

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Cognition

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

Distributional learning of speech sound categories is gated by sensitive periods[☆]

Rebecca K. Reh^a, Takao K. Hensch^{b,c}, Janet F. Werker^{a,*}

^a Department of Psychology, University of British Columbia, Canada

^b Department of Molecular and Cellular Biology, Harvard University, USA

^c International Research Center for Neurointelligence, University of Tokyo, Japan

ARTICLE INFO

Keywords:

Statistical learning
Sensitive period
Phoneme
Infancy
Electroencephalography (EEG)
ERP oddball task

ABSTRACT

Perceptual attunement to the native phonetic repertoire occurs over the first year of life: an infant's discrimination of non-native phonetic contrasts declines while their discrimination of native phonetic contrasts improves, with the timing of change consistent with sensitive periods. The statistics of speech sound distributions is one source of input used to collapse non-native phonetic category boundaries, while sharpening native ones. Distributional learning can be a domain-general mechanism, yet given the timing of perceptual attunement, we hypothesized that this learning mechanism may be maturationally delimited in the content domain of phonetic categories. Here, we assessed whether sensitivity to the distribution of speech sounds in the environment declines as the period of perceptual attunement closes. We used electroencephalography (EEG) to investigate whether neuronal responses to native 'ra' and 'la' phones are modulated differently in older vs young infants by exposure to either a bimodal or unimodal sound distribution spanning the [r] ~ [l] phoneme space. The native contrast, ra-la, is discriminable at all three ages, ensuring that we were testing the distributional learning mechanism, rather than confounding it with a decline in discrimination to a non-native distinction. English monolingual infants ($n = 131$) at 5-, 9- and 12-months-old were familiarized to either a unimodal or bimodal distribution of /ra-/la/ speech sounds. Immediately following familiarization, an ERP oddball task was used to assess discrimination. Results showed that brief exposure to a bi- vs uni-modal distribution is sufficient to alter neuronal responses to subsequent /ra/ vs /la/ speech sounds at 5-months and 9-months, but not at 12-months. These results are the first to capture a progressive decline in sensitivity to distributional statistics in the environment. A potential mechanistic explanation based on critical period biology is discussed.

1. Introduction

Since the seminal study by Eimas, Siqueland, Jusczyk, and Vigorito (1971), infants are known to begin life with perceptual sensitivities that support discrimination of similar-sounding syllables such as /ba/ -/pa/, and do so better at boundaries that are typically used across the world's languages to distinguish one phoneme from another. Moreover, these sensitivities even extend to non-native speech sound contrasts that infants have never before heard (Streeter, 1976; Trehub, 1976; Werker, Humphrey, & Tees, 1981). Importantly, infants attune to the native language phonemic repertoire across the first year of life, with a decline in discrimination of non-native phonetic distinctions (e.g., Werker & Tees, 1984) and an improvement of native ones (e.g., Kuhl, Williams,

Lacerda, Stevens, & Lindblom, 1992). Over the past four decades, many theories and models of speech perception development have been proposed (for a review see Houston, 2011) to account for this age-related attunement (also called 'perceptual narrowing'). One such mechanism is statistical learning, whereby infants are influenced by the distributional properties of their phonetic input. The first experimental evidence of distributional learning in infants was published by Maye, Werker, and Gerken in *Cognition* in 2002, with Jacques Mehler as Editor of the journal and of this particular paper.

Since this publication, a number of studies have confirmed and extended distributional learning of phonetic categories in infants (Maye, Weiss, & Aslin, 2008; Wanrooij, Boersma, & van Zuijlen, 2014a; Yoshida, Pons, Maye, & Werker, 2010; Liu & Kager, 2017; Cristia, 2018).

[☆] This paper is a part of special issue "Special Issue in Honour of Jacques Mehler, *Cognition's* founding editor".

* Corresponding author at: Department of Psychology, 2136 West Mall, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

E-mail address: jwerker@psych.ubc.ca (J.F. Werker).

<https://doi.org/10.1016/j.cognition.2021.104653>

Received 27 July 2020; Received in revised form 28 February 2021; Accepted 1 March 2021

0010-0277/© 2021 Elsevier B.V. All rights reserved.

However, it is still unclear whether the strength of distributional learning remains constant across development, or changes with age. Indeed, two recent papers reached opposite conclusions regarding age-related changes in the strength of distributional learning (Liu & Kager, 2017; Cristia, 2018). There is considerable evidence (as reviewed below) that phonetic perception is refined during an early sensitive period. Distributional learning, in contrast, is often considered a domain-general mechanism that is capable of supporting learning at any age. If the efficacy of distributional learning declines with age along the same time frame as perceptual attunement, a reexamination is warranted of whether this type of learning is necessarily always domain-general, and why statistical learning can be effective across the lifespan for some modalities/stimuli but not others. In the current study, we provide direct evidence that distributional learning is itself gated by plasticity, providing convergent evidence that distributional learning is at least one of the learning mechanisms that accounts for perceptual attunement across the first year of life.

1.1. Perceptual attunement

While infants begin life with a preference for speech over non-speech sounds (Vouloumanos & Werker, 2007) and for well-formed syllables (Gómez et al., 2014), their perception is also already influenced by experience. Newborns show a preference for listening to speech sounds heard in utero (Moon, Lagercrantz, & Kuhl, 2013), and by 4–6 months start showing differential processing of familiar, native speech sounds (Kuhl et al., 1992; Polka & Werker, 1994; Yeung, Chen, & Werker, 2014). By 10 to 12 months of age, infants show a decline in discrimination of many non-native consonant distinctions (e.g. Werker & Tees, 1984, see Werker & Hensch, 2015, for a review) and an improvement in native ones (e.g., Kuhl et al., 2006; Sato, Ito, & Mazuka, 2010; Narayan et al., 2010), a process we call “perceptual attunement”. While there are some non-native distinctions that remain discriminable across the lifespan, for many others, discrimination of non-native distinctions is poor after 12-months, suggesting that plasticity may have begun to abate by this age.

The timing of declining plasticity is set by both environmental and biological constraints. In infants born prematurely, the time period for perceptual attunement corresponds with their gestational age rather than their chronological age since birth, suggesting that additional language experience can only impact the system when it is maturationally ready (Peña et al., 2012). Moreover, when pharmacological agents prematurely open circuit plasticity, the period of perceptual attunement shifts forward in time (Weikum, Oberlander, Hensch, & Werker, 2012). In contrast, bilingual infants show a protracted period of perceptual attunement, such that the ability to discriminate non-native contrasts remains open well past the age at which monolingual infants fail to do so (Petitto et al., 2012). Thus, total exposure in combination with maturational openness to input appears necessary to tune phonetic categories.

Perhaps the strongest evidence that perceptual attunement is gated by biological sensitive periods comes from studies showing that early linguistic experience has a lasting impact on adult perception. Take the case of international adoption, where early experience is vastly outweighed by subsequent language input. These studies show that even when an individual has no exposure to their first language past the infancy or toddler period, adult relearning of difficult phonological contrasts from the first language is faster and more complete than it is in individuals who never had such early exposure (Choi et al., 2017; Singh, Liederman, Mierzejewski, & Barnes, 2011; Oh, Au, & Jun, 2009). More recently, Norrman and colleagues compared adults who were adopted from China to Sweden before 30 months of age (mean age, 18.5 months) to native Mandarin and native Swedish adults on their discrimination of a vowel distinction used only in Swedish /y/-/ʏ/ with a lexical tone distinction used only in Mandarin (high flat vs high rising). ERP results showed the adoptees discriminated the Chinese tone as well as Chinese

native speakers, but were no better than the Chinese native speakers on the Swedish vowel distinction, and significantly worse than Swedish adults (Norrman, Bylund, and Thierry, 2019). Neuroimaging studies reveal that this early experience alters the way linguistic information is processed years later. In adults who were adopted from China to Quebec as infants with no further use of a Chinese language, fMRI activation to a Chinese tone distinction matched neuronal activity patterns observed in Chinese-French bilinguals who had moved to Quebec at the same age, but had maintained their Chinese (Pierce, Klein, Chen, Delcenserie, & Genesee, 2014). These results provide striking evidence that early experience indelibly alters phonetic sensitivity, and provide additional evidence that the openness to experience for phonetic learning has begun to close early in life.

1.2. Distributional learning

As noted above, there is increasing evidence that distributional learning may be a mechanism by which infants attune to the phonetic categories of their native language (see Swingley, 2019; Werker, 2018, for other possible learning mechanisms). In any given language, the distribution of produced speech sounds contains acoustic cues that reflect phonemic categories (see Lotto, Sato, & Diehl, 2004; Werker et al., 2007 for analyses of selected corpora, and Schatz, Feldman, Goldwater, Cao, & Dupoux, 2019 for an acoustic analysis of a more comprehensive corpus). Spoken syllables differ from one another along a number of different acoustic/phonetic dimensions (e.g. loudness, voice onset time (VOT), formant structure, pitch contour), but only some of these are critical to phonemic category distinctions. For any particular minimally different syllable pair, the criterial differences tend to cluster in a bimodal distribution along the relevant acoustic dimension (Fig. 1). For example, native English speakers’ productions of /r/ and /l/ fall into a bimodal distribution along the trajectories of the 2nd and 3rd formant frequency onsets (Lotto et al., 2004). In contrast, if phonetic variation is not used phonemically, i.e., to contrast meaning in a given language, then speech sound production falls into a more flat or unimodal distribution along the same acoustic dimension. This is seen in Japanese speakers’ productions of the liquid flap (illustrated with the word ‘ringo’ in Fig. 1), the Japanese consonant that falls roughly between the English /r/ and /l/ (Lotto et al., 2004).

The logic of distributional learning is that infants growing up in an English environment will be exposed to a bimodal distribution of native

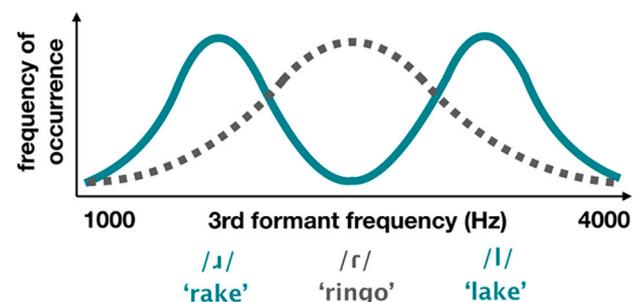


Fig. 1. Schematic illustrating the distribution of speech sounds infants of different language environments experience.

