An acoustic perspective on 45 years of infant speech perception, Part 1: Consonants[1]

Chandan R. Narayan

York University, Toronto

Published in *Languages and Linguistics Compass* (2019)

https://doi.org/10.1111/lnc3.12352

[1] This paper is dedicated to my *gurvyau*, Patrice Speeter Beddor and Janet Werker.

Abstract

In this two-part review we examine major results from infant consonant (Part 1), vowel, and suprasegmental (Part 2) discrimination research over the past forty-five years from an *acoustic perspective*—an exegesis of the developmental speech perception literature that appeals to both acoustic aspects of speech contrasts and historically relevant typological facts about the 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 review highlights challenges presented by four decades of literature, identifies broad patterns in infant consonant 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.

Keywords: Infant speech perception, acoustics, phonetics, phonological typology, historical phonology

1. INTRODUCTION

In 1971, Eimas and his colleagues showed that very young infants discriminate English-like voicing categories, catalyzing an explosion of infant speech perception research in the decades that followed. Subsequent to their pioneering study, a large body of the infant speech perception literature has been devoted to understanding whether phonetically diverse speech contrasts are discriminated by infants and how (if at all) their ability changes over the course of development. This research has led the fields of language acquisition and developmental psychology to converge upon theoretical norms that have, up until recently, been *de rigeur* for the discipline. An important conceptual thread that runs throughout much of the literature is the notion that young infants discriminate speech contrasts in a language-general fashion (discriminating contrasts both present and absent in their ambient language) and that this ability changes dramatically over the course of their first year, reflecting the phonology of their soon-to-be native language. This is perhaps best exemplified by adults' poor perception of certain non-native contrasts which young infants' from the same ambient language environment successfully discriminate (e.g., Werker and Tees, 1984). As infants' perception of increasingly diverse phonetic contrasts was tested however, the developmental paths of changing discrimination abilities (cf. Aslin and Pisoni, 1980) were likewise shown to be more diverse (e.g., Polka and Sundara, 2001; Narayan et al., 2010). That is, the literature now reflects the fact that 1) not all linguistically significant phonetic contrasts are discriminated equally well by infant speech perceivers and 2) that the acoustic-phonetic nature of the speech contrast, apart from its presence or absence in the ambient speech, affects discrimination patterns in infancy. While developmental psychologists are beginning to assemble a comprehensive and increasingly complex understanding of infants' speech perception abilities in their first year, important, necessarily linguistic questions remain unanswered: Why are infants more successful at discriminating some types of naturally occurring phonetic contrasts[2] over others regardless of whether or not the contrast exists in their ambient-language phonology? What are the acoustic and psychoacoustic factors that underlie differential discrimination performance? Are patterns of infant speech perception related to larger phonological patterns (both synchronic and historical) observed within languages? Some early research (such as the work of Rebecca Eilers in the late 1970s) and especially work in the past twenty years has provided us with descriptions of a variety of perceptual patterns that deviate from canonical assumptions made in the early infant speech perception literature. As such, we feel that the infant speech discrimination literature can be reexamined from a point of view that lends cohesion to our understanding of these questions in light of these varied patterns. In this two-part review we offer an *acoustic perspective* on infant speech perception. This approach to the literature grounds explanations for infants' varying perceptual behavior in the acoustics of speech sounds, and integrates insights from both the linguistic and psychological literatures on perception, phonetics, phonological typology and history. We suggest phonologies take shape (in part) as a result of the intrinsic acoustic-perceptual properties of speech contrasts and are mirrored in infants' early discrimination abilities.

1.2 An acoustic perspective: Background and Proposal

Three milestones in the infant speech perception literature have guided the intellectual trajectory of subsequent research for a number of years. Eimas et al. (1971) showed that 1- to 4-

[2] We differentiate "naturally occurring" from artificial contrasts as the latter is often used in the developmental speech perception literature to uncover certain types of acoustic dimensions which infants may be sensitive to. The focus of our review is to elucidate infants' perception of phonetic contrasts that naturally occur in the world's languages.

month old English-hearing infants discriminate English-like VOT contrasts. Soon thereafter, Trehub (1976) showed that English-hearing one- to three-month old infants, but not adults, discriminate both a Czech fricative contrast and a French oral/nasal vowel contrast. Werker and Tees (1984) elegantly demonstrated that by the time infants are 10-12 months old they begin perceiving non-native consonant contrasts like adults, that is, initially discriminable contrasts are no longer discriminable by their first birthday. These three papers inspired a wealth of infant speech perception research from the 1980s through the 2000s, cataloging the types of phonetic contrasts that are discriminable by infants and how their perception changes over the course of development, with the general finding that non-native phonetic contrasts are discriminable by young infants, and this ability diminishes towards the end of the first year.

An acoustic perspective on the infant speech discrimination literature takes as its starting point a notion of *acoustic salience* as a gradient description of the relative discriminability of a speech contrast both within and between languages. That is, acoustic salience can be understood as a cline along which all phonetic contrasts lie, with endpoints representing either complete acoustic separation (in a perceptually relevant acoustic dimension) or acoustic overlap, predicting either poor or good discrimination, respectively. Our approach to understanding infant speech perception patterns relies on both raw acoustic details of speech contrasts as well as their perception by adults for whom the contrasts are phonologically significant. As detailed below, adults perceive certain phonological contrasts better than others (e.g., [ʃa]-[sa] is perceived better than [fa]-[θa] by English-speaking adults, and [ma]-[na] is perceived better than [na]-[ŋa] by Filipino-speaking adults). We propose that by appealing to the acoustic salience of speech contrasts, native-speaking adults' perception, and the linguistic history of speech contrasts across language families, the patterns observed in the infant speech perception literature, especially those that are inconsistent with the canonical language-general to language-specific development, can be best captured.[3] We hope to provide support for a recalibration of the infant speech perception research program such that future work can situate questions, and explanations of results within linguistically informed patterns observed in languages.

This two-part review (Part I: Consonants; Part II: Vowels and Suprasegmentals) highlights important findings (though not exhaustive) from the infant speech discrimination literature with an eye towards acoustic explanations for discriminatory patterns. The contrasts we highlight are defined along acoustic dimensions (e.g., voice-onset time, fast formant transitions, noise, etc.) and capture a wide variety of speech sound distinctions that are employed across the world's languages.

2. Infants' discrimination of consonants[4]

2.1 Oral obstruent voicing: Voice-onset time

The results of numerous studies, subsequent to Eimas et al. (1971), examining infants' perception of voicing distinctions in a variety of languages and using a host of methodologies, are far from uniform in their findings (Narayan, 2013). The primary acoustic correlate of voicing

[3] For the most part, the acoustic perspective on infant speech perception is a *post-hoc* analysis of developmental patterns. We might predict, however, that where adults exhibit poor discrimination performance on native consonant contrasts, so too will infants show patterns of perception differing from more typical perceptual reorganization (see §2.2).

