Received: 5 April 2019 | Revised: 22 November 2019 | Accepted: 22 December 2019

DOI: 10.1111/lnc3.12369

ARTICLE

WILEY

An acoustic perspective on 45 years of infant speech perception. II. Vowels and suprasegmentals

Chandan R. Narayan

Speech and Psycholinguistics Lab, Department of Languages, Literatures, and Linguistics, York University

Correspondence
Chandan R. Narayan, Speech and Psycholinguistics Lab, Department of Languages, Literatures, and Linguistics, York University, Canada.
Email: chandann@yorku.ca

Abstract

In this two-part review we examine the major results from infant consonant (Part 1), vowel, and suprasegmental (Part 2) discrimination research over the past 45 years from an *acoustic perspective*—an exegesis of the developmental perception literature that appeals to both acoustic aspects of speech contrasts and historically relevant typological facts about sound systems of the world's languages. We argue that infants' speech discrimination abilities are best viewed through a lens that considers both synchronic and diachronic aspects of the particular speech contrast. The key to this approach is the notion that acoustic–perceptual *salience*, or the relative separation of speech categories along perceptually relevant acoustic dimensions and corresponding discrimination performance in adults, is reflected in both infant's perceptual performance and patterns observed in phonological typology and history. The present review highlights challenges offered by four decades of literature, identifies broad patterns in infant vowel perception according to the acoustic properties of speech contrasts, and offers linguistically motivated explanations and directions for future research into the nature of young infants' discrimination abilities.

1 | OVERVIEW

In Part 1 of this two-part review we outlined an *acoustic perspective* on the major findings in infant speech perception research over the past 45 years, with a specific focus on infants' discrimination of consonant contrasts. An acoustic perspective on the infant speech perception literature seeks explanations for the various patterns observed in development that are rooted in the acoustic characteristics and historical and typological reflexes of speech contrasts in the world's languages. Below we focus on infants' discrimination of vocalic and suprasegmental (tone and duration) contrasts.

2 | INFANTS' DISCRIMINATION OF VOWELS

Owing in part to their acoustic prominence and relative ease in synthesizing for experimental settings, infant discrimination of vowel quality contrasts makes up the overwhelming majority of infant speech perception literature. Critical for vowel perception is the relationship between the first two formants (F1 and F2) and to some extent fundamental frequency (e.g., Barreda & Nearey, 2012; Miller, 1953), though recent research suggests that higher-dimensional representations of the entire spectrum perform as well as formant-based models of vowel perception (Molis, 2005).

Vowel quality discrimination has been explored across the history of infant speech perception. Trehub's (1973) study used the high-amplitude sucking procedure[1] to examine naturally produced [a]-[i] and [i]-[u] contrasts (in both isolation and CV contexts), both of which 1- to 4-month old infants were able to discriminate. We expect this result from an acoustic perspective, as [a], [i], and [u] lie at the extremes of the F1 × F2 acoustic space. Further, nearly all languages contrast these three vowels (Maddieson, 1984), making these distinctions both typologically ubiquitous and acoustically salient. Soon after Trehub's (1973) study, Swoboda, Morse, and Leavitt (1976) examined 2-month-old infants' perception of a subtler contrast, the tense-lax high-front vowel distinction, [i]-[ɪ], again in a high-amplitude sucking task with synthesized stimuli with the same duration. The contrast is primarily characterized by [ɪ] having higher F1 and lower F2 than [i]. The infants in Swaboda et al.'s study discriminated the contrast in what the authors describe as both *within* (contrasts involving two instances of either [ɪ] or [i] differing minimally in F1 and F2) and *between* (contrasts involving two phonemes [ɪ] and [i]) phonetic categories, that is, equally spaced (in F1 × F2 space) stimuli were discriminated equally well. Small acoustic differences in the synthesized vowels resulted in discrimination, suggesting that the vowel perception in early infancy is continuous.