Infants growing up exposed to English (teal) will hear variability in the pronunciation of liquid phones which supports the distinction between ‘ra’ and ‘la’, whereas infants exposed to Japanese (dashed grey) will hear variability in pronunciation that does not support this distinction. So, for example, each time English-learning infants hear the word ‘rake’ there is some variability in the onset frequency of the 3rd formant of the /r/, however this variability falls in a narrow range, which is separable from the variability which occurs around an English speaker’s production of the /l/ in the word ‘lake’. In contrast, when Japanese learning infants hear the Japanese word ‘ringo’ (apple), the variability in the onset frequency of the 3rd formant of the /r/ spans the English /r/ and /l/ boundary.

phonemes, e.g., /ra/ vs /la/, that supports maintenance or sharpening of /ra/-/la/ discrimination, whereas infants growing up hearing Japanese will be exposed to statistics that support collapsing the distinction. While a distributional learning experimental study has not, to our knowledge, previously been done with /ra/-/la/, studies of distributional learning of other consonant (Maye et al., 2008; Maye, Werker, & Gerken, 2002; Yoshida et al., 2010), vowel (Wanrooij et al., 2014a, b), and tonal (Liu & Kager, 2014; Liu & Kager, 2017) contrasts show that experimentally manipulating distributional exposure does indeed support enhanced discrimination or collapse of phonetic distinctions in young infants.

Distributional learning has been assessed experimentally by first briefly familiarizing infants to a set of syllables that – although presented in random order – bear the statistics of either a unimodal or bimodal frequency distribution, prior to testing discrimination of the phonetic category difference. Following familiarization to syllables drawn from a bimodal distribution, infants aged 6–8 months and younger were found to discriminate the contrast, while those exposed to a unimodal distribution did not (e.g., Maye et al., 2002). Moreover, bimodal exposure has been shown to generalize improved discrimination of one contrast to another (e.g. da/ta to ga/ka and vice versa) if the relevant acoustic/phonetic dimension (in this case voicing) is held constant (Maye et al., 2008). Generalization indicates that infants are learning higher-level phonetic regularities, as opposed to simple repetition of specific speech sounds, suggesting distributional learning to be a potentially powerful learning mechanism for explaining perceptual attunement.

Distributional learning, then, is both a theoretically plausible and experimentally verified method through which infants may learn the phonetic categories of their native language. However, statistical learning is widely considered a domain-general mechanism that supports learning throughout life. From this perspective, despite the stable representations accrued over a lifetime of daily language listening experience, the brain should still be open to changing perception if sufficient learning time is provided. Such a view, however, fails to consider the developmental consolidation – particularly of specialized sensory and perceptual systems (such as language) that occur during critical windows in development (Hensch, 2004; Reh et al., 2020; Takesian and Hensch, 2013). These changes in circuit properties which are organized through a combination of innate, genetic instructions; spontaneous activity; and environmentally driven neuronal activity may govern the developmental time period over which experience in general, and distributional learning in particular, might be most effective in driving perceptual attunement (see Werker & Hensch, 2015).

Studies of distributional learning provide direct evidence that it is more effective in early infancy than in childhood and adulthood. In illustration, Maye and Gerken (2001) investigated the impact of distributional learning on English speaking adults' discrimination of the English unaspirated [t] from voiced [d]. This contrast is interesting because while it is native to English, the unaspirated [t] is not naturally found in a syllable initial position, and is difficult for adults and older infants to distinguish from a voiced [d] (Pegg & Werker, 1997). Despite this, bimodal exposure did not enhance discrimination; instead, discrimination declined following familiarization to the unimodal distribution (Hayes-Harb, 2007; Maye & Gerken, 2001). Studies have also examined the effect of distributional exposure on adult perception of foreign vowel contrasts (Escudero, Benders, & Wanrooij, 2011; Wanrooij & Boersma, 2013; Wanrooij, Escudero, & Raijmakers, 2013). Native Spanish speakers, living in the Netherlands and learning Dutch as a second language, were exposed to a bimodal distribution of the Dutch /a/-/a:/ vowel contrast, which is difficult for Spanish speakers to perceive. While there was improvement in discrimination of this acoustically salient vowel length difference following bimodal exposure, music was used as the control condition, as opposed to either a flat or unimodal distribution. Therefore, it is impossible to ascertain whether the overall effect was driven by simple exposure to the sounds, or whether the distributional information played a role. A follow-up study that contained a

unimodal distribution condition showed that adult discrimination of the contrast improved by the same amount regardless of exposure condition, arguing that simple exposure to the speech sounds, and not the distributional information, was driving the effects (Wanrooij, Boersma, & Benders, 2015). Moreover, such second-language learners were being exposed to the Dutch vowel contrast in their everyday lives, possibly boosting their sensitivity to the contrast. Indeed, length of residence in the Netherlands was correlated with the degree of learning following distributional exposure (Wanrooij et al., 2013).

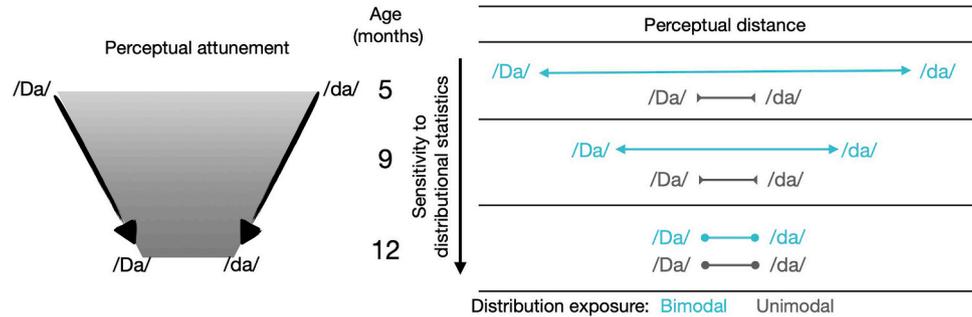
We are aware of only one study that directly compares the effect of distributional learning in young infants and adults. Wanrooij, Boersma, and van Zuijlen (2014b) tested Dutch listeners' perception of a Southern British English vowel contrast using EEG. While they found that infants as young as 2–3 months were sensitive to the distributional information, adults were not, even following 12 min of familiarization to the distributions. The authors further found no evidence of improvement in discrimination driven by distributional exposure on a behavioral level (Wanrooij, De Vos, and Boersma, 2015). Taken together, these studies indicate adult perception may be improved to some degree by repeated exposure to different speech sounds, but appears to be much less sensitive to the statistics of the presented tokens. Studies with older infants indicate that the decline in efficacy of distributional learning in adults begins around the end of the first year of life, in accordance with the timing of perceptual attunement. Twice as much exposure to a bimodal distribution was required to alter 10-month-olds' discrimination of a non-native plosive contrast, when compared with 6-month-olds (Yoshida et al., 2010). Similarly, bimodal exposure to a tonal contrast improved Dutch infants' discrimination of a non-native tonal contrast, but only at 11-months, not 14-months of age (Liu & Kager, 2017).

The studies reviewed above showing a decline in distributional learning around the end of the first year of life and beyond have asked whether bimodal exposure can improve discrimination of a non-native phonetic contrast. A potential confound, however, is that ongoing attunement to the native phonemic repertoire in infancy leads to a declining ability to discriminate non-native phonetic contrasts (Fig. 2a). Therefore, past the period of perceptual attunement, infants and adults may no longer be able to even perceive the differences along the continuum of the distributional information presented. As such, testing distributional learning by using a non-native contrast conflates the efficacy of the learning mechanism itself with the listeners' possible insensitivity to the acoustic/phonetic information to be learned. Here, we tested the efficacy of distributional learning by exploring whether familiarization to a unimodal distribution could collapse discrimination of a phonemic contrast that is native for our population of infants (see Fig. 2b) but that is known to attune across the first year of life in infants for whom the contrast is non-native (see Fig. 2a).

Theoretically, distributional learning serves both to enhance discrimination (if exposed to a bimodal distribution of speech sounds), or interfere with discrimination (if exposed to a unimodal distribution of speech sounds). Thus, testing age-related changes on collapsing a native contrast in native listeners eliminates the confound outlined above (Fig. 2b). If they are open to the effects of distributional learning, the presentation of a unimodal distribution along a native phoneme continuum would be predicted to interfere with infants' discrimination of the phonemic contrast, whereas the presentation of a bimodal distribution would be predicted to maintain or strengthen it. Moreover, the strength of this effect can be assessed at any age, since sensitivity to the relevant acoustic dimensions should be maintained for native phonemic contrasts. To our knowledge, no study has yet assessed how distributional learning alters discrimination of a native phoneme distinction across age. Thus, the current study is the first to examine age-related differences in the efficacy of distributional learning in collapsing a native contrast.

