# 1

# Developmental perspectives on phonological typology and sound change

**CHANDAN NARAYAN**

## 1.1 Introduction

The relationship between first language acquisition and phonologization lies at the crossroads of developmental psychology and historical phonology—disciplines not often considered in the same breath when addressing the nature of sound patterns and change. Despite these traditional boundaries I believe that this combined research program can make significant contributions to a more nuanced understanding of why sound systems look the way they do and change in particular directions. The present chapter deals with the relationship between the earliest stages of language acquisition and the shape of phonological systems and phonological processes including sound change. The term “developmental” encompasses both the dynamic nature of the cognitive mechanisms underlying infants’ and very young children’s emerging organization of their acoustic-phonetic environment as well as the nature of the linguistic environment itself. Of particular significance here is the *potential* contribution of developmental processes (infant speech perception and caregiver speech production) to the phonologization of acoustic variance in the input. The scope of these developmental contributions is not limited to the infant and her abilities but also characteristics of the unique register used by caregivers when interacting with infants. This research program asks two questions:

1. Is there a relationship between patterns in developmental speech perception and the relative rarity in sounds in phonological systems of the world’s languages?
2. What is the role of caregiver-infant interactions in providing acoustic conditions which could potentially lead to sound change?

My approach to these questions are guided throughout by the notions that phonological inventories reflect, to some degree, sufficient acoustic-perceptual salience between contrasts (Liljencrants and Lindblom, 1972;

Lindblom and Maddieson, 1988; de Boer, 2001; Oudyer, 2006; Narayan, 2008), that misperception can lead to sound change (Ohala, 1981), and that the adaptive nature of speech production (Lindblom, 1990) has the potential to create less-than-ideal learning situations for infants. I argue that (1) those contrasts for which infants require language experience in order to discriminate are those that are rare in phonological systems and often the targets of change and (2) that the relationship between developmental speech perception performance and the shapes of phonological systems is mediated by infants' initial psychoacoustic biases and the acoustic salience of the contrasts in question. With respect to sound change, it must be made clear, at the outset, that the present discussion does not deal with the diffusion of sound change across a community of learners (children), but rather with what some have termed the "seeds" of sound change (Hombert *et al.*, 1979). In examining the structured variability inherent in the acoustic signal available to very young children (e.g., Foulkes *et al.* 2005) and the perceptual biases children bring to the language-learning table, I explore possible developmental influences on the directions of phonologization.

### 1.1.1 *Children's productive phonology*

History has provided us with numerous examples of how linguists have approached the connection between development, phonological typology and sound change (Herzog, 1904; Sweet, 1913; Baudouin de Courtenay, 1972; Grammont, 1933; Jakobson, 1968; Jakobson and Waugh, 1979). The approach in much of this literature examines performance factors in development rather than children's competence. That is, linguists have asked whether children's emerging productive phonology resembles well-known sound changes. For the most part, the answer to this question has been *no*, for the types of phonological changes that are reflected in children's productions are too varied to be reflections of real, phonetically motivated sound changes. An early appraisal of this confound is provided by Baudouin de Courtenay (1895/1972):

> ...when the child has not yet begun to talk but is already aware of the properties of the native language and can understand it within certain limits, that is, when the child has reached the state of advanced audition and perception, but without phonation, there naturally cannot be any question of neophonetic alternations or divergences, since these depend on *individual* pronunciation

Others have viewed the relationship between children's productive phonology and phonological change more sympathetically and directly (e.g., Labov 1989). Grammont (1933) suggests that children's productions are a "microcosm" of historical change, while Stampe went a step further in suggesting that children are the prime agents in phonological change (Stampe, 1972). Greenlee and Ohala 1980 argue that both children and adults are responsible for the type of phonetic variation that can lead to sound change. Under the rubric of Ohala's misperception-based sound change (Ohala, 1981), where physical constraints on articulatory and perceptual dynamics lead to the phonologization of variation, Greenlee and Ohala (1980) outline the shifts of child phonology that are similar to diachronic processes (e.g., French $\tilde{V}$ > French loans in Vietnamese and children learning French *Vŋ*).

More recently, linguists have suggested that the relationship between child phonology and historical phonology should be played down precisely because "typical or potential sound changes" do not match observed phonological states in children's production (Kiparsky, 1988). Blevins (2004) argues that the mismatch between children's productive phonology and sound changes (i.e., the types of production mistakes made by children do not always look like typical sound changes) is a non-problem, as the enterprise of child phonology does not necessarily assess competence (in the form of perceptual acuity) but rather performance factors very likely governed by physiological development (see also Hale and Reiss 1998). She outlines children's productions, described in terms of phonological rules, as falling under two categories: those resulting from immature articulatory development, or secondly true "mini-sound changes" which may spread through a community of speakers. As Blevins and others have argued, the problem with looking to children's productive phonology for clues to directions in phonologization is that articulatory and perceptual capacities in the first few years of life mature along differing time scales, with motor control and oral tract development lagging behind the shaping of perceptual competence. At the earliest stages of language acquisition, production is not necessarily a reflection of competence (perceptual discrimination and categorization), with perceptual acuity becoming honed well before infants' production of their first word at around 12 months.

The clearest demonstration of the connection, and perhaps influence, of children's productions and phonological phenomena can be seen in typological inventories. Sound change aside, linguists have recognized the connection between the age of productive acquisition of phonologically relevant phones and the relative rarity of these sounds in phonological systems (Ferguson, 1973). In general, age of successful production can be described as exponentially related to frequency of occurrence in the world's sound systems, that is, the more rare the consonant, the later its productive acquisition.[1] Figure 1 plots the age of productive “mastery”[2] of consonants by American English-, Putonghua Chinese-, and Jordanian Arabic-learning children against the consonants’ frequency of occurrence in the UPSID (Maddieson, 1984). The plots show a general cross-linguistic trend with simple, oral obstruents being produced very early in productive development, intermediate acquisition of voiceless sibilants, and late production of interdentals, dorsals, affricates, and obstruents with secondary articulations (e.g., pharyngealized stops in Arabic). This relationship suggests that languages are less likely to exhibit sounds which are articulatorily more *difficult* for children (as measured by the relative lateness of their mastery).

We turn next to the flip side of the production/typology connection. In the next section, I ask whether typological generalizations can be derived from the performance of infants in speech perception tasks, that is, is there a relationship between the types of contrasts that infants *fail* to discriminate and contrasts that are rare in the world’s sound systems?

## 1.2 Infant speech perception and phonological typology

The literature on infant speech perception presents remarkable evidence of the capacity for infants to discriminate non-native phonetic contrasts across a host of genetically unrelated languages. Infants as young as 1 month have been shown to discriminate non-native contrasts that adults find difficult to discriminate (Eimas *et al.*, 1971; Trehub, 1976). For example, Trehub (1976) showed that English-hearing infants aged 5-17 weeks successfully discriminated a French oral-nasal vowel contrast ([pa]-[pã]) and a Czech fricative-place contrast ([za]-[ʝa]). When English-speaking adults were asked to discriminate the Czech contrast, they performed at chance levels. This phenomenon is perhaps best captured by the work of Werker

---

[1] While there is certainly a relationship between the frequency of occurrence and the emergence of certain phonological structures (see Levelt *et al.* 1999; Demuth and Johnson 2003; Rose 2009), the relationship between accurate production of individual phones and the frequency of those phones in the ambient language of the child is less clear than the overall typological frequency across languages. Appendix A provides a table of the frequency of consonants in the Brent corpus of infant-directed speech (Brent and Siskind, 2001).

[2] Mastery of English consonants in Templin’s (1957) study is described as 75% accuracy, while a more strict criterion of 90% is used in Hua & Dodd’s (2000) and Amayreh & Dyson’s (1998) studies.

