Environmental Influences on Infants' Native Vowel Discrimination: The Case of Talker Number in Daily Life

Christina Bergmann
LSCP
Département d'Etudes Cognitives
ENS, EHESS, CNRS
PSL Research University and
Language Development Department
Max Planck Institute for Psycholinguistics

Alejandrina Cristia
LSCP
Département d'Etudes Cognitives
ENS, EHESS, CNRS
PSL Research University

Both quality and quantity of speech from the primary caregiver have been found to impact language development. A third aspect of the input has been largely ignored: the number of talkers who provide input. Some infants spend most of their waking time with only one person; others hear many different talkers. Even if the very same words are spoken the same number of times, the pronunciations can be more variable when several talkers pronounce them. Is language acquisition affected by the number of people who provide input? To shed light on the possible link between how many people provide input in daily life and infants' native vowel discrimination, three age groups were tested: 4-month-olds (before attunement to native vowels), 6-month-olds (at the cusp of native vowel attunement) and 12-month-olds (well attuned to the native vowel system). No relationship was found between talker number and native vowel discrimination skills in 4- and 6-month-olds, who are overall able to discriminate the vowel contrast. At 12 months, we observe a small positive relationship, but further analyses reveal that the data are also compatible with the null hypothesis of no relationship. Implications in the context of infant language acquisition and cognitive development are discussed.

Correspondence should be sent to Christina Bergmann, Language Development Department, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. E-mail: chbergma@gmail.com

Forming discrete categories based on continuous sensory input is a fundamental human skill across many domains. A key instance of category formation takes place during early language acquisition, when infants tune into the sound system of their native language. There are many possibilities to divide up the acoustic space into units that can be combined to form distinct word forms. Infants have to discover which sound contrasts (such as /r/ and /l/, which change word meaning in English, but not Japanese) to pay attention to. This process has its onset early in infancy, as infants begin to show indicators of tuning into their native vowel system at the age of six months (see the meta-analysis by Tsuji & Cristia, 2014).

Given this timeline, mechanisms that can account for infants' ability to build sound categories must work from the signal, involving limited or no top-down information. Consequently, all currently proposed mechanisms of sound category acquisition rely heavily on the computation of statistics over some input representations. This includes distributional learning (Maye, Werker, & Gerken, 2002); the perceptual magnet effect (Kuhl, 2004); and proposals that allow for interactions with other sources of information, such as the developing proto-lexicon (e.g., Elsner, Goldwater, Feldman, & Wood, 2013; Feldman, Griffiths, Goldwater, & Morgan, 2013; Swingley, 2009; Yeung & Werker, 2009). Infants must both draw the lines between native sound categories and discover how many categories actually exist, clearly a difficult joint learning problem.

It is possible that talker characteristics are a hurdle during language processing, and specifically sound categorization, because talkers vary a great deal in their productions along the same acoustic dimensions that distinguish speech sounds. As a result, one person's /a/ (as in "cot") might be indistinguishable from another's /ɔ/ (as in "caught" on the acoustic level. Talker differences can be traced to a number of factors ranging from physical differences to idiosyncratic ways of articulating sounds (Hillenbrand, Getty, Clark, & Wheeler, 1995). As a consequence, talker variation proves an interesting test case of how exactly infants begin tuning into their native vowel system, as current theories make mutually exclusive predictions on how infants must deal with variation in their input. We first lay out the three theoretical predictions, then discuss empirical evidence, and finally outline our approach in the remainder of the Introduction.

If infants were to track the statistics of raw acoustic realizations across salient dimensions, they might be misled as to their native category system in the presence of multiple input talkers. Talker-specific variation can mask linguistically relevant information, and the overlap between categories increases dramatically when introducing multiple talkers (Hillenbrand et al., 1995; specifically Fig. 4; see also Kuhl, 2004).

Alternatively to computing possibly misleading statistics over input from all talkers in the environment, infants might be able to separate talkers (Kleinschmidt & Jaeger, 2015) using contextual information, so that statistics for each talker are tracked independently. This should lead to less information being available for the formation of each talker-dependent category when being exposed to multiple people (assuming that equal amount of speech input is available across low- and high-talker-variability scenarios). The resulting prediction in both scenarios—computing confusing input statistics and separating input by talker—is that increased talker variation has *negative* consequences for sound category development.

Not all researchers who work on the topic of early language acquisition agree with these negative predictions. A sizable community believes that talker variability could play no role at all. This position rests on the assumption that, from birth, infants are

able to ignore the difference between talkers, by either abstracting away features that are confusing, or computing talker-invariant representations early on during processing (Dehaene-Lambertz & Peña, 2001; Kuhl, 1979, 2004). With this ability in place, talker variation should have *no impact* on sound category development.

Yet others argue that talker variability might be *helpful* during sound category acquisition. This is the case if variability leads infants to focus on linguistically informative dimensions of the speech signal (Rost & McMurray, 2009; Seidl, Onishi, & Cristia, 2014). The core assumption is that infants base their learning on a representation where talker-dependent information is not (or only weakly) correlated with linguistically important information. As a consequence, infants who are exposed to more variable input should benefit, because they will start learning which aspects of the speech signal are important for linguistic processing, and which are not, at an earlier age than peers with less variable input.

Empirical support exists for all three stances, and we will give a brief overview here (for a more extensive review focusing on sound discrimination, see Bergmann, Cristia, & Dupoux, 2016; for a broad discussion of talker variation refer to van Heugten, Bergmann, & Cristia, 2015). For example, Jusczyk, Pisoni, and Mullennix (1992) showed a negative effect of multiple talkers for two-month-olds in a sound discrimination task, especially when introducing a short delay that likely increases task demands. In contrast, Kuhl (1979) found that six-month-olds succeeded in distinguishing two vowels, both in the presence and absence of multiple talkers. The third, positive scenario was supported by a study showing that, when learning phonotactic rules, four- and eleven-month-old infants benefit from hearing multiple talkers during the learning phase (Seidl, Onishi, et al., 2014). While the diversity in the findings just mentioned could be due to methodological and/or age differences, across the literature it is not possible to discern a clear pattern, which is modulated by those factors (Bergmann et al., 2016).