[4] Our discussion of infant speech perception must come with certain caveats. The tendency in much of the literature is to describe infant behavior using the dichotomous outcome of successful or unsuccessful discrimination. The acoustic perspective on infant speech perception does not suggest a binary "perceptible/imperceptible" distinction, but rather a cline along which speech contrasts are more or less discriminable. The literature does not always lend itself to a graded interpretation of perceptual behavior, which is further complicated by the fact that researchers use many different methodological paradigms (see Kuhl, 1985), making the comparison of their results not necessarily straightforward. Our review generalizes broad discriminatory patterns from authors' own conclusions on their results. Finally, many studies, especially the early literature, suffer from being statistically underpowered due to small sample sizes and noisy measurements, thereby making false positives or false negative more likely (Bergmann et al., 2018).

in consonants is *voice-onset time* (VOT),[5] or the time elapsed between the release of an oral constriction and the onset of periodic oscillation of the vocal folds initiating the beginning of the post-consonantal vowel. Languages that exhibit the same number of voicing categories may show differing divisions of the VOT continuum. For example, while speakers of American English, a two-voicing-category language, produce voiced stops with a VOT between ~ -50ms and +25ms and voiceless stops between +25ms and 100ms, speakers of Dutch, another two-voicing-category language, produce voiced stops at around -100ms and voiceless stops around 0ms VOT (Lisker and Abramson, 1964). The VOT continuum is divided into broadly defined categories in the linguistics literature: *lag* VOTs fall on the positive side of the continuum, that is, voicing begins after the release of the oral constriction and *lead* VOTs fall on the negative side, with voicing beginning before the release of the constriction. Within the lag category, languages might differentiate *short-lag* sounds, with VOTs roughly between 0-50ms, and *long-lag* sounds with VOTs even higher.

The early infant speech perception literature was focused on pressing the nature of infants' inborn discriminative abilities against the influences of ambient language input. While the VOT continuum proved to be the perfect test case for exploring these various influences on speech perception, the highly variable results precluded a clear interpretation in the literature. Lasky, Streeter, Eimas, Eilers and colleagues all showed noteworthy infant perception patterns in lead/lag versus short/long-lag VOT contrasts. In the most general terms, this research showed an asymmetry in infants' perception of the two types of distinctions. In all studies where stimuli mimicked a short-lag vs. long-lag VOT distinction (such as what is found in North American English), infants succeeded in discriminating the contrast (Eimas et al., 1971; Lasky et al., 1975;

[5] The other acoustic correlates are the timing and frequency of *F*1 and the fundamental frequency immediately following the release of the stop.

Streeter, 1976; Eilers et al., 1979). Interestingly, infants whose native language background did not contrast short vs. long-lag also successfully discriminated the distinction. Gikuyu (Bantu, Kenya)-learning infants, for example, discriminated a +10/+40 ms VOT contrast (Streeter, 1976) and Spanish-learning infants discriminated a +20/+60 ms contrast (Lasky et al., 1975). On the other hand, the results of infants' perception of lead vs. short-lag VOT (as in the Dutch voicing contrast) is quite different. The majority of these studies suggested that infants' discrimination is quite poor (Eilers et al., 1979; Eimas et al, 1971; Lasky et al., 1975). Only two studies (Eimas, 1974; Streeter, 1976) showed infants' successful discrimination of the lead/short-lag contrast. Gikuyu-learning infants discriminated both a lead/simultaneous (-30/0 ms) VOT distinction as well as the short/long-lag distinction. It remains unclear, however, whether the lead VOT discrimination results from experience with Gikuyu or the psychophysical salience of the contrast, for English-learning infants do not show discrimination of a similar distinctions (Eimas et al., 1971; Eilers et al., 1979). These studies suggest that the lead/short-lag implementation of voicing is disadvantageous, relative to the short/long-lag contrast, from the infant's point of view. An acoustic perspective suggests that contrasts in the lag region of the VOT continuum are most likely privileged by the perceptual system for psychophysical reasons (Pisoni, 1977) as it provides more robust acoustic cues to a voicing contrast than does the lead VOT region. A critical acoustic correlate resulting from the momentary asynchrony of the release of the oral closure and onset of vocal fold oscillation is aspiration noise, the amplitude of which, relative to vocalic amplitude is an indication of voicing (Repp, 1979). The lack of aspiration in lead VOT, owing in part to a smaller glottal aperture relative to lag VOTs (Browman and Goldstein, 1986), is perceptually unfavorable.

The issue of the acoustic salience of the phonetic implementation of voicing becomes further complicated when we consider the multiple acoustic cues involved. Lag VOTs are characterized by the nature of the excitation of the first formant (F1) from the acoustic energy of laryngeal vibration. When voicing occurs much later than the release of an oral stop (as with long-lag VOTs), the resulting acoustic pattern is characterized by so-called F1 cutback (Liberman et al., 1958), where F1 is not excited until very late in the CV transition. Adult listeners are sensitive to F1 cutback with higher onset F1 frequency associated with voiceless stop perception (Stevens and Klatt, 1974; Kluender, 1991). Sensitivity to F1 cutback has been found even in listeners whose native language exhibits a lead/short-lag voicing contrast (Spanish), where F1 cutback is much less evident than in the short/long-lag contrast (Benkí, 2005). From an acoustic perspective, the inconsistent discrimination of lead/short-lag VOTs by infants and their more consistent discrimination of lag contrasts can be explained by the confluence of robust acoustic cues available when long-lag VOTs are contrasted.

Note that the perceptual advantage afforded to the lag VOT distinctions in infancy has an analogue in production, where mastery of lead VOT occurs relatively late compared to short-lag VOT in languages like Spanish (Eilers et al., 1984), French (Allen, 1985), and Thai (Gandour et al., 1986) (but see Whalen et al., 2007, for VOT in babbling). We would suggest that the lack of discrimination success shown by infants for the lead/short-lag distinction would be reflected in the types of VOT distinctions used in the world's languages. While the typological patterns remain unclear, owing to a lack a comprehensive cross-linguistic survey of voicing implementation along VOT, it is certain that all languages that utilize the VOT continuum have some sort of short-lag category (Keating et al., 1983) with which to contrast either lead or long-lag VOT.

