Fifty Years of Infant Vowel Discrimination Research: What Have We Learned?

Sho TSUJI*** and Alejandrina CRISTIA***

Key words: Infant development, speech perception, vowel discrimination

1. Introduction

The study of infant vowel perception over the last decades has spawned 60 papers, among which a diverse range of insights on early perceptual abilities and their development become evident (see Tsuji and Cristia submitted, for details on how this corpus was reconstructed). The aim of the present article is, therefore, to review these experimental results to provide a broad and current picture of infant vowel perception.

This picture is organized into 4 key axes, bearing respectively on methodological considerations; the format of representations prior to, or independent from, language exposure; changes (or lack thereof) with development; and a comparison of “standard” (monolingual, typically developing) infants with other populations. Within each axis, we group studies as they relate to key theoretical and empirical questions. Each of these subsections begins with one paragraph stating the general question, and the overall answer that has been achieved. This article can be read from beginning to end, to gain a complete overview of the field, or it can be used as a reference piece, to consult the literature on one specific topic.

2. Methodological considerations

2.1 Characteristics of vowels

Vowels can differ phonemically in quality or quantity (length). Vowel quality differences are typically represented by their articulatory or acoustic characteristics. The articulatory features can be specified along the following dimensions: Backness represents the horizontal tongue position relative to the back of the mouth; height represents the vertical position relative to the roof of the mouth; roundedness encodes lip rounding; and nasalization indicates whether air flows through the nose during vowel articulation. Tenseness has been difficult to define articulatorily, but it is generally associated with fronting and raising.

As to acoustic characteristics, vowel contrasts are captured through the first three formants (commonly...
referred to as F1, F2, F3), in combination with their duration (Hillenbrand et al. 1995). Changes in formant values reflect different positions of lips, tongue and jaw and are thus linked to articulation. In particular, F1 is associated with height, F2 with backness, and F3 with roundedness.

2.2 Methods for assessing infant speech sound perception

For reasons of space, we cannot provide a full methodological review. Nonetheless, it suffices for our goals to state that a variety of behavioral methods have been used, including High-amplitude sucking (HAS; for details, see Jusczyk 1985); Central Fixation (CF; Werker et al. 1998); Headturn Preference Paradigm (HPP; Kemler Nelson et al. 1995); Conditioned Head-Turn (CHT; Werker et al. 1997); Anticipatory Eye-Movement (AEM; Albareda-Castellot et al. 2010); and, more rarely, infant body movement (Weir and Lamb 1990). Psychophysiological methods used include heart rate (Clarkson and Berg 1983), electroencephalography (EEG; Cheour et al. 2000), and magnetic encephalography (MEG; Kujala et al. 2004). Neuroimaging methods are so far limited to Near-infrared spectroscopy (NIRS; Minagawa-Kawai et al. 2008). These methods have been combined with a variety of paradigms, including simply spontaneous responses to specific stimuli; responses to novelty (after familiarization for a fixed duration or habituation; or the detection of an infrequent sound over the background of more frequent ones); and trained change detection or classification.

The possibility that different methods and paradigms can lead to different results is commonly recognized, because some require an overt response and attention to the task but not others, and some rely on instinctive responses whereas others rely on trained ones. To provide an integrated picture across all findings, we do not group studies by method, but instead simply indicate key methodological choices. Readers can download a spreadsheet containing all methodological details (including infant age and language) from sites.google.com/site/InPhonDB.

2.3 How do experimental manipulations affect discrimination?

Not only do methods differ between studies, but even within one method, stimuli can be presented under different conditions, for instance with different interstimulus intervals (ISI). Several studies have explored the impact of such differences by combining the study of vowel characteristics with temporal aspects of stimulus presentation. Briefly, longer ISI and longer stimuli promote better discrimination.

Clarkson and Berg (1983) presented newborns with changes in vowel quality between [a] and [i] tokens that were either separated by 0 or 500 ms of silence, and measured their cardiac response. Newborns only reacted to changes in the latter case, suggesting that silences were necessary to avoid loading infants’ short term memory. This result was replicated by Byrne et al. (1994) with 3- and 6-month-old infants and using variants of the diphthong [ai]. The role of ISI in the more specific context of informational masking was further explored by Cowan et al. (1982), who documented that infants reacted to a vowel change in a backward-masking condition (e.g., a change from a vowel pair like [a, a] to a different one like [e, a]) if the ISI was 300 ms, but not if it was 150 ms. In consonance with adults’ results, infants succeeded in a forward-masking condition (e.g., a change from [a, a] to [a, e]) at both ISI.