By 6 months, infants' perception of vowels is transformed, consistent with the theory of perceptual reorganization (Werker & Tees, 1984), reflecting ambient language vowel categories (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Kuhl et al. (1992) explain this warping as a *perceptual magnet effect* and develops the Native Language Magnet model, where prototypical vowels (identified by adults as being ideal representatives of a particular vowel category) from the native vowel space essentially attract surrounding acoustically similar vowels, rendering them perceptually indistinguishable from the prototype. In Kuhl et al.'s (1992) study, English- and Swedish-hearing infants were presented with synthesized prototypical [i] (found in both English and Swedish) and [y] (the rounded [i] found in Swedish) and equally spaced (in F1 × F2 space) variants in a conditioned head-turn task. As predicted by the magnet effect, English-hearing infants accepted acoustic variants of [i] as being nondifferent from the prototype [i], while discriminating variants of [y] more readily. The opposite was true of Swedish-hearing infants, where prototypical [y] attracted its variants more than [i]. While the perceptual magnet effect is itself an acoustically motivated theory, examining the results from a typological and a more detailed acoustic perspective complicates their interpretation. Kuhl et al. (1992) do not mention that the phonological inventories of the two languages are not the same, as both [y] and [i] are native to Swedish-hearing infants. If we probe their results with this in mind, we notice that, although acoustically less-good variants of both [y] and [i] are attracted to the prototypes for Swedish infants, [i] variants are not attracted as readily as [y] variants to their respective prototypes. That is, English-like (nonprototypical) [i] is perceived differently than [y] for Swedish-hearing infants with respect to amount of acoustic variability that is accepted before it is treated as a *different* speech sound. Kuhl et al.'s (1992) data suggest that the [y] prototype against which variants are assessed acts as a stronger *magnet* than does the English-like [i] prototype. The [i] category tolerates less variability than does the [y] category in their results. How might we inform the varying language-specific behavior with an acoustic perspective? We would suggest that the design of the stimulus set belies the acoustic complexity of the [i]-[y] contrast by rendering it as a two-dimensional problem. The effects of lip rounding are seen most directly in frequency bands at and above F2. With an effectively increased vocal tract length from lip rounding, F3 is expected to be dramatically lower compared to [i] (e.-g., Fant, 1964). Indeed more recent research suggests that Swedish [i] and [y] may differ by less than one Bark in F1 and F2, with F3 being the distinctive acoustic feature separating the categories (Kuronen, 2001). We suggest that the differing discrimination patterns for [i] and [y] for Swedish-hearing infants' may reflect the incomplete acoustics offered to infants in the experiment: the infants may have been sensitive to the absence of F3 lowering in the stimuli and considered stimuli with a low F2 as not different from [y].

One of the more significant developments in the infant vowel perception literature is the finding that many contrasts are discriminated in an asymmetrical fashion (see Polka & Bohn, 2003 for a review), that is, infants often discriminate contrasts when a particular vowel is used as a background stimulus and changed to another, but not the other way around. This was first noticed by Polka and Werker (1994), who found that there was an order-of-presentation effect in a conditioned head-turn task for English-hearing infants' discrimination of the German [u]-[y] and [ʊ]-[Y] (the phonetically lax counterpart to [u]-[y]). Infants (6–12 months old) performed better when the direction of change was from [Y] to [ʊ] (rather than [ʊ] to [Y]), and when the direction of change was from [y] to [u] (rather than [u] to [y]). English-speaking adults rated the German stimuli in terms of their similarity to English vowels. German vowels [y], [Y], [ʊ], and [u] were rated similar to English [u] and [ʊ], suggesting the German back vowels are perceived like English back vowels than front vowels. The authors took this result as evidence that the back vowel in each contrast pair ([u]-[y] and [ʊ]-[Y]) is prototypical. As discrimination occurred most readily when the background stimulus in the conditioned head-turn task was from the nonprototype category, Polka and Werker's (1994) interpreted their results as consistent with Kuhl's perceptual magnet model. This interpretation was soon abandoned in a follow-up study examining English and German infants' perception of [ɛ]-[æ] (English) and [u]-[y] (German) (Polka & Bohn, 1996). Consistent with Polka and Werker (1994), the [y] to [u] change was easier than [u] to [y] for both German- and English-hearing infants. Further, the [ɛ] to [æ] change was easier than the reverse presentation for all infants in the study. That the presentation asymmetry effect was evident in German *and* English (which does not have a contrastive /y/) infants suggested that there is a language-general explanation for the findings. Polka and Bohn (1996, 2003) and Bohn and Polka (2001)) noticed that in nearly all cases of vowel perception asymmetries in infancy (Best & Faber, 2000; Bohn & Polka, 2001; Desjardins & Trainor, 1998; Polka & Bohn, 1996; Polka & Werker, 1994; Swoboda et al., 1976) the weaker or more challenging discrimination has been when a vowel from the periphery of $F1 \times F2$ space is used as the background stimulus. This observation has led to Polka and Bohn formalizing the Natural Referent Vowel framework (Polka & Bohn, 2011). The Natural Referent Vowel framework suggests that the periphery of the vowel space is indeed privileged in perception, owing in part to the natural consequences of the auditory system's resolution of closely occurring formants. That is, vowels that occur on the periphery of the vowel space are often characterized by formants in close proximity to one another, allowing a mutual amplification of the acoustic energy in a narrower spectral region. This results in a perceptual integration or increased salience of the formant known as focalization (e.g., Vaissière, 2011). In order to capture the focal aspect of the privileged vowels (i.e., those showing the asymmetry when used as a background stimulus in discrimination tasks) Polka and Bohn (2011) recognize that spectral dimensions beyond F1 and F2 must be considered. Without considering F3 and F4, the peripheral status of vowels such as [y] cannot be assessed. Further, Polka and Bohn (2011) allude to the typological nature of vowel systems as reflecting this periphery. We agree that the peripheral status of vowels such as [i], [u], and [a] lends itself to being *natural* at a linguistic level. Confer the fact that languages with the most minimal inventories will contrast at least these three vowels (Maddieson, 1984). Coupled with research in the infant-directed speech literature that indicates that focal vowels are more extremely produced (in F1 × F2 space) by caregivers, we can understand them being part of our acoustic endowment (e.g., Kuhl et al., 1997).