2.2 Oral obstruent place of articulation: Formant transitions & burst structure

The literature is rife with examples of infants' discrimination of (egressive) oral-place contrasts in consonants (Eimas, 1974; Moffit, 1971; Morse, 1972; Jusczyk, 1977; Jusczyk and Thompson, 1978; Werker et al., 1981; Werker and Tees, 1984; Bertoncini et al., 1987). Articulatorily and acoustically, place-of-articulation contrasts in onset oral obstruents are best characterized by a momentary noisy broadband burst at release (reflecting the size of the oral cavity in front of the constriction) and a fast (~ 20-40ms) movement of F2 and F3 of the following vocalic segment. Such contrasts are universal in the world's languages. The earliest demonstration of infants' discrimination of oral obstruent place was by Moffit (1971) who, using a heart-rate dishabituation procedure, showed that 5-month-old English-hearing infants were able to discriminate between synthetic [ba] and [ɡa] stimuli. As the age of successful discrimination of place was pushed earlier and earlier in development (e.g., Morse, 1972 with 2-month-olds using the [ba]-[ɡa] contrast), the rhetoric of the literature began to focus on the likely innateness of this discrimination ability. For example, Eimas (1974) showed that English-hearing 2-3 month-old infants dishabituated (in a high-amplitude sucking procedure)[6] when a background stimulus changed from an adult-identified [dæ] to [ɡæ]. Synthetic CV syllables with onset F2 and F3 values comparable to [d] and [ɡ] were presented to infants. Very young infants discriminated stimuli that crossed a perceptual category boundary. The very young infants' discrimination of the discontinuity suggested that place-of-articulation perception is innate. The innateness hypothesis was challenged by research suggesting that non-human primates perceive place-of-

[6] High-Amplitude Sucking relies on infants becoming habituated to a background speech stimulus, indexed by the rate and amplitude of sucking on a pacifier. Once the sucking rate falls below a preset criterion rate, the infant is said to be habituated at which point the auditory stimulus is changed to a different category. Dishabituation, or an increase in sucking rate relative to the rate at habituation, indicates discrimination.

articulation (more specifically [b]-[ɡ]) in a way similar to adult humans (Sinnott et al., 1976; Kuhl and Padden, 1983).

At the most basic level, an acoustic perspective on infants' perception of the fast formant transitions associated with obstruent place contrasts suggests that the types of contrasts for which infants showed discrimination in these early studies belies the acoustic complexity of the variety of linguistically significant contrasts found in the world's languages. To highlight this tendency in the literature, notice that the types of syllable contrasts that researchers have investigated can be broadly classified as acoustically salient. Consider the fact that the early studies by Morse, Moffitt, and Eimas used stimuli tokens that were maximally different along perceptually relevant acoustic dimensions. For example, the contrast between velar and alveolar places are robust before low vowels, like those seen in the stimuli used in early demonstrations of infants' perception of place of articulation. The relatively low F2 of low vowels provides more of an F2 transition in velar contexts, which have a high F2 onset. Patterns of confusion among consonant places of articulation suggest that the [ba]-[ɡa] contrast is rarely misperceived by adults, even when minimal noise is added to the signal (Miller and Nicely, 1955; Phatak et al., 2008). Acoustically, the formant transitions for bilabial ([p b]) and velar ([k ɡ]) oral stops are maximally distinctive along the perceptually relevant dimension, the F2 transition. Suppose infants in these early demonstrations were presented with an acoustically less salient contrast, such as [ti]-[ki], we might expect reduced performance (or at least discrimination worse than for [ba]-[ɡa]). Reduced discriminability would be predicted from studies with older children showing their greater reliance on spectrally distinctive cues for discrimination than adults (Sussman, 2001; Mayo and Turk, 2005). The reasoning here is that F2 transitions between the velar stop ([k]) and the following front vowel ([i]) would resemble, acoustically, the transition between [t] and [i] (Chang et al., 2001).[7]

One of the most famous demonstrations of oral place-of-articulation discrimination of stops in infancy sought to uncover the time course of perceptual development from "language-general" to "language-specific" listening. By the late 1970s, psychologists had demonstrated that adult listeners often exhibit diminished perception of non-native phonetic contrasts (e.g., Miyawaki et al., 1975; Larkey et al., 1978). Coupled with research from infant speech perception literature (e.g., Trehub 1976; Werker et al., 1981) showing young infants' successful discrimination of non-native contrasts, Werker and Tees (1984) asked when infants begin listening to speech as native-language listeners. English-hearing infants in their study discriminated the Hindi (Indo-Aryan) dental-retroflex contrast ([t̪a]-[ʈa]) and the Nlaka'pamux (Salish) velar-uvular ejective contrast [k'i]-[q'i] using the Conditioned Head Turn paradigm at 6-8 and 8-10 months of age. But only the infants from Hindi- and Nlaka'pamux-speaking homes discriminated their respective native contrast at 10-12 months. English-speaking adults failed to discriminate both the Hindi and Nlaka'pamux contrasts, presumably because experience with the ambient language strengthens phonological contrasts while obscuring contrasts that not functional. Their choice of the Hindi [t̪a]-[ʈa] and Nlaka'pamux [k'i]-[q'i] contrasts is significant from an acoustic perspective. While the onset of F2 and F3 transitions certainly contribute to the distinction of retroflex stops, associated with a very low F3 and high F2 (Dave, 1977; Hamman, 2003), so too does the amplitude of the burst following the release of the stop (Ohala and Ohala, 2001). Perceptually, the distinction is relatively salient. Ahmed and Agarwal's (1969) confusion

[7] Historically, many languages (e.g., Old English, Italian) have undergone processes where [k], followed by a front vowel like [i] or [e] is fronted to the affricate [t͡ʃ]. In essence, [k] becomes more [t]-like when the articulation is pulled forward by the front vowel. This type of velar fronting is also common in early word productions.

matrix of Hindi consonants shows voiceless dentals and retroflexes being mistaken for each other on average 8% of the time. Similarly, when the contrast is at the end of the syllable, the confusion is again relatively low (~10%) in varying vowel contexts (Ohala and Ohala, 2001). Linguistically, the status of the dental-retroflex contrast in oral stops is stable, remaining unchanged in the histories of Dravidian (Krishnamurti, 2003) and Indo-Aryan languages (Masica, 1993) as well as Australian languages that exhibit the contrast (e.g., Nunggubuyu) (Hamilton, 1996).

The velar-uvular ejective (glottalic egressive) contrast used in Werker and Tees (1984) is advantageous from an acoustic perspective. The high-front vowel context [-i] allows maximal acoustic contrast of the dorsal gestures due to coarticulation. That is, the velar constriction in [k'i] is more front than in the context of a back vowel, allowing for higher peak spectral frequencies (Guion, 1997). We hypothesize that the resulting palatalized velar and high burst frequency likely increases the perceptual salience of the velar-uvular ejective contrast. The acoustics of the velar-uvular distinction must be considered when assessing the fact that English-hearing infants discriminate the Nlaka'pamux contrast. A test of this hypothesis would be to examine infants' perception of the same phonemic place contrast in a vocalic environment that did not enhance acoustic salience, such as [k'a]-[q'a].