To shed light on the relation between early language development, specifically phonetic category learning, and input taker variability, we adopt an approach that does not rely solely on short-term learning and performance in the laboratory: individual variation among infants. This line of research relies on studying differences between (groups of) infants by tracking both diverging features of infants' environments and their performance in fairly standard laboratory-based tasks. Correlations between the two measures are compatible with the interpretation that there may be an underlying association. This method has prominently been applied when investigating the possible link between language abilities and the quantity of input (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991) as well as the role of home environment quality (Melvin et al., 2017).

On a more subtle level, qualitative aspects of infants' input also shape their path into language. For example, mothers' vowel space size when talking to their infant aged 6–12 months correlated with the infants' performance in a speech perception task at the same age (Liu, Kuhl, & Tsao, 2003). In another study documenting a more specific association, differences in how distinctly /s/-/ʃ/ (as in "sip" versus "ship") were pronounced by caregivers predicted their respective infants' ability to discriminate the same sound pair beyond overall speech rate or pitch (Cristia, 2011). The latter study suggests that the more confusing infants' input, the less likely they are to perform well when having to distinguish two similar sounds in laboratory studies. In other words, this line of research establishes that the effects of characteristics of infants' everyday

environment on language development can be measurable via laboratory tasks in the individual infant.

Our study adopts the previously established approach to assessing individual variation to address the key question of whether talker variation in infants' everyday life could potentially affect early language acquisition. If all infants received input from just one main caregiver, it would not be possible to observe any differences. However, in many cultures, different daycare models and diverse household structures (e.g., with multiple siblings or including extended family) lead to natural variation in the number of people who habitually talk to an infant.

We sought to shed light on the possible link between how many people provide input and how infants tune into their native vowel system. We chose a vowel contrast because talkers are expected to differ more in vowel than consonant articulation, as the former are greatly impacted by variation in vocal tract length and structural configuration (Hillenbrand et al., 1995). We specifically selected a vowel pair previously described as difficult to distinguish for young infants (Pons, Albareda-Castellot, & Sebastián-Gallés, 2012), which also varies greatly across French adult talkers (Gendrot & Adda-Decker, 2005, Table 5). Our focus was on native vowel discrimination (rather than other abilities potentially affected by experiences with talker variability, such as cross-talker normalization); consequently, we used the implementation of this contrast in a single talker. To estimate talker variability in infants' input, we asked parents about their child's schedule during a typical week. We tested three age groups to track development, namely 4-month-olds (before attunement to native vowels), 6-month-olds (at the cusp of native vowel attunement), and 12-month-olds (well attuned to the native vowel system). As outlined above, there are three mutually exclusive predictions on the impact of talker variability. If talker variation hinders acquisition of native vowel categories, we should observe a negative association between the number of talkers in the environment, and infant performance in a vowel discrimination task. A positive association would be consistent with views stating that infants use talker variation to determine what acoustic changes are linguistically relevant. Finally, no association would be most compatible with views that infants ignore, or automatically compensate for, talker variation.

EXPERIMENTS

The present study was preregistered in two steps on Open Science Framework prior to data collection. The preregistrations are available on the project website, along with all stimuli, analysis scripts, anonymized data, and supplementary materials: https://osf.io/q9cpa/. Details on the preregistration, all exploratory analyses conducted, and documentation of all changes in data processing with respect to preregistration can be found on the project website. The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the CERES (Conseil d'évaluation éthique pour les recherches en santé) under IRB 2015140001072.

Participants

Birth records and contact information are provided by official sources to the babylab; about three months after their child's birth, parents receive a letter describing the babylab. Parents who respond to this letter are added to the database, and those whose child matched preset age criteria were then contacted for this experiment. All children were, according to parental reports, monolingual learners of French (at least 90% exposure) and born full-term.

In each age group, we included 46 participants in our final sample (girls per age group: 27 4-month-olds; 21 6-month-olds, 19 12-month-olds). The mean ages (along with minimum and maximum age) in days were 138 (121–151) for the 4-month-olds, 180 (154–211) for the 6-month-olds and 358 (339–392) for the 12-month-olds.

To arrive at the final sample, we had to test 75 additional infants, who were excluded for the following reasons: audible crying (17 4-month-olds; 10 6-month-olds; 8 12-month-olds), fussiness (14 4-month-olds; 4 6-month-olds; 1 12-month-olds), <1.5s looking time in any of the test trials (4 4-month-olds; 3 6-month-olds) and parental interference (4 4-month-olds; 8 6-month-olds; 2 12-month-olds).

The children in our final sample came from a homogeneous socioeconomic group; all were growing up in an urban setting and with mothers who had all completed 12 years of high school and, with two exceptions, attended university for several years (number of years in formal education: $M = 17.2$, $SD = 1.95$).

Stimuli

The acoustic stimuli were two native vowels /i/ and /e/, presented as nuclei in syllables with /g/ as the onset. The resulting monosyllabic minimal pair is probably low frequency in infants' input (/gi/ is equivalent to the name "Guy," none of our participants in the final sample was called Guy; /ge/ means "gay"). The resulting CV syllables /gi/ and /ge/ were recorded by a native female speaker of French in a lively manner, as if speaking to a young infant, and thus contained variable intonation patterns. We created three syllable lists, one for habituation and two for test, each containing seven unique tokens separated by a one second interstimulus interval for a total duration of 20 sec (vowel duration was matched across the three lists). The test syllables were matched across lists for their intonational pattern so that there was equal intonational variation in the /i/ and /e/ lists.