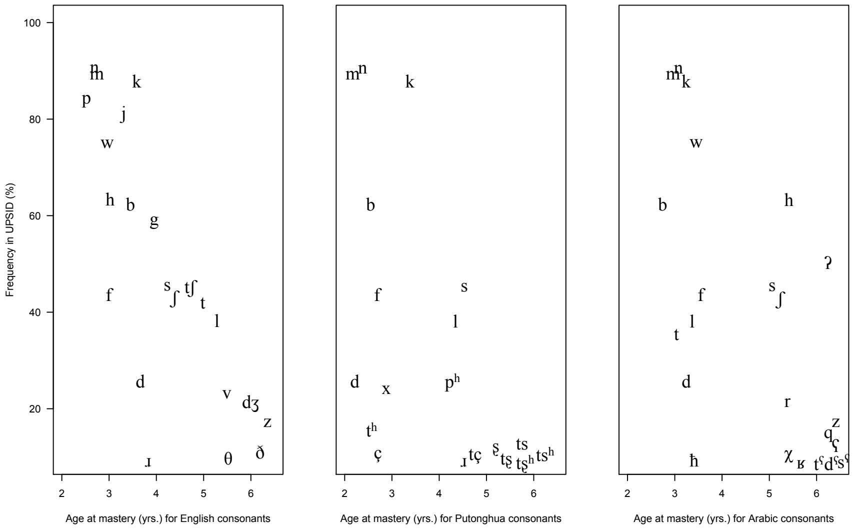

FIG. 1.1 Age of production mastery according to frequency in the UPSID (Maddieson, 1984) in American Englsh (Templin, 1957), Putonghua (Hua and Dodd, 2000), and Jordanian Arabic (Amayreh and Dyson, 1998). Values are jittered within each year.

and colleagues, who showed that discrimination of non-native contrasts follows a distinctive developmental pattern. Werker and Tees (1984) showed that English-learning infants aged 6-8 months discriminated non-native consonant contrasts (Hindi voiceless dental/retroflex [t̪a]-[ʈa]; Hindi dental voiceless/voiceless aspirated [t̪ʰa]-[d̪ʰa]; Nlaka’pamux velar-uvular ejective [k’i]-[q’i]). By 10-12 months, however, English-learning infants failed to reliably show discrimination of the same contrasts. At 10-12 months, infants from both Hindi-speaking and Nlaka’pamux-speaking homes discriminated their native contrasts. This same pattern of perceptual *reorganization* was subsequently found for other consonants as well (e.g., /ɹ/-/l/ with Japanese- and English-learning infants, Tsushima *et al.*, 1994). The perceptual reorganization has been shown to occur earlier for vowels, with infants’ non-native discrimination abilities declining by 6 months (Kuhl *et al.*, 1992; Polka and Werker, 1994). Results like these, showing the effects of experience, led infant speech perception research to converge on the idea that infants come into the world with language-*general* speech perception, which becomes tuned to the particular phonetic characteristics of their native language by the end of their first year. While the generalization that infants are born *citizens of the world* is compelling from a neural plasticity point of view (e.g., Huttenlocher, 2002), it is certainly not the complete picture of the nature of infant speech perception.

Infants' performance on speech perception tasks generally lies on a cline of more or less discriminability, with developmental profiles being mediated by the intersection of innate perceptual bias, psychoacoustic salience, and language experience (Aslin and Pisoni, 1980; Narayan *et al.*, 2010). Recent work has demonstrated that certain contrasts follow a path of *facilitation*, whereby initially poor discrimination is enhanced with native language experience (Polka *et al.*, 2001; Kuhl *et al.*, 2006; Narayan *et al.*, 2010). The facilitation of discrimination highlights the fact that initial speech perception abilities are poor (or undetectable using behavioral methods) for certain contrasts; not all phonetic contrasts are perceptually equal for the infant and the relative difficulty is mediated by acoustic salience. I outline below some instances in the infant speech perception literature that reflect a connection between relatively poor discrimination, acoustic salience, and typological frequency.

### 1.2.1 *Case studies*

**Nasal place of articulation**—In Narayan (2008) I examined the relationship between nasal-place acoustics and nasal place typology in the world's languages. In general, languages are more likely to exhibit a two-way /m/-/n/ contrast than a three-way /m/-/n/-/ŋ/ contrast in syllable-initial position (Maddieson, 1984; Anderson, 2008). I argued, based on static (F2 x F3 frequencies at the onset of the NV transition) and dynamic acoustic properties (RMS energy change from nasal murmur to V) of the three nasal places in Filipino (Figure 2) and corresponding discrimination tests with adult Filipino-speaking listeners, that the acoustic-perceptual salience of the /m/-/n/ distinction is more robust than /n/-/ŋ/. Both static and dynamic acoustic measurements showed better classification (with discriminant analyses) of the /m/-/n/ and /m/-/ŋ/ distinctions than the /n/-/ŋ/ contrast where tokens showed significant overlap along the critical acoustic dimensions. Consequently, the /n/-/ŋ/ distinction is disproportionately affected by adverse listening conditions. In the noisiest listening condition (-5dB SNR), discrimination of the [na]-[ŋa] contrast fell to chance while discrimination of both [ma]-[na] and [ma]-[ŋa] remained near ceiling.

In a follow-up study I examined the perception of nasal place contrasts in Filipino- and English-learning infants (Narayan *et al.*, 2010) using the Visual Habituation technique.[3] Following on from the typological and acoustic-perceptual results from Narayan (2008), the [na]-[ŋa] contrast

---

[3] See Werker *et al.* 1998 for details regarding infant speech perception methods.

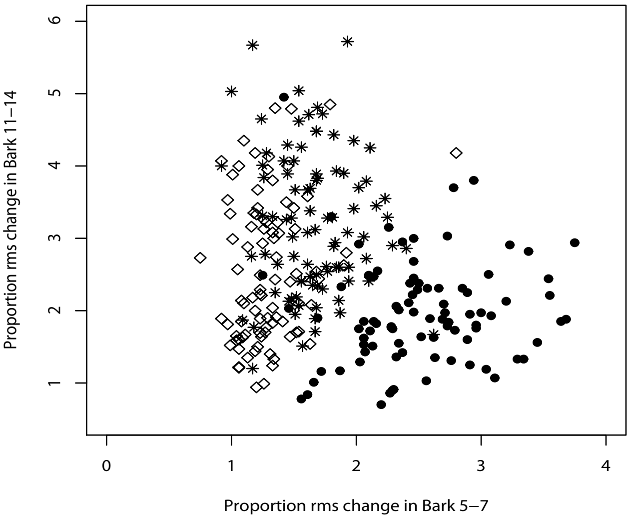

FIG. 1.2 Proportion RMS energy change from nasal murmur to post-nasal vowel in Bark 5-7 and 11-14 for [ma](•), [na](∗), and [ŋa](⋄) in Filipino. Used with permission from Narayan (2008).

proved difficult for both groups of infants. English-hearing infants at 10-12 and 6-8 months discriminated the acoustically robust, and typologically common [ma]-[na] contrast. English-learning infants did not reliably discriminate the acoustically fragile [na]-[ŋa] contrast, even at 6-8 months, an age when other non-native (oral) consonant contrasts are successfully discriminated. Even very young English-learning infants (4-5 months) were unable to discriminate [na]-[ŋa] while they successfully discriminated the acoustically robust [ma]-[na] contrast. When 10-12 and 6-8-month-old Filipino-learning infants' discrimination of the native [na]-[ŋa] contrast was tested, only the older group showed discrimination.

The results from (Narayan *et al.*, 2010) are suggestive of a role for acoustic salience in developmental speech perception. The [na]-[ŋa] contrast, which is acoustically fragile (relative to the robust [ma]-[na] contrast), is poorly discriminated in early infancy and only successfully discriminated with appropriate language experience by the end of the first year. I would suggest that infants' difficulty discriminating the perceptually similar syllable-initial [n]-[ŋ] contrast contributes to the typological restrictions on nasal onsets and the directions of sound change patterns observed in nasals in the world's languages (i.e., Pr.Austronesian syllable-initial *m,n,ŋ > Thao, Malagasy, Tetun, Hawaiian, Tahitian *m, n*).

