11 Acoustic Characteristics of Infant-directed Speech as a Function of Prosodic Typology

Yuanyuan Wang, Amanda Seidl, and Alejandrina Cristia

11.1 Introduction

11.1.1 IDS as a Unique Register

When interacting with infants, caregivers, regardless of sex, age, or social status, modify the prosodic and phonetic features of their speech by using infant-directed speech (IDS; e.g., Fernald 1993, Papoušek & Papoušek 1981, Snow 1977, Soderstrom 2007). This is a speech style that infants prefer to adult-directed speech (ADS; Cooper & Aslin 1990, Fernald 1985). During the past forty years, a rather extensive body of research has investigated the acoustic differences between IDS and ADS in various languages. Early results suggest that both suprasegmental and segmental acoustic cues differ between the two registers cross-linguistically (e.g. Bernstein Ratner & Luberoff 1984, Fernald et al. 1989, McMurray, Kovack-Lesh, Goodwin & McEchron 2013, Soderstrom 2007).

The aforementioned and other studies have identified distinctive characteristics of IDS, such as slower tempo, higher pitch, wider pitch range, shorter utterances, and larger vowel spaces. Indeed, in a recent systematic review of this literature, 85 percent of studies reported slower tempo and/or longer vowel duration; 92 percent found higher pitch (typically measured through fundamental frequency (F0); we will use pitch and F0 interchangeably); and 82 percent documented larger vowel spaces (Cristia 2013). However, these studies have not explored whether IDS and ADS differ to the same extent or along the same acoustic dimensions in typologically diverse languages; nor have they examined the related question of whether the contrasts along the relevant prosodic dimensions (stress, tone, and pitch accent) are maintained in IDS. These questions are important since they may inform us about why IDS in general displays exaggerated acoustic characteristics as compared to ADS in the first place. They may also explain the similarities and differences that we see in various languages. A further benefit of this research is that it provides a much-needed insight into our understanding of the input to infants. Thus, it is worthwhile to foster this line of research by providing a review of IDS acoustic...
characteristics in languages of different typologies, such as stress languages, tone languages, and pitch-accented languages.

In addressing these research questions, it will be borne in mind that IDS-ADS differences may also vary as a function of infant development. For example, Bernstein Ratner and Luberoff (1984) report different degrees of vowel space expansion depending on whether the child is babbling, at the one-word, or at the two-word stage. In addition, the way in which the speech corpus has been elicited also explains differences in acoustic divergence between the registers. IDS-ADS divergence depends for instance on whether the speech is conversational or read (Shute & Wheldall 1995), or whether the adult addressee is a family member or an experimenter (Johnson, Lahey, Ernestus & Cutler 2013).

While extralinguistic factors such as age of the infant and elicitation procedure may impact the quality of IDS, we might expect that the language’s phonology to have an equally strong influence on the quality of IDS. However, many studies point to a large degree of similarity in IDS across languages. Indeed, Kuhl et al. (1997) report similar degrees of vowel space expansion in three prosodically distinct languages (stress-timed English and Russian, pitch-accented Swedish). Similarly, Fernald et al. (1989) document heightened pitch in six languages as diverse as English and Japanese with little discussion of how these key typological differences might impact the implementation of IDS. While this chapter aims to examine prosodic differences, it also takes into account other aspects of the phonology of a language and/or cultural differences that may explain some cross-linguistic divergences in the acoustic profile of IDS in typologically distinct languages.

Before moving on to the actual review, it should be emphasized that there has been substantial debate as to whether IDS-ADS differences are functional (i.e. whether they promote linguistic and cognitive development), and whether they are linguistically relevant (i.e. whether caregivers specifically enhance linguistic contrasts, see a recent discussion in Cristia 2013). In the present chapter, we will remain agnostic on this matter. Instead our goal is simply to evaluate whether acoustic differences across the registers occur along the same dimensions, and to a similar extent in languages with different prosodic profiles, and to what extent are prosodic contrasts similarly cued across registers.

11.1.2 Research Questions

This review seeks to provide a systematic, rather than an exhaustive overview of the literature on IDS-ADS along both suprasegmental and segmental dimensions in order to address the following two questions:

1. Do the two registers differ qualitatively and quantitatively in the same way in languages with distinct lexical prosodic systems?
Are the relevant lexical prosodic contrasts (stressed vs. unstressed, toned vs. un-toned, pitch-accented vs. unaccented) similarly cued in IDS and ADS?