Taken together the Natural Referent Vowel framework and Native Language Magnet model are consistent with an acoustic perspective on infant speech perception, yet some questions remain. The Natural Referent Vowel framework does not offer a mechanism underlying the unidirectionality of the asymmetry. That is, why is *peripherality* favored by the perceptual system? One possibility is that focal or peripheral vowels have a masking effect, or at the neural level there is an inhibition of response to nonfocal stimuli. In this sense, we might marry an idea central to the Native Language Magnet model with the Natural Referent Vowel framework, namely *neural commitment*. It may be the case that peripheral vowels are privileged as a result of the warping of the organization of the auditory mechanism around the frequencies where we find focal vowels. Another question following from these models is how peripherality is quantified. This becomes important when we consider vowels in opposition where there is no clear peripheral vowel in formant space. For example, in Polka and Werker's (1994) study, the German [ʊ]-[Y] contrast has no clear peripheral member. Polka and Werker's reporting of the acoustics of their stimuli suggest that [ʊ] is closer to the lower limit of F2 than [Y] is to the higher limit, making it more peripheral. But more comprehensive descriptions of the German vowel space suggest that it may not always be the case. In the Kiel Corpus of Spontaneous German, the peripherality of [ʊ] or [Y] (Figure 1) cannot be determined by simply appealing to formant space alone (Pätzold & Simpson, 1997). We suspect that central vowels in contrast pose a problem for the Natural Referent Vowel framework, but perhaps one that can be resolved if framework were to expand the notion of peripherality to include additional, higher band, formant dimensions (F3, F4) (Vaissière, 2011).[2]

3 | INFANTS' DISCRIMINATION OF SUPRASEGMENTAL CONTRASTS

3.1 | Tone: F0

Infants' perception of linguistically relevant fundamental frequency distinctions in monosyllables had not been investigated from a developmental perspective until relatively recently. Using

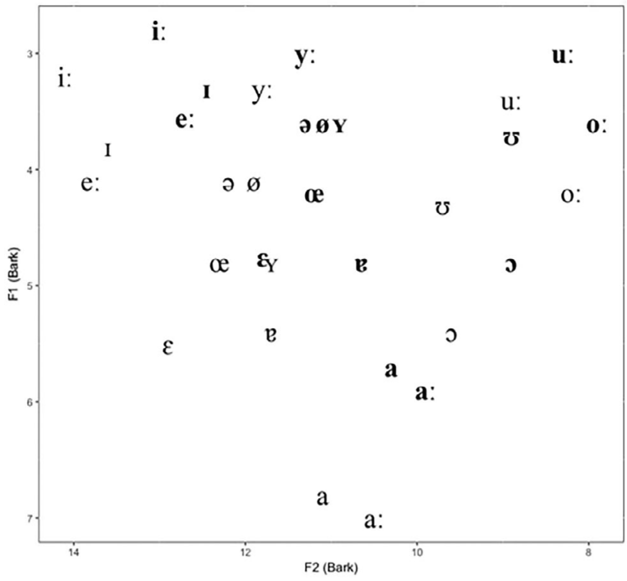

FIGURE 1 Vowel space for read German by female (plain) and male (bold) speakers. *Source:* Figure adapted from Pätzold and Simpson (1997)