2.3 Ingressive place of articulation

A welcome follow-up to the perceptual reorganization findings of Werker and Tees (1984) addressed a rare phonetic contrast and its perception in infants. Click sounds are created with two closures in the oral cavity (one velar and one more forward) and the lowering of the tongue body to create a pocket of low pressure. When the more forward constriction is released, air rushes in to the mouth to equalize pressure resulting in a relatively loud burst. Clicks are extremely rare, occurring in about 2% of the world's languages (Maddieson, 1984), mostly in southern and eastern Africa. Best et al. (1988) found that adult English-speakers' discrimination of click place is extremely good (98% correct for apical vs. palatal on the high end and 81% on the low end for apical vs. lateral). Best and colleagues followed up the adult finding with a series of infant studies. Recall that the typical motivation for perceptual reorganization in infant speech perception was adults' lack of perceptual sensitivity to non-native phonetic contrast (like English speakers' poor discrimination of Hindi [t̪a]-[ʈa]). That adult English speakers discriminated click places of articulation well above chance and near ceiling in some cases suggests that infants' non-native perception would be equally good. Using a Visual Habituation[8] task, Best et al. (1988) found English- and Zulu-hearing infants (and adults) successfully discriminated a Zulu alveolar-lateral click contrast [|a]-[||a] at 6, 8, 10, and 12 months of age.

Based on the typological rarity of clicks, an acoustic perspective would suggest that they are weak with respect to acoustic salience, though this is clearly not the case. Rather, clicks present a case whereby salience and ubiquity/rarity in sound systems do not overlap and can be attributable to a more basic source. Their rarity can perhaps be attributed to a historical accident from sequences of egressive consonants (Ladefoged, 1968; see Fuchs et al., 2007 for evidence from German word boundaries) with a resulting ingressive airstream mechanism with audibly distinct acoustics.[9] There is mounting evidence from the neuroscience literature that clicks are

[8] Much like High-Amplitude Sucking, Visual Habituation relies on the infant habituating to background audio stimuli, in this case indexed by looking time to a plain visual stimulus. When looking time falls below a preset criterion looking time, the infant is said to be habituated, at which point two types of test trials are presented: a "same" trial where the audio stimuli are from the same phonetic category as the habituating stimuli, or a "change" trial where the audio stimuli are from a phonetic category different from the habituating stimuli. If looking time to change trials are longer than same trials, infants are said to have discriminated the contrast.

[9] It is worth noting here that we might attribute the typological rarity of clicks to a notion of "articulatory complexity." While certain consonant manners of articulation appear, at the outset, to involve more or less coordination of articulators, it may not necessarily be the case that biomechanical effort increases with the more

processed by the brain in a way different form egressive speech sounds. Best and Avery (1999) report that native Zulu speakers, and not native English speakers, show a left hemisphere advantage in the processing of clicks in a dichotic listening experiment. More recently, Agnew et al. (2011) showed that the pattern of cortical response, in an fMRI task, is different for egressive versus ingressives sounds for English-speaking adults. Only with experience with the native click inventory do Zulu speakers treat otherwise non-speech sounds as linguistically relevant. Taken together, we view click contrasts as a special case of perceptual acuity driven by acoustic salience.

2.3 Fricative place of articulation: Broadband noise

Infants' perception of phonetically significant noise contrasts, or broad-band aperiodic acoustic energy found in fricative articulations like [s] or [f], has been contested since the 1970s. High-frequency noise results from air forced through a small aperture opening in the oral cavity. The constriction location (specifically the length of the oral cavity in front of it) determines where in the frequency spectrum the aperiodic energy is centered. For example, the center of the aperiodic energy created by [s] is higher in the frequency (~7KHz) spectrum than [ʃ] (~4KHz) (Jongman et al., 2000). For the most part, the perception of place of articulation in fricatives relies upon discriminating the difference between these concentrations of noise at various frequency bands. Consequently, the smaller this acoustic difference, the less salient and more confusable the contrast. Eilers et al. (1977), using a variant of the conditioned head turn paradigm, showed that the [f]-[θ] contrast proved difficult to discriminate for 6-8 month-old infants. Only at 12-14

---

articulators involved in producing a speech sound. That is to say, while articulatory complexity is a compelling concept for explaining certain typological patterns, it is simply not as tractable as acoustic (and corresponding perceptual) complexity in capturing *both* infant speech perception and typological patterns.

months did the majority of infants in their study show discrimination of [fi]-[θi], while failing to show discrimination of [fa]-[θa]. That same year in a conference report, Holmberg et al. (1977) showed 6 month-olds discriminated the [fa]-[θa] contrast but noted that infants required twice as many trials to achieve criterion (a proxy measurement of perceptual difficulty) in a conditioned head turn paradigm than they did for the [s]-[ʃ] contrast. Contrast these results with those of Levitt et al.'s (1988) study, where 2-month-olds discriminated the [fa]-[θa] contrast using the High-Amplitude Sucking paradigm. It is difficult to compare the results of Holmberg et al. (1977) with Leavitt et al. (1988) as neither explicitly describes the nature of the fricative noise used in their stimuli. What can be said, however, is that discrimination of the [f]-[θ] contrast in infancy is far from straightforward with results varying according to age, linguistic context, and experimental paradigm. An acoustic perspective on these varied results would point to the similarity in spectral peak frequency of frication for the phones in question. The length of the oral cavity in front of the labiodental constriction for [f] and interdental constriction for [θ] is very similar, resulting in similarly shaped noise. Jongman et al. (2000) showed that the spectral peak of the frication noise significantly characterized place of articulation in fricatives. Spectral peak locations were nearly identical for [f] and [θ] in their study (~7.7KHz), resulting in a misclassification rate of ~25% in a discriminant analysis (with other predictors combined with spectral peak). The linguistic import of the weak acoustic salience of the contrast is evident in synchronic mergers of [f] and [θ] in several varieties of English (e.g., Cockney, Newfoundland, African American, etc.).