ISI is also acutely relevant when measuring change detection using EEG, at least in newborns. In Leppänen et al. (1999), larger responses to both a repeated [ka] and an infrequent [ka] were registered with 855 ms compared to 425 ms ISI.

Finally, the comparison of two studies assessing 2-month-old infants suggests that sheer stimulus duration may affect discriminability (Swoboda et al. 1976, 1978). Infants were able to notice within-category contrasts when vowels were 250 ms long, but not when they were 60 ms long.

2.4 What acoustical and multimodal cues affect infant vowel representation?

Vowels in natural speech mostly occur within strings of speech sounds, and are connected with visual talker information. This section shows that infants track and use multiple sources of information, both in the auditory signal and in other modalities.

Bohn and Polka (2001) investigated which aspects of the acoustic signal in CVC syllables infants relied upon in order to distinguish vowel categories. German 6- to 12-month-old infants were tested on their discrimination ability of native vowels in a [dVt] context which was either unmodified or had various kinds of information removed. Infants were equally able to discriminate the original syllables in conditions which preserved both onset and offset formant transitions, and those where only the stable vocalic portion was present. In contrast, onset and offset formant transitions on their
own were insufficient for discrimination.

The possibility that infants represent vowels in a modality-independent way has captured researchers’ attention. American English-learning infants as young as 2 months tend to look longer at faces with open mouths while hearing [a], and faces with closed, spread lips while hearing [i] (e.g., Kuhl and Meltzoff 1984). In fact, in terms of mismatch responses, visual and auditory information cueing the large vowel quality contrast [a-i] may have a similar informational value (Bristow et al. 2008). Indeed, EEG revealed that brain responses to a vowel change occurring in a crossmodal condition (i.e., background trials cued through a silent video, test trials cued auditorily) did not differ from those occurring in a unimodal condition (both background and test trials were auditory). Finally, it has been reported that already by 3–5 months of age American English-learning infants tend to approach vowel qualities observed on an audiovideo in their own productions (Kuhl and Meltzoff 1996), suggesting that audiovisual vowel information can bias infants’ vocalizations.

3. The format of representation

3.1 Do vowel quality contrasts have a privileged status?

Vowels in infants’ natural input show variability on multiple dimensions. A set of studies demonstrate that infants attend to variation in pitch, talker, and vowel quality. However, their discrimination of vowel quality changes elicits the largest responses and is particularly robust in the face of variation in other dimensions, with syllables as the basic encoding units.

Trehub (1973) reported that 1–4-month-old learning infants in an HAS study discriminated the contrasts in [pa-pi], [ta-ti], [a-i] and [u-i], but not tonal square wave or siren wave contrasts. Similarly, newborns presented with either a change of vowel quality ([α-i]) or pitch (low-high) responded with increased motor movements to both, but responses to the former were stronger (Weir and Lamb 1990).

Sensitivity to vowel quality may organize perception, and thus resist competing variation. Most 6-month-old infants tested with CHT generalized from a trained male-spoken [a-i] contrast to tokens with different voice (female, child) and pitch (rising, falling) characteristics with little training (Kuhl 1979), and some could do so for the more difficult [a-ɔ] contrast (Kuhl 1983). Even younger infants prioritize vowel quality over other contrasts, as shown with 1- to 4-month-old infants who could discriminate [a-i] both in the presence and in the absence of pitch variation (but had difficulties discriminating between pitch contrasts in the presence of vowel variation, Kuhl and Miller 1982); and with 2-, 3-, and 6-month-old infants who responded more strongly to a stimulus where both voice and vowel category changed, than when only voice changed (Marean et al. 1992). The primacy of vocalic over pitch changes is even reflected in the ease with which change detection can be localized within the infant brain. In a unique MEG study, Kujala et al. (2004) could localize the neural source for the detection of an infrequent [i] over the background of repeated steady-pitched [a] for all of the 10 newborns tested, whereas the detection of an infrequent rising-pitched [a] could only be localized in 6 of them.