The average formant values in Bark for the habituation and test lists are as follows: Habituation list /gi/ F1 = 3.37, F2 = 14.70; Test list /gi/ F1 = 3.36, F2 = 14.70; Test list /ge/ F1 = 5.13, F2 = 14.29. T-tests confirmed that the habituation and test items of /gi/ did not differ significantly ($p = .95$). We further confirmed that voice onset time and prevoicing duration in [g] across both test lists were not significantly different with two t-tests (both $p > .7$).

Procedure

Discrimination was tested using a habituation–dishabituation task implemented in a central fixation procedure, which has been previously used to measure individual variation among infants in discrimination tasks (Cristia, 2011; Houston, Horn, Qi, Ting, & Gao, 2007; Seidl, French, Wang, & Cristia, 2014). Infants sat in a sound-proof booth

on their parent's lap in front of a wall-mounted 27" monitor (type Iiyama PROLITE E2773HS-GB1). Sound was presented from two loudspeakers next to the monitor (model JBL Control 1Pro). Parents were instructed not to talk or point at the screen and listened, via noise-canceling headphones, to masking music which was overlaid with the experimental stimuli in random order and varied intensity. Via a camera above the screen, infants' reactions were observed by the experimenter, who controlled the experiment outside the sound-proof booth without being aware of the trial type currently being presented. The experimenter wore noise-canceling headphones to further ensure blinding.

We presented our experiment using the Lincoln Lab LOOK software (Meints & Woodford, 2008). The experimenter recorded online whether the infant looked to the screen or not via a button press. Trials ended early when infants looked away from the screen for more than two consecutive seconds, and a new trial only started when infants looked back toward the screen for at least one second. On the screen, infants saw a bull's eye on a gray background during the trials, and moving colorful shapes on a black background between trials to redirect their attention to the screen. All infants were tested by the same experimenter.

The habituation criterion was set to 50% of the average of the three trials with the longest looking times. The maximal trial duration was the duration of the files, 20 sec. After completing the habituation phase with the background syllable /gi/, which could last up to 24 trials, infants heard one trial with the novel syllable /ge/, followed by two trials with new tokens of the habituated syllable /gi/ and a final trial with the novel syllable /ge/. No trials were repeated.

Before the first habituation trial and after the last test trial, we presented an attention trial to establish an independent criterion for infants' fussiness. To that end, infants saw a smiling baby's face on a gray background accompanied by "coucou" (*hi*) for maximally 20 sec. The same criteria (trial onset after looking to the screen for one second; trial end criterion after looking away for two consecutive seconds) were applied.

Questionnaire

To assess infants' input variability in daily life, we used a questionnaire which parents filled out in the laboratory (available via the project website). On this questionnaire, parents were asked to complete a schedule of a "typical week" and fill in for each day's morning, afternoon, and evening who talked to the child for more than 20 consecutive minutes. This criterion aimed to exclude brief encounters to simplify the parents' task. If there were regular visitors or caretakers, but they could vary across weeks, parents marked this as well. A second sheet asked for details on each talker or group of changing talkers, including sex, age (range), highest educational degree, and if they spoke with a nonlocal or non-native accent. If a change in the infant's routine had taken place in the last two months, parents filled out a second schedule and ensured that all talkers were mentioned in the sheet noting their details. The main analyses focus on the number of talkers older than two years the infant is currently exposed to. We included children as input talkers due to emerging evidence that infants preferably listen to children's voices and presumably learn from them as well (Polka, Masapollo, & Ménard, 2014). Further, Shneidman, Arroyo, Levine, and Goldin-Meadow (2013) showed that all speech directed to children, including from their siblings (and presumably other children), predicts vocabulary development better

than input from the main caregiver alone (but see Shneidman & Goldin-Meadow, 2012; for conflicting results with Mayan children). The number of children in infants' input was comparatively low, with 79 of 138 participants not hearing speech from children at all. Interested readers can find additional analyses and the raw data on the project website.

RESULTS

All analyses were done in R (R Core Team, 2016). Preliminary analyses of the 6-month-olds' data suggested that we had not considered a number of factors in the pre-registration, most saliently the shape of the distribution of number of talkers. For the results of the preregistered analyses and detailed discussion, see the supplementary materials on the project website. We report here analyses that deviate from the preregistered ones but were more appropriate.

The number of people that parents reported as talking to the child in a typical week ranged from two to 21, with a median of six for the 4-month-olds and eight for the two older age groups, as depicted in Figure 1. As the distributions were not normal, this predictor was log-transformed for subsequent analyses. We provide analyses with the raw data as preregistered in the supplementary materials.

Turning to vowel discrimination as measured in the laboratory, infants needed on average 12.17 (SD = 5.67) trials to habituate, with a range of 4 (the preset minimum) to 24 trials (the preset maximum). A linear regression with number of trials to habituation as dependent measure and age group and log number of talkers in the input as
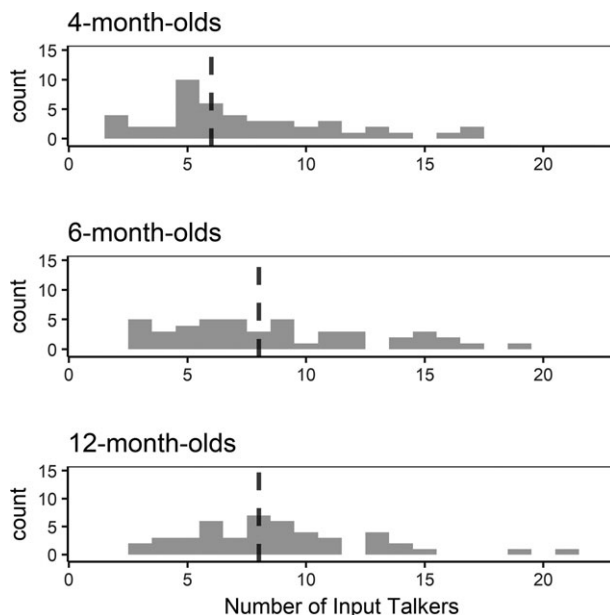

Figure 1 Histogram of number of talkers. The distribution of the number of talkers in infants' input, separated by infant age group. The dashed line indicates the median number of talkers.

predictors only revealed a main effect of age groups, but no effect of the number of talkers or an interaction (cf. supplementary materials on the project website).