Infants' perception of the voiceless alveolar-postalveolar sibilant distinction [s]-[ʃ] is likewise variable in the literature. While Holmberg et al.'s (1977) report indicates English-hearing infants' successful discrimination of the contrast at 6 months, Nittrouer (2001) showed more nuanced results. She tested infants between 6-14 months on vowel contrasts preceded by either [s] or [ʃ], the sibilant place-of-articulation contrasts, and a stop-voicing contrast. Six of the 15 infants in her study who could discriminate vowel quality (either [sa]-[su] or [ʃa]-[ʃu]) could also discriminate [sa]-[ʃa], while out of 8 infants who could distinguish a stop-voicing contrast ([ta]-[da]), none discriminated the sibilant place. Like other fricatives, the [s]-[ʃ] contrast is acoustically characterized by a host of features relating to frication noise (spectral and temporal). Unlike [f]-[θ], however, the varied results in infants' perception of [s]-[ʃ] does not have a clear acoustic source. Alveolars and postalveolars are generally well distinguished in terms of spectral peak frequency ([s] higher than [ʃ]) and overall amplitude but comparable in terms of frication duration (Jongman et al., 2000). Adult listeners' perception of the distinction is governed by both frication properties as well as transitions (Whalen, 1991). Consonant confusion studies similarly show rare misperceptions of alveolar and alveopalatal sibilants (e.g., Miller and Nicely, 1955 in [-a] context). Discriminant analysis using relevant acoustic parameters likewise shows low rates of sibilant misclassification (Jongman et al., 2000). Historically, /s/ and /ʃ/ appear to merge in different directions, even in neighboring languages. For example, Sanskrit /s/ > Bengali [ʃ], while Sanskrit /ʃ/ > Oriya [s]. Similarly, children's early productions do not show definitive patterns favoring one sibilant over the other (Smit et al., 1990). It is useful to point out that from an acoustic perspective, infants' variable discrimination of the [s]-[ʃ] contrast in a low vowel context (as in Holmberg et al., 1977 and Nittrouer, 2001) would likely be even more affected in front vowel contexts. For example, we would predict that the [si]-[ʃi] contrast is acoustically less salient than the same sibilant contrast preceding a low vowel as a result of palatalization, or the backing of the tongue blade of [s] in anticipation of the high front vowel. As a result, [si] is likely to be acoustically similar to [ʃi]. In an identification study, Lawrence and Byers (1969) found that hearing-impaired listeners, while identifying [s] and [ʃ] at an overall rate of 85%, most often confused the two when followed by front vowels. Recently, Li and Zhang (2017) showed that the Mandarin prevocalic dental-palatal sibilant contrast (/s/-/ɕ/, /ts/-/tɕ/, /ts^{h}/-/tɕ^{h}/) in a high-front vowel context ([i]) was less distinct (longer reaction times in an AX task) than in low and back vowel contexts for both English- and Mandarin Chinese speakers. They also showed that the historical consequence of the lack of acoustic salience between dental and palatal sibilants in high front vowel contexts was that many Chinese dialects do not exhibit the contrast when a third sibilant place (retroflex) was not present in the phonology. Indeed the [s]-[ʃ] contrast is famously neutralized before front vowels in languages like Japanese and Korean.

Cristià et al. (2011) examined English-hearing infants' perception of the Polish voiceless alveopalatal-retroflex sibilants ([ça], [şa]) within a paradigm designed to test the effect of exposure to various distributions of acoustic features. The Polish sibilant contrast is characterized by the centroid frequency of frication noise and the onset spectral characteristics of the post-sibilant vowel. Crista et al. (2011) presented 4-6-month-old English-hearing infants with distributions of modified real [şa] and [ça] tokens varying along continua of the two critical acoustic features. After initial exposure to either flat or bimodal distributions of the stimuli grid, infants demonstrated discrimination only for natural combinations of retroflex frication and vowel and not alveopalatal combinations. These results indicated that given exposure to a bidimensional distribution of fricative place features, learning is restricted to the retroflex place of articulation. From an acoustic perspective, that infants in Cristià et al.'s study treated alveopalatal sibilants across the stimuli continuum as roughly equivalent suggests that it is acoustically more variable and perceptually a more flexible category. Listeners in McGuire's (2007) study of adults' labeling of Polish [ça] and [şa] showed less consistency at the alveopalatal end. Indeed, the center of gravity for alveopalatal frication is more widely distributed (Zygis and Padgett, 2010) and has higher amplitude (Nowak, 2006) than the retroflex.

2.4 NASAL PLACE OF ARTICULATION: MURMUR TO FORMANT TRANSITIONS

Some place-of-articulation contrasts rely not only on static F2 or F3 frequencies, but a dynamic change in energy across the lower bands of the spectrum. Nasal consonants (like [m] and [n]), so-called because air escapes the nose while being obstructed in the oral cavity, are generally voiced, meaning the vocal folds are set into oscillation creating a low-frequency resonance in the coupled oral and nasal cavities (Stevens, 1998). Acoustically, place of articulation of nasals is correlated with a change in energy between the low-frequency resonance (or *murmur*) and onset of the following broadband vowel (Kurowski and Blumstein 1987; Narayan, 2008). Hillenbrand (1984) showed that English-hearing infants, at 5.5-6.5 months discriminate the bilabial ([ma])-alveolar ([na]) contrast across various vowel and speaker contexts in a conditioned head turn task. Narayan et al. (2010) followed with a study examining English- and Filipino-hearing infants' perception of [ma]-[na] and [na]-[ŋa] at 4-5, 6-8, and 10-12 months in a visual habituation (VH) paradigm. The velar nasal [ŋ] is found at the beginning of syllables in the languages of the Philippines. In an earlier study, Narayan (2008) showed that the Filipino [na]-[ŋa] contrast was less salient, than the [ma]-[na] contrast, both acoustically, with the dynamic energy change of [ŋ] resembling that of [n], and perceptually, with native Filipino-speaking adults confusing [na] and [ŋa] more than [ma] and [na]. Correspondingly, English-hearing infants showed difficulty discriminating the contrast at 4-5, 6-8, and 10-12 months, while successfully discriminating the [ma]-[na] contrast at all ages. Filipino-hearing infants in the study only discriminated the [na]-[ŋa] contrast at 10-12 months, suggesting that despite the presence of the contrast in the ambient language, its acoustic salience renders it perceptually weak to infants (Narayan et al., 2010). Only with enough native-language exposure to the phonetic contrast, and its subsequent linguistically contrastive usage, are infants able to redirect their attention to the otherwise weak acoustic salience. Typologically, the [n]-[ŋ] contrast is rarer than [m]-[n] in the world's languages (Narayan, 2008). Additionally, the contrast has merged to [n] in some Austronesian languages (e.g., Hawai'ian, Tahitian). Recently Sundara et al. (2018) have challenged the results of Narayan et al. (2010). Using exactly the same Filipino nasal tokens from Narayan et al. (2010), Sundara et al. showed that English- and French-hearing 6-month-olds discriminated the [na]-[ŋa] under a stricter habituation criterion, longer habituation trials, and shorter ISI than used in the original study. In Narayan et al.'s (2010) study, infants were considered habituated to background acoustic stimuli when their looking time to a plain visual stimulus on a particular trial decreased to 40% of the longest looking time on a previously presented trial. In Sundara et al.'s study, habituation was registered when looking time decreased by 50% of the longest looking time. As a result of these methodological changes, young infants were shown to discriminate the non-native Filipino nasal contrast. While it is not clear why the English- and French-hearing infants in Sundara et al.'s study, who were exposed to habituating stimuli for a longer time than the English- Filipino-hearing infants in Narayan et al.'s study, exhibited discrimination of [na]-[ŋa], we argue that their result is nonetheless consistent with an acoustic perspective. Relative to the typologically ubiquitous [na]-[ma] contrast, the [na]-[ŋa] contrast required that infants in Sundara et al.'s study have more exposure in order to be discriminated, suggesting that the [na]-[ŋa] contrast is less salient than the [na]-[ma] contrast.[10]