This is not to say that pitch variation is completely orthogonal to vowel discrimination. On the contrary, a CHT study varying pitch height and pitch contour parametrically suggests that discrimination improves when the pitch changes over the course of the vowel, but is not overly high (Trainor and Desjardins 2002). This appears to ensue from tradeoffs in perceptibility of the formants (which are blurred by high pitch) and increased attention (captured by the changing pitch contour).

Other studies assessed to what extent infants pay attention to the vowel separately from the consonantal context. Miller and Eimas (1979) habituated 2- to 4-month-olds with a pair of syllables and tested them on a change of vowel ([ba, da] versus [be, de]) or a recombination ([ba, de] versus [be, da]). Since the strengths of dishabituation did not differ significantly across conditions, the authors propose that infants may encode the stimuli holistically, a conclusion repeated in Jusczyk and Derrah (1987).

3.2 Can infants perceive within-category variation?

The format of representations has been a key interest in the field. This strand of research suggests that infants are sensitive to within-category variation in vowels, but their sensitivity might be organized in a non-linear way depending on prototypes and category boundaries, as follows.

Early reports had concluded that vowels were not perceived categorically, based on evidence that 2-month-old American English-learning infants in a HAS study were equally able to discriminate within- and between-category contrasts drawn from a continuum between [i] and [ɪ] (Swoboda et al. 1976). However, other work documented greater sensitivity for between-than within-category differences (although acoustic
distance is not always matched across the two types of contrasts). First, in a follow-up with shortened vowels, infants discriminated [i] from [ɪ], but not from the intermediate token (Swoboda et al. 1978). Similarly, EEG reveals that, over the background of a repeated [y], the detection of a categorically different [i] is significant whereas that of an ambiguous [iy] is not, both for infants tested shortly after preterm birth (Cheour-Luhtanen et al. 1996), fullterm birth (Cheour-Luhtanen et al. 1995), and at about 3 months of age (Cheour et al. 1997). Using NIRS, Minagawa-Kawai et al. (2007a) documented that Japanese infants’ brain responses to within- versus between-category vowel length contrasts indeed differed at certain points of development (cf. Section 4.2).

Another body of literature suggests that the sensitivity to within-category changes depends on the extent to which vowel exemplars are prototypical of a sound category, as shown by two groups of results. First, infants seem to respond more vigorously to sounds that are prototypical of a given vowel category in newborns, as documented for both native [i, u] and non-native [y, u] through HAS (Aldridge et al. 2001). Other work suggests that within-category sensitivity around prototypical vowel exemplars is reduced for native language categories in 6-month-olds (as in the CHT studies Grieser and Kuhl 1989; Kuhl 1991; Kuhl et al. 1992, cf. Section 4.1), or possibly by birth (in a HAS study by Moon et al. 2013) using the native [i] and the non-native [y].

Second, an EEG study with 6- to 12-month-old infants exposed to American English found greater and more synchronous brain responses to more extreme [i] exemplars (Zhang et al. 2011). Naturally, it is unclear to what extent these results are due to innate biases towards certain vowel regions, or to the effects of exposure.

3.3 Can infants initially discriminate all linguistic contrasts?

It is often assumed that infants, before language exposure shapes their perception, are able to hear all contrasts that are used in any language. Although more evidence is needed to systematically map out early discrimination abilities, it is clear that infants do not require experience to discriminate vowels, but they do not discriminate all contrasts and all contrast directions equally well.

Indeed, while infants sometimes succeed with even small non-native contrasts (e.g., [pa-pə] in 1- to 4-month-olds, Trehub 1976; [buk-bɪk] in 4.5-month-old Japanese infants using CF, Mazuka et al. in press), failures have been recorded even for large contrasts (Nittouer 2001 found that 35% of 6- to 14-month-old American English-learning infants failed to discriminate native [a-u] in CHT; Lacerda 1991 found no statistical evidence for discrimination in a group of Swedish 1- to 6-month-olds for [a-u], or the more difficult [ba-bæ], for which a marginal effect was found in Miller and Eimas 1979, also HAS; see also Mazuka et al. in press, for failures to discriminate [buk-bʊk] and [bɪk-bɪk]).