A related but distinct concern is the accumulation of language experience across different ages. It could be argued that a decline in distributional learning with age represents not a decline in circuit

A. Nonnative



B. Native

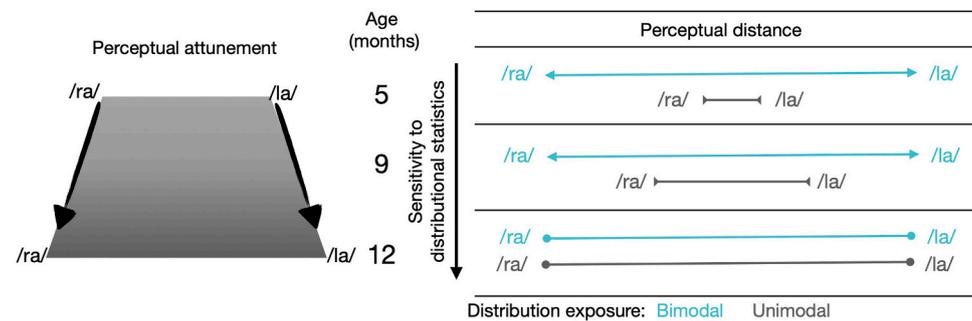


Fig. 2. Diagram outlining the predicted impact of distributional exposure across age. A. Over the period between 5 and 12 months of age, infants’ phonetic perception attunes to their native language environment, such that they become less adept at discriminating certain sound contrasts that are not present in their native language. At 5 months of age, exposure to a bimodal distribution is predicted to enhance infants’ discrimination of this contrast, while exposure to a unimodal distribution would be expected to collapse it. If infants are indeed less sensitive to distributional information, then by 12 months of age, discrimination should be reduced (due to perceptual attunement), and neither condition should show measurable discrimination between the phones. B. In the case of native phones, infants’ ability to discriminate the sound contrast is maintained or even increases over the period from 5 to 12 months. At 5 months, distributional exposure leads to the same predictions as those made if a non-native sound contrast is used, with an increase in perception observed in the bimodal condition, and a decline in perception observed in the unimodal condition. However, the 12-month-olds show a very different pattern if a native contrast is used, with discrimination of the native contrast maintained in both conditions.

plasticity, but instead the inability of a brief, recent exposure to overwhelm the accumulation of native language statistics. Apart from strong evidence in the literature (above) — particularly that from international adoptees — the conflation of declining statistical learning efficacy with strong priors is also addressed by using a native language contrast. In this case, even young infants will have accumulated multiple times more exposure to the distinction in everyday listening than will be provided in a brief distributional learning study.

Here, we investigate the effectiveness of distributional information to alter perception of a native phonemic sound contrast by testing infants whose ages span the period of perceptual attunement. Specifically, we examine distributional learning of the English /r/-/l/ phonemic contrast in native English-learning infants. This /r/-/l/ (liquid) contrast is considered a difficult contrast both for perception and production, and is one of the hardest distinctions for non-native listeners (for example Japanese speakers) to learn to discriminate and produce (McCandliss et al., 2002; Yamada & Tohkura, 1990; Iverson et al., 2003). Indeed, multiple training studies with Japanese adults show only limited improvement even across many hours and days of training (e.g., Ingvalson, Holt, & McClelland, 2012). Moreover, perception of this speech sound distinction is influenced by experience over the first year of life. Discrimination declines for Japanese-learning infants (Tsushima et al., 1994) and improves for infants growing up in English-speaking homes between 6–8 and 10–12 months of age (Kuhl et al., 2006). Therefore, we reasoned this contrast may be sensitive to brief distributional exposure.

While the majority of distributional learning studies to date in both infants and adults have used a discrete distribution, with a limited number of steps, this is not an accurate reflection of the language environment. In any given speaker’s production of different speech sounds, no single token is ever acoustically identical. In addition, variability has been found to promote category formation in a number of other learning paradigms (Quinn & Bhatt, 2010; Rost & McMurray, 2009), therefore the narrow acoustic variability presented in traditional distributional learning studies may actually limit their effectiveness. To introduce at least some of the variability that would be encountered

naturally, we created a continuous distribution of 200 speech sounds spanning the /r/-/l/ continuum, from which the bimodal and unimodal distributions were drawn. While the continuum clearly spanned the /r/-/l/ phonetic space as rated by English adults (see Fig. 3), not all tokens drawn were equally consistently labeled as /r/ or as /l/, ensuring some perceptual variability as well. A similar approach has been used with infants previously (Wanrooij, De Vos, & Boersma, 2015), and has been found to be effective for distributional learning. In addition, this multi-step continuum allows the use of acoustically unique tokens during the test phase, which infants did not hear during familiarization.

In the current study, we tested infants at the ages of 5, 9, and 12 months to investigate changes in the strength of distributional learning of a liquid speech sound continuum. These ages were chosen in order to span the period of phonetic attunement to the native language consonant repertoire. Five months was chosen as a time point of active perceptual narrowing, when infants are sensitive to non-native phonemic contrasts but beginning to show evidence of some sensitivity to native information (Yeung, Chen, & Werker, 2013; Choi et al., 2017), have yet to improve on native phonemic contrasts, and should be

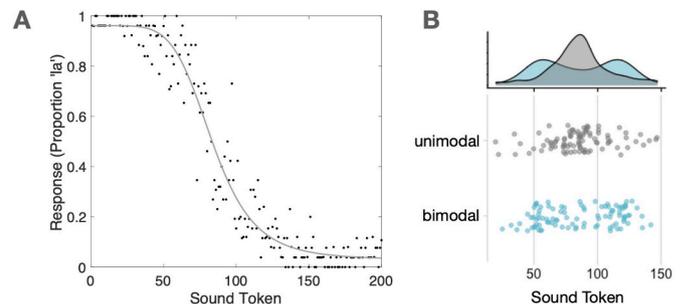


Fig. 3. Stimuli set. A. Proportion of time each token was rated as ‘la’ by a group of 10 native English speaking adults. B. The sound tokens played during the bimodal (teal) and unimodal (grey) condition, with accompanying density plot of the distributions.

sensitive to the distributional information in the language environment (Kuhl et al., 2006; Werker & Tees, 1984). The 9-month age was chosen as a point when infant perception of phonetic contrasts has begun to narrow to the native language environment, as reflected in a reduction in their ability to perceive non-native sound contrasts (Werker & Tees, 1984). By 12 months, infant perceptual reorganization to the native phonemic repertoire should be complete; by this age most infants no longer demonstrate discrimination of non-native phonemic contrasts, and their perception of native sound contrasts has improved (Kuhl et al., 2006; Werker & Tees, 1984).

Following familiarization, infants were tested on their ability to detect a change from /ra/ to /la/ or vice-versa using an event-related potential (ERP) oddball detection task. We hypothesized that distributional learning would decrease in effectiveness across the three ages tested, as the period of phonetic attunement closed. Specifically, we expected to see a reduced effect with age (between 5 and 12 months) in the collapsing of a native /r/-/l/ contrast following familiarization to a unimodal distribution.

2. Methods

2.1. Participants

Data from 45 five-month-old (20 female, age in days: median = 148, min = 138, max = 167), 44 nine-month-old (22 female, age in days: median = 274, min = 258, max = 289), and 42 twelve-month-old (22 female, age in days: median = 365, min = 351, max = 418) infants are included in the present analysis. Families were recruited from the participant database of the Early Development Research Group at the University of British Columbia. As previous studies have reported an extended period of perceptual narrowing for bilingual infants (Petitto et al., 2012), we restricted our subject pool to infants growing up in a monolingual English family environment (for which English comprised at least 90% of infant language input). Language exposure was assessed using a modified version of the language exposure questionnaire (Bosch and Sebastián-Gallés, 1997; see also Byers-Heinlein et al., 2020). The mean English language exposure was 97%. An additional 94 infants were tested but excluded from analysis due to failure to complete the study ($n = 30$), parental interference during the study ($n = 1$), equipment error ($n = 9$), parental report of over 10% exposure to a language other than English ($n = 2$), or failure to meet EEG data quality criteria ($n = 52$).

2.2. Stimuli

We recorded natural tokens of the syllables [ra] and [la] produced by a male native English speaker. Each syllable was 356 msec in duration. One token of each syllable was chosen to serve as the end point of the distribution, and a continuum of 200 speech sounds spanning the acoustic space between the end points was created using the program TANDEM-STRAIGHT (2008 ICASSP). Due to the non-linearities of acoustic perception we recruited ten adult native-English speakers (8 female) to determine the perceptual boundary along the /ra/-/la/ continuum. Participants listened to nine repetitions of each sound token, presented in random order, and were asked to categorize the sound as either a [ra] or a [la]. Results from this study are shown in Fig. 3a, and illustrate a clear perceptual boundary along the speech sound continuum.

Using this sound continuum, we created both a unimodal and bimodal distribution of speech sounds, modelled after those used in Maye et al. (2002). To create the familiarization distributions, the continuum was divided into eight bins, based on adults' perceptual ratings. A polynomial was fit to the perceptual curve, and bin boundaries determined based on the token at which the proportion of 'la' responses fell between 1 and 0.9 (bin 1), 0.9–0.8 (bin 2), 0.8–0.7 (bin 3), 0.7–0.5 (bin 4), 0.5–0.3 (bin 5), 0.3–0.2 (bin 6), 0.2–0.1 (bin 7), and 0.1–0 (bin

8). Bins 4 and 5 span slightly larger perceptual space due to the rapid perceptual shift from 'la' to 'ra' along the continuum, resulting in a limited number of tokens within this range. Once bin boundaries were defined, tokens were randomly selected from each bin to form the unimodal and bimodal distributions, according to Table 1. The resulting distributions are shown in Fig. 3b.

2.3. Design

Infants were randomly assigned to one of two familiarization conditions. During familiarization, infants were presented with syllables drawn from the liquid speech sound continuum; the syllables were drawn from either a bimodal (bimodal condition) or a unimodal (unimodal condition) distribution (Fig. 3b). Immediately following familiarization, infants were tested on their ability to discriminate between two speech sounds taken from near the ends of the continuum (token 18 and token 154). These tokens were not presented to the infants during the familiarization phase. Discrimination was assessed using an ERP oddball paradigm.