---

[10] A reviewer raises an important question regarding the effect of lab exposure to a speech contrast versus ambient language exposure. Although the Filipino-hearing infants in Narayan et al.'s study had accumulated more exposure to [na]-[ŋa] than the English-hearing infants at 6-8 months of age, they nonetheless did not exhibit discrimination of

This interpretation is bolstered by the successful discrimination of the [na]-[ma] contrast in Narayan et al.'s (2010) study under a less-strict habituation criterion (cf. Holmberg's, 1977 results with the [fa]-[θa] contrast).[11]

2.5 LIQUID PLACE OF ARTICULATION: F3

The North American English contrast between either a bunched or retroflex /ɹ/ and alveolar lateral /l/ is perhaps the most well-known example of the adults' difficulty discriminating a non-native consonantal contrast (Goto, 1971; Miyawaki et al., 1975). The general finding is that listeners (most often Japanese-speaking adults) exhibit difficulty discriminating American English [ɹa]-[la], while native listeners perform categorically (as with many other consonantal acoustic features). Eimas (1975) showed that 2-3 month-old infants discriminated synthetic syllable onset [ɹ l] stimuli differing in steady-state F3 in a High-Amplitude Sucking task. Tsushima et al. (1994) showed in a conference report that [ɹ]-[l] perception follows a developmental trajectory similar to other oral consonants. The authors tested Japanese-hearing infants at 6-8 and 10-12 months with a native [w]-[j] and non-native [ɹ]-[l] contrast in a VH paradigm. Their results showed infants discriminated both native and non-native contrasts at 6-8 months, but only the native contrast at 10-12 months. More recently, infants' perception of the [ɹ]-[l] contrast has been shown to be more nuanced than shown by Tsushima et al. (1994). Kuhl et al. (2006) show that English-hearing infants' perception of the contrast improves with age. In a conditioned head turn task, 6-8 month-old English- and Japanese-hearing infants discriminated

---

the native nasal contrast. One possible explanation would be that the 10-12 month period is a necessarily sensitive period in development when the effects of the accumulation of native language exposure becomes evident.

[11] Sundara et al. (2018) also showed English- and French-hearing infants' successful discrimination of a dental-retroflex nasal contrast ([n̪a]-[ɳa]). We are unaware of any language which contrasts dental with retroflex nasals in word-initial position (Hamann, 2003). Tamil, like all other Dravidian (and Indic), only exhibits the contrast intervocalically.

the synthetically generated contrast at a rate of 65%, well below native adult levels of discrimination (Miyawaki et al., 1975). By 10-12 months, however, English-hearing infants' perception of the contrast improved to approximately 75%. Kuhl et al.'s (2006) results also show a directional asymmetry, where infants' perception of the native contrast improves only when conditioned to discriminate a background [ɹ] stimulus to [l]. The authors offer an acoustic explanation for its basis. Kuhl et al. (2006) suggest that [l] in [a] context masks the higher formats of the [l] transition. That is, in pre-[a] position, higher formant (F2 and F3) transition information is lowered, similar to [ɹa]. The American-English /ɹ/ is articulatorily complex, with some speakers producing a bunched tongue and other a retroflex version with comparable acoustic effects (Zhou et al., 2008). Languages with similar articulatory configurations in the lateral/rhotic space show various paths to resolving their acoustic similarity. For example, speakers of Puerto Rican Spanish produce /r/ as an approximant and its contrast with /l/ is neutralized in coda position (Simonet et al., 2008). Neutralization of liquids is also seen in Dravidian. Proto-Dravidian exhibits five liquids, two laterals (dental and retroflex) and three rhotics (pre-, post-alveolar, and palatal) (Christdas, 1988). Acoustically, the Tamil palatal rhotic [ɻ] has F2/F3 characteristics similar to the retroflex lateral [ɭ] (Narayanan et al., 1999). While some Dravidian languages have retained the contrast (e.g., some Tamil caste dialects), [ɻ] has merged with [ɭ] in Kannada, and disappears all together in Toda and Kota (Krishnamurti, 2003). The fragility of the liquid/rhotic space, as demonstrated by asymmetries in their perception by infants, highlights the idea that infants' initial reception of liquid contrasts is guided by sensitivity to the acoustic similarity between phones differing in higher formant distinctions.

2.6 Cross-manner distinctions (stops-glides, stops-fricatives)

The literature provides us with a few examples of infants' perception of cross-manner distinctions. The [b]-[w] contrast, which is found in many of world's languages often with alternations between the labio-velar glide [w] and the voiced labio-dental fricative [v], is arguably weak in salience from an acoustic perspective. The critical acoustic characteristic for the [b]-[w] contrast is the spectro-temporal behavior of F2 and F3. When the slope of the F2 is steep (temporally short) adult listeners tend to hear stimuli as [b]. When the F2 transition is more gradual (reflecting the slower jaw movement from the initial [w] position to the following vowel), listeners report [w] (Miller and Liberman, 1979). The perceptual effect of formant transition duration is offset by the duration of the following vowel, leading to a complex normalization that listeners automatically process when discriminating the contrast. Early reports of this normalization in infants was taken as evidence for an innate linguistic endowment (Eimas and Miller, 1980). Oller et al. (1993) compared adult and infant discrimination performance using the same methodological paradigm with the [b]-[w] distinction. They found that while English-hearing adults discriminated the short-long F2 transition regardless of following vowel duration, infants failed to show discrimination. The weak acoustic salience of the contrast is further evident in historical patterns of genetically disparate languages where labial glides often changed to bilabial stops (e.g., Sanskrit > Bengali; Old Tamil > Kannada; Latin > Spanish) and vice-versa (e.g., Ancient Hebrew > Modern Hebrew; Italian > Neopolitan).

The cross-manner contrast, [d]-[ð], differs along both the place-of-articulation dimension (alveolar vs. interdental) as well as manner (stop vs. fricative). In English, the articulation of the interdental fricative seems to vary considerably with a stop-like realization (most likely with the tongue tip resting behind the upper front teeth rather than between the upper and lower front teeth). Polka et al. (2001) showed a very interesting developmental profile in infants' and adults' (from French- and English-speaking homes) perception of the contrast. While young infants (6-8 months) from both language backgrounds discriminated the contrast, discrimination by English-hearing infants, for whom the contrast is present in the ambient language, improved with age between 10-12 months and adulthood. For French infants and adult listeners, the discriminability of the contrast remained unchanged across development. Interestingly, the proportion of subjects reaching criterion (in a conditioned head turn task) was above chance in every group tested. This result has significant import from an acoustic perspective, as it confirms the more general trend for the [d]-[ð] to be reliably separable along various acoustic dimensions as well as discriminable in a variety of listening conditions for adults. Zhao (2010) examined the acoustics of the distinction and found that stop-like [ð] is separable from [d] along parameters of burst amplitude, burst spectrum, and onset of F2 among others. Further, consonant confusion studies show relatively low misidentification of [ð] as [d] and vice-versa in both quiet and noisy settings (Miller and Nicely, 1955; Wang and Bilger, 1973). That Polka et al.'s (2001) results showed a facilitation effect in the perception of [d]-[ð], with listeners becoming *better* at discriminating the contrast as they got older, is consistent with the idea that as the contrast becomes *phonologized* (or part of the abstract linguistic system) in the listener, the acoustic parameters underlying the distinction are reinforced.