The Natural Referent Vowel (NRV) framework (Polka and Bohn 2011) has been proposed to capture a phenomenon called perceptual asymmetry: peripheral vowels (cf. Section 2.1) can act as anchors, rendering discrimination from these vowels towards central ones more difficult than the reverse. Numerous reports are consistent with this prediction, namely: [i-ɪ] in 2-month-olds (Swoboda et al., 1978); [dɛt-det], [dʌt-dyt] and [dʊt-dyt] in both Canadian English- and German-learning 6-8- and 10-12-month-olds (CHT, Polka and Werker 1994, Polka and Bohn 1996); and [i-e] in Spanish- and Catalan-learning 4- and 6-month-olds (Pons et al. 2012). A more complex pattern of results is documented in Polka and Bohn (2011), where Danish-learning 6- to 12-month-olds were tested on the non-native [dɛt-dɔt], or the native [dɛt-det] or [dɛt-dɛt]. For the non-native English contrast, both the younger and the older half of infants showed an asymmetry in the predicted direction. For the [det-det] contrast, only the younger half of the infants showed this asymmetry. Finally, for the [det-dot] contrast, again only the younger age-group showed an asymmetry; however, the asymmetry was in the direction opposite from the prediction. Asymmetries may disappear with age and/or experience (e.g., Polka and Werker 1994, Pons et al. 2012).

4. Changes with age: A developmental perspective

4.1 How does vowel quality perception develop?

Infant vowel perception changes during the first year of life. Language exposure affects both the internal organization of sound categories (cf. Section 2.4 and 3.3), and the discrimination between categories. It should be noted that, although a recent study challenges the assumption that newborns’ perception is not yet shaped by language (Moon et al. 2013; cf. Section 3.2), there is ample evidence that infants’ perception is shaped by language exposure during the first year of life beyond potential in utero modulations. Moreover, a meta-analysis places the key age at about 6 months (Tsuji and...
Cristia 2013). Additional findings suggest that maturation and experience could enhance certain sensitivities, and lead to differential neural processing.

The process by which exposure to a given language influences speech sound perception has been captured in several models of early speech perception, most prominently the Native Language Magnet model (NLM; Kuhl 1994, Kuhl et al. 2008), the Perceptual Assimilation Model (PAM; Best 1994), and the developmental framework for Processing Rich Information from Multidimensional Interactive Representations (PRIMIR; Werker and Curtin 2005). NLM and PRIMIR in particular assume experience-based perceptual reorganization in the first year, which leads to decreases in sensitivity to non-native contrasts, and increases in sensitivity to native contrasts. PAM, in contrast, does not bear on the process of attunement, but rather explains how non-native contrasts are processed by reference to native ones.

Early evidence for language-specific vowel perception relied on non-linearities in the detection of within-category changes, demonstrated by infants’ better ability to discriminate vowels in the direction from a non-prototypical to a prototypical native exemplar of [i] than vice versa (Grieser and Kuhl 1989, Kuhl 1991; cf. Section 3.2). In a subsequent seminal CHT study, Kuhl et al. (1992) suggested that this perceptual pattern was tied to native language perception. American English-learning 6-month-olds failed to detect vowel changes around the prototypical [i] in their language but were sensitive to the same acoustic distances centered around non-native [y], while Swedish infants tested with the same stimuli readily heard such changes around the non-native [i] and missed them around native [y].

Other studies have focused on infants’ between-category discrimination. Declines for non-native contrasts have been recorded repeatedly: Polka and Werker (1994) found it for Canadian English-learning infants’ performance on non-native [u-y] and [u-y] using both CF (between 4 and 6 months) and CHT (between 6–8 and 10–12 months); Mazuka et al. (in press) for Japanese infants’ discrimination of German [bük-byk] between 4.5 and 10 months of age; and Jansson-Verkasalo et al. (2010) for Finnish 6- versus 12-month-olds responding to the non-native [y-e], as detected with EEG. A comparable decline has been found for the discrimination of American English dialectal variants of [a] between 7 and 11 months of age (using CF; Phan and Houston 2008). Other work combines cross-linguistic and cross-sectional data. Indeed, while both Spanish- and Catalan-learning infants were able to discriminate the Catalan contrast [deði-duði] at 4 months of age, only the latter did so at 8 months (Bosch and Sebastián-Gallés 2003, tested with HPP).