2.4. Experimental procedure

Infants were seated on a parent's lap in a dimly lit, sound-attenuated chamber. The experiment was presented via Psychtoolbox-3 in Matlab (Mathworks, Inc.), running on a Dell laptop outside the sound-attenuated testing booth (IAC Acoustics). During the familiarization phase, 96 tokens from the /ra/-/la/ continuum were played at a rate of 1.4 Hz. Similar to the Maye et al. (2002) paradigm, 48 filler syllables were also presented during this phase, in this case [da] syllables recorded from a native Hindi speaker. The familiarization phase was broken into 6 blocks of 24 s duration, with 4 s of silence between each block, and lasted approximately 2.5 min. Immediately following familiarization, an ERP oddball paradigm was used to assess infants' discrimination, and lasted approximately 5 min. A single token, serving as the standard, was presented in repetition at a rate of 1 Hz. A new token, the deviant, semi-randomly occurred ~17% of the time. The semi-random design did not allow for a deviant to immediately follow another deviant token, and specified the repetition of at least 4 standard tokens at the start of the paradigm. In order to control for possible directionality effects in discrimination, the deviant was counterbalanced such that half the infants heard token 18 as the deviant, and half heard token 154. A total of 300 stimuli were presented during this phase. Auditory stimuli were presented to the infant through a speaker (Fostex 6301NE) at a volume level of 70 dB. A silent video showing screen savers and a short cartoon (which contained no human characters or mouth movements) was presented on a Samsung 24" LCD monitor positioned in front of the infants for the duration of the study to help them stay still and engaged. Parents listened to masking music through headphones. During the assessment phase, if infants attempted to pull at the EEG cap, or excessive movement was disrupting the recording, a research assistant blew bubbles to distract the infant. Any intervention by the researcher during the familiarization phase led to the infant being excluded from the analysis. The experiment was recorded using a Panasonic Color

Table 1

Table showing the number of tokens randomly selected from each bin for the unimodal and bimodal distributions respectively.

Bin number	Unimodal	Bimodal
1	4	4
2	4	16
3	8	8
4	16	4
5	16	4
6	8	8
7	4	16
8	4	4

CCTV camera and SHURE SLX1 microphone to facilitate data analysis.

2.5. EEG recording and analysis

Infant electrophysiological responses were measured using a Hydrocel Geodesic 64-channel cap, acquired with the Net Amps 400 amplifier (Philips-Electrical Geodesics, Inc.). Infants' head circumference and distance from nasion toinion and ear to ear were measured in order to select and correctly place the EEG cap. Electrode impedances were measured for each channel and kept at or below 50 k Ω . Any channels reading above 50 k Ω were marked as bad in subsequent analysis. During recording, the EEG signal was sampled at a rate of 1000 Hz and referenced to the vertex (Cz).

Preprocessing analysis was done in NetStation (Philips-Electrical Geodesics, Inc). The EEG signal was band-pass filtered from 0.3 to 30 Hz, segmented into 1-s-long epochs beginning 200 msec prior to syllable onset and baseline corrected for the first 200 msec. Automatic artifact detection identified channels with a voltage change of over 200 μ V within an 80 ms sliding window. Bad channels were interpolated and the signal re-referenced to the average reference. If over 10% of channels were identified as bad, data for that infant were excluded from analysis. Automatic artifact detection also identified segments containing a blink artifact, identified by a change in voltage of greater than 140 μ V over a sliding 80 ms window. These artifacts were reviewed by an experimenter blind to infant condition. Segments containing an eye blink or other movement artifact were marked as bad and excluded from analysis. If over 50% of the deviant epochs were marked as bad, data for that infant were excluded from analysis. If an infant did not make it through the entirety of the assessment phase, data from that infant was included if the infant contributed at least 20 good deviant trials. The total number of epochs included for each condition and age group are shown in Table 2. Data were exported to a .mat file for further analysis and visualization in MATLAB (Mathworks, Inc.) and MNE (Gramfort et al., 2013; Gramfort et al., 2014).

ERPs to the standard and deviant were generated for each infant by averaging together all good trials from each condition. For the maturational comparison, the ERP amplitude and latency of the first positive peak (P1) were calculated from the response to the standard trials using MNEs EvokedArray.get_peak method (Gramfort et al., 2014). Since the get_peak method selects the channel with the largest peak voltage in order to calculate amplitude and latency, and slight variations in peak latency have been observed at different scalp locations (Hamburger, and Burgt, 1991), this method was not used to calculate the peak latency for the comparison between the standard and the deviant in each individual. Instead, the ERP from 6 central-frontal channels (3, 4, 6, 7, 9, 54) were averaged together. P1 latency was calculated by finding the time of the peak positive voltage response within a 200 ms window centered on the mean ERP latency of each age group.

The mismatch response (MMR) was calculated by subtracting the average ERP to the standard stimuli from the average ERP to the deviant stimuli. The MMR was assessed using a cluster-based sensor-by-time permutation *F*-test to determine the relevant time windows and electrodes over which the ERP to the deviant stimuli significantly differed from the response to the standard. This approach addresses the multiple comparisons problem inherent in EEG data by binding together related spatiotemporal features (Maris & Oostenveld, 2007). Clustering was thresholded at an alpha level of 0.005. The cluster-based test statistic

was then compared to the permutation distribution, obtained by randomly sorting trials into the standard and deviant conditions, calculating the test-statistic on each random draw, and repeating 1000 times.

3. Results

3.1. Maturation of ERP response with age

The ERP response to both the standard and deviant stimuli at all ages was characterized by a frontocentral positive deflection followed by a negative deflection (Fig. 4). With age, there was a significant reduction in peak latency (5-month: 209 ms \pm 0.021, 9-month: 186 ms \pm 0.017, 12-month: 185 ms \pm 0.024; 1 way ANOVA p < 0.0001, F = 17.34), as well as peak amplitude (5-month: 8.91 μ V \pm 2.62, 9-month: 6.50 μ V \pm 2.00, 12-month: 6.11 μ V \pm 2.74; 1 way ANOVA p < 0.0001, F = 16.01). Tukey's multiple comparison post-test confirmed that the major driver of the effect is the drop in both amplitude and latency occurring between 5 and 9 months of age. Similar changes have been reported in previous infant ERP studies (Morr et al., 2002; Choudhury & Benasich, 2011; Ortiz-Mantilla, Hämäläinen, Realpe-Bonilla, & Benasich, 2016), reflecting maturation of the neuronal processes underlying this response. While ideally data from all three ages would be analyzed in an omnibus analysis, given the difference in amplitude and latency of the ERP response at 5-months in comparison to 9- and 12-month-old infants, only the ERP data at 9- and 12 months was combined in an ANOVA, with the 5-month-old infants analyzed separately.

3.2. 5-month-old analysis

We confirmed that distributional learning was able to alter phonetic perception in a group of 5-month-old infants. We first assessed whether infants in the bimodal group were able to discriminate the /ra-/la/ contrast. A cluster based sensor-by-time permutation *F*-test was run to assess whether the response to the standard differed from the response to the deviant. A single significant cluster was identified in the central-frontal region, from 354 to 619 ms following stimulus onset (Monte Carlo p = 0.013; Fig. 5). This time window and region is consistent with the time-course of the mismatch response found in the literature at this age group (Dehaene-Lambertz, 2000).

To determine whether familiarization to the unimodal distribution altered infant perception, we repeated this analysis on the unimodal data set. The permutation *F*-test revealed no significant cluster, with the smallest Monte Carlo p value at 0.589.

In order to more directly assess differences between the conditions, we compared the mean mismatch response for the electrodes and time window identified in the bimodal condition to that of the unimodal condition. The mean voltage for this time period in the bimodal group was 2.50 \pm 2.31 μ V, while the mean voltage over the same time period in the unimodal group was -0.66 \pm 3.40 μ V. The two groups were significantly different (ttest, p = 0.0006, tstat = 3.65), corroborating the results found using the permutation cluster approach.

It should be noted that at this age we observed a positive MMR, as opposed to the more traditional negativity observed in adults. This positive response to the deviant has been observed previously in studies investigating infant phonetic perception (Dehaene-Lambertz, 2000). Changes in the MMR from a positivity to a negativity over development

Table 2

Number of epochs included in the analysis for each group, showing the mean, min and max values.

	5 month		9 month		12 month	
	Unimodal	Bimodal	Unimodal	Bimodal	Unimodal	Bimodal
Standard	175 (127, 241)	150 (94, 227)	145 (78, 220)	142 (97, 195)	143 (98, 212)	155 (81, 216)
Deviant	39 (25, 56)	34 (25, 45)	34 (20, 52)	34 (25, 43)	33 (24, 47)	36 (20, 48)

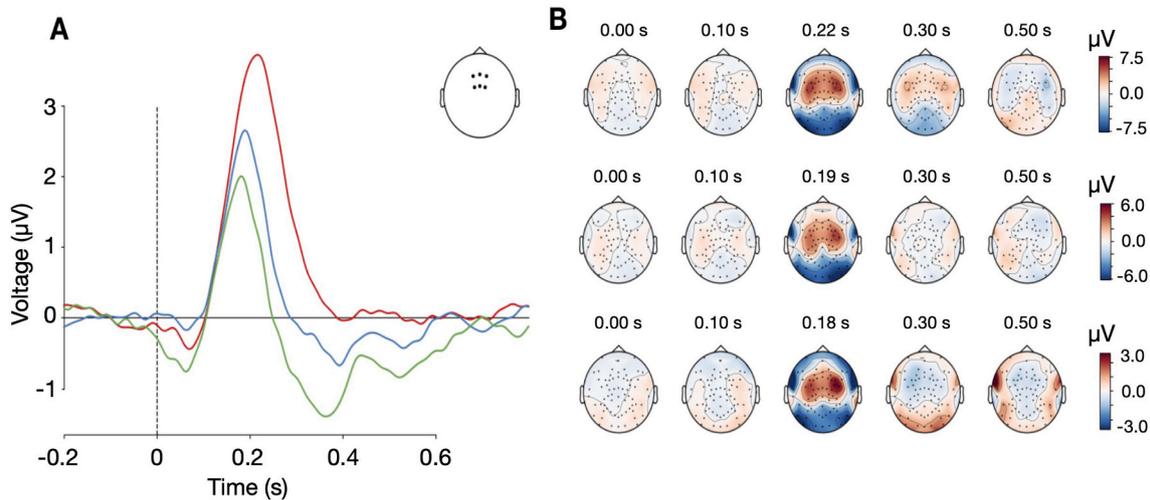


Fig. 4. Development of ERP response.

A. Neuronal response to the standard stimuli over the frontal central electrodes for 5- (red), 9- (blue), and 12- (green) month-olds.

B. Neuronal response to the standard showing the topographical distribution of the voltage change over the scalp, in 5- (top), 9- (middle), and 12- (bottom) month-olds.