The dependent variable in the laboratory experiment was a discrimination score, infants' listening time to novel test trials divided by their total listening time to all (novel and habituated) test trials. This score can range between 0 and 1, with .5 indicating no difference between listening times to novel and previously habituated test trials. The question of why it is preferable to use ratios, rather than difference scores, when attempting to focus on sound discrimination skills precisely has been discussed in more detail elsewhere (e.g., Cristia, 2011); in a nutshell, difference scores reflect individual variation at other levels, such as speed of processing. In the present experiment, the mean discrimination score across all age groups is .533, which is significantly above .5 (one-sample t-test against chance of .5 $t(137)$ = 3.019, two-tailed p = .003, 95% CI [.511, .554], Cohen's d = 0.52).

To answer our key research question, we fit a linear regression declaring discrimination score as dependent measure, and age group and log-transformed number of talkers in the input, as well as their interaction, as predictors. Significance at the factor level was assessed with a type II ANOVA in the package "car" (Fox & Weisberg, 2011). Age group was a significant predictor with $F(2,132)$ = 5.265, p = .006. The effect of age group emerged because younger infants performed better than older ones (all t-tests are one-sample against .5 chance, all p's are two-tailed): 4-month-olds $t(45)$ = 2.326, p = .025, 95% CI [.505, .576], Cohen's d = 0.69; 6-month-olds $t(45)$ = 4.175, p < .001, 95% CI [.535, .601], Cohen's d = 1.25; 12-month-olds $t(45)$ = −0.533, p = .596, 95% CI [.448, .530], Cohen's d = −0.16. Neither the log-transformed number of talkers ($F(1,132)$ = 1.595, p = .21) nor the interaction of the two predictors ($F(2,132)$ = 2.414, p = .093) reached significance. Bayesian analyses[1] confirmed that age group alone predicts discrimination scores (BF_{10} = 4.026), and that log number of input talkers does not add explanatory power.

Although no interaction was found, in the interest of fully describing our data for readers, we carried out correlations of discrimination scores with log-transformed number of talkers within each age group, as depicted in Figure 2. For the 4-month-olds, this correlation is not significant and positive (r = .106, CI: [−.189, .384], p = .482, BF_{10} = 0.147), for the 6-month-olds nonsignificantly negative (r = −.126, CI: [−.402, .170], p = .403, BF_{10} = 0.163), and for the 12-month-olds significantly positive (r = .303, CI: [.014, .545], p = .041, BF_{10} = 0.920). All BF_{10} for correlations were calculated based on Wetzels and Wagenmakers (2012).

In addition to the analyses described here, we carried out further exploratory analyses, which were in part suggested by anonymous reviewers. All analyses can be found on the project website at https://osf.io/q9cpa/. The main pattern of results presented here remain, including when considering only adults (defined as aged 13 or older, following Shneidman et al., 2013) as input talkers.

---

[1]Bayes factors (BF) indicate how much more probable the data were under one hypothesis compared to a second hypothesis. We report BF_{10}, comparing the data to their compatibility with the alternative hypothesis of an (undirected) correlation. The BF_{01} indicating the probability of the observed data under the null hypothesis can be found in the supplementary materials, along with explanatory figures. All BF were computed in R using the package BayesFactor (Morey & Rouder, 2015), supplementary computations and illustrations used JASP (JASP Team, 2016).
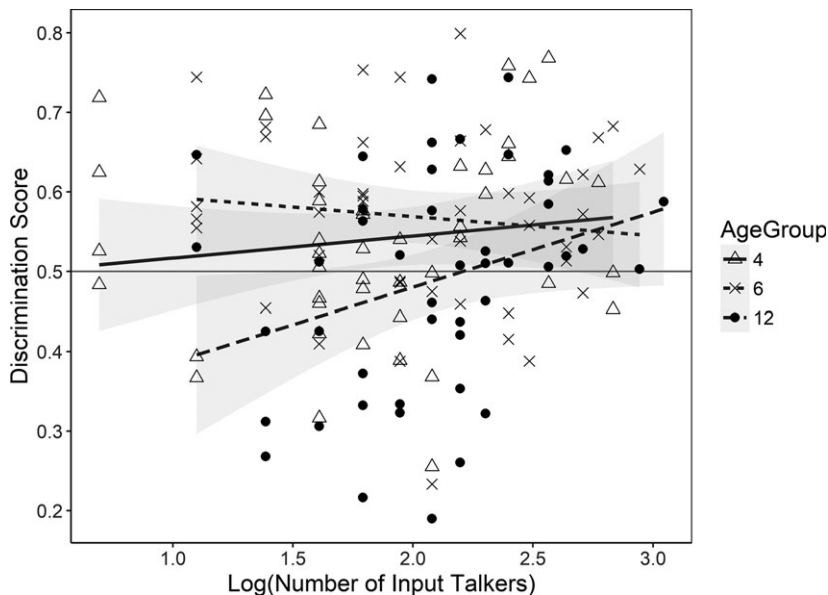
Figure 2 Discrimination scores as a function of log-transformed number of talkers in infants' input. Linear regression lines are superimposed, along with the respective 95% confidence intervals. Each line style and symbol corresponds to one age group; every symbol indicates one participant. The horizontal gray line marks a discrimination score of .5 (no looking preference).

DISCUSSION

We set out to investigate whether the number of talkers in daily life, a proxy of talker variation in the input, predicts infants' developing ability to discriminate native vowels, measuring potential long-term effects of variation in infants' natural experience. In the Introduction, we laid out three possibilities of input variability influencing language development, each supported by a theoretical framework and by experimental findings that pertained to infants' ability to deal with talker variation in the laboratory in various learning and discrimination tasks. Our goal was thus to disentangle three mutually exclusive predictions regarding the impact of talker variability in daily life, which covered all possible outcomes: a positive, negative, or no relationship. Which theoretical prediction is best supported by the data? To answer this question, we must pause to integrate across multiple views of the data. We believe our data are best described by considering not only null-hypothesis significance tests, but also the size of the correlation coefficient, which provides a strength of association metric, and the Bayes factor, which compares the strength of the evidence for one versus another hypotheses (Wagenmakers, Morey, & Lee, 2016).

