlenth/emmeans>

Makowski, D., Ben-Shachar, M. S., & Lüdtke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, 4(40), 1541. <https://doi.org/10.21105/joss.01541>

Nam, Y., & Polka, L. (2016). The phonetic landscape in infant consonant perception is an uneven terrain. *Cognition*, 155, 57–66. <https://doi.org/10.1016/j.cognition.2016.06.005>

Oakes, L. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, 11(3), 255–268. <https://doi.org/10.1080/15248371003699977>

- Oakes, L. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *Infancy*, 22(4), 436-469. <https://doi.org/10.1111/inf.12186>
- Oakes, L., Sperka, D., DeBolt, M., & Cantrell, L. (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. <https://doi.org/10.3758/s13428-019-01244-y>
- Ohde, R. N., & Stevens, K. N. (1983). Effect of burst amplitude on the perception of stop consonant place of articulation. *The Journal of the Acoustical Society of America*, 74, 706–714. <https://doi.org/10.1121/1.389856>
- Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41(1), 221-231. [https://doi.org/10.1016/S0167-6393\(02\)00105-X](https://doi.org/10.1016/S0167-6393(02)00105-X)
- 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>
- Polka, L., Ruan, Y., & Masapollo, M. (2019). Understanding vowel perception biases: a meta-analytic approach. In A. Nyvad, M. Hejná, A. Højen, A. Jespersen & M. Hjortshøj Sørensen (Eds.), *A sound approach to language matters – In honor of Ocke-Schwen Bohn* (pp. 561-582). Dept. of English, School of Communication & Culture, Aarhus University.
- Repp, B. H., & Lin, H.-B. (1989). Acoustic properties and perception of stop consonant release transients. *The Journal of the Acoustical Society of America*, 85(1), 379–396. <https://doi.org/10.1121/1.397689>
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Schwartz, J.-L., Abry, C., Boë, L. J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the dispersion-focalisation theory. *Speech Communication*, 45(4), 425–434. <https://doi.org/10.1016/j.specom.2004.12.001>
- Sundara, M., Ngon, C., Skoruppa, K., Feldman, N. H., Onario, G. M., Morgan, J. L., & Peperkamp, S. (2018). Young infants' discrimination of subtle phonetic contrasts. *Cognition*, 178, 57–66. <https://doi.org/10.1016/j.cognition.2018.05.009>

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>

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>

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1-48. <https://doi.org/10.18637/jss.v036.i03>

Werker, J. F., Shi, R., Desjardins, R., Pegg, J., Polka, L., & Patterson, M. (1998). Three methods for testing infant speech perception. In A. M. Slater (Ed.). *Perceptual development: Visual, auditory, and speech perception in infancy* (pp. 389–420). UCL Press.

Winitz, H., Scheib, M. E., & Reeds, J. A. (1972). Identification of stops and vowels for the burst portion of /p, t, k/isolated from conversational speech. *The Journal of the Acoustical Society of America*, *51*(4, Pt. 2), 1309–1317. <https://doi.org/10.1121/1.1912976>

Wright, R. A. (2001). Perceptual cues in contrast maintenance. In E. V. Hume & K. Johnson (Eds.), *The Role of Speech Perception in Phonology* (pp. 251-277). Academic Press.

Wright, R. A. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 34-57). Cambridge University Press.

Zue, V. (1976). *Acoustic Characteristics of Stop Consonants: A Controlled Study*. Doctoral dissertation, M.I.T.

Data, Code, and Materials Availability Statement

The preregistration, stimuli, acoustic measurements, data, and analysis code are available on OSF at <https://osf.io/rew4p/>.

Ethics Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of University of California, Los Angeles (#10-001562; 10 December 2010—to date) and by the Comité d’Ethique de la Recherche of University

Paris-Cité, Paris (#00012022-127).

Authorship Statement

Sharon Peperkamp and **Megha Sundara** designed the study, acquired funding, oversaw the experiment, and wrote the manuscript. **Megha Sundara** performed the acoustic analyses of the stimuli, **Sharon Peperkamp** analyzed the data.

Contributorship Statement

Sho Tsuji carried out the meta-analysis. Anne-Caroline Fiévet, Minqi Liu and Evelyn Cortess-Gress recruited participating families. Calypso Anquetil, Juliette Dessertine, Minqi Liu, Evelyn Cortess-Gress and Ferhat Karaman tested the infants.

Acknowledgments

We are most grateful to Sho Tsuji for sharing her stimuli with us and for carrying out the meta-analysis, as well as for stimulating discussion. Thanks also to Alex Cristia for feedback. This work was supported by research grants ANR-24-CE28-2285-01 and NSF BCS-2214017.

Supplementary Material

A. Additional non-preregistered analyses

We carried out two more non-preregistered analyses. First, Bayesian mixed effects modeling allows for convergence regardless of the complexity of the model structure. Here, we used *brms* (Bürkner, 2017) to analyze our data with an additional factor Block (trials 1 and 2 vs. trials 3 and 4) and a slope for Trial Type under participant:

```
brm(looktime ~ Language * TrialType * Condition * Block + (1 + TrialType | participant)
```

Probabilities of direction (pd), which index the existence of an effect and range from 50% to 100%, were obtained using the *bayestestR* package (Makowski et al., 2019). The complete model output is available on OSF.