3. Infant consonant perception from an acoustic perspective: Summary

Our review of the major findings in the infant consonant discrimination literature highlights the notion that not all consonant 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 not fixed, and reflect the acoustic similarity and dissimilarity of the speech sounds in question. Our review identifies four types of contrasts in the literature: 1) those that infants can perceive but which non-native adults show difficulty perceiving (e.g., [t̪]-[ʈ] by English speakers, [ɹ]-[l] by Japanese speakers); 2) those that young infants *and* adults show difficulty discriminating (e.g., [n]-[ŋ] by Filipino-hearing infants and adults, [f]-[θ] by English-hearing infants and adults); 3) those that young infants *and* adults discriminate well (e.g., clicks by English-hearing infants and adults); 4) those that adults perceive better than infants (e.g., [s]-[ʃ] by English speakers). An acoustic perspective does not necessarily speak to contrasts of type (1). We suggest that adults' difficulty in discriminating certain non-native contrasts is a reflection of the developmental process of perceptual reorganization which is driven by phonological patterns in the ambient environment rather than fine-grained acoustic salience of the contrast. That English-speaking adults no longer successfully discriminate the Hindi dental-retroflex contrast is a result of insufficient Hindi exposure and not the acoustic nature of the dental-retrofelx contrast. The speech contrasts for which infants show inconsistent or poor discrimination (types 2 and 4) are those that are fragile in acoustic salience (e.g., Burnham 1986; Narayan et al., 2010; Liu and Kager, 2014) and require greater experience with the language environment to perceptually segregate. An infants' model of linguistically relevant (to her native language) acoustic features are either well aligned with acoustically separable categories, or require refinement from experience.

Another broad generalization we can distill from our review is that transient distinctions (noisy bursts and rapid formant movement) lend themselves more to being successfully discriminated in early infancy and later subject to the perceptual reorganization exemplified by the work of Werker and Tees (1984). Speech contrasts that are characterized by temporally longer and louder acoustic features are more problematic in terms of identifying a definitive developmental trajectory. That is, speech sounds like nasals, fricatives, and liquids show patterns of perception in infancy that deviate from more typical *language-general* to *language-specific* reorganization. In linguistic terms we would predict that the more sonorous the consonant (with the exception of clicks), the more likely its perception in infancy to be characterized by enhancement (contrasts becoming more discriminable over time), asymmetries (where discrimination is found only in particular directions of stimuli change), and sustained maintenance (non-native contrasts remaining discriminable through infancy and beyond). This distinction mirrors the literature in categorical perception where consonant contrasts are more likely to be perceived categorically than vowel or tone contrasts. We might speculate that this reflects a more general trend in human cognition to treat transient and sustained events differently.

Part 2 of our review considers infants' discrimination of vowel and suprasegmental (tone vocalic duration) contrasts in the world's languages from an acoustic perspective. We argue that by appealing to acoustic-salience, phonological typology, and history, we can best understand infants' varied perception of vowels and suprasegmentals.

REFERENCES

Agnew, Z. K., McGettigan, C., & Scott, S. K. (2011). Discriminating between auditory and motor cortical responses to speech and nonspeech mouth sounds. *Journal of Cognitive Neuroscience*, *23*(12), 4038-4047.

Ahmed, R., & Agrawal, S. S. (1969). Significant features in the perception of (Hindi) consonants. *The Journal of the Acoustical Society of America*, *45*(3), 758-763.

Allen, G. D. (1985). How the young French child avoids the pre-voicing problem for word-initial voiced stops. *Journal of Child Language*, *12*(01), 37-46.

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

Benkí, J. (2005). Perception of VOT and first formant onset by Spanish and English speakers. In *Proceedings of the Fourth International Symposium on Bilingualism* (pp. 240-248).

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristià, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, *89*(6), 1996-2009.

Bertoncini, J., Bijeljac-Babic, R., Blumstein, S. E., & Mehler, J. (1987). Discrimination in neonates of very short CVs. *The Journal of the Acoustical Society of America*, *82*(1), 31-37.

Best, C.T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345.

Best, C. T., & Avery, R. A. (1999). Left-hemisphere advantage for click consonants is determined by linguistic significance and experience. *Psychological Science*, *10*(1), 65-70.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, *3*(1), 219-252

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

Chang, S.C., Plauché, M., Ohala, J., (2001). Markedness and consonant confusion asymmetries. In Hume and Johnson (Eds.), *The Role of Speech Perception in Phonology*. San Diego: Academic Press.

Christdas, P. (1988). *The Phonology and Morphology of Tamil, Cornell University* (Doctoral dissertation, Ph. D. dissertation).

Cristià, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of phonetics*, *39*(3), 388-402.

Dave, R. (1977). Retroflex and dental consonants in Gujarati. A palatographic and acoustic study. *Annual Reports of the Institute of Phonetics, University of Copenhagen*, *11*, 27-155.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research*, *20*(4), 766-780.

Eilers, R. E., Gavin, W., & Wilson, W. R. (1979). Linguistic experience and phonemic perception in infancy: A crosslinguistic study. *Child Development*, 14-18.

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.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, *16*(3), 513-521.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r l] distinction by young infants. *Perception & Psychophysics*, *18*(5), 341-347.

Eimas, P. D., and Miller, J. L. (1980). Contextual effects in infant speech perception. *Science*, *209*(4461), 1140-1141.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303-306.

Fuchs, S., Koenig, L. L., & Winkler, R. (2007). Weak clicks in German. In *Proceedings of 16th ICPhS–International Congress of Phonetic Sciences, Saarbrücken* (pp. 449-453).

Gandour, J., Petty, S. H., Dardarananda, R., Dechongkit, S., & Mukngoen, S. (1986). The acquisition of the voicing contrast in Thai: a study of voice onset time in word-initial stop consonants. *Journal of Child Language*, *13*(03), 561-572.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, *9*(3), 317-323.

Guion, S. G. (1997). The role of perception in the sound change of velar palatalization. *Phonetica*, *55*(1-2), 18-52.

Hamilton, P. (1996). *Phonetic Constraints and Markedness in the Phonotactics of Australian languages* (Doctoral dissertation, University of Toronto).