**Fricative contrasts: /f/-/θ/ and /s/-/z/**—Dental, non-sibilant fricatives are rare in the world’s sound systems, occurring in only 3.99% of the languages surveyed in the UPSID (Maddieson, 1984). In the WALS database of 567 genetically diverse languages, they occur in just 43 (7.6%)(Maddieson, 2008). Correspondingly, contrasts involving dental fricatives have been shown to pose problems for infants in speech discrimination tasks. In a series of studies in the 1970s, Eilers and colleagues showed that English-hearing infants at both 6-8 months and 10-12 months fail to accurately discriminate the English labiodental-interdental fricative place distinction ([fa]-[θa])(Eilers, 1977) using the Conditioned Head Turn procedure (CHT). The older group showed discrimination of the contrast only when the fricative was followed by [i]. This result proved highly controversial and led to two subsequent studies, both of which showed English-learning infants discriminating the [fa]-[θa] contrast. While Holmberg *et al.* (1977) showed that 6-month-olds discriminated the contrast, they noted that subjects required twice as many trials to achieve criterion (an indirect measurement of perceptual difficulty) than they did to reach criterion on the /s/-/ʃ/ contrast. Further, at 2 months, infants were shown to successfully discriminate [fa]-[θa] using the High-amplitude sucking (HAS) procedure (Levitt *et al.*, 1988). The conflicting reports of labiodental-interdental fricative discrimination in English-hearing infants suggests the perceptual difficulty of the contrast relative to plosive obstruent place contrasts.

I suggest that there is an acoustic source for infants’ difficulty discriminating /f/-/θ/, which potentially contributes to the relative rarity of the contrast in sound systems. In a recent acoustic study of 20 American English speakers, the fricative noise in both sounds was shown to have similar duration (165 ms), spectral peak locations (8 kHz), mean spectral moments (5.1 kHz), kurtosis and skewness (Jongman *et al.*, 2000), all of which contribute to place perception in fricatives (Behrens and Blumstein, 1988; Jongman, 1988; Hedrick and Ohde, 1993). In Jongman et al.’s (2000) study, when 21 acoustic predictors were used in a discriminant analysis classification, 27% of labiodental tokens were classified as interdentals, and 26% of interdentals as labiodentals. This rate of confusion is consistent with human perceptual confusions between the two fricative places. In Miller and Nicely 1955, at the highest signal-to-noise ratio and with the broadest band of frequency information (+12 dB SNR, 200-6500 Hz), listeners identified /θ/ as /f/ at a rate of 26%. Further, in several varieties of English (e.g., working-class London speech), /f/ and /θ/ are merging.

Another fricative contrast that has proved difficult for infants to discriminate is the alveolar voicing contrast. In a HAS procedure, English-learning infants (1-4 mos.) failed to discriminate [sa]-[za] contrast (Eilers and Minifie, 1975; Eilers *et al.*, 1977). There is corresponding asymmetry in the distribution of voiced and voiceless alveolar fricatives in the world's languages as well. In UPSID 69% of alveolar fricatives are voiceless. While there is a clear articulatory/aerodynamic reason behind the preference for voiceless (over voiced) fricatives[4] (Ohala, 1983) and corresponding devoicing of /z/ (Smith, 1997), there is no clear acoustic-perceptual reason for infants' failure to discriminate /s/-/z/ at such an early age.[5] Indeed English-speaking adults' perception of the contrast leads to little confusion (Miller and Nicely, 1955), perhaps owing in part to differences in voice onset time and fricative duration and amplitude (Jongman *et al.*, 2000).

**VOT**—Discrimination of voice onset time (VOT) distinctions provided the earliest demonstration of infants' ability to perceive speech-like sounds categorically (Eimas *et al.*, 1971) and laid the groundwork for research in the 1970s and 1980s testing the limits of infants' perceptual abilities. A series of studies by Lasky, Streeter, Eimas, Eilers and colleagues revealed interesting patterns with respect to distinctions between lead/lag VOT contrasts versus short/long lag contrasts. In general, there is 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 (similar to the English implementation of voicing), infants succeeded in discriminating the contrast (Eimas *et al.*, 1971; Lasky *et al.*, 1975; Streeter, 1976; Eilers *et al.*, 1979). Interestingly, infants whose native language background did not contrast short vs. long lag also successfully discriminated the distinction. Kikuyu-learning infants discriminated a +10/+40 ms VOT contrast (Streeter, 1976) and Spanish-learning infants discriminated a +20/+60 ms contrast (Lasky *et al.*, 1975).

The results of infants' perception of lead vs. short-lag VOT is quite different, however. The overwhelming majority of studies investigating this distinction suggest that infants' discrimination is quite poor. Only two studies (Eimas, 1974; Streeter, 1976) have shown infants' successful discrimination of the prevoicing/short-lag contrast (Table 1). Kikuyu-learning infants discriminated both a prevoiced/simultaneous (-30/0 ms) VOT distinction as well as the short/long-lag distinction. It remains unclear, however, whether the prevoiced discrimination results from experience with Kikuyu or the psychophysical salience of the contrast, for English-learning

---

[4] The turbulent noise of fricatives requires a high volume velocity of pulmonic airflow, which is necessarily impeded by the oscillating vocal cords during voicing.

[5] A drawback in interpreting early infant speech perception results is that precise acoustic measurements of stimuli are often unavailable, thus precluding robust comparisons across studies. Eilers's studies likewise provided minimal acoustic data. Data which were provided suggest that /s/ and /z/ differed along the perceptually-critical parameters of voice onset time and fricative duration.

TABLE 1.1 Summary of infants' discrimination of lead vs. short-lag VOT contrasts

| VOT contrast (ms) | Discrimination successful? | Age (mos.) | Lg. background | Author |
|---|---|---|---|---|
| -30/0 | + | 2 | Kikuyu | Streeter (1976) |
| -30/0 | - | 6 | English | Eilers *et al.* (1979) |
| -60/+10 | - | 6 | English | Eilers *et al.* (1979) |
| -20/+10 | - | 6 | English | Eilers *et al.* (1979) |
| -20/+20 | - | 4-6.5 | Spanish | Lasky *et al.* (1975) |
| -40/+20 | - | 2,3 | English | Eimas (1974) |
| -20/0 | - | 1,4 | English | Eimas *et al.* (1971) |
| -70/+10 | + | 2,3 | English | Eimas (1974) |
| -60/+10 | - | 6 | English | Eilers *et al.* (1979) |

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 from the infant's point of view (but see Aslin *et al.* (1981)). The lag region of the VOT continuum is most likely privileged by the perceptual system for psychophysical reasons (Pisoni, 1977) as it provides more robust cues to a voicing contrast (aspiration, F1 onset) than does the lead/short-lag distinction.

The perceptual advantage afforded to the short/long lag distinction in infancy has an analogue in production as well, where mastery of prevoicing occurs relatively late compared to short-lag VOT in languages like Spanish (Eilers and Benito-Garcia, 1984), French (Allen, 1985), and Thai (Gandour *et al.*, 1986) (but see Whalen *et al.*, 2007, for VOT in babbling). The connection between infants' greater success at discriminating short/long-lag contrasts versus lead/short-lag contrasts and typological patterns remains unclear, owing to a lack a comprehensive cross-linguistic survey (similar to UPSID or WLS) of voicing implementation along VOT. Keating *et al.* (1983)'s survey of 51 languages shows that a voicing contrast always utilizes a "voiceless unaspirated" (short lag) stop. Keating (1984) suggests that in contrast to the short-lag implementation of VOT, languages which feature stop voicing contrasts are equally likely to use "fully voiced" (lead) or "voiceless aspirated" stops. The perceptual patterns of infants would predict, however, that languages more often utilize a short/long-lag implementation of voicing than lead/short-lag VOT.