   With these considerations in mind, we recap our results. In the two younger age groups, we found that the *p*-values for our key factor, talker number, did not reach the significance threshold. From that alone, however, we cannot conclude that there is evidence for no relationship between variability in the input and vowel discrimination. Bayes factors can help illuminate whether the measure was insensitive (due to noisy data and/or lack of power) or whether the observed data support the null hypothesis—in the present case implying that there is no relationship between the

number of talkers in infants' input and their ability to discriminate native vowels. The Bayes factors for the 4- and 6-month-olds provided evidence for the hypothesis that the amount of talkers an infant is exposed to does *not* affect vowel discrimination in our laboratory task (at least for the vowels we have chosen and in the way we tested infants). This is consistent with the small correlation coefficients. As for the oldest age group, we found a significant positive correlation among the 12-month-olds. While the *p*-value was below the alpha threshold of .05, the correlation coefficient was low, and the Bayes factor was near 1. This level of Bayes factor means that the evidence is inconclusive: The actual level of association observed is equally compatible with the hypothesis of a positive relationship, and with the hypothesis of no relationship.

Before moving on, we must point out two additional reasons why we are not convinced by the one significant result in the 12-month-olds. This result seems to be driven partially by below-chance performance found among children hearing few talkers (see Figure 2). While this pattern could be interpreted in a fixed-familiarization study (as a familiarity preference), below-chance performance is impossible to interpret in a habituation study like ours. Moreover, as a group, 12-month-olds did not show above-chance performance, suggesting that a task that was reasonably easy at four and six months was seemingly not as easily solved at twelve months. There might be several reasons for the low overall performance of the 12-month-olds, and our data do not allow us to disentangle what exactly this age group was doing; for the sake of completeness, we discuss briefly three reasonable explanations. First, it might have been the case that for the oldest children in our study the task was too boring. However, the low dropout rates in this age group speak against this possibility: Unlike in the younger two groups, we did not need to exclude any participants due to low looking times during test and only one participant due to an overall drop of attention. Second, the task might be intrinsically too hard. Yet, the younger age groups succeeded, and in general children's ability to distinguish native vowel contrasts is thought to increase as they mature (Tsuji & Cristia, 2014). Third, it is possible that around their first birthday, infants begin employing different strategies to solve the task at hand. A subset, namely those showing the familiarity preference, might, for example, have considered the bull's eye that was used as visual target as an object. Consequently, they were not reacting to a change in acoustics, but a change in label. In such mispronunciation tasks, one would indeed expect longer looking to the target when hearing the correct label compared to a mispronunciation, even with vowel changes (Mani & Plunkett, 2008). However, other implementations of the task find longer looks to the wrong object-label pair, and those might be considered more similar to our task as there was only one object displayed during learning and test (Stager & Werker, 1997). We might thus expect the same novelty preference independent of infants' particular strategy (phone discrimination or word object-label learning). Given our current data, it is not possible to disentangle these possibilities, and none seems very likely. On the whole, the correlation between performance and talker number in this age group is significant, but there are sufficient reasons, both conceptual (group effect at chance, some individuals' performance inexplicably below chance) and empirical (small effect size and a Bayes factor near 1 and thus inconclusive evidence for either the null or alternative hypotheses), to disfavor a strong interpretation whereby talker number and speech perception are robustly related.

Theoretical implications

Taking all age groups together, what are the implications of our results? We found no negative impact of talker variability in daily life on vowel discrimination in all three age groups, which would have been predicted by accounts basing language acquisition on statistical computations over the raw acoustic input (Maye et al., 2002). There are three possible interpretations compatible with these results.

First, perhaps the abstractionist account (Dehaene-Lambertz & Peña, 2001) is correct. If innate abilities allow infants to store and process talker-specific information separately from linguistic information, then the finding of no negative impact is easily accounted for. However, it cannot explain a positive impact of talker variability in the input, a possible shortcoming if the significant correlation in our 12-month-olds is taken to be true. Even more damning, this theory fails to account for the well-established fact that even adults are affected in their sound processing abilities by talker variation and change (Creel, Aslin, & Tanenhaus, 2008).

Second, it is possible that by 4–6 months infants gather sufficient experience with variable speech to overcome between-talker differences to a certain (but not perfect) extent. This could be accomplished by exploiting any source of variability, or specifically by using talker variability. Indeed, it has been proposed that the variability found within infant-directed speech, even in a single talker, may allow infants to prepare for between-talker differences (Kuhl et al., 1997).

A third possible interpretation is that infants might be able to *adapt* to talkers' voices, and learn sounds and words from them. Adaptation is a process by which representations and expectations are adjusted to the current situation. This process can take place on many different linguistic levels and can be driven by various cues. In adult speech processing, the most common account holds that listeners use the lexical context to infer the target sounds, and adjust for differences between expected and observed sounds on the phonological and/or lexical levels (Creel et al., 2008; Dahan, Drucker, & Scarborough, 2008; McQueen, Cutler, & Norris, 2006). Which cues adult listeners employ, what level of representation is being adjusted in which way, and what factors modulate adaptation are subject to ongoing debate. This is also the case for toddlers, who are thought to employ a range of more or less linguistically informed heuristics to adapt to unexpected pronunciations, using lexical context where available (Mulak, Best, Tyler, Kitamura, & Irwin, 2013), but also simply becoming more accepting of variable pronunciations when prompted with socially diverse visual cues (for a discussion, see Schmale, Seidl, & Cristia, 2015).