a modified conditioned head-turn paradigm, Mattock and Burnham (2006) found that English- and French-hearing infants were able to discriminate Thai rising versus low tones at 4 and 6 months, but lost this sensitivity at 9 months. Mandarin-hearing infants were able to discriminate the contrast at both 6 and 9 months of age. These results were then replicated in English- and French-hearing infants using an alternating/nonalternating preference paradigm (Mattock, Molnar, Polka, & Burnham, 2008), where discrimination is indexed by looking time to a plain visual stimulus when background audio stimuli is either alternating (between a familiarization and test tone) or not. Taken together, Mattock et al.'s (2006, 2008) results indicated that infants' perception of tone reorganizes along a time-course similar to vowels, and earlier than consonants. Using an alternating/nonalternating preference procedure, Yeung, Chen, and Werker (2013) found a similar decline in discrimination performance from 4 to 9 months for English infants' discrimination of the Cantonese rising- (25) and mid-level (33) tones. With looking time to a plain visual stimulus as the dependent measure, infants in this task are familiarized with background stimuli from one category, then three types of test stimuli that are the same as the familiarization (nonalternating), different from the familiarization (nonalternating), or alternating. Chinese (both Cantonese- and Mandarin-hearing) infants in their study showed an asymmetry in looking time to test trials depending upon the tone of the familiarization stimulus (25 or 33). Only when infants were familiarized to tone 25 did they show discrimination. Further, the asymmetry was language-specific, with Mandarin-hearing infants showing a familiarity preference nonalternating tone 25 test stimuli when familiarization stimuli were from the same category. The authors suggest that, while tone 25 is non-native for Mandarin speakers, it is readily identified with the Mandarin rising tone by Mandarin-speaking adults (Francis, Ciocca, Ma, & Fenn, 2008). Cantonese-hearing infants showed no such preference for nonalternating test stimuli. These studies suggest that language-specific perception of tones can occur as early as 4 months. They also argue for a similar decline in the discrimination of lexical tones between 6 and 9 months among infants learning a nontone language as older English-hearing infants no longer showed discrimination. Contra the perceptual decline findings of Mattock et al. (2008) and Yeung et al. (2013), Liu and Kager (2014) found that Dutch-hearing infants succeeded at discriminating the Mandarin high-level (T1) and high-falling (T4) tone contrast at 5–6, 8–9, 11–12, 14–15, and 17–18 months, with no decline in sensitivity using a visual habituation method. The T1 and T4 contours in their study began at roughly the same F0 value, with T4 dropping by around 60% over the course of the syllable. In the same study, Liu and Kager tested the hypothesis that the maintained discriminability of the contrast was related to the acoustic salience of the T1/T4 contrast by decreasing the acoustic distance between the tones. A second experiment presented Dutch-hearing infants with the same T1/T4 contrast with a reduced F0 difference between stimuli. Infants presented with weaker acoustic salience version of the T1/T4 contrast showed discrimination at 5–6 and 17–18 months, but not intervening periods of development. The authors interpret these results as indicating that acoustic salience mediates the strength of (non-native) tonal contrast discrimination, with weaker salience resulting in less robust discrimination after 6 months. The authors suggest that the rebound in infants' perception at 17–18 months may reflect an accumulated knowledge and sensitivity to fundamental frequency contours in the form of intonation in Dutch. Importantly, infants' recovery of sensitivity to a non-native tone contrast weak in acoustic salience suggests that the window of tone acquisition is not closed until the second year.