**/l/-/ɹ/**—Perception of the [la]-[ɹa](English "r") contrast has recently been shown to be be facilitated by native language experience (Kuhl *et al.*, 2006). Kuhl (2006) investigated English- and Japanese-learning infants' perception of the (naturally produced) contrast at 6-8 and 10-12 months of age using the CHT procedure. At 6-8 months, both English- and Japanese-learning infants discriminated the contrast at a rate of 65% correct, well below native levels of discrimination ability (approximately 80% correct for synthetic stimuli in Miyawaki *et al.* 1975). By the end of their first year, English-learning infants' perception of the contrast improved to approximately 75% correct. Further supporting the relative difficulty of /l/-/ɹ/ discrimination in infancy, Kuhl's (2006) results revealed a directional asymmetry, where facilitation of the contrast occurs only when infants are conditioned to discriminate a change from /ɹ/ to /l/.

The English "r" is rare among the world's sound systems (occurring in roughly 2% of the languages in UPSID compared with 39% of languages with /l/) and notoriously difficult to produce and perceive for non-native speakers (e.g., Goto 1971; Miyawaki *et al.* 1975; Polka and Strange 1985). Acoustically, /ɹ/ and /l/ have very similar spectral profiles, differing primarily in $F3$, which is characteristically low in /ɹ/ (Fant, 1960; Dalston, 1975; Espy-Wilson, 1992).

**/d/-/ð/**—Typologically, interdental fricatives are rare, relative to their nearest plosive counterparts. For example, 44% of UPSID languages exhibit alveolar stops, while 7% show interdental fricatives. The asymmetry becomes even more apparent when compared to similar stop-fricative contrasts at other places of articulation: 96% of languages have bilabial stops vs. 44% labiodental fricatives; 97% with velar stops vs. 28% velar fricatives; 70% with dental or alveolar stops vs. 7% with interdental fricatives. L2 speakers of a language exhibiting the fricative often resort to substitution of /ð/ with /d/ or /z/ (e.g., Dutch speakers' production of English *the* as [dʌ] and French speakers' production [zʌ]). Similar substitutions for /ð/ are found in non-standard Englishes: /ð/> *African American Vernacular English* [d]; *Cockney English* [v].

Infants' discrimination of the /d/-/ð/ contrast was the first to show the developmental profile of *facilitation* (Polka *et al.*, 2001). Two groups of infants were tested: French-learning infants, for whom the /d/-/ð/ contrast is nonnative, and English-learning infants, for whom the contrast is native to their ambient phonology. At 6-8 months, both French- and English-learning infants showed mean *A'*[6] only slightly better than chance, scores less than that of the control contrast (/b/-/v/), which is native to both groups. In addition, infants' *A'* scores showed more variation for /d/-/ð/ contrast than the control. By 10-12 months, mean *A'* scores for English-learning infants increased slightly, while remaining unchanged for

---

[6] *A'* is a nonparametric index of sensitivity similar to *d'*, the difference between $z$-scored proportion of *hits* and *false alarms* in discrimination.

French-learning infants. Adult English speakers showed *A'* scores reflecting ceiling levels of discrimination. Adult French speakers' *A'* scores remained unchanged from the infant groups. These results are suggestive of the interpretation that language experience serves to *facilitate* (or improve) native contrast discrimination. Further, they also show that the initial state of /d/-/ð/ perception is less accurate than a similar stop-fricative (here /b/-/v/) contrast.

In clean listening conditions, English-speaking adults discriminate the /d/-/ð/ quite well (Polka *et al.*, 2001), but with the addition of additive noise, confusion patterns result that are consistent with both the infants' relatively poor discrimination as well as the substitution patterns observed in L2 speakers attempting to produce /ð/ (Miller and Nicely, 1955). Taken together, results from native infant and adult perception are suggestive of a low-level acoustic source for relatively poor discrimination of the /d/-/ð/ and the substitutions observed.

### 1.2.2 *Implications*

Infants' perceptual sensitivities are far from *language universal.* The outline presented above, highlighting instances where infants' perceptual performance falls short of the language-general perceptual specification often cited by linguists and psychologists, corresponds to the typological regularities found across the world's languages. I would suggest that these contrasts, which are *fragile* in terms of their acoustic distinctiveness, are prone to misperception at the earliest stages of phonological development.

Another stage in the acoustics/development/typology story is type frequency in the lexicon and token frequency in ambient speech. It is often the case that phones in a weak acoustic-perceptual salience relationship are rather infrequent exemplars in the lexicon (such as /ð/ restricted to demonstrative articles in English) as well as token frequency (as in syllable-initial /ŋ/in Filipino) in a language (Narayan, 2008). I would suggest that if infants have minimal evidence (in terms of a stochastic mechanism for category formation)(Johnson, 1997; Pierrehumbert, 2001; Maye *et al.*, 2002) for an already acoustically weak contrast, which is then coupled with a low functional load (Martinet, 1933), they have the potential to affect misperception-based change (Greenlee and Ohala, 1980). This argument is further bolstered by the fact that in some children, production patterns suggest an effect of perception on early lexical representations (Macken, 1980; Rose, 2009).

## 1.3 Infant-directed speech and phonologization

The *natural* state of the acoustic-phonetic input to infants (or infant-directed speech or IDS) and its relation to emerging phonology has been the focus of a growing body of literature on speech category learning (Kuhl *et al.*, 1997; Andruski *et al.*, 1999; Liu *et al.*, 2003; Werker *et al.*, 2007). Much of this work is driven by recent models of category learning as a function of the frequency of the input, where infants are shown to discriminate phonetic categories when familiarized to tokens comprising different modes in an artificially created stimulus continuum (Maye *et al.*, 2002, 2008). Researchers have found such modally distributed cues in the acoustic input to infants. For example, Werker et al. (2007) showed that Japanese- and English-speaking mothers, when teaching new words to their young infants consistently produced acoustically distinct modes of vowel quality (/ɪ/ vs. /ɛ/ and /ɛ/ vs. /e/ for English) and vowel duration (/i/ vs. /iː/ and /ɛ/ vs. /ɛː/ for Japanese). Much of the research examining IDS has highlighted its *enhancing* hallmarks, where categorical phonetic distinctions are exaggerated (i.e., vowel duration, vowel quality, tone – Tang and Maidment 1996; Liu *et al.* 2007). The present section considers an often overlooked acoustic consequence of the IDS register, namely the reduced clarity of contrast in the speech to very young infants (Baran *et al.*, 1977; Malsheen, 1980; Sundberg and Lacerda, 1999), and its implications for the directions of sound change.

### 1.3.1 *From emotional and social to linguistic function*

Much like the developing perceptual system, caregivers’ speech changes over the course of an infant’s first year. In early infancy (before infants’ one-word stage) IDS is very much a biologically relevant acoustic signal, serving to assuage, arouse, and regulate infants’ attentional state (Sachs, 1977; Fernald, 1992), stimulating them to “calm awareness” (Cooper and Aslin, 1989). By the time infants begin producing their first words, the communicative intent underlying IDS is said to take on a more linguistic function. Psychologists and linguists have arrived at this conclusion by examining the changing acoustic clarity (the distance between phonemes in some acoustic space) of IDS over the course of development. For example, early research on the prosodic quality of English IDS showed that the distinctive pitch excursions characteristic of stereotypical IDS decreased as the child’s age increased (Garnica, 1977). Fernald *et al.* (1989) examined prosodic characteristics of IDS in English, French, German, Italian, and Japanese, and found similar results: higher mean $f_0$ with wider range, longer pauses, shorter utterances, and more repetitions compared to adult-directed speech (ADS). These acoustic characteristics were more exaggerated in speech to very young infants and decreased as children became more skilled in language use (Fernald, 1992). Exaggerated prosodic features such as intonational patterns and syllable duration contribute to the IDS affect, which has been shown to be preferred by young infants. Infants younger than 6 months show more attentional and affective response to IDS than do infants at 9 months (Werker and McLeod, 1989). More recently IDS has been shown to facilitate word segmentation, which has implications for other aspects of language learning. Infants (7 months old) were exposed to either IDS or ADS nonsense sentences where the statistical structure of the syllables served as the only cue to word boundaries. Only infants exposed to the IDS input were able to distinguish words from part words (Thiessen *et al.*, 2005).