Hamann, S. (2003). *The Phonetics and Phonology of Retroflexes* (Doctoral dissertation, Universiteit Utrecht).

Hillenbrand, J. (1984). Speech perception by infants: categorization based on nasal consonant place of articulation. *The Journal of the Acoustical Society of America*, *75*(5), 1613-1622.

Holmberg, T. L., Morgan, K. A., & Kuhl, P. K. (1977). Speech perception in early infancy: Discrimination of fricative consonants. *The Journal of the Acoustical Society of America*, *62*(S1), S99-S99.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252-1263.

Jusczyk, P. W. (1977). Perception of syllable-final stop consonants by 2-month-old infants. *Perception & Psychophysics*, *21*(5), 450-454.

Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception & Psychophysics*,*23*(2), 105-109.

Keating, P., Linker, W., & Huffman, M. (1983). Closure duration of stop consonants. *Journal of Phonetics*, 11, pp.277-290.

Kluender, K. R. (1991). Effects of first formant onset properties on voicing judgments result from processes not specific to humans. *The Journal of the Acoustical Society of America*, *90*(1), 83-96.

Krishnamurti, B. (2003). *The Dravidian languages*. Cambridge: Cambridge University Press.

Kuhl, P. K. (1985). Methods in the study of infant speech perception. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 223-251). Westport, CT, US: Ablex Publishing.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *The Journal of the Acoustical Society of America*, *73*(3), 1003-1010.

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.

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), F13-F21.

Kurowski, K., & Blumstein, S. E. (1987). Acoustic properties for place of articulation in nasal consonants. *The Journal of the Acoustical Society of America*, *81*(6), 1917-1927.

Ladefoged, P. (1968). *A phonetic study of West African languages: An auditory-instrumental survey*. Cambridge University Press.

Lasky, R.E., Syrdal-Lasky, A., Klein, R.E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, *20*, 215–225.

Larkey, L. S., Wald, J., & Strange, W. (1978). Perception of synthetic nasal consonants in initial and final syllable position. *Perception & Psychophysics*, *23*(4), 299-312.

Lawrence, D. L., & Byers, V. W. (1969). Identification of voiceless fricatives by high frequency hearing impaired listeners. *Journal of Speech, Language, and Hearing Research*, *12*(2), 426-434.

Levitt, A., Jusczyk, P. W., Murray, J., & Carden, G. (1988). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 361.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1, 153-167.

Li, M., & Zhang, J. (2017). Perceptual distinctiveness between dental and palatal sibilants in different vowel contexts and its implications for phonological contrasts. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1), 18.

Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422

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.

Masica, C. P. (1993). *The Indo-Aryan Languages*. Cambridge University Press.

Mayo, C., & Turk, A. (2005). The influence of spectral distinctiveness on acoustic cue weighting in children’s and adults’ speech perception. *The Journal of the Acoustical Society of America*, *118*(3), 1730-1741.

McGuire, G. L. (2007). *Phonetic category learning* (Doctoral dissertation, The Ohio State University).

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, *27*(2), 338-352.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, *25*(6), 457-465.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331-340.

Moffitt, A. R. (1971). Consonant cue perception by twenty-to twenty-four-week-old infants. *Child Development*, 717-731.

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, *14*(3), 477-492.

Narayan, C. R. (2008). The acoustic–perceptual salience of nasal place contrasts. *Journal of Phonetics*, *36*(1), 191-217.

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.

Narayan, C. (2013). Developmental perspectives on phonological typology and sound change. In A.C.L. Yu (Ed.), *Origins of Sound Change: Approaches to Phonologization*, (p.128). Oxford: Oxford University Press.

Narayanan, S., Byrd, D., & Kaun, A. (1999). Geometry, kinematics, and acoustics of Tamil liquid consonants. *The Journal of the Acoustical Society of America*, *106*(4), 1993-2007.

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

Nowak, P. M. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics*, *34*(2), 139-152.

Ohala, M. and Ohala, J. J. (2001). Acoustic VC transitions correlate with degree of perceptual confusion of place contrast in Hindi. *Travaux du Cercle Linguistique de Copenhague*. 31.265-284.

Oller, D. K., Eilers, R. E., Burns, R., & Urbano, R. (1993). Perception of the stop/glide contrast in infancy. *Phonetica*, *50*(1), 1-14.

Phatak, S. A., Lovitt, A., & Allen, J. B. (2008). Consonant confusions in white noise. *The Journal of the Acoustical Society of America*, *124*(2), 1220-1233.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *The Journal of the Acoustical Society of America*, *61*(5), 1352-1361.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/– /ð/perception: evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, *109*(5), 2190-2201.

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, *22*(2), 173-189.

Simonet, M., Rohena-Madrazo, M., and Paz, M. (2008). “Preliminary Evidence for Incomplete Neutralization of Coda Liquids in Puerto Rican Spanish.” In *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, ed. Laura Colantoni and Jeffrey Steele, 72-86. Somerville, MA: Cascadilla Proceedings Project.

Sinnott, J. M., Beecher, M. D., Moody, D. B., & Stebbins, W. C. (1976). Speech sound discrimination by monkeys and humans. *The Journal of the Acoustical Society of America*, *60*(3), 687-695.

Smit, A. B., Hand, L., Frieilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, **55**, 29–36.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America*, *55*(3), 653-659.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3-45.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*, *259*(5538), 39.

Sundara, M., Ngon, C., Skoruppa, K., Feldman, N. H., Onario, G. M., Morgan, J. L., & Peperkamp, S. (2018). Young infants' discrimination of subtle phonetic contrasts. *Cognition*, *178*, 57-66.

Sussman, J. E. (2001). Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions? *The Journal of the Acoustical Society of America*, *109*(3), 1173-1180.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 466-472.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 466-472.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., ... & Best, C. T. (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 *The 3rd International Conference on Spoken Language Processing, ICSLP 1994, Yokohama, Japan, September 18-22, 1994*.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *The Journal of the Acoustical Society of America*, *54*(5), 1248-1266.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 349-355.

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.

Whalen, D. H. (1991). Perception of the English/s/–/ʃ/distinction relies on fricative noises and transitions, not on brief spectral slices. *The Journal of the Acoustical Society of America*, *90*(4), 1776-1785.

Whalen, D. H., Levitt, A. G., & Goldstein, L. M. (2007). VOT in the babbling of French-and English-learning infants. *Journal of Phonetics*, *35*(3), 341-352.

Zhao, S. Y. (2010). Stop-like modification of the dental fricative /ð/: an acoustic analysis. *The Journal of the Acoustical Society of America*, *128*(4), 2009-2020.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English/r. *The Journal of the Acoustical Society of America*, *123*(6), 4466-4481.

Żygis, M., & Padgett, J. (2010). A perceptual study of Polish fricatives, and its implications for historical sound change. *Journal of Phonetics*, *38*(2), 207-226.