From an acoustic perspective, the extant literature on tone perception in infancy is interesting for the types of tonal contrasts that have *not* been tested in the laboratory setting. For the most part, the demonstrations of non-native discrimination ability early in infancy and subsequent perceptual decline have focused on contrasts that are decidedly robust (Burnham, 1986) in acoustic salience (cf., rising or falling vs. level). Typologically, such contrasts are likely common in the world's tone languages. We hypothesize that the general finding of infants' early sensitivity and later perceptual decline would be complicated by the acoustic salience of tone stimuli with similar fundamental frequency contours. Indeed, the results of Liu and Kager (2014) predict this as such, but a more direct test is possible. Consider that it is perhaps rarer for a language to contrast contour tones in the same direction (two or more rising or falling tones) than contour tones varying in slope of opposite direction (rising vs. falling vs. level). For example, a survey of over 1,100 Chinese topolects (Wurm, Li, & Baumann, 1987) indicates that all Chinese languages exhibit at least one rising tone. Only 25% of these languages, however, contrast two rising tones (most commonly mid-rising vs. high-rising). Cantonese, which is rare in that it contrasts six lexical tones differing in relative fundamental frequency and change over time, provides an interesting potential test of for the salience hypothesis about tone. Perceptually, rising tones are more often confused with each other than contour versus level tones. Khouw and Ciocca (2007) examined adult Cantonese listeners' perception of Cantonese lexical tones produced by 10 talkers. Tones 25 and 23 (both mid-rising), with 25 having a higher terminal fundamental frequency than 23, were misidentified as the other rising tone at a rate of 22%. On the other hand, level tones (55, 33, 22) were confused with contour tones at a rate of 8%. Mok, Zuo, and Wong (2013) report that rising tone pairs (25 and 23) and level tone pairs (33 and 22) are indeed merging in both perception and production in Hong Kong Cantonese. Such language-internal effects of tone contrast salience suggest that tone inventories are mediated by tone acoustics and that infants may likewise show variable effects in their perception.

3.2 | Vowel duration

Duration is a phonetic feature that can be used contrastively in both vowels and consonants. Vocalic duration contrasts are implemented by extending the period of vocalization relative to a shorter variant. Though duration is intrinsically linked to vowel quality (e.g., tense vowels are longer than lax vowels) (Peterson & Lehiste, 1960), and quality differences often accompanies duration contrasts (like in Dutch, Booij & Booij, 1995), it is more or less used without quality distinctions in languages like Finnish, Estonian, and Japanese. While the UPSID database (of 451 languages) suggests 11% of the world's languages exhibit vowel length distinctions, a more comprehensive corpus, the World Phonotactics Database (of over 3,500 languages),[3] suggests that roughly one third of languages contrast length. Languages that contrast long versus short vowels generally have a 2:1 duration ratio (Lehiste, 1970). Perceptually, difference limens have been found to vary. In general, the longer the reference duration, the longer the difference limen required for detection (Henry, 1948; Ruhm, Mencke, Milburn, Cooper, & Rose, 1966; Stott, 1935). Further, signal amplitude, but not frequency, is related to the perception of duration differences. Perceived duration differences for louder sounds are smaller than those obtained for more quiet sounds (Ruhm et al., 1966).

The earliest exploration of infants' perception of vocalic duration differences was Eilers, Oller, and Benito-Garcia's (1984) study of infants' discrimination of final-syllable vowel length in preconsonantal position ([(CV)mad] vs. [(CV)maːd]). Using a modified conditioned head-turn task and resynthesized stimuli, the authors found that 5–11-month-old English-hearing infants marginally discriminated vowel duration contrasts differing by at least 100 ms. As the duration difference increased (from 100 ms to 300 ms), discrimination performance improved, but in all cases, infants' discrimination was considerably poorer than English-speaking adults' perception of the same contrast. The authors conclude that infants require at least a 1:3 duration ratio to be perceptible (greater than the average duration differences found in languages that contrast vowel length). Using a 100 ms versus 200 ms duration difference in a visual habituation paradigm, Mugitani et al. (2009) found U-shaped developmental results in their examination of Japanese vowel length. While English-hearing 18-month-olds discriminated naturally produced [taku] versus [taːku], Japanese-hearing 18-month-olds exhibited an asymmetry in perception, discriminating the contrast only when habituated to the long vowel. Japanese-hearing 10-month-olds and Japanese-speaking adults, however, showed no asymmetry in their discrimination of the contrast. The authors suggest that the distribution of vowel length in spoken Japanese privileges short vowels, allowing them to serve as perceptual anchors, resulting in long vowels being treated as longer examples of short vowels. Infants habituated to short vowels do not treat longer vowels differently, but when habituated to a long vowel a shorter vowel is distinctive. The authors suggest that at 18-month-olds vowel length is becoming phonologized in Japanese infants, with younger infants exhibiting discrimination based on lower-level acoustic properties of the duration difference. They argue that adults' symmetric discrimination of the duration contrast results from their fully phonemic perception of the length contrast. Sato, Sogabe, and Mazuka (2010) examined even younger Japanese-hearing infants' perception of the Japanese vowel length contrast and found evidence for acquisition of the contrast after 9.5 months. Using the visual habituation paradigm, 4- and 7.5-month-old did not show evidence of discrimination of a contrast where short vowel tokens were less than 50% in duration of long vowel tokens. The authors suggest that infants' perception of vowel duration is enhanced over the course of their first year (much like as suggested with nasal place perception) such that younger infants are less sensitive to durational differences than are infants toward 10 months.