Independent of which mechanisms listeners employ, adaptation requires at least some short exposure phase to variation along relevant dimensions for the listener to be able to adjust. As yet, this ability has rarely been assessed in young infants. Indeed, studies aiming to investigate how talker or other acoustic differences affect infants' perception aimed at blocking adaptation. Typically, syllables or words spoken by multiple talkers are presented in short succession (Dehaene-Lambertz & Peña, 2001; Jusczyk et al., 1992; Kuhl, 1979; Polka et al., 2014; Seidl, Onishi, et al., 2014) or a familiarization phase with very little variation is followed by abrupt talker or other acoustic change is introduced (van Heugten & Johnson, 2012; Houston & Jusczyk, 2000). It would be interesting for future work to explore to role of adaptation in infants' sound discrimination, as other experimental evidence suggests that all prerequisites are in place from early on: To be able to adapt to talkers, infants must at least be able to

pick up on visual and environmental cues and/or be sensitive to different voice charac-
teristics that do not contain linguistic information but transport talker-identifying
information. As to visual cues, infants can already match male and female voices to
faces at two months (Patterson & Werker, 2003). Additionally, visual cues, such as lip
movements, influence infants' sound learning already at six months (Teinonen, Aslin,
Alku, & Csibra, 2008), showing the role of environmental information and infants'
abilities to exploit it during and for language processing. Moreover, even top-down,
lexically informed adaptation might be available within the first year of life as both
empirical findings and emerging theories suggest an interplay between word-level and
sound-level processing throughout language acquisition. Infants as young as six
months of age have been shown repeatedly to know a few common words (Bergelson
& Swingley, 2012; Mandel, Jusczyk, & Pisoni, 1995; Tincoff & Jusczyk, 1999), which
could serve as anchor. Emerging formal accounts show that discrimination of a sound
contrast is facilitated when the two sounds are used in different wordforms (Feldman
et al., 2013) or associated with different objects (Gogate, Prince, & Matatyaho, 2009;
Yeung & Nazzi, 2014; Yeung & Werker, 2009).

   All three interpretations, then, could explain why there is no evidence for negative
effects on sound discrimination at any of the ages tested. How about the oldest age
group, where a positive, albeit small, correlation was found? Although, in our view,
this result needs to be taken with great caution, we must point out that this precise
pattern was predicted by a framework where infants use variability to learn to weigh
linguistic and talker-specific information in the acoustic signal differently (Rost &
McMurray, 2009; Seidl, Onishi, et al., 2014). As we only find a small, positive correla-
tion at 12 months, and not at younger ages, it is tempting to propose a developmental
change, enabling older infants to harness the beneficial properties of variable input
which points to the reliable and important aspects of the signal (as proposed by, e.g.,
Rost & McMurray, 2009). Multiple studies have remarked a sharp increase in infants'
word recognition abilities around their first birthday (most saliently Bergelson &
Swingley, 2012), and the common factor might lie in an advancement in categorization
skills and the ability to weigh cues differently (Younger & Cohen, 1986). Nonetheless,
in the absence of a significant interaction of talker number with age group in our data,
and given the multiple problems we raised with the interpretation of results in this age
group, we do not pursue this possibility further.

Strengths, limitations, and future work

Our main conclusion is for the absence of a negative effect of talker number on sound
discrimination. We explained above that this should not be viewed as the preternatu-
rally ambiguous null result within the null-hypothesis significance testing paradigm,
but instead, in the two younger of our three age groups, we find evidence in favor of
the hypothesis of no effect (evidenced by the Bayes factors). Some readers may
nonetheless wonder, would we have measured evidence for an effect with statistical
reliability with increased power? We believe the answer is no. Our sample size of 46
infants in each age group is comparable to those found in previous work looking at
individual variation in infant sound processing (e.g., 44 infants in Altvater-Mackensen
& Grossmann, 2015; who looked at intermodal matching as a function of caregiver
social behavior; 42 in Cristia, 2011; looking at caregivers' pronunciation of /s/ as a
predictor of infants' /s-ʃ/ discrimination; 32 in Liu et al., 2003; tying parental vowel

size to infants' conditioned head-turn learning; 75 in Melvin et al., 2017; who corre-
lated sound discrimination and quality of home life; see also Cristia, Seidl, Junge,
Soderstrom, & Hagoort, 2014, for a meta-analysis of infant sound processing studies
as predictors of concurrent and later vocabulary). Instead, we will conclude that, given
the task and talker number estimations we have used, there is no evidence for a nega-
tive effect and, in the two younger infants, evidence for no effect. We now discuss
whether this conclusion may or may not generalize broadly beyond the task and talker
number measure.

One may wonder whether the stimuli might have been too easy or too difficult, lead-
ing to ceiling or floor effects. Previous research pointed to /i-e/ being acquired between
four and twelve months, with the former age failing to discriminate it (Pons et al.,
2012). As we wanted to be able to measure variance in infants' discrimination abilities
without a ceiling effect, we selected this contrast. Contrary to what one would have
expected following previous work, even 4-month-olds showed an overall discrimination
response. The fact that our 4-month-olds were able to discriminate the two sounds
speaks against this task being overly difficult, and thus prevents an interpretation of
lack of correlation due simply to floor effects. Might there have been ceiling effects,
thus restricting the performance range? The answer is again no: As Figure 2 shows,
there is considerable variation in responses measured within each age group, variation
that speaks against a ceiling interpretation.

Alternatively, it stands to reason that the instrument used as a proxy of infants'
input variability could be to blame. The measure of infants' environment and input,
derived from a questionnaire parents filled out in the laboratory, was specifically
designed for this study. Thus, it lacks extensive validation, for example, with observa-
tions at home. Validation would depend on data that are time-consuming to generate,
either for the parents, who could complete a dense diary, or for the experimenter who
analyzes daylong recordings, where trained listeners annotate talkers in a process that
takes typically at least as long as the recording itself (here, we would require several
full days of recordings; cf. Bergelson et al., submitted). Without prior evidence that
talker variation plays an important role for infant language acquisition, it may not be
justified to allocate substantial human resources to this task. Therefore, we chose to
rely on parents' subjective judgment and provided detailed written instruction and
assistance from the experimenter while filling out the questionnaire. Although
responses may vary between participants, the questions asked were clear and general
enough that it is unlikely they biased parents' responses. We are thus confident that
we capture the typical input of a given infant in the time frames we analyzed (a typical
week, based on slots of several hours per day).