The effects of Trial Type and Block were near-credible (Trial Type: $\beta = -0.38$, 95% CI = [-0.80, 0.04], pd = 96.5%; Block: $\beta = 0.37$, 95% CI = [-0.04, 0.78], pd = 96.7%). There was also a near-credible interaction of Trial Type and Block ($\beta = -0.34$, 95% CI = [-0.74, 0.05], pd = 95.8%). Post-hoc analyses using the *emmeans* package revealed a credible effect of Trial Type in the first but not the second block (Block 1: $\beta = -1.43$, HPD = [-2.64, -0.31], pd = 99.3%; Block 2: $\beta = -0.09$, HPD = [-1.21, 1.05], pd = 56.1%). Finally,

there was a credible triple interaction between Language, Condition, and Block ($\beta = -0.57$, 95% CI = $[-0.99, -0.16]$, pd = 99.8%). Post-hoc analyses revealed that this triple interaction was due to the fact that for the English infants, the interaction of Condition and Block was credible ($\beta = -2.89$, HPS = $[-5.28, -0.59]$, pd = 99.1%), while for the French ones it was near-credible but went in the other direction ($\beta = 1.65$, 95% HPD = $[-0.55, 3.93]$, pd = 92.2%). Indeed, English infants had overall longer looking times in the first block when habituated to labial /*ompa*/, but overall longer times in the second block when habituated to coronal /*onta*/; for French infants, this pattern was reversed.

Second, it was brought to our attention during the write-up of this paper that outlier detection should be done based on the residuals of the regression model, rather than based on the raw looking times (see, e.g., Kutner et al., 2004). We thus reran the analyses, starting now with the original sample. As before, Language did not interact in any way with Trial Type; this factor and its interactions were hence not included in the model. Similarly, there was again no slope for Trial Type as it yielded a singular fit. Using the outlierTest function in the *car* package (Fox & Weisberg, 2019) on the model output, we detected one outlier. After replacement of this infant in the same counterbalancing group, the same analytic procedure yielded once more a model with Trial Type, Condition, and its interaction as fixed factors and a random intercept for Participant. This model revealed an effect of Trial Type ($\beta = -0.44$, $SE = 0.21$, $t = -2.05$, $\chi^2 = 4.17$, $p = .041$), with longer looking times to switch than to same trials. Neither the effect of Condition nor the Trial Type \times Condition interaction were significant (both $t < 1$).

B. Additional tables

Table S1. Comparison of looking times in the experiment in Tsuji et al. (2015) and the present one.

	Tsuji et al. (2015)	present study
pretest	4-month-olds: 13.7 sec. ⁱ	13.1 sec.
	6-month-olds: 13.6 sec. ⁱ	
habituation	/ <i>ompa</i> /: 163 sec.	/ <i>ompa</i> /: 109 sec.
	/ <i>onta</i> /: 193 sec.	/ <i>onta</i> /: 100 sec.
test		
habituation to labial	$M_{\text{same}} = 7.61$ sec.	$M_{\text{same}} = 5.84$ sec.
	$M_{\text{switch}} = 9.68$ sec.	$M_{\text{switch}} = 7.01$ sec.
habituation to coronal	$M_{\text{same}} = 8.42$ sec.	$M_{\text{same}} = 6.28$ sec.
	$M_{\text{switch}} = 7.99$ sec.	$M_{\text{switch}} = 6.62$ sec.
post-test	4-month-olds: 14.0 sec. ⁱ	6.50 sec.
	6-month-olds: 13.9 sec. ⁱ	

ⁱ Personal communication from S. Tsuji.

Table S2. Comparison of methods used in Tsuji et al. (2015) and in the present study.

	Tsuji et al. (2015)	present study
infants' main language of exposure	Dutch or Japanese	English or French ⁱ
number of infants	64 (32 per language group)	
infants' age	4 months and 6 months (for each language)	5 months
infants' exposure to other languages	Japanese: NA Dutch: all probably at most 10% ⁱⁱ	English: all at most 10% French: 19 at most 10%, 12 at most 50%, 1 80%
checkerboard	dynamic	static
habituation criterion	sliding window of 4 trials; drop to 60% compared to first 4 trials	sliding window of 4 trials; drop to 60% compared to longest 4 consecutive trials
maximum number of habituation trials	28	
maximum trial duration	14 seconds	
ISI	1 second	
order of test stimuli	same-switch-same-switch	counterbalanced: same-switch-same-switch or switch-same-switch-same
pre- and post-test	visual: video of colorful, rolling smiley ⁱⁱⁱ audio: multiple tokens of /ni:m/ spoken by English speaker ⁱⁱⁱ	visual: same static checkerboard as in habituation and test phase audio: multiple tokens of /pɔk/ spoken by French speaker
minimum time looking at screen for trial not to be repeated	1 second ⁱⁱⁱ	2 seconds
looking time during false starts included in total looking time of repeated trials?	no ⁱⁱⁱ	
maximum time looking away before trial ends	2 seconds ⁱⁱⁱ	1 second

time looking away included in total looking time?		no ⁱⁱⁱ	
false habituators		kept	
outliers	not examined ⁱⁱⁱ		replaced
no post-test dishabituation	not examined ^{iv}		kept

ⁱ One French-learning infant had less than 50% exposure to French.

ⁱⁱ All infants were raised in monolingual families; the 10% threshold for the recruitment criterion is estimated from memory (personal communication from P. Fikkert).

ⁱⁱⁱ Personal communication from S. Tsuji.

^{iv} Tsuji et al. (2015) did not test for dishabituation at the individual level; they compared looking times in pre- to post-test in an ANOVA by age and language, observing no main effect or interaction.

License

Language Development Research (ISSN 2771-7976) is published by TalkBank and the Carnegie Mellon University Library Publishing Service. Copyright © 2026 The Author(s). This work is distributed under the terms of the Creative Commons Attribution-Non-commercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits any use, reproduction and distribution of the work for non-commercial purposes without further permission provided the original work is attributed as specified under the terms available via the above link to the Creative Commons website.