A great deal of work documents maintenance for native contrasts, namely EEG-recorded responses to the native [i-y] contrast in Finnish newborns and 3-month-olds (Cheour et al. 1998); [a-i] in 2-, 3-, and 6-month-old American English learners (Marean et al. 1992); [doði-duði] and [deði-duði] in both Catalan- and Spanish-learning infants at 4 and 8 months of age (Sebastián-Gallés and Bosch 2009); [ø-e] in Finnish learners tested at 6 or 12 months (Jansson-Verkasalo et al. 2010); and [sak-sak] in Dutch 11- and 15-month-olds (Benders submitted). Interestingly, a training study showed equally good discrimination of the vowel contrast embedded in [tb-teb] after exposure to a monomodal or a bimodal distribution in 8-month-old Canadian English learners, suggesting that when a contrast is discriminated well in the first place, vowel perception is resilient to short-term distributional learning (Pons et al. 2008).

Another group of studies has actually found sensitivity can increase with age. In a rare longitudinal CHT study on vowel discrimination, Cardillo (2010) documented an improvement between 7 and 11 months in American English learners’ discrimination of Finnish [u-y]. Similarly, Mazuka et al. (in press) reported enhancement for the German-spoken [i-ɛ] contrast between 4.5 and 10 months. These improvements could be due simply to maturation, or they could be due to infants learning about their native categories, and later mapping these non-native sounds into two separate native categories, as predicted by PAM.

Finally, a recent trend has been to assess to what extent vowel contrasts lead to left-dominant activations, in the context of the hypothesis that asymmetric processing is a sign of the emergence of brain networks that have specialized for the ambient language. Minagawa-Kawai et al. (2011) provide a comprehensive review of results using NIRS. They conclude that vowel contrasts are processed by a largely bilateral network early on in development, which becomes increasingly left-lateralized with age. The precise age at which this occurs appears to vary across different contrasts. For example, stable left-dominance in temporal brain areas associated with auditory discrimination was evident as early as 7–8 months (but not yet at 3–4 months) in Japanese infants for native [i-u], but not the non-native [u-u] (Minagawa-Kawai et al. 2007b). In contrast, left-dominance was only evident from 11–12
months onwards but not yet between 7–10 months in Japanese infants presented with native [ita-itte] (Sato et al. 2003). Recent research by Arimitsu et al. (2011) suggests that left-dominance can already be present in newborns, albeit in more posterior brain regions and possibly reflecting auditory short-term memory activation.

To our knowledge, no systematic review has sifted through EEG evidence for lateralization. Certainly, spatial localization with EEG cannot be undertaken lightly, and even using source localization methods can be challenging for infant data because accurate localization requires good signal-to-noise ratios as well as precise knowledge of the physical properties of the system in which the signal is traveling, the infant head (e.g., Hämäläinen et al. 2011, Whittingstall et al. 2003). Moreover, until the emergence of high-density EEGs, hemispheric asymmetry descriptions were done at the sensor level, where source is particularly uncertain given that it depends on the precise characteristics of cortical folding, which changes greatly with age. Not surprisingly, reports of lateralization in such vowel discrimination work are sparse. Only Bristow et al. (2008) have reported localization results using dipole modeling, finding that vowel repetition resulted in greater reconstructed amplitudes in left temporal cortices, and smaller reconstructed amplitudes in right frontal cortices in a group of 2-month-old infants. Future work could exploit this strategy to shed further light on this question, and complement the poor temporal resolution of NIRS.

4.2 How does vowel quantity perception develop?

The perception of vowel quantity contrasts has been studied much less than that of quality, despite the fact that length plays a key phonemic or acoustical role in many languages. An overview of this literature reveals a somewhat mixed pattern of results in terms of infants’ sensitivity to this kind of contrast and in terms of changes with age/language exposure.

Indeed, some behavioral studies reported that at 4-6-month-old English-learning infants discriminate vowels differing only in length (100, 200 or 300 ms vowels in items such as [mad-maid], Eilers et al. 1984; and English learners tested with 88 ms versus 180 ms vowels in items like [teki-teki], even after hearing monomodal distributions of vowel length, Pons et al. 2006). In contrast, Sato et al. (2010) found little evidence of discrimination in Japanese 4-month-olds tested with a vowel length contrast between 100 and 200 ms in [mana] and [mama]. By the end of the first year, discrimination is certainly in place (English 11-month-olds in Eilers et al. 1984; Japanese 10-month-olds in both Sato et al. 2010 and Mugitani et al. 2009).