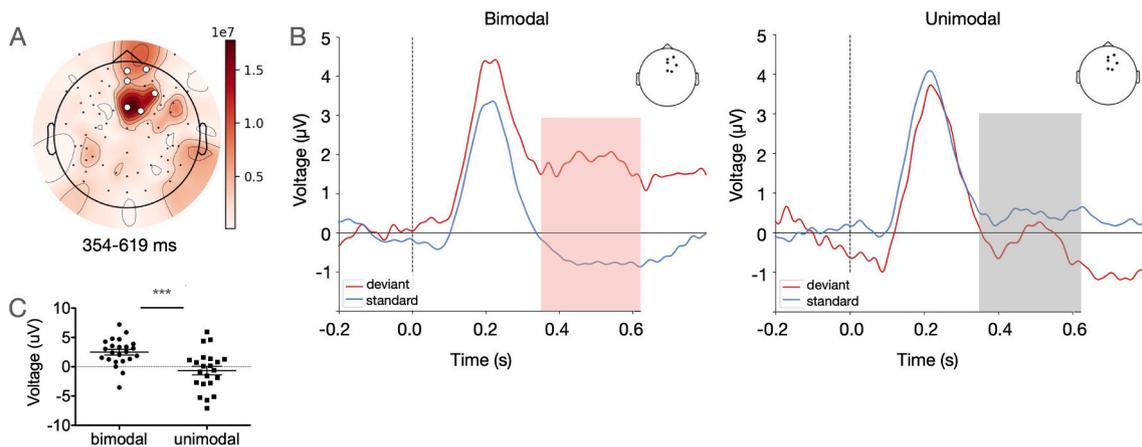


Fig. 5. 5-month-old infants are sensitive to distributional exposure during familiarization.

A. Topographic map of the averaged F -values from the sensor-by-time permutation cluster analysis on the 5-month bimodal data, showing a significant cluster over the central electrodes, between 354 and 619 ms following stimulus onset.

B. ERP elicited by the standard (blue) and deviant (red) stimulus for infants in the bimodal (left) and unimodal (right) condition. The time window for the cluster identified in the bimodal analysis is shaded in both panels, with pink used to highlight Monte Carlo $p < 0.05$. Topographic map indicates electrodes included in the generation of the ERP waveform.

C. Mean MMR voltage over the electrodes and time window identified in the permutation analysis, comparing bimodal and unimodal groups.

have been documented, and vary depending on the difficulty of the sound contrast (Morr et al., 2002; Maurer, Bucher, Brem, & Brandeis, 2003; Cheng & Lee, 2018). Therefore, it is possible that the unimodal group does not represent a failure to discriminate the sound, but instead the averaging together of two groups, one showing a positive MMR and one showing a negative MMR. If this were indeed the case, we would predict the distribution of MMR to fall into two groups, one negative and one positive. To test this, we performed D'Agostino & Pearson omnibus normality test, which indicated that both the unimodal and bimodal data were normally distributed (bimodal $K2 = 2.823$, unimodal $K2 = 0.025$), arguing against this interpretation.

3.3. 9- and 12-month-old analysis

Next, we examined how distributional learning changes with age. We first confirmed that infants from 9 to 12 months of age were able to discriminate the /ra-/la/ contrast in the bimodal condition. A cluster

level sensor-by-time permutation F -test was run to assess whether the response to the standard differed from the response to the deviant. A single significant cluster was identified in the central-frontal region, from 241 to 329 ms following stimulus onset (Monte Carlo $p = 0.045$; Fig. 6a). While infants at this age still showed a mismatch positivity as opposed to a negativity in response to the deviant sound, evidence of maturation was seen. While a significant MMR was not observed in the 5-month-olds until after 350 ms, in 9- and 12-month-olds the response occurred more rapidly and was localized to a left-central cluster, possibly reflecting lateralization of language responses. Next, we ran the same cluster analysis on infants exposed to the unimodal condition. No significant clusters were identified (Fig. 6b).

Given our hypothesis that the response to the unimodal condition may change with age, we ran a cluster analysis on the 9- and 12-month-olds separately. The 9-month-olds did not show a significant response to the deviant token (with the smallest Monte Carlo p value = 0.555), indicating that unimodal exposure did collapse discrimination. In

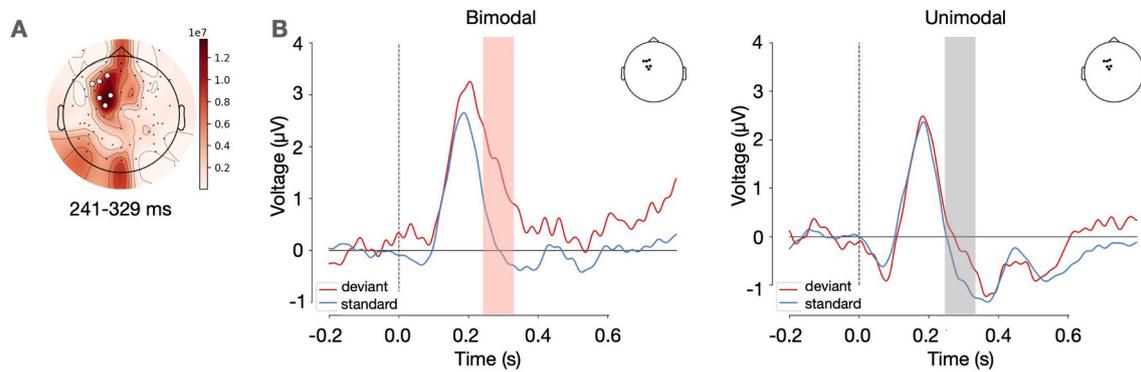


Fig. 6. Infants at 9- and 12-months discriminate the contrast following exposure to the bimodal distribution.

A. Topographic map of the averaged F -values from the sensor-by-time permutation cluster analysis on the 9- and 12-month bimodal data, showing a significant cluster over the left central electrodes, between 241 and 329 ms following stimulus onset.

B. ERP elicited by the standard (blue) and deviant (red) stimulus for 9- and 12-month-old infants in the bimodal (left) and unimodal (right) condition. The time window for the cluster identified in the bimodal analysis is shaded in both panels, with pink used to highlight Monte Carlo $p < 0.05$. Topographic map indicates electrodes included in the generation of the ERP waveform.

contrast, the 12-month-olds exhibited a mismatch response very similar to that seen in the bimodal group, with a positive voltage response over the left-frontal-central channels occurring from 201 to 414 ms (Monte Carlo $p = 0.039$; Fig. 7).

In order to directly compare the unimodal and bimodal conditions across the two ages, we calculated the mean voltage of the mismatch response for the sensors and time window identified in the bimodal cluster permutation analysis. A two-way ANOVA showed no significant interaction between condition and age ($p = 0.974$, $F = 0.001$), and close but not significant effects of both condition ($p = 0.090$, $F = 2.944$) and age ($p = 0.059$, $F = 3.640$). As the lack of a significant effect in this analysis makes it difficult to interpret the data, we turned to a different analysis approach.

Latency of the ERP response has also been used as an indication of detection of the deviant at these ages, and captures similar information to the mismatch response (Ortiz-Mantilla et al., 2016), with the ERP to the deviant showing a longer latency than the ERP to the standard. A 3-way mixed ANOVA on the 9- and 12-month-old data revealed a significant interaction between condition (bimodal vs unimodal), token type (standard vs deviant) and age ($p = 0.047$, $F = 5.889$). We then followed this with 2-way mixed ANOVAs on the 9- and 12-month-old infants

separately to determine how the effects differed at each age. At 9 months, there was a significant condition \times token type interaction ($p = 0.004$, $F = 9.23$), with infants in the bimodal condition showing the expected increase in ERP latency to the deviant stimulus, but not the unimodal condition (Fig. 8). In contrast, a 2-way mixed ANOVA on the 12-month-old data showed a significant effect of token ($p = 0.017$, $F = 6.22$), but no interaction ($p = 0.541$, $F = 0.38$), and no effect of condition ($p = 0.3$, $F = 1.10$). These results provide further support that infants at 9 months remained sensitive to the familiarization exposure, while infants at 12 months were able to discriminate the /ra-/la/ stimulus regardless of condition.

4. Discussion

Statistical learning is often framed as a domain-general mechanism (Thiessen & Saffran, 2007). Diverse sensory modalities are clearly sensitive to environmental statistics. However, differences in the innate organization and connectivity of specific brain regions and their rate of maturation may all impact how statistics from the environment are processed (Reh et al., 2020). In addition, differences in the stimulus properties themselves may be expected to influence how the information is processed across different modalities. We argue that the sensitivity to all statistics present in the environment is not static across development. The distributional statistics of some kinds of stimuli are undoubtedly learnable by some brain systems throughout development. However, for other kinds of stimuli – with speech sounds, and in particular consonants being a prototypical example – there may be delimited time-periods during which the brain areas specialized for processing those sounds can be influenced by distributional statistics. Indeed, the essence of critical periods is that distinct brain areas are shaped by specific information from the environment during discrete time windows in development (Werker & Hensch, 2015). The mechanisms via which environmental statistics are learned may differ between different sensory modalities, and also across developmental time, reflecting differences in both brain capacity and stimulus properties.

In the current study, we explored how the efficacy of distributional learning of phonetic categories changes over the first year of life, using the age range during which perceptual attunement to the native phonetic repertoire has been shown to occur. Our results support the hypothesis that the strength of distributional learning declines with age, following the timeline of perceptual attunement to the native language. Specifically, we find that at 5 months of age, infants' perception is highly influenced by the presented statistical distribution. At this age, infants exposed to the bimodal condition show a clear mismatch response to the deviant stimulus. In contrast, we do not find any significant difference

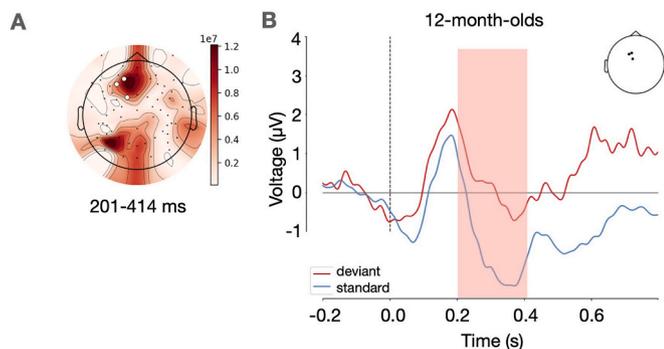


Fig. 7. Infants exposed to the unimodal distribution discriminate the contrast at 12-months.