The dependent variable chosen to capture infants input, the number of talkers in a
typical week, might not be covering the time frame that most impacts infants' language
acquisition on the sound level. Other possibilities have been explored in supplementary
analyses based on the same questionnaire data that were used for the main results, and
none added explanatory power to the statistical models. We tested, among other things,
whether the presence of a stable talker and the average number of people present at any
one time have an effect. Therefore, we can rule out other plausible measures of input
variability that could predict a difference in infants' language development. More
fine-grained measures, for example, based on voice similarity, would rely on daylong
recordings like those just described, which are arguably too costly to obtain for a study
exploring a new, and hitherto unstudied, factor in language acquisition.

Taking all three age groups together, it is clear that our data provide no support for the interpretation that talker variation *negatively* impacts vowel discrimination to an extent similar to other real-life predictors documented in previous work, such as parental input quality (Altvater-Mackensen & Grossmann, 2015; Cristia, 2011; Liu et al., 2003). Nonetheless, the evidence from speech technology for variability degrading unsupervised category clustering (Bergmann et al., 2016) is so compelling, that we believe it is worth exploring it further. It is possible that more fine-grained predictors could possibly uncover a negative link between input variability and the development of infants' discrimination ability at a level that was not possible in the present study. Ideally, this would be done with daylong recordings that sample over several days or weeks (as in Koorathota, Morton, Amatuni, & Bergelson, 2016). From these, one could extract voice characteristics of all talkers in infants' input as well as measurement of the times different talkers spoke in succession, which could, for example, influence adaptation. Gathering such data was not feasible in the scope of the present study, but others may also find this prospect interesting, as these data could further enrich our understanding of the impact of infants' linguistic environment on their language development. Similarly, the strongest evidence in favor of interactions between talker variability and speech perception emerges from studies where generalization is studied (e.g., Houston & Jusczyk, 2000). We purposefully chose to use a standard within-talker discrimination task to tap native vowel categorization rather than generalization skills, and future research might address the interesting prediction that infants exposed to few talkers are less able to generalize to novel talkers than their more experienced peers.

Before closing this article, we want to address an important issue: Our study builds on the assumption that the discrimination score extracted from a habituation task is a continuous and graded measure of individual infants' sensitivity, or put otherwise, an index of how well individual infants discriminate the categories. There are many empirical and conceptual arguments supporting such an assumption, of which we will mention two. First, such graded scores are widely used in individual variation work both in general information processing and more specific speech tasks; for instance, the Fagan test is a habituation–dishabituation task based on visual stimuli, in which graded discrimination scores gathered from infants at six to twelve months predict IQ and academic achievement in adulthood better than standardized cognitive tests (e.g., Fagan, Holland, & Wheeler, 2007). As for speech, a number of studies have used a measure that is essentially the same as ours, assuming that greater dishabituation indicates greater sensitivity (e.g., Cristia, 2011; Houston et al., 2007; Melvin et al., 2017). This assumption has recently received independent support (albeit in terms of group discrimination performance) in a meta-analysis of vowel discrimination using mostly habituation tasks, where effect sizes were found to correlate with the spectral distance between vowels (Tsuji & Cristia, 2017). Second, from a conceptual viewpoint, we do not see merit in the alternative interpretation whereby a dishabituation task can only be used to discriminate infants who do distinguish the categories from infants who do not. Indeed, the only cognitive model compatible with this view is one in which discrimination at the individual level is directly determined by the presence of symbolic categories: Only some children will "have" the categories and will be able to tell them apart, and then will do so with as much ease as if distinguishing between a tree and a dog. We think most developmentalists would disagree with such a cognitive model, and thus, the only alternative is to imagine that in infants, as in other perceivers,

individual performance is graded because individual sensitivity is graded. We have indirect evidence that this is a common assumption from the statistical tools used: If the pass/fail view of dishabituation were prevalent, we should see widespread use of chi-squared tests on discrete groups, whereas the use of t-tests or ANOVA assumes that outcomes provide a continuous measure of dishabituation. An astounding majority of recent vowel discrimination studies uses the second type of statistics (Tsuji & Cristia, 2014). In short, we do not believe that discrimination scores such as those employed in this paper can only be read as dichotomic pass/fail judgments, although we emphasize that further work is necessary to establish to what extent such measurements vary due to random noise or unrelated capabilities versus reliable individual differences in the linguistic domain (see also Cristia, Seidl, Singh, & Houston, 2016; for test–retest reliability estimates).

CONCLUSIONS

To sum up, the present study sought to assess a theoretically crucial question: To what extent do infants' real-life experiences with multiple talkers shape their vowel discrimination skills? Our findings do not align with expectations drawn from a line of theories of infant perceptual category learning, which assume that categories are acquired on the basis of input statistics. In the case of speech categories, those statistics become confusing as more talkers provide the input. Instead, our data support views where talker variation does not greatly impact early vowel category acquisition before the first birthday, thus inviting further experimental and theoretical work. We suggest exploring specifically mechanisms of adaptation in the first year of life.

ACKNOWLEDGMENTS

FUNDING

CONFLICT OF INTEREST

Both authors declare no conflict of interest.

CONTRIBUTIONS

CB and AC designed the study, CB prepared stimuli, set up the experiment, handled the preregistration, and tested the participants. CB and AC performed the main analyses, conducted supplementary analyses, and wrote the paper.

REFERENCES

Altvater-Mackensen, N., & Grossmann, T. (2015). Learning to match auditory and visual speech cues: Social influences on acquisition of phonological categories. *Child Development*, *86*(2), 362–378.