Interestingly, the picture is even more complex when ERP and NIRS evidence is taken into account. Friederici et al. (2002) reported that German 2-month-olds detected a change from a frequent short vowel (202 ms) to an infrequent long one (341 ms), whereas they did not in the opposite direction (notice that the timing, complexity, and distribution of responses varied depending on infants’ awake status). The same asymmetry was reported in Friederici et al. (2004), who were able to find some index of change detection in the short to long condition provided that infants were fully awake during testing. Friederici et al. (2008) also found a larger response for a long deviant in a sequence of short vowels in German 1-month-old infants (note that no reliable response was observed for male infants with high testosterone levels). Whereas none of the behavioral studies noted above reported significant asymmetries, one has been documented in Japanese toddlers of about 18 months of age, but in the opposite direction: they detected a change from long to short, but not short to long (Mugitani et al. 2009). Since no such asymmetry was found in a group of English-learning 18-month-olds tested with the same stimuli, Mugitani et al. (2009) conclude that the Japanese 18-month-olds’ behavior reflected language exposure.

The richest longitudinal sample to date comes from a cross-sectional NIRS study that suggests that the development of discrimination for vowel length is not linear. In Minagawa-Kawai et al. (2007a), the brain responses of Japanese infants in several age groups between 3 and 28 months of age to vowel length contrasts drawn from the same category (both classified as e.g. short by adult Japanese listeners), or from different categories (one short and the other long). While brain responses during change trials did not differ significantly in within-category versus between-category change blocks at 3–4 and 10–11 months, they did at 6–7, 13–14, and 25–27 months. Interestingly, left-lateralization was evident later than in vowel quality studies (see previous Section): only the last group exhibited significantly left-dominant responses.

Given the diversity in methods, future longitudinal work may be able to shed light on the contributions of age, experience, and experimental choices in evoking asymmetrical patterns of discrimination for vowel quantity. In addition, it is of interest to study how vowel quality and quantity changes interact, particularly in languages like German or Dutch, where vowels often
5. Special populations

5.1 Assessing language development in at-risk populations

Since vowel discrimination can be studied very early, it could potentially provide an index of language development of individual infants to inform clinical application, and it could be applied to group differences that highlight the effects of maturation and/or experience on vowel perception. A more extensive review of the predictive value of infant speech perception measures has been undertaken elsewhere (see Cristia, Scidl, Soderstrom, and Hagoort, in press). Briefly, in terms of vowel processing, group comparisons between at-risk and typically-developing infants appear more robust and interpretable than the prediction of individual variation among the latter population.

In an early study, Swoboda et al. (1976, 1978) investigated vowel discrimination in 2-month-olds with a history of perinatal complications (such as low Apgar scores, premature birth, or extremely low birthweight). Like control participants, these infants reacted to both within-category and between-category contrasts when instanti- ated in long (250 ms) vowels; however, they differed from control infants in the weakness of the between-category response when short (60 ms) vowels were employed. That early study also reported a host of other differences between groups that extended beyond the presence or absence of a response to a change, and pertained learning of the contingent sound presentation and the effect of ISI in the strength of the discrimination response.

A sizable EEG literature has extended our knowledge well beyond these initial results by comparing brain responses in infants at familial risk of language impairments and controls not in the amplitude of the positive deflection to the deviant, but in its longer latency (Friedrich et al. 2004). Making precise claims is, however, difficult, given that not all experimental manipulations allow such group differences to emerge (Leppänen et al. 1999, Pihko et al. 1999).

Another strand in this literature investigates prematurely born infants, who tend to experience language delays in childhood (see Bosch 2011 for an insightful discussion). Comparing fullterms and maturation-matched preterms, both poorer discrimination early on and poorer language-specific tuning have been reported. Figueras-Montiu and Bosch (2010) document, with HPP, poorer discrimination of native [doði-duði] at 4 months (with no differences across groups at 8 months). As for tuning, Jansson-Verkasalo et al. (2010)’s EEG work revealed no difference across fullterms and preterms in detection of native vowel changes measured at 6 and 12 months, nor on detection of a non-native contrast measured at 6 months. However, responses were greater for the non-native contrast in preterms than fullterms at 12 months. This was interpreted as poor neural commitment to the native language patterns, a conclusion strengthened in this study through a longitudinal investigation, which revealed smaller vocabularies at 24 months the larger the brain response to the non-native contrast at 12 months.

Results are more variable for the prediction of individual variation among typically developing infants. Whereas Tsao et al. (2004) report that poorer performance in Finnish [y-u] discrimination predicts slower language development, this result did not replicate in the same lab (Cardillo, 2010).