A. The topographic map of the averaged F -values from the sensor-by-time permutation cluster analysis on the 12-month unimodal data, showing a significant cluster over the left central electrodes, between 201 and 414 ms following stimulus onset.

B. ERP elicited by the standard (blue) and deviant (red) stimulus for infants in the unimodal condition at 12 months. The time window identified in the permutation cluster analysis is shaded in pink indicating a Monte Carlo $p < 0.05$. Topographic map indicates which electrodes were included in the generation of the ERP waveform.

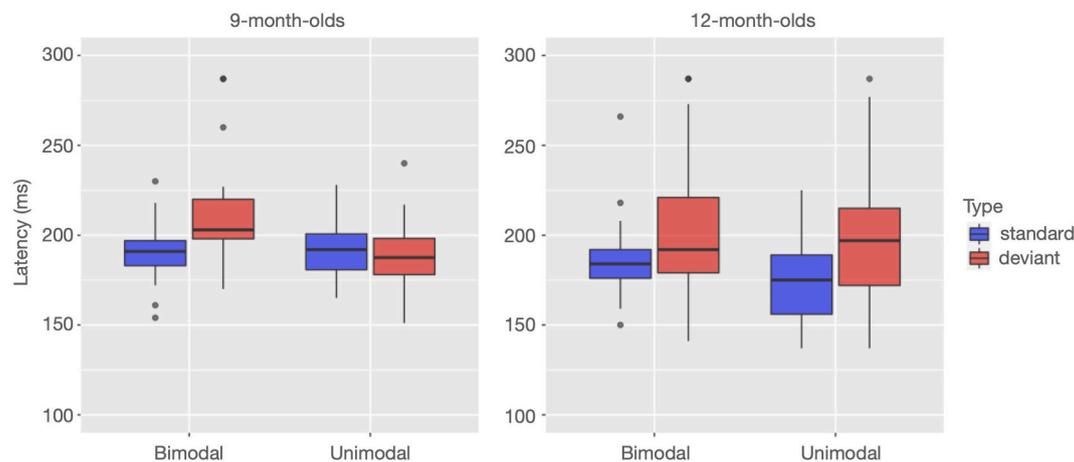


Fig. 8. ERP latency indicates a decline in the sensitivity to distributional information with age. The latency of the P1 ERP in 9- and 12-month-olds to either the standard (blue) or the deviant (red) stimulus.

between the response to the standard and the deviant stimulus after infants have been exposed to the unimodal distribution.

Between 9 and 12 months, the efficacy of distributional learning declines. At 9 months, infants in the unimodal condition fail to show a significant mismatch response to the deviant stimuli, and there is no difference in the latency of the ERP regardless of whether the standard or deviant stimulus is presented. Thus, distributional exposure is still effective in changing phonetic sensitivity at this age. In contrast, 12-month-olds exposed to the unimodal condition still show a mismatch response to the presentation of the deviant stimuli, and the latency of the response is longer to the deviant than to the standard, an indication that the change in stimulus has been detected. Due to the fact that we are exploring how distributional exposure alters perception of a *native* phonetic contrast, the fact that infants at 12 months of age continue to discriminate the /ra/-/la/ contrast even after exposure to the unimodal distribution indicates a decline in the efficacy of distributional learning at this age.

With the stimuli and design used in the current study, we can be confident that this age-related decline in the efficacy of distributional learning is not driven by a reduction in the brain's ability to detect the differences among the stimuli over which the statistics operate. While previous studies support the idea that distributional learning declines by the end of the sensitive period for phonetic attunement (Liu & Kager, 2017; Yoshida et al., 2010), the use of non-native phonetic contrasts made it impossible to disentangle whether this decline was due to a true change in the efficacy of distributional learning with age, or an inability of the infants even to perceive the two categories presented. We addressed this concern by using the native English /ra/-/la/ phonetic contrast in English-learners. Therefore, the decline in distributional learning we observe is not driven by the older infants' inability to perceive the native syllables presented, but rather by a decline in their ability to use distributional statistics to reduce sensitivity to a phonemic contrast by collapsing the phonetic category distinction.

As mentioned above, previous studies (e.g., Liu & Kager, 2017; Yoshida et al., 2010) have also reported a maturational decline in the effectiveness of distributional learning, however a recent meta-analysis suggested that distributional learning actually increases with age (Cristia, 2018). While Liu and Kager (2014) reports a large effect for a tonal contrast at 11–12 months, Liu and Kager (2017) had not yet been published and was not included in the meta-analysis. Thus, the meta-analysis may have caught the rising phase of plasticity during an early sensitive period, but not the subsequent decline. Additionally, the meta-analysis assumed that the time course of sensitivity to distributional information would be the same for different types of phonetic contrasts. While necessary in order to gather together sufficient data to run a meta-analysis, these assumptions are not supported by our current

understanding of the differences in perceptual salience of vowels, consonants, and tones. The majority of (the admittedly few) studies on distributional learning which use consonant phonetic contrasts, show the greatest efficacy of distributional learning around 6–8 months, with a decline by 10 months. While Liu and Kager (2014, 2017) report greater distributional learning at 11–12 months, their use of a tonal contrast makes it unlikely that the time course of sensitivity follows the same trajectory as that of studies using consonants. Additionally, their data support the hypothesis that the efficacy of distributional learning is not linear, but first increases then declines after 11–12 months (Liu and Kager, 2017), in line with the opening and closing of sensitive periods for phonetic category attunement (Werker & Hensch, 2015). Our data further support the hypothesis that the efficacy of distributional learning changes dynamically over development, and is most efficacious during sensitive periods of development, when infant perception of a particular phonetic contrast shows heightened sensitivity to environmental input.

Our study is not able to determine to what extent the perception of the /ra/ - /la/ contrast was enhanced in the bimodal condition versus collapsed in the unimodal condition. However, previous research suggests that some level of discrimination is present as young as 6 months independent of previous experience hearing the distinction contrastively (Kuhl et al., 2006). Therefore, our data suggest that the changes brought about by distributional learning involved collapsing of a category distinction driven by unimodal exposure. This is similar to what was seen in the original Maye et al. (2002) study, which also employed a contrast that prior research (Pegg & Werker, 1997) had suggested could be discriminated by infants at that age. The capacity of statistical information to drive a decline in perception across a category distinction provides an important mechanism by which the perceptual decline for certain phonetic contrasts may occur.

The observed decline in plastic response to environmental input could arise via a number of different underlying processes. An increase in the stability of the neuronal representation for the /ra/-/la/ phonetic contrast, for example, would make it more difficult to alter perception via short distributional exposures. In other words, the very establishment of native categories may render the neuronal representation of the phonetic contrast resistant to environmental perturbation. Alternatively, the actual strength of the neuronal systems driving circuit plasticity may decline with age, reducing the adaptive circuit response to the same environmental manipulation. These two alternatives are not mutually exclusive, and indeed both may play a role.

The first possibility outlined above is consistent with, but does not require, a critical period explanation. In the context of critical periods, this circuit stabilization occurs during a unique window of time during development, and is set regardless of the amount of exposure to novel sound contrasts after this time. However, it is also possible that circuit

reorganization which occurs during the first year of life can be overcome with enough novel language experience. Our study does not directly address this possibility. While we show that sensitivity to distributional statistics declines with age, it is possible that a longer exposure would have been able to disrupt 12-month-olds' perception of the /ra-/la/ contrast. Regardless, the fact that the same amount of exposure did not disrupt discrimination at 12 months as did disrupt discrimination at 5 and 9 months does represent a decline in the efficacy of distributional learning by this age and argues for a decline in plasticity. First, while the liquid consonant does not occur frequently, even 5-month-olds are expected to have encountered substantially more than two minutes of exposure to this contrast – thus, if discrimination was based solely on the total amount of accumulated information, distributional collapse of a native contrast would not be expected even at 5 months and certainly not at 9 months. Secondly, the studies with international adoptees reviewed in the Introduction provide strong evidence that even in cases where second language experience after infancy vastly outweighs the language experience of the first year of life, the brain retains the imprint of this first language experience.

While the exact nature of how the brain represents phonemes remains unclear, recent work in adults implanted with electrocorticography (ECoG) electrodes has shown that neurons in the Superior Temporal Gyrus (STG) are responsive to specific spectral properties in the auditory input which signal the manner and place of consonant articulation (Mesgarani, Cheung, Johnson, & Chang, 2014). Thus, for example, some neurons show selective activation in response to phones with short VOT, while others show selective activation for longer VOT (Mesgarani et al., 2014). These non-linearities in responsiveness to specific auditory cues theoretically support categorical perception of phones in adults. While we do not have similar data in infants, typically neuronal response properties become more selective over development, showing increased precision in their tuning for specific auditory properties (Zhang, Bao, & Merzenich, 2001). This increased selectivity is likely driven by a combination of refining thalamocortical and intracortical connectivity (Sun, Liu, Tao, & Zhang, 2019) and a developmental remodeling of inhibitory input, which acts to balance excitatory activity within the network (Dorrn, Yuan, Barker, Schreiner, & Froemke, 2010). These processes may increase the categorical perception of native language phonemes in adults, but may also act to limit circuit plasticity in the face of changing input after a sensitive period in early life.

Even if the physical connections remain in adulthood to support some plasticity, work in a variety of model species has revealed a number of molecular brakes which actively limit circuit plasticity. While many of these mechanisms were first described in primary sensory areas, and involve fundamental perceptual processes such as binocular matching and tuning of the tonotopic map (Takesian & Hensch, 2013), work in songbirds illustrates the generalizability of these principles to vocal-motor learning as well (Yazaki-Sugiyama, 2019). Some of these brakes are structural, such as the formation of perineuronal nets around inhibitory cells and increases in myelin. Others are functional, such as the protoxin Lynx1, which limits plasticity by dampening the ability of neuromodulators such as acetylcholine to disinhibit cortical circuits (Morishita, Miwa, Heintz, & Hensch, 2010; Takesian, Bogart, Lichtman, & Hensch, 2018).