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (submitted). What do North American babies hear? A large-scale cross-corpus analysis.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences USA*, *109*(9), 3253–3258.

Bergmann, C., Cristia, A., & Dupoux, E. (2016). Discriminability of sound contrasts in the face of speaker variation quantified. In A. Papafragou, D. Grodner, D. Mirman & J. Trueswell (Eds.), *Proceedings of the 38th annual meeting of the cognitive science society* (pp. 1331–1336). Austin, TX: Cognitive Science Society.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*(2), 633–664.

Cristia, A. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *The Journal of the Acoustical Society of America*, *129*(5), 3271–3280.

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development*, *85*(4), 1330–1345.

Cristia, A., Seidl, A., Singh, L., & Houston, D. (2016). Test–retest reliability in infant speech perception tasks. *Infancy*, *21*(5), 648–667.

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, *108*(3), 710–718.

Dehaene-Lambertz, G., & Peña, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *NeuroReport*, *12*(14), 3155–3158.

Elsner, M., Goldwater, S., Feldman, N., & Wood, F. (2013). A joint learning model of word segmentation, lexical acquisition, and phonetic variability. In *Proceedings of empirical methods in natural language processing* (pp. 42–54).

Fagan, J. F., Holland, C. R., & Wheeler, K. (2007). The prediction, from infancy, of adult IQ and achievement. *Intelligence*, *35*(3), 225–231.

Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review*, *120*(4), 751.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*, 2nd edn. Thousand Oaks, CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Gendrot, C., & Adda-Decker, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: An automatic analysis of large broadcast news corpora in French and German. *Variations*, *2*, 2–4.

Gogate, L. J., Prince, C. G., & Matatyaho, D. J. (2009). Two-month-old infants' sensitivity to changes in arbitrary syllable–object pairings: The role of temporal synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 508–519.

van Heugten, M., Bergmann, C., & Cristia, A. (2015). The effects of talker voice and accent on young children's speech perception. In S. Fuchs, D. Pape, C. Petrone & P. Perrier (Eds.), *Individual differences in speech production and perception* (Vol. 3, pp. 57–88). Berlin, Germany: Peter Lang.

van Heugten, M., & Johnson, E. K. (2012). Infants exposed to fluent natural speech succeed at cross-gender word recognition. *Journal of Speech, Language, and Hearing Research*, *55*(2), 554–560.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Houston, D. M., Horn, D. L., Qi, R., Ting, J. Y., & Gao, S. (2007). Assessing speech discrimination in individual infants. *Infancy*, *12*(2), 119–145.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236–248.

JASP Team. (2016). JASP Version 0.7.5.5, Computer software.

Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, *43*(3), 253–291.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203.

Koorathota, S., Morton, S., Amatuni, A., & Bergelson, E. (2016). 6 & 7-month-olds' Noun Input: Human and Automated Corpus Analyses. (Presented at the International Conference on Infant Studies)

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America*, *66*(6), 1668–1679.

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

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., . . . Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686.

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, *6*(3), F1–F10.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound pattern of their own names. *Psychological Science*, *6*(5), 314–317.

Mani, N., & Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science*, *11*(1), 53–59.

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

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126.

Meints, K., & Woodford, A. (2008). Lincoln Infant Lab Package 1.0: A new programme package for IPL, Preferential Listening, Habituation and Eyetracking. Retrieved from http://www.lincoln.ac.uk/psychology/babylab.htm (WWW document: Computer software & manual)

Melvin, S. A., Brito, N. H., Mack, L. J., Engelhardt, L. E., Fifer, W. P., Elliott, A. J., & Noble, K. G. (2017). Home environment, but not socioeconomic status, is linked to differences in early phonetic perception ability. *Infancy*, *22*(1), 42–55.

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes Factors for Common Designs [Computer software manual]. Retrieved from https://CRAN.R-project.org/package = BayesFactor (R package version 0.9.12-2)

Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., & Irwin, J. R. (2013). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent. *Child Development*, *84*(6), 2064–2078.

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191–196.

Polka, L., Masapollo, M., & Ménard, L. (2014). Who's talking now? Infants' perception of vowels with infant vocal properties. *Psychological Science*, *25*(7), 1448–1456.

Pons, F., Albareda-Castellot, B., & Sebastián-Gallés, N. (2012). The interplay between input and initial biases: Asymmetries in vowel perception during the first year of life. *Child Development*, *83*(3), 965–976.

R Core Team. (2016). *R: A language and environment for statistical computing [Computer software manual]*. Vienna, Austria: R Core Team. Retrieved from https://www.R-project.org/

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349.

Schmale, R., Seidl, A., & Cristia, A. (2015). Mechanisms underlying accent accommodation in early word learning: Evidence for general expansion. *Developmental Science*, *18*(4), 664–670.

Seidl, A., French, B., Wang, Y., & Cristia, A. (2014). Toward establishing continuity in linguistic skills within early infancy. *Language Learning*, *64*(s2), 165–183.

Seidl, A., Onishi, K. H., & Cristia, A. (2014). Talker variation aids young infants' phonotactic learning. *Language Learning and Development*, *10*(4), 297–307.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*(03), 672–686.

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, *15*(5), 659–673.

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.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1536), 3617–3632.

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850–855.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*(2), 172–175.

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, *56*(2), 179–191.

Tsuji, S., & Cristia, A. (2017). Which acoustic and phonological factors shape infants' vowel discrimination? Exploiting natural variation in InPhonDB. In *Annual conference of the international speech communication association* (pp. 2108–2112). https://doi.org/10.21437/interspeech.2017-1468

Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, *25*(3), 169–176.

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057–1064.

Yeung, H. H., & Nazzi, T. (2014). Object labeling influences infant phonetic learning and generalization. *Cognition*, *132*(2), 151–163.

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*(2), 234–243.

Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development*, *57*(3), 803–815.