5.2 Vowel perception in bilinguals

A growing literature investigates the development of speech perception skills in bilinguals, who may follow a different developmental path than monolinguals by virtue of their different life experience (for a recent review, see Curtin et al. 2011). While Spanish-Catalan bilinguals at 4 months of age performed equally to monolingual peers, at 8 months they had difficulties in distinguishing the contrasts [deði-deði] and [doði-duði] (Bosch and Sebastián-Gallés 2003, Sebastián-Gallés and Bosch 2009). However, they succeeded if the contrast was acoustically more distinct ([deði-duði]). This finding suggests that an interaction of bilingual status and acoustic characteristics influences discrimination. Differential distribution of attention in monolinguals versus bilingual infants may also play some role in
explaining the difference in performance. The studies above were carried out with CF, a paradigm which relies on infants' recovery of attention. A newer method, AEM, calls on a different suite of cognitive skills, namely on infants' anticipatory eye movements to a previously learned association between vowel and side of screen; thus, this method operates like a forced-choice categorization. When tested with the latter method, 8-month-old bilinguals’ performance in the acoustically less distinct contrasts was not significantly worse compared to their monolingual peers (Albareda-Castellot et al. 2011).

On the basis of such statements, it would seem that electrophysiological methods that do not load on attention could shed clearer light on whether monolingual and bilingual infants differ in vowel discrimination. In the preliminary results from a large study, which unfortunately does not include any statistical analyses, Shafer et al. (2011) recorded EEG in Spanish-English bilinguals and English monolinguals, in age groups ranging from 3 to 36 months; here, we concentrate on their results from the first year (3, 6, and 12 months). Infants were presented with repeated [e], which were interspersed with infrequent [ɪ] near the middle and the end of the sequence. Overall, infrequent vowels elicited brain responses with a positive mismatch detection response, whose latency decreased with age in the monolingual group, whereas it did not seem to do so in the bilingual group. Additionally, the amplitude of this response did not differ considerably across the monolingual and bilingual group, except possibly in the 12-month-old group. Unlike other groups, female bilinguals tended to show responses with an inverse polarity, which is in fact the typical polarity in the adult literature.

In a follow-up with the same stimulation, Shafer et al. (2012) concentrated on a group of 6-month-old Spanish-English bilinguals and another of 6-month-old monolinguals to explore the hypothesis that this polarity difference indicated increased sensitivity to the vowel change in female bilinguals. Based on the idea that final syllables attract more attention than medial syllables, the authors hypothesized that separate analyses of medial and final deviants would shed light on the polarity of the mismatch response. In consonance with their predictions, there was a negative response to the deviant in final position, but a positive one in the medial position, with the strength of both responses being inversely correlated when individual data were inspected. It is unclear to what extent this explains the group and sex differences described in the 2011 article, since neither language group nor sex appeared to interact with the polarity shift dependent on position. Nonetheless, there was an interaction with gender within the final position, as females showed relatively more negative responses than males. No interaction with group was reported in this paper, suggesting that at 6 months, mismatch responses are not significantly different in monolingual and bilingual infants.

6. Conclusion

Pooling all available published studies, this review has revealed a rich and diverse literature on infant vowel discrimination. To date, a similar up-to-date review on the early perception of consonants is missing. Evidence from a variety of studies suggests that some aspects of vowels and consonants are not processed in the same way (e.g., Bonatti et al. 2004). Although these differences are not restricted to discrimination, in order to gain a comprehensive picture of early speech sound perception it will be important to assess in how far the conclusions from the present review would hold for vowels specifically, or for speech sounds in general.

Working towards this end, we have created the open-access, updatable online resource InPhonDB (sites.google.com/site/InPhonDB). It currently contains the majority of studies reviewed in this article, and it may be extended to studies on consonants. Currently, the spreadsheet-format resource is a useful complement to this qualitative review in providing quantitative information on relevant independent and dependent variables in each experiment.

Acknowledgements

We thank Minna Huotilainen, Bob McMurray, and Koki Ikeda for helpful discussion. Portions of this work have been done with the support of ANR-10-LABX-0087 and ANR-10-IDEX-0001-02.

References


Tsui, S. and A. Cristia (submitted) “Community-augmented meta-analyses: The example of infant vowel discrimination.”


(Received Dec. 1, 2013, Accepted Dec. 29, 2013)