Interestingly, acoustic features of IDS at the level of the segment also seem to change over the course of an infant's development. Malsheen (1980) examined voicing in a longitudinal study of English IDS spoken to children ranging from 6 months to 5 years of age and found that only when infants were 15-16 months old did mothers significantly separate the voiced and voiceless categories along VOT. At 15-16 months, mothers implemented longer VOTs for voiceless tokens than in their voiceless tokens to younger infants. Baran *et al.* (1977) found no significant differences in VOT between IDS and ADS when infants were 12 months old. Sundberg and Lacerda (1999) found that in the IDS addressed to 3-month-old Swedish infants, VOT was significantly shorter in both voiced and voiceless stops than in ADS. This resulted in more overlap between the voicing categories in IDS. The authors provide a developmental account of their findings by suggesting that acoustic properties of obstruents are less "specified" in the IDS to young infants and gradually reach adult-directed VOT values at around the time infants produce their first word. More recently, in a study of Norwegian IDS, Englund (2005) found that alveolar and velar stops have longer VOTs during infants' first six months than in ADS. While there were no differences in the voiced/voiceless distinction along VOT between the two registers, the developmental profile of the data suggested that VOT in IDS becomes more like ADS as infants get older. The developmental account is consistent with studies of IDS vowel production as well, where acoustic clarity is found only in those lexical categories used by the child (Bernstein Ratner, 1984).

What I argue in the case study below is that the not-so-careful speech to very young infants has acoustic consequences which have the potential to become phonologized by infants in this perceptually sensitive stage of development (Werker and Tees, 1984). The interaction between the socially driven imperatives of early IDS and contrastive phonetic salience can provide the learner with the kind of structured acoustic variability associated with misperception-based sound changes (Ohala, 1981).

### 1.3.2 *Modeling voicing in English IDS and ADS*[7]

The covariation between voice-onset time (VOT) and post-consonantal $f_0$ is a well-known source of tone in languages that have historically lost voicing contrasts (Matisoff, 1973; Hombert *et al.*, 1979). As a result of naturally conditioned pitch perturbation, where voiced consonants exert a lower $f_0$ than do voiceless consonants on a following vowel (Abramson and Lisker, 1985), a relatively low tone develops on vowels following previously voiced consonants and a relatively higher tone develops on vowels following previously voiceless consonants. Hombert and Ohala (1979) note: "The historical development of tones (tonogenesis) can result from the reinterpretation by listeners of a previously intrinsic cue after recession and disappearance of the main cue." While the primary cue to voicing in English (VOT) has not "disappeared" as happened in many cases of tonogenesis, I would argue that the IDS register contributes to acoustic ambiguity in voicing that is consistent with the development of tone.

Previous studies have shown that the distribution of voiced and voiceless tokens along VOT are more similar in the IDS to infants under 12 months than in the IDS to older infants or in ADS (in American English and Swedish). Voiceless VOTs are generally shorter in IDS resulting in more overlap with voiced VOTs (Baran *et al.*, 1977; Malsheen, 1980; Sundberg and Lacerda, 1999) compared to ADS or IDS to older infants. In a recent study of word-initial voicing in American English IDS and ADS we (myself together with Kyle Gorman and Daniel Swingley from the University of Pennsylvania) examined VOT and post-consonantal $f_0$ in the hope of understanding 1) the regularity of the acoustic features of voicing available to young infants and 2) the relative weights of VOT and $f_0$ in predicting voicing in IDS and ADS. In examining the covariation of VOT and $f_0$ in voicing in two different registers we hope to shed light upon the history of the interaction between these features as providing potentially ambiguous and ultimately misinterpretable cues.

### 1.3.3 *Methodological approach: Logistic regression modeling*

Voicing in English IDS and ADS was modeled using binary logistic regression (Hosmer and Lemeshow, 1989; Gelman and Hill, 2007). Logistic regression is a linear modeling technique that generates coefficients ($\beta$) for predictor variables that contribute to a classification of binary data (here voiced or voiceless). The predictors of voicing were the VOT and $f_0$

---

[7] Resarch reported in this section was conducted in close collaboration with Kyle Gorman and Daniel Swingley at the Institute for Research in Cognitive Science at the University of Pennsylvania.

characteristics for word-initial plosives in the speech of eight women: four from the Brent corpus of infant-directed speech (Brent and Siskind, 2001) speaking to their infants at 9 months and four from the Buckeye corpus of conversational (adult-directed speech)(Pitt *et al.*, 2007). The speakers from the Buckeye corpus were selected on the basis of their being new mothers or soon-to-be mothers. Forced-phoneme alignment of the audio from the Brent corpus (Quam *et al.*, 2008) allowed us to examine the acoustic characteristics of word-initial consonants and following vowels. The Buckeye corpus includes a phoneme-aligned parse. A trained phonetician (CRN) measured VOTs by hand for all speakers. Five hundred utterances per Brent speaker were examined, and approximately 20 minutes of speech from each Buckeye speaker were examined.

VOT was calculated as the time between the word-initial stop burst, characterized by a brief high-frequency noise, and the onset of periodic laryngeal vibration of the post-stop vowel measured at the first zero crossing. The VOT of prevoiced tokens was calculated as the time between the onset of periodic vibration during stop closure and the release of the stop into the following vowel. In general, the release of the stop was simultaneous with the onset of periodic voicing of the vowel. In keeping with the literature on VOT, prevoiced tokens were assigned a negative value (e.g. Keating *et al.* 1981). In order to control for varying speech rate, which is known to be slower in IDS compared to ADS (Kuhl *et al.*, 1997), VOT was normalized by dividing the raw VOT measurement (ms) by the duration of the following vowel. This ratio has been shown to serve as a perceptual criterion for voicing category affiliation (Boucher, 2002).

Voiced regions inside the post-stop vowel region were extracted and pitch tracks obtained (at 1 ms time steps) using `SWIPE'` (Camacho, 2007). The pitch extraction algorithm required that the voiced region be at least 10ms. Tokens with less than 10ms of post-stop voicing were discarded. The procedure yielded 1200 IDS and 1058 ADS CV tokens. A visual inspection of all the pitch tracks confirmed that there were no obvious halving errors in the extraction. In order to control for individual speakers' pitch ranges, raw $f_0$ measurements were normalized by speaker using the standard $z$ calculation. Following Umeda (1981), peak (or maximum) $f_0$ (in the first half of the post-stop vowel) was computed for analysis.

### 1.3.4 *Results*

Analyses of mean VOTs according to register and voicing were consistent with previous reports—there was a voicing x register interaction suggesting that voiced and voiceless stops in IDS showed more overlap along VOT than in ADS [$F(1,\ 2258)=1552$, $p<0.0001$], that is, the modes of VOT were more separable for voiced and voiceless tokens in ADS relative to IDS.