The varied results of infants' perception of vowel duration can be interpreted from an acoustic perspective. Eilers et al.'s (1984) study suggests that difference thresholds (67% increase) for infants are different from adult listeners. Difference ratios in both Mugitani et al. (2009) and

Sato et al. (2010) are 1:2 which perhaps contributes to the differing patterns of results. It is also telling that the 1:2 ratio, though perhaps an average in languages that contrast vowel length, is smaller than the reported difference in normal spoken Japanese (Hirata, 2004). Infants' inconsistent perceptual patterning with respect to vocalic duration may reflect immature perceptual tuning to these windows, which roughly correspond to durational differences in languages with contrastive vowel length.

That infants, whose ambient language exhibits phonemic vowel length, nonetheless show difficulty in discriminating vocalic duration contrasts early in infancy is telling from an acoustic perspective. We suggest that the variety of linguistically relevant vowel length contrasts in the world's languages are more nuanced than those demonstrated in the infant speech perception literature, and that when this variety is addressed, vocalic duration contrasts will emerge as a challenge for universalist theories suggesting that infants are born with ability to discriminate all the phonetic contrasts exhibited in the world's sound systems.

4 | DEVELOPMENT FROM AN ACOUSTIC PERSPECTIVE

Our review of the major segmental and suprasegmental findings in the infant speech perception literature makes obvious the notion that not all speech contrasts are equal from the infant's perspective. That is, it is difficult to characterize a general trajectory regarding changes in infants' discrimination ability over time without giving careful consideration to the acoustic nature of the stimuli in question, and by proxy, the ways in which the cline of acoustic salience for speech contrasts are resolved in the languages of the world. The most basic generalization the literature suggests is that infants' perception becomes honed, reflecting native-language acoustic-phonetic distinctions at around the first and into the second year. Infants' perceptual boundaries in their first 10–12 months are less regular, and reflective of the acoustic similarity and dissimilarity of the speech sounds in question. The speech contrasts for which infants show inconsistent or poor discrimination are those that are fragile in acoustic salience (e.g., Burnham, 1986; Liu & Kager, 2014; Narayan, Werker, & Beddor, 2010) and require greater experience with the language environment to perceptually segregate. In this respect, an acoustic perspective on infant speech perception is in keeping with other acoustic approaches (e.g., Kuhl's Native Language Magnet) in that a direct appeal is made to acoustic distinctiveness in early infancy. An infants' model of linguistically relevant (to her native language) acoustic features are either well aligned with acoustically separable categories, or require refinement with experience.

In Part 1 of our review we offered a general characterization of infant discrimination performance as reflecting the nature of the underlying acoustic characteristics of the contrasts. Transient distinctions (noisy bursts and rapid formant movement) lend themselves to being successfully discriminated in early infancy and later subject to the perceptual reorganization, while contrasts that are characterized by temporally longer and louder acoustic features are more problematic in terms of identifying a definitive developmental trajectory. For the most part, our review of vowel and suprasegmental discrimination conforms to this general tendency.