Throughout this paper we have argued for sensitive periods in early life for phonetic learning. Yet, this does not preclude circuit plasticity after sensitive period closure. Rather, the circuit changes that occur during the sensitive period delineate the boundaries within which future plasticity can occur. This plasticity is evident in speech processing: adults adjust on a second-by-second basis to the phonetic properties of encountered native speakers (see Clayards, Tanenhaus, Aslin, & Jacobs, 2008 for experimental evidence), as well as to accents more or less well, depending upon our experience with them. Similarly, children beyond 1–2 years of age can improve their perception of many non-native speech sound contrasts, as do adults across the lifespan. There are, however, enormous individual differences in attainment (Díaz, Baus,

Escera, Costa, & Sebastián-Gallés, 2008; Golestani, Molko, Dehaene, LeBihan, & Pallier, 2007), and most often, late learning does not result in native levels of performance (see Abrahamsson et al., 2018, for a review). And, as the work with international adoptees reviewed in the Introduction shows, for many other contrasts, there is a residual imprint of early experience, revealing a limit to plasticity after the first year of life. Latent anatomical traces of early life experience have been reported in animal models (DeBello et al., 2001), which may broaden the repertoire of usable maps with no further need for dramatic adult plasticity (Knudsen, 1998).

Moreover, there are lifelong consequences on the integration and lexical use of phonetic categories learned or absent during the first year of life. Even toddlers performing word recognition and word learning tasks are unable to use some acoustically salient non-native distinctions that they can still discriminate (Dietrich, Swingley, & Werker, 2007; Liu & Kager, 2018). This suggests that their discrimination may be based on acoustic salience and not linguistic in nature. Similarly, adults show deficits in the lexical use of phonetic distinctions learned after early infancy, even for those non-native contrasts that are acoustically salient. For example, fluent Spanish-Catalan adult bilinguals who acquired Catalan in early childhood (but importantly, after first learning Spanish) show reaction time deficits on lexical decision tasks when compared to Catalan monolinguals or sequential Catalan first then Spanish bilinguals, despite daily Catalan use (Pallier et al., 1997; Pallier et al., 2001). Similarly, both bilingual Spanish-Swedish adults whose families moved to Sweden when the children were between the ages of 3–8 years, as well as adults who were adopted as children into Sweden from a Spanish speaking country at these same ages (3–8 years) perform more poorly than native Swedish-only adults in lexical decision tasks involving Swedish-only vowel length distinctions (Norrman & Bylund, 2016).

What these findings suggest is that while there is potential for lifelong plasticity, including phonological learning, it likely does not rely on precisely the same neural circuits or connections sculpted by early experience. Any complex behaviour, such as language, arises from the interaction of multiple neuronal circuits. Irrevocable changes at one level in the pathway may be compensated at the level of behaviour. We have addressed this by focusing on one mechanism – distributional learning of consonant categories – and demonstrate that, by one year of life, phonetic perception becomes less malleable in the face of altered language statistics.

Author contributions

Rebecca K. Reh: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft. **Takao K. Hensch:** Conceptualization, Methodology, Writing - review & editing, Funding acquisition. **Janet F. Werker:** Conceptualization, Resources, Methodology, Writing - original draft, Funding acquisition.

Declaration of Competing Interest

The authors have no competing interests to declare.

Acknowledgements

We would like to thank Dr. Mark Scott for the creation of the stimuli set; Maegan Mitchell for her assistance with data collection and analysis; Savannah Nijeboer and Jacqueline Cloake for their general lab and family recruitment support; and all of the parents and infants who participated in the research. We also gratefully acknowledge support from the Canadian Institute for Advanced Research to RR, JFW and TKH [XBC-WERK-162441-R], the Natural Sciences and Engineering Research Council of Canada grant to JFW [RPGIN-215-03967], and WPI-IRCN to TKH [JSPS].

Supplementary material

The data collected and analyzed in this manuscript can be found on Open Science Framework at https://osf.io/y6x5d/?view_only=a56b90e7247e4d44a5c72807686fa2bb

References

- Abrahamsson, N., Høyenstam, K., & Bylund, E. (2018). Age effects on language acquisition, retention, and loss: Key hypotheses and findings. In K. Høyenstam, I. Bartning, & L. Fant (Eds.), *High level language proficiency in second language and multilingual contexts* (pp. 16–49). Cambridge University Press.
- Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, 65(1), 33–69. [https://doi.org/10.1016/s0010-0277\(97\)00040-1](https://doi.org/10.1016/s0010-0277(97)00040-1).
- Byers-Heinlein, K., Schott, E., Gonzalez-Barrero, A. M., Brouillard, M., Dubé, D., Jarda, A., ... Tamayo, M. P. (2020). MAPLE: A multilingual approach to parent language estimates. *Bilingualism: Language and Cognition*, 23(5), 951–957. <https://doi.org/10.1017/S1366728919000282>.
- Cheng, Y. Y., & Lee, C. Y. (2018). The development of mismatch responses to mandarin language tone in 12- to 24-month-old infants. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00448>.
- Choi, J., Broersma, M., & Cutler, A. (2017). Early phonology revealed by international adoptees birth language retention. *Proceedings of the National Academy of Sciences*, 114(28), 7307–7312. <https://doi.org/10.1073/pnas.1706405114>.
- Choudhury, N., & Benasich, A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: Prediction to 3 and 4 year language and cognitive abilities. *Clinical Neurophysiology*, 122, 320–338. <https://doi.org/10.1016/j.clinph.2010.05.035>.
- Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3), 804–809. <https://doi.org/10.1016/j.cognition.2008.04.004>.
- Cristia, A. (2018). Can infants learn phonology in the lab? A meta-analytic answer. *Cognition*, 170, 312–327. <https://doi.org/10.1016/j.cognition.2017.09.016>.
- DeBello, W. M., Feldman, D. E., & Knudsen, E. I. (2001). Adaptive axonal remodeling in the midbrain auditory space map. *The Journal of Neuroscience*, 21(9), 3161–3174. <https://doi.org/10.1523/JNEUROSCI.21-09-03161.2001>.
- Dehaene-Lambertz, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience*, 12(3), 449–460. <https://doi.org/10.1162/089892900562264>.
- Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences*, 105(42), 16083–16088. <https://doi.org/10.1073/pnas.0805022105>.
- Dietrich, C., Swingle, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences*, 104(41), 16027–16031. <https://doi.org/10.1073/pnas.0705270104>.
- Dorrn, A. L., Yuan, K., Barker, A. J., Schreiner, C. E., & Froemke, R. C. (2010). Developmental sensory experience balances cortical excitation and inhibition. *Nature*, 465(7300), 932–936. <https://doi.org/10.1038/nature09119>.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303–306. <https://doi.org/10.1126/science.171.3968.303>.
- Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, 130(4). <https://doi.org/10.1121/1.3629144>.
- Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, 17(3), 575–582. <https://doi.org/10.1093/cercor/bbk001>.
- Gómez, D. M., Berent, I., Benavides-Varela, S., Bion, R. A., Cattarossi, L., Nespor, M., & Mehler, J. (2014). Language universals at birth. *Proceedings of the National Academy of Sciences*, 111(16), 5837–5841. <https://doi.org/10.1073/pnas.1318261111>.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. S. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7. <https://doi.org/10.3389/fnins.2013.00267>.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, 86, 446–460. <https://doi.org/10.1016/j.neuroimage.2013.10.027>.
- Hamburger, H. L., & Burgt, M. A. (1991). Global field power measurement versus classical method in the determination of the latency of evoked potential components. *Brain Topography*, 3(3), 391–396. <https://doi.org/10.1007/BF01129642>.
- Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, 23(1), 65–94. <https://doi.org/10.1177/0267658307071601>.
- Hensch, T. K. (2004). Critical period regulation. *Annual Review of Neuroscience*, 27, 549–579. <https://doi.org/10.1146/annurev.neuro.27.070203.143437>.
- Houston, D., & Tharpe, M. A. (2011). Infant speech perception. In R. Seewalk (Ed.), *Comprehensive handbook of pediatric audiology* (2nd ed., pp. 47–62). Plural Publishing.
- Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate /r-l/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition*, 15(2), 255–274. <https://doi.org/10.1017/S1366728912000041>.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1). [https://doi.org/10.1016/s0010-0277\(02\)00198-1](https://doi.org/10.1016/s0010-0277(02)00198-1).
- Knudsen, E. I. (1998). Capacity for plasticity in the adult owl auditory system expanded by juvenile experience. *Science*, 279(5356), 1531–1533. <https://doi.org/10.1126/science.279.5356.1531>.
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2). <https://doi.org/10.1111/j.1467-7687.2006.00468.x>.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606–608. <https://doi.org/10.1126/science.1736364>.
- Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, 133(2), 385–394. <https://doi.org/10.1016/j.cognition.2014.06.004>.
- Liu, L., & Kager, R. (2017). Statistical learning of speech sounds is most robust during the period of perceptual attunement. *Journal of Experimental Child Psychology*, 164, 192–208. <https://doi.org/10.1016/j.jecp.2017.05.013>.
- Liu, L., & Kager, R. (2018). Monolingual and bilingual infants' ability to use non-native tone for word learning deteriorates by the second year after birth. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2018.00117>.
- Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case for Japanese acquisition of /r/ and /l/. In J. Slička, S. Manuel, & M. Matthies (Eds.), *Proceedings of the from sound to sense: 50+ years of discoveries in speech communication* (pp. C381–C386). Massachusetts, MA: Massachusetts Institute of Technology.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>.
- Maurer, U., Bucher, K., Brem, S., & Brandeis, D. (2003). Altered responses to tone and phoneme mismatch in kindergartners at familial dyslexia risk. *NeuroReport*, 14(17), 2245–2250. <https://doi.org/10.1097/00001756-200312020-00022>.
- Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us? In A. H. J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th annual boston university conference on language development* (pp. 480–490). Cascadia Press.
- Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11(1), 122–134. <https://doi.org/10.1111/j.1467-7687.2007.00653.x>.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3). [https://doi.org/10.1016/s0010-0277\(01\)00157-3](https://doi.org/10.1016/s0010-0277(01)00157-3).
- McCandliss, B. D., Fiez, J. A., Protopoulos, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience*, 2(2), 89–108. <https://doi.org/10.3758/cabn.2.2.89>.
- Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, 343(6174), 1006–1010. <https://doi.org/10.1126/science.1245994>.
- Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica*, 102(2), 156–160. <https://doi.org/10.1111/apa.12098>.
- Morishita, H., Miwa, J. M., Heintz, N., & Hensch, T. K. (2010). Lynx1, a cholinergic brake, limits plasticity in adult visual cortex. *Science*, 330(6008), 1238–1240. <https://doi.org/10.1126/science.1195320>.
- Morr, M. L., Shafer, V. L., Kreuzer, J. A., & Kurtzberg, D. (2002). Maturation of mismatch negativity in typically developing infants and preschool children. *Ear and Hearing*, 23(2), 118–136. <https://doi.org/10.1097/00003446-200204000-00005>.
- Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, 13(3), 407–420. <https://doi.org/10.1111/j.1467-7687.2009.00898.x>.
- Norrman, G., & Bylund, E. (2016). The irreversibility of sensitive period effects in language development: Evidence from second language acquisition in international adoptees. *Developmental Science*, 19(3), 513–520. <https://doi.org/10.1111/desc.12332>.
- Norrman, G., Bylund, E., & Thierry, G. (2019). How early language development of international adoptees stands the test of time. In *Poster presented at the Eleventh Annual Meeting of the Society for the Neurobiology of Language, Helsinki, Finland*.
- Oh, J. S., Au, T. K.-F., & Jun, S.-A. (2009). Early childhood language memory in the speech perception of international adoptees. *Journal of Child Language*, 36, 1–10. <https://doi.org/10.1017/S0305000909990286>.
- Ortiz-Mantilla, S., Hämäläinen, J. A., Realpe-Bonilla, T., & Benasich, A. A. (2016). Oscillatory dynamics underlying perceptual narrowing of native phoneme mapping from 6 to 12 months of age. *Journal of Neuroscience*, 36(48), 12095–12105. <https://doi.org/10.1523/jneurosci.1162-16.2016>.
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64(3), B9–B17. [https://doi.org/10.1016/S0010-0277\(97\)00030-9](https://doi.org/10.1016/S0010-0277(97)00030-9).
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, 12(6), 445–449. <https://doi.org/10.1111/1467-9280.00383>.
- Pegg, J. E., & Werker, J. F. (1997). Adult and infant perception of two English phones. *The Journal of the Acoustical Society of America*, 102(6), 3742–3753. <https://doi.org/10.1121/1.420137>.