TABLE 1.2 Register-specific logistic regression models of voicing as a function of VOT and $f_0$. The table shows the significant contribution of the VOT x $f_0$ interaction only in the IDS model of voicing.

| **IDS** | $\beta$ | Std. Error | 95% CI | $z$ | Sig. |
|---|---|---|---|---|---|
| (Intercept) | 0.01 | 0.09 | (-0.16, 0.17) | 0.11 | 0.91 |
| VOT | -6.39 | 0.40 | (-7.21, -5.64) | -15.91 | <0.0001 |
| $f_0$ | -.59 | 0.10 | (-0.79, 0.40) | -5.82 | <0.0001 |
| VOT x $f_0$ | -2.03 | 0.44 | (-2.91, -1.17) | -4.59 | <0.0001 |
| **ADS** | | | | | |
| (Intercept) | 0.35 | 0.11 | (0.15, 0.56) | 3.32 | <0.0005 |
| VOT | -6.82 | 0.47 | (-7.79, -5.95) | -14.58 | <0.0001 |
| $f_0$ | -0.31 | 0.12 | (-0.54, 0.08) | -2.62 | <0.01 |
| VOT x $f_0$ | 0.22 | 0.52 | (-0.86, 1.18) | 0.41 | 0.61 |

There was also a general pitch perturbation effect (with no interaction of register) suggesting that voiced stops were followed by a lower pitch than voiceless stops.[8]

The logistic regression models of IDS and ADS were fit using VOT, $f_0$, and their interaction as predictors of voicing. Table 2 presents regression models of voicing in IDS and ADS. Both registers show a significant main effect of VOT, with a negative slope ($\beta$) indicating that an increase in VOT results in less voiced prediction. $f_0$ is significant in both registers as well, again with a negative slope confirming the pitch perturbation effect.

The interaction between VOT and $f_0$ is significant in only the IDS model. The interactions (plotted in figure 3) suggest that as VOT increases, $f_0$ has a greater effect on voicing prediction. In IDS, $f_0$ becomes more and more predictive of voicing as VOT increases. As a result, where VOT is most ambiguous in the signal (VOT ratio between 0 and 0.5), $f_0$ becomes more useful as an indicator of voicing. No such effect is present in ADS.

[8] Both the VOT and $f_0$ analyses were conducted using 2 (register: IDS, ADS) x 2 (voicing: voiced, voiceless) x 3 (place of articulation: velar, apical, bilabial) ANOVAs. There were considerably more prevoiced tokens in the ADS sample compared to tokens from IDS. We explored the possibility that the interaction between voicing and register was driven by the more negative mean VOT for voiced tokens in the ADS sample. This interpretation was not supported, as the interaction was also significant when prevoiced tokens were removed from the analysis. There was an expected effect of place on VOT, with velars having the longest, followed by alveolars, then bilabials.

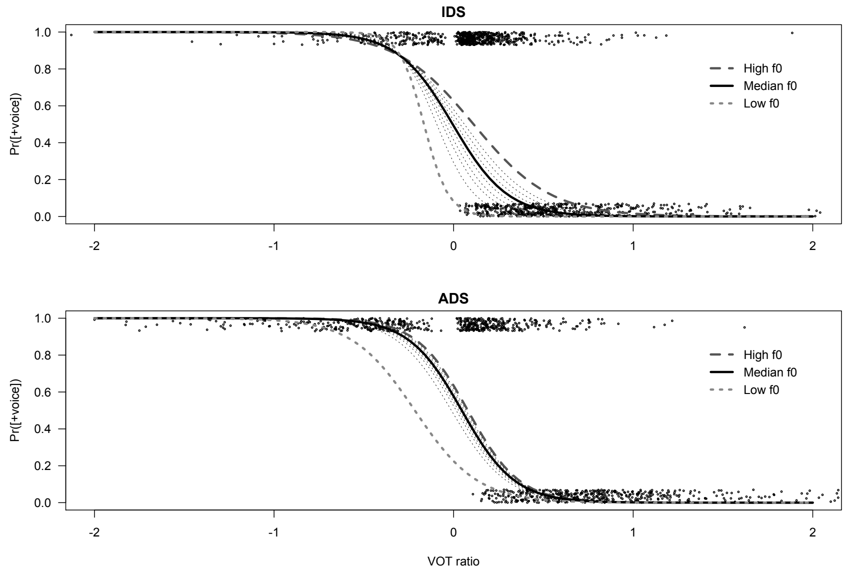

FIG. 1.3 VOT x $f_0$ interaction in voicing prediction in IDS and ADS. The two panels show probability curves overlaid on a jittered VOT x voicing scatterplot. The curves represent quantile values of the distributions of $f_0$. For example, the median curve (solid line) represents the $f_0$ value below which 50% of the data lie in each register. “High” and “Low” $f_0$ represent the highest and lowest $f_0$ values in each corpus.

For example, given a VOT ratio of 0.25 (where there is significant overlap between voiced and voiceless tokens) and an $f_0$ value at the 10% quantile (a relatively low pitch), the probability of the token being voiced is approximately 55% in IDS. Given the same VOT and a median (50% quantile) $f_0$ value, the probability of the token being voiced is approximately 30%. In the ADS model, similar shifts in $f_0$ do not substantially change voicing predictions.[9] Voicing is therefore more consistently implemented with VOT in the ADS corpus, thus minimizing the effect of $f_0$ as a cue to voicing.

---

[9] Voicing implementation was also modeled by considering individual speaker variation using hierarchical logistic regression (Gelman and Hill, 2007; Gorman, 2009). Results similar to the models presented in table 2 were obtained, with 3 out of the 4 IDS speakers showing a significant interaction between VOT and $f_0$. The interaction was not significant for any of the ADS speakers.

### 1.3.5 *Implications*

At nine months, infants in English-speaking environments are being exposed to highly variable VOT, with considerable overlap between categories, promoting the regularity of pitch perturbation to be a reliable cue to voicing. While we do not expect English-learning infants to reinterpret the regularity of pitch perturbation as tone, what this study points up is the acoustic instability of the IDS to young infants, particularly at an age when they are thought to be developing and honing their perceptual sensitivity (Werker and Tees, 1984; Narayan *et al.*, 2010). The models suggest that, at this young age, consistency of the pitch perturbation effect essentially prevents the learner from making incorrect predictions of voicing.

VOT as the primary cue to voicing in English is salvaged, however, for by the time infants produce their first words, VOT as a cue to voicing in IDS resembles the consistency of ADS, thereby precluding infants' reinterpretation of $f_0$ as the primary cue to voicing.

## 1.4 Conclusions

Linguists have long speculated on the potential role for infants and children in historical phonology processes (Baudouin de Courtenay, 1972). For the most part, these speculations have been relegated to the domain of speech production and the nuances of child phonology. As Blevins (2004) has noted, the enterprise of locating phonetically motivated historical processes in child phonology is inherently confounded with inter- and intra-speaker variability associated with the quickly maturing vocal apparatus. That is, language-specific phonological patterns may be obscured in children's productions, by changing physical constraints. Unfettered by these non-cognitive limitations, behavioral studies of speech perception offer a window into the earliest sensitivities of infants, which in turn allow us to assess infant biases and connections to directions of phonological change and typological regularities.

Despite the remarkable cognitive abilities exhibited by young infants, the earliest stages of their speech perception are less than ideal. Infants, rather than being "citizens of the world" are more likely members of the *majority party*, with initial perceptual abilities reflecting acoustically robust, and typologically common, phonetic contrasts. I argued that ubiquitous contrasts in the world's languages reflect, to some degree, the natural perceptual biases infants bring to the language-learning table. Conversely, contrasts which are typologically rare reflect the relative difficulty of their discrimination by young infants. The consistency between patterns in developmental speech perception and phonological typology is consistent with *functionally* based approaches to phonetics-phonology interface such as Lindblom’s dispersion theory (Liljencrants and Lindblom, 1972; Lindblom, 1986; Johnson *et al.*, 1993) which proposes that phonological contrasts are sufficiently distinctive perceptually (in order to be learned and remain stable). While the question of how and why certain perceptually difficult contrasts remain in sound systems cannot be answered within the present research program, we appeal to general learning mechanisms (i.e., statistical learning in terms of frequency of occurrence, cf. Maye *et al.*, 2002) for their persistence.