The various developmental trajectories taken by speech contrasts in an infants' early perception are eloquently characterized by Aslin and Pisoni (1980). Aslin and Pisoni describe the possible routes speech contrasts may take toward a mature perception. Each of their four theoretical approaches posit differing initial states and development of infant speech perception: (a) Perceptual Learning—learning accounts for discriminatory behavior, with no contrasts being represented at birth; (b) Attunement—infants are born with the ability to perceive certain basic speech sounds, while other sounds require experience with language to become perceptible; (c) Universal theory—infants are born with the ability to perceive all speech sounds in the world's languages and contrasts that are present in the native language persist while others become less perceptible; and (d) Maturational theory—suggests that perceptual ability unfolds according to a biologically determined schedule. While Aslin and Pisoni (1980) clearly state that no one theory will account for the development of all speech contrasts, there has been considerable effort to characterize these four possible accounts under a dominant *Universal Theory* (Nittrouer, 2001; but see Aslin, Werker, & Morgan, 2002). Our point in interpreting the infant speech perception data with an eye toward acoustic details is to highlight the role of acoustic salience (as a basic psychophysical phenomenon) as a mediating factor in determining the developmental trajectory for the perception of speech contrasts. Further, in our estimation, the directions in which speech contrasts change, are enhanced, or are neutralized in the course of histories of natural languages are suggestive of their varying positions along the cline of distinctiveness, and as such, inform our understanding of developmental patterns in speech perception.

5 | CONCLUSION

An acoustic perspective on the results of infant speech perception point up the import of acoustic–perceptual relationships and linguistic history to our understanding of why different speech contrasts exhibit varied patterns of discrimination in infancy. We believe that many instances in the infant speech perception literature suggest a role for acoustic salience and phonological typology in the interpretation of why infants exhibit variable discrimination of speech contrasts. In outlining an acoustic perspective, we hope that future research into young infants' discrimination abilities address issues of relative discriminability of speech contrasts by adults (via confusion studies or speech in noise tasks for example), context effects of contrast enhancement of individual phones (i.e., whether stimuli are more or less discriminable in varying contexts), and sound inventory and change patterns in language histories. In detailing these additional perceptual and linguistic phenomena together with an infants' perception of a particular speech contrast, we arrive at a comprehensive understanding of the dual influences of natural auditory–acoustic endowments and linguistic experience in shaping the directions of an infant's developing perceptual biases.

ACKNOWLEDGEMENT

This article is dedicated to my *gurvyau*, Patrice Speeter Beddor and Janet Werker.

ORCID

Chandan R. Narayan https://orcid.org/0000-0003-0782-3467

ENDNOTES

[1] Please see footnotes 5 and 7 in Part 1 (https://doi.org/10.1111/lnc3.12352) for an explanation of major infant speech discrimination methods.

[2] Thus far we have addressed an acoustic perspective on vowel perception in infants being raised in monolingual environments. The literature on bilingual vowel perception (and speech perception in general) is vast and deserves serious attention from the linguistics and phonetics communities. While we will not discuss the

bilingual literature in detail in the present review, we would like to acknowledge an interesting developmental profile found in bilingual infants' perception of vowel contrasts in Spanish and Catalan. In two studies using the familiarization–preference procedure, Bosch and Sebastián-Gallés (2003) and Sebastián-Gallés and Bosch (2009) describe a U-shaped development of vowel perception in infants raised in bilingual Catalan–Spanish environments. Eight-month-old infants from bilingual households exhibited difficulty discriminating vowels the tense/lax [e]-[ɛ] Catalan contrast (Bosch & Sebastián-Gallés, 2003) and the [o]-[u] contrast found in both Spanish and Catalan (Sebastián-Gallés & Bosch, 2009). In both studies, younger (4-month olds) and older (12-month olds) infants succeeded at discriminating the contrasts. The authors account for the poor discrimination in 8-month olds by offering an acoustic explanation. That is, both the mid-front tense/lax and the back vowel contrasts are closer together in F1 × F2 space than the reference contrasts ([e]-[u]), which were successfully discriminated across development.

[3] http://phonotactics.anu.edu.au

REFERENCES

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. *Child Phonology*, *2*, 67–96.

Aslin, R. N., Werker, J. F., & Morgan, J. L. (2002). Innate phonetic boundaries revisited (L). *The Journal of the Acoustical Society of America*, *112*(4), 1257–1260.

Barreda, S., & Nearey, T. M. (2012). The direct and indirect roles of fundamental frequency in vowel perception. *The Journal of the Acoustical Society of America*, *131*(1), 466–477.