- Peña, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. *Journal of Neuroscience*, 32(33), 11159–11163. <https://doi.org/10.1523/jneurosci.6516-11.2012>.
- Petitto, L., Berens, M., Kovelman, I., Dubins, M., Jasinska, K., & Shalinsky, M. (2012). The “perceptual wedge hypothesis” as the basis for bilingual babies’ phonetic processing advantage: New insights from fNIRS brain imaging. *Brain and Language*, 121(2), 130–143. <https://doi.org/10.1016/j.bandl.2011.05.003>.
- Pierce, L. J., Klein, D., Chen, J., Delcenserie, A., & Genesee, F. (2014). Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences*, 111(48), 17314–17319. <https://doi.org/10.1073/pnas.1409411111>.
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2), 421–435. <https://doi.org/10.1037/0096-1523.20.2.421>.
- Quinn, P. C., & Bhatt, R. S. (2010). Learning perceptual organization in infancy: The effect of simultaneous versus sequential variability experience. *Perception*, 39(6), 795–806. <https://doi.org/10.1068/p6639>.
- Reh, R. K., Dias, B. G., Nelson, C. A., Kaufer, D., Werker, J. F., Kolb, B., Levine, J. D., & Hensch, T. K. (2020). Critical period regulation across multiple timescales. *Proceedings of the National Academy of Sciences*, 117(38), 23242–23251. <https://doi.org/10.1073/pnas.1820836117>.
- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12(2), 339–349. <https://doi.org/10.1111/j.1467-7687.2008.00786.x>.
- Sato, Y., Ito, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology*, 46, 106–119. <https://doi.org/10.1037/a0016718>.
- Schatz, T., Feldman, N., Goldwater, S., Cao, X. N., & Dupoux, E. (2019). Early phonetic learning without phonetic categories – Insights from large-scale simulations on realistic input. <https://doi.org/10.31234/osf.io/4w4wh>.
- Singh, L., Liederman, J., Mierzejewski, R., & Barnes, J. (2011). Rapid reacquisition of native phoneme contrasts after disuse: You do not always lose what you do not use. *Developmental Science*, 14(5), 949–959. <https://doi.org/10.1111/j.1467-7687.2011.01044.x>.
- Streeter, L. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*, 259(5538), 39–41. <https://doi.org/10.1038/259039a0>.
- Sun, Y. J., Liu, B. H., Tao, H. W., & Zhang, L. I. (2019). Selective strengthening of intracortical excitatory input leads to receptive field refinement during auditory cortical development. *Journal of Neuroscience*, 39(7), 1195–1205. <https://doi.org/10.1523/JNEUROSCI.2492-18.2018>.
- Swingle, D. (2019). Learning phonology from surface distributions, considering Dutch and English vowel duration. *Language Learning and Development*, 15(3), 199–216. <https://doi.org/10.1080/15475441.2018.1562927>.
- Takesian, A. E., Bogart, L. J., Lichtman, J. W., & Hensch, T. K. (2018). Inhibitory circuit gating of auditory critical-period plasticity. *Nature Neuroscience*, 21(10), 218–227. <https://doi.org/10.1038/s41593-017-0064-2>.
- Takesian, A. E., & Hensch, T. K. (2013). Balancing plasticity/stability across brain development. *Progress in Brain Research*, 207, 3–34. <https://doi.org/10.1016/b978-0-444-63327-9.00001-1>.
- Thiessen, E., & Saffran, J. (2007). Domain-general learning capabilities. In E. Hoff, & M. Shtatz (Eds.), *Blackwell handbook of language development* (pp. 68–86). Wiley-Blackwell.
- Trehub, S. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 47(2), 466–472. <https://doi.org/10.2307/1128803>.
- Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6–12 months: Language-specific developmental changes in speech perception abilities. In *International conference on spoken language processing, Yoko-hama, Japan*.
- Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, 10(2), 159–164. <https://doi.org/10.1111/j.1467-7687.2007.00549.x>.
- Wanrooij, K., & Boersma, P. (2013). Distributional training of speech sounds can be done with continuous distributions. *The Journal of the Acoustical Society of America*, 133(5). <https://doi.org/10.1121/1.4798618>.
- Wanrooij, K., Boersma, P., & Benders, T. (2015). Observed effects of “distributional learning” may not relate to the number of peaks. A test of “dispersion” as a confounding factor. *Frontiers in Psychology*, 6, 1241. <https://doi.org/10.3389/fpsyg.2015.01341>.
- Wanrooij, K., Boersma, P., & van Zuijen, T. L. (2014a). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology*, 5, 77. <https://doi.org/10.3389/fpsyg.2014.00077>.
- Wanrooij, K., Boersma, P., & van Zuijen, T. L. (2014b). Distributional vowel training is less effective for adults than for infants. A study using the mismatch response. *PLoS One*, 9(10), Article e109806. <https://doi.org/10.1371/journal.pone.0109806>.
- Wanrooij, K., De Vos, J., & Boersma, P. (2015). Distributional vowel training may not be effective for Dutch adults. In *Proceedings of the 18th international congress of phonetic sciences (ICPhS 2015)*. Glasgow, UK: University of Glasgow.
- Wanrooij, K., Escudero, P., & Raijmakers, M. E. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics*, 41(5), 307–319. <https://doi.org/10.1016/j.wocn.2013.03.005>.
- Weikum, W. M., Oberlander, T. F., Hensch, T. K., & Werker, J. F. (2012). Prenatal exposure to antidepressants and depressed maternal mood alter trajectory of infant speech perception. *Proceedings of the National Academy of Sciences*, 109(Supplement 2), 17221–17227. <https://doi.org/10.1073/pnas.1121263109>.
- Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics*, 39(4), 703–728. <https://doi.org/10.1017/s0142716418000152>.
- Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, 66(1), 173–196. <https://doi.org/10.1146/annurev-psych-010814-015104>.
- Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52(1), 349–355. <https://doi.org/10.2307/1129249>.
- Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, 103(1), 147–162. <https://doi.org/10.1016/j.cognition.2006.03.006>.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63. [https://doi.org/10.1016/s0163-6383\(84\)80022-3](https://doi.org/10.1016/s0163-6383(84)80022-3).
- Yamada, R. A., & Tohkura, Y. (1990). Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese. In *Proceedings of the 1990 international conference on spoken language processing*.
- Yazaki-Sugiyama, Y. (2019). Neuronal mechanisms regulating the critical period of sensory experience-dependent song learning. *Neuroscience Research*, 140, 53–58. <https://doi.org/10.1016/j.neures.2018.11.002>.
- Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68(2), 123–139. <https://doi.org/10.1016/j.jml.2012.09.004>.
- Yeung, H. H., Chen, L. M., & Werker, J. F. (2014). Referential labeling can facilitate phonetic learning in infancy. *Child Development*, 85(3), 1036–1049. <https://doi.org/10.1111/cdev.12185>.
- Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, 15(4), 420–433. <https://doi.org/10.1111/j.1532-7078.2009.00024.x>.
- Zhang, L. I., Bao, S., & Merzenich, M. M. (2001). Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nature Neuroscience*, 4(11), 1123–1130. <https://doi.org/10.1038/nn745>.