Directions for future research might include exploring infant biases in the perception of typologically asymmetric distributions of suprasegmental features such as tone. Recent work suggests that infants’ discrimination of certain tone contrasts follows the typical profile of perceptual reorganization (Werker and Tees, 1984). Mattock and Burnham (2006) showed that English-learning infants discriminated Thai rising vs. falling and rising vs. low tones more accurately at 6 months than at 9 months, suggesting that infants’ perceptual sensitivities have reorganized in the direction of privileging native contrasts. Given the connections between the earliest biases in infant speech perception and typological patterns, we might next ask if infants discriminate acoustically similar tones (e.g., tones 22 vs. 33 from Cantonese) as well as tone contrasts with a robust salience (e.g., Cantonese 21 vs. 25) (Khouw and Ciocca, 2007).

Finally, the connection between development, typology and phonologization is not limited to child behavior. This chapter also outlined the type of variability associated with infant-directed speech in English and the similarity between its acoustic characteristics and the phonetic conditions giving rise to tone from the loss of voicing contrasts. While this analysis does not claim to capture a sound change in progress, it provides evidence for treating infant-directed speech as critical input to the developing speech perception system, particularly when emotional affect results in either hyper- and hypo- articulation (Lindblom, 1990), which can potentially be reinterpreted as phonologically different from the intended *linguistic* gesture.

## Appendix: Frequency of Consonants in Brent Corpus

Table 1.3 Frequency of consonants within the Brent corpus of infant-directed speech. Proportions are rounded to two decimal places. Consonants are listed in decreasing order of frequency.

| **Consonant** | **Proportion in Brent** |
|---|---|
| t | 0.13 |
| n | 0.10 |
| ɹ, s | 0.07 |
| k, d, l | 0.06 |
| m, j, ð, w | 0.05 |
| b, g, h | 0.04 |
| z, p | 0.03 |
| f, ŋ | 0.02 |
| v, θ, ʃ, ʧ, ʤ | 0.01 |
| ʒ | <0.01 |

# *Bibliography*

Abramson, A. and Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In *Phonetic Linguistics* (ed. V. A. Fromkin), pp. 25–33. Academic Press, San Diego.

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

Amayreh, M.M. and Dyson, A.T. (1998). The acquisition of Arabic consonants. *Journal of Speech, Language, and Hearing Research*, **41**, 642–653.

Anderson, G.D.S. (2008). The velar nasal. In *The World Atlas of Language Structures Online* (ed. M. Haspelmath, M. Dryer, D. Gil, and B. Comrie). Max Planck Digital Library, Munich.

Andruski, J.E., Kuhl, P.K., and Hayashi, A. (1999). The acoustics of vowel in Japanese women's speech to infants and adults. In *Proceedings of the 14the International Congress of Phonetic Sciences*, Berkeley, pp. 2177–2179. University of California.

Aslin, R.N. and Pisoni, D.B. (1980). Some developmental processes in speech perception. In *Child Phonology: Perception* (ed. G. Yeni-Komishan, J. Kavanaugh, and C. Ferguson), Volume 2, pp. 67–96. Academic Press, New York.

Aslin, R.N., Pisoni, D.B., Hennessy, B.L., and Perey, A.J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development*, **52**, 1135–1145.

Baran, J.A., Zlatin Laufer, M., and Daniloff, R. (1977). Phonological contrastivity in conversation: a comparative study of voice onset time. *Journal of Phonetics*, **5**, 339–350.

Baudouin de Courtenay, J. (1895/1972). *A Baudouin de Courtenay Anthology: The Beginnings of Structural Linguistics.* Indiana University Press.

Behrens, S. and Blumstein, S. (1988). On the role of amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *Journal of the Acoustic Society of America*, **84**(3), 861–867.

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother-child speech. *Journal of Child Language*, **11**, 557–578.

Boucher, V.J. (2002). Timing realtions in speech and identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception & Psychophysics*, **64**(1), 121–130.

Brent, Michael and Siskind, J. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, **81**, B33–B44.

Camacho, Arturo (2007). *SWIPE': A sawtooth waveform inspired pitch estimator for speech and music.* Ph.D. thesis, University of Florida.

Cooper, R.P. and Aslin, R.N. (1989). The language environment of the young infant: implications for early perceptual development. *Canadian Journal of Psychology*, **43**, 247–265.

Dalston, R.M. (1975). Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America*, **57**(2), 462–469.

de Boer, B. (2001). *The Origins of Vowel Systems.* Oxford University Press.

Demuth, K. and Johnson, M. (2003). Truncation to subminimal words in early French. *Canadian Journal of Linguistics*, **48**(3/4), 211–241.

Eilers, R.E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. *Journal of the Acoustic Society of America*, **61**(5), 1321–1336.

Eilers, R.E. and 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**, 313–336.

Eilers, R.E. and Minifie, F. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research*, **18**, 158–167.

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

Eilers, R.E., Wilson, W.R., and Moore, J.M. (1979). Speech discrimination in the language-innocent and the language-wise: a study in the perception of voice onset time. *Journal of Child Language*, **6**, 1–18.

Eimas, P.D. (1974). Linguistic processing of speech by young infants. In *Language perspectives: acquisition, retardation and intervention* (ed. R. Schiefelbusch and L. Lloyd). University Park Press.

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

Englund, K. (2005). Voice onset time in infant directed speech over the first six months. *First Language*, **25**(2), 219–234.

Espy-Wilson, C. (1992). Acoustic measures for linguistic features distinguishing the semivowels /wjrl/ in american english. *Journal of the Acoustic Society of America*, **92**(2), 736–751.

Fant, G. (1960). *Acoustical Theory of Speech Production.* Mouton, 's-Gravenhage.

Ferguson, C.A. (1973). Fricatives in child language acquisition. *Papers and Reports on Child Language Development* (6), 61–85.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In *The Adapted Mind* (ed. J. Barkwo, L. Cosmides, and J. Tooby), pp. 391–428. Oxford University Press, New York.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., and Fukui, I. (1989). A cross-language study of prosodic modifications in mothers’ and fathers’ speech to preverbal infants. *Journal of Child Language*, **16**(3).

Foulkes, P., Docherty, G., and Watt, D. (2005). Phonological variation in child-directed speech. *Language*, **81**(1), 177–206.

Gandour, J., Petty, S.H., Dardarananda, R., Dechongkit, S., and 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**, 561–572.

Garnica, O.K. (1977). Some prosodic and paralinguistic features of speech to young children. In *Talking to Children* (ed. C. Snow and C. Ferguson), pp. 63–88. Cambridge University Press, Cambridge.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York.

Gorman, K. (2009, March). Hierarchical regression modeling for language research. Technical report, Institute for Research in Cognitive Science, University of Pennsylvania.

Goto, H. (1971). Auditory perception by normal japanese adults of the sounds “l” and “r”. *Neuropsychologia*, **9**, 317–323.

Grammont, M. (1933). *Traité de phonétique*. Delagrave, Paris.

Greenlee, M. and Ohala, J.J. (1980). Phonetically motivated paralles between child phonology and historical sound change. *Language Sciences*, **2**(2), 283–308.

Hale, M. and Reiss, C. (1998). Formal and empirical arguments concerning phonological aquisition. *Linguistic Inquiry*, **29**, 656–683.

Hedrick, M. and Ohde, R.N. (1993). Effect of relative amplitude of frication on perception of place of articulation. *Journal of the Acoustic Society of America*, **94**(4), 2006–2026.

Herzog, E. (1904). *Streitfragen der romanischen Philologie*. Halle M. Niemeyer.