Best, C. T., & Faber, A. (2000). Developmental increase in infants' discrimination of nonnative vowels that adults assimilate to a single native vowel. In *International Conference on Infant Studies, Brighton, UK* (pp. 16–19).

Bohn, O. S., & Polka, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *The Journal of the Acoustical Society of America*, *110*(1), 504–515.

Booij, G. E., & Booij, G. E. (1995). *The phonology of Dutch*. Oxford: Clarendon.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, *46*(2–3), 217–243.

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied PsychoLinguistics*, *7*(03), 207–239.

Desjardins, R. N., & Trainor, L. J. (1998). Fundamental frequency influences vowel discrimination in 6-month-old infants. *Canadian Acoustics*, *26*(3), 96–97.

Eilers, R. E., Oller, D. K., & Benito-Garcia, C. R. (1984). The acquisition of voicing contrasts in Spanish and English learning infants and children: A longitudinal study. *Journal of Child Language*, *11*(02), 313–336.

Fant, G. (1964). *Formants and cavities*. Proceedings of the Fifth International Congress of Phonetic Sciences, pp. 120–141. Münster.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268–294.

Henry, F. M. (1948). Discrimination of the duration of a sound. *Journal of Experimental Psychology*, *38*(6), 734–743.

Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, *32*(4), 565–589.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, *35*(1), 104–117.

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

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.

Kuronen, M. (2001). *Acoustic character of vowel pronunciation in Sweden-Swedish and Finland-Swedish* (Working papers-Lund University Department of Linguistics). 94-97.

Lehiste, I. (1970). *Suprasegmentals*. Oxford, England: Massachusetts Inst. of Technology Press.

Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, *133*(2), 385–394.

Maddieson, I. (1984). *Patterns of sounds*. New York: Cambridge University Press.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, *10*(3), 241–265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, *106*(3), 1367–1381.

Miller, R. L. (1953). Auditory tests with synthetic vowels. *The Journal of the Acoustical Society of America*, *25*(1), 114–121.

Mok, P. P., Zuo, D., & Wong, P. W. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change*, *25*(3), 341–370.

Molis, M. R. (2005). Evaluating models of vowel perception. *The Journal of the Acoustical Society of America*, *111* (2), 2433–2434.

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

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.

Nittrouer, S. (2001). Challenging the notion of innate phonetic boundaries. *The Journal of the Acoustical Society of America*, *110*(3), 1598–1605.

Pätzold, M., & Simpson, A. P. (1997). Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. *Arbeitsberichte des Instituts für Phonetik und digitale Spachverarbeitung Universität Kiel*, *32*, 215–247.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, *32*(6), 693–703.

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

Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231.

Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467–478.

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.

Ruhm, H. B., Mencke, E. O., Milburn, B., Cooper, W. A., & Rose, D. E. (1966). Differential sensitivity to duration of acoustic signals. *Journal of Speech, Language, and Hearing Research*, *9*(3), 371–384.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology*, *46*(1), 106–119.

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: Is the distributional account all there is to it? *Developmental Science*, *12*(6), 874–887.

Stott, L. H. (1935). Time-order errors in the discrimination of short tonal durations. *Journal of Experimental Psychology*, *18*(6), 741–766.

Swoboda, P. J., Morse, P. A., & Leavitt, L. A. (1976). Continuous vowel discrimination in normal and at risk infants. *Child Development*, *47*, 459–465.

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, *9*(1), 91–96.

Vaissière, J. (2011). On the acoustic and perceptual characterization of rewference vowels in a cross-language perspective. In *The 17th International Congress of Phonetic Sciences (ICPhS XVII)* (pp. 52-59).

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.

Wurm, S. A., Li, R., & Baumann, T. (1987). *Language atlas of China*. Far East: Australian Acad. of the Humanities; Longman Group.

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.

AUTHOR BIOGRAPHY

Chandan Narayan is Associate Professor of Linguistics at York University in Toronto. He is interested in the intersections of acoustic phonetics, speech perception, historical phonology, and language acquisition.

How to cite this article: Narayan CR. An acoustic perspective on 45 years of infant speech perception. II. Vowels and suprasegmentals. *Lang Linguist Compass*. 2020;e12369. https://doi.org/10.1111/lnc3.12369