Holmberg, T.L., Morgan, K.A., and Kuhl, P.K. (1977, December). Speech perception in early infancy: Discrimination of fricative consonants. Presented at the 94th Meeting of the Acoustical Society of America.

Hombert, J-M., Ohala, J.J., and Ewan, W.G. (1979). Phonetic explanations for the development of tones. *Language*, **55**, 37–58.

Hosmer, D.W. and Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley and Sons, New York.

Hua, Z. and Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, **27**(1), 3–42.

Huttenlocher, P.R. (2002). *Neural Plasticity:The Effects of Environment on the Development of the Cerebral Cortex*. Harvard University Press.

Jakobson, R. (1968). *Child Language Aphasia and Phonological Universals*. Walter de Gruyter.

Jakobson, R. and Waugh, L. (1979). *The Sound Shape of Language*. Indiana University Press, Bloomington.

Johnson, K. (1997). Speech perception without speaker normalization: an exemplar model. In *Talker Variability in Speech Processing* (ed. K. Johnson and J. Mullenix), pp. 145–166. Academic Press.

Johnson, K., Flemming, E., and Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, **69**, 505–528.

Jongman, A. (1988). Duration of frication noise required for identification of English fricatives. *Journal of the Acoustic Society of America*, **85**(4), 1718–1725.

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

Keating, P.A. (1984). Phonetic and phonological respresentations of stop consonant voicing. *Language*, **60**(2), 286–319.

Keating, P.A., Linker, W., and Huffman, M. (1983). Patterns of allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, **11**, 277–290.

Keating, P.A., Mikoś, M.J., and Ganong III, W.F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustic Society of America*, **70**(5), 1261–1271.

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

Kiparsky, P. (1988). Phonological change. In *Linguistics: The Cambridge Survey* (ed. F. Newmeyer), Volume 1: Theoretical foundations, pp. 363–415. Cambridge University Press, Cambridge.

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

Kuhl, P.K., Stevens, E., Hayashi, A., Degushi, T., Kiritani, S., and Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, **9**, F13–F21.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., and Lindblom, B. (1992, January). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255**(5044), 606–608.

Labov, W. (1989). The child as lingusitic historian. *Language Variation and Change*, **1**, 85–97.

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

Levelt, C., Schiller, N., and Levelt, W. (1999). The acquisition of syllable types. *Language Acquisition*, **8**, 237–264.

Levitt, A.G., Jusczyk, P.W., Murray, J., and 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), 361–368.

Liljencrants, J. and Lindblom, B. (1972). Numerical simulation of vowel quality: The role of perceptual contrast. *Language*, **48**(4), 839–862.

Lindblom, B. (1986). *Phonetic universals in vowel systems.* Academic Press, Orlando.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In *Speech Production and Speech Modelling* (ed. W. Hardcastle and A. Marchal), pp. 403–439. Kluwer Academic, The Netherlands.

Lindblom, B. and Maddieson, I. (1988). Phonetic universals in consonant systems. In *Language, Speech and Mind* (ed. L. Hyman and C. Li). Routledge.

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

Liu, H.M., Tsao, F.M., and Kuhl, P.K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, **43**(4), 912–917.

Macken, M.A. (1980). The child's lexical representation: The 'puzzle-puddle-pickle' evidence. *Journal of Linguistics*, **16**, 1–17.

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

Maddieson, I. (2008). Presence of uncommon consonants. In *The World Atlas of Language Structures Online* (ed. M. Haspelmath, M. Dryer, D. Gil, and B. Comrie), Chapter 19. Max Planck Digital Library, Munich.

Malsheen, B.J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In *Child Phonology* (ed. G. Yeni-Komishan, Kavanaugh, and C. Ferguson), Volume 2: Perception. Academic Press, San Diego.

Martinet, A. (1933). Remarques sur le Systême Phonologique du Francais. *Bulletin de la Société de Linguistique de Paris*, **34**, 191–202.

Matisoff, J. (1973). Tonogenesis in Southeast Asia. In *Cosonant types and tone* (ed. L. Hyman), pp. 71–95. University of Southern California.

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

Maye, J., Weiss, D.J., and Aslin, R.N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, **11**(1), 122–134.

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

Miller, G.A. and Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustic Society of America*, **27**(2), 338–352.

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

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

Narayan, C.R., Werker, J.F., and 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.

Ohala, J.J. (1981). The listener as a source of sound change. In *Papers from the Parasession on Language and Behavior* (ed. C. Masek, R. Hendrik, and M. Miller), pp. 178–203. Chicago Linguistic Society.

Ohala, J.J. (1983). The origin of sound patterns in vocal tract constraints. In *The Production of Speech* (ed. MacNeilage, P.F.), pp. 189–216. Springer.

Oudyer, P-Y. (2006). *Self-Organization in the Evolution of Speech.* Studies in the Evolution of Language. Oxford University Press.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In *Frequency Effects and the Emergence of Linguistic Structure* (ed. J. Bybee and P. Hopper), pp. 137–157. John Benjamins.

Pisoni, D.B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustic Society of America*, **61**(5), 1352–1362.

Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech, (2nd release).

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

Polka, L. and Strange, W. (1985). Perceptual equivalence of acoustic cues that differentiate /r/and /l/. *Journal of the Acoustic Society of America*, **78**(4), 1187–1197.

Polka, L. and Werker, J.F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 421–435.

Quam, C., Yuan, J., and Swingley, D. (2008). Relating intonational pragmatics to the pitch realizations of highly frequent words in English speech to infants. In *CogSci2008*.

Rose, Y. (2009). Internal and external influences on child language productions. In *Approaches to Phonological Complexity* (ed. I. Chitoran, F. Pellegrino, and E. Marsico), pp. 329–351. Mouton de Gruyter.

Sachs, J. (1977). The adaptive significance of linguistic input to prelinguistic children. In *Talking to Children* (ed. C. Snow and C. Ferguson). Cambridge University Press.

Smith, C.L. (1997). The devoicing of /z/ in American English: effects of local and prosodic context . *Journal of Phonetics*, **25**(4), 471–500.

Stampe, D. (1972). *A Dissertation on Natural Phonology.* Ph.D. thesis, University of Chicago.

Streeter, L. (1976). Language perception of two-month-old infants show effects of both innate mechanisms and experience. *Nature*, **259**, 39–41.

Sundberg, U. and Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica*, **56**, 186–199.

Sweet, H. (1913). *Collected papers of Henry Sweet.* Clarendon Press, Oxford.

Tang, J.S-Y. and Maidment, J.A. (1996). Prosodic aspects of Cantonese child-directed speech. *Speech, Hearing and Language*, **9**, 257–276.

Templin, M.C. (1957). *Certain language skills in children: Their development and interrelationships.* Greenwood, Westport, Conn.

Thiessen, E., Hill, E., and Saffran, J. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, **7**(1), 53–71.

Trehub, S.E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, **47**(2), 466–472.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., and Best, C. (1994). Discrimination of english /r-l/ and /w-y/ by japanese infants at 6-12 months: Language-specific developmental changes in speech perception abilities. In *Proceedings of the International Conference on Spoken Language Processing*, Volume 4, pp. 1695–1698. Acoustical Society of Japan.

Umeda, N. (1981). Influence of segmental factors on fundamental frequency in fluent speech. *Journal of the Acoustical Society of America*, **70**(2), 350–355.

Werker, J.F. and McLeod, P.J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responses. *Canadian Journal of Psychology*, **43**(2), 230–246.

Werker, J.F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (2007). Infant-directed speech supports phonetic category learning in english and japanese. *Cognition*, **103**, 147–162.

Werker, J.F., Shi, R., Desjardins, R., Pegg, J.E., Polka, L., and Patterson, M. (1998). Three methods for testing infant speech perception. In *Perceptual development: Visual, auditory, and speech perception in infancy* (ed. A. Slater), pp. 389–420. UCL Press.

Werker, J.F. and Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, **7**, 49–63.

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