In order to address these two questions, we review previous work comparing IDS to ADS in three prosodically distinct languages: stress languages (Section 11.2 IDS in Stress Languages), tone languages (Section 11.2: IDS in Tone Languages), and pitch-accented languages (Section 11.3: IDS in Pitch-accented Languages). Finally, the implications of the cross-linguistic similarities and differences observed in the review are discussed and areas for future study are considered (Section 11.5 Implications and Future Directions).

11.2 IDS in Stress Languages

11.2.1 Definition of Stress and Stress Languages

According to a classic definition, stress is the linguistic manifestation of rhythmic alternation of strong and weak beats (Hayes 1995). In stress languages, one or more syllables in each word or phrase are relatively “more prominent” than others. That is, stressed syllables appear at a constant rate, while the vowels in unstressed syllables undergo shortening or reduction (Delattre 1966) in order to accommodate this rhythmic requirement. The stress placed on words is lexical stress. Cross-linguistically, stressed syllables have higher pitch, longer vowel duration, higher amplitude, and non-reduced vowel quality (Fry 1958; Lehiste 1970). Any of these properties may contribute to the perception of stress, although some of these properties may be more dominant; for example, in a seminal study conducted by Fry (1958), F0 was found to be the most important cue for stress perception in English adult listeners. However, Fry’s study investigated words only in the focal position of an utterance, a position where phrase-level prominence usually occurs. Thus, results from this study concerning the salience of pitch characteristics can only be assumed to be associated with phrasal-level prominence (see Gordon, 2014 for relevant discussion). Subsequent work showed that most languages rely on a combination of cues, such as duration, intensity, and/or F0, for the perception of stress (e.g. Gordon 2004; Gordon & Applebaum 2010). Stress may contribute to faster processing of speech and better recognition of linguistic elements (Cutler & Foss 1977; Grosjean & Gee 1987), although people from different language backgrounds may rely on different cues to facilitate speech processing (e.g. Boll-Avetisyan & Kager 2014). In addition, it is found that stress may facilitate English-learning infants’ word segmentation (Cutler & Norris 1988; Jusczyk, Houston & Newsome 1999). The next sections examine the acoustic characteristics of IDS in stress languages.
11.2.2 The Acoustic Properties of IDS in Stress Languages

Although a substantial body of research has examined the acoustic profiles of IDS in stress languages, most of them focus on utterance length, speech rate, as well as focus (See a review by Cristia 2013). Given the narrow scope of the present review, we focus only on describing the two most relevant studies to stress.

Albin and Echols (1996) conducted an experimental study to compare the word-level acoustic properties between IDS and ADS. In this study, speech samples were collected in a laboratory setting from speech to eight 6-month-old and eight 9-month-old infants elicited from caregivers by using 7 target objects/words. The 7 target words included different word length and stress patterns, for example, monosyllabic (e.g. fish), final-stressed (e.g. kangaroo), and final-unstressed words (e.g. bracelet, elephant, alligator, and hippo). Average duration, pitch peak, and peak amplitude were measured for whole words and for each of the target syllables (the final syllables). The results showed that for monosyllabic words (fish), the duration in IDS was significantly greater than in ADS, while pitch and amplitude were not significantly different across registers. For final-stressed syllables (as in kangaroo), all the three measures (duration, pitch, and amplitude) were significantly greater in IDS than that in ADS. Finally, for final-unstressed syllables (as in bracelet, elephant, alligator, and hippo), both duration and amplitude were found to be significantly different in IDS than in ADS. In sum, while duration, amplitude, and pitch of final-stressed syllables were all greater in IDS compared to ADS, only the measures of duration and amplitude were greater in IDS than in ADS in final-unstressed syllables.

Wang, Cristia, and Seidl (2015) extended these findings in three ways. First, they collected IDS and ADS from two groups of participants at two points in development. One group included 10 mother-infant dyads (in which the infants were around 4 months of age), and the other 10 mother-infant dyads (in which the infants were around 11 months of age). Second, they examined the target words in two different positions, utterance-medial and utterance-final. Twenty-three different pairs of objects, each representing a bisyllabic trochaic word (e.g. teddy, bacon, basil, etc.), were used in the elicitation. Finally, in addition to the duration and pitch measures already breached in Albin and Echols (1996)’s work, they investigated vowel peripherality, which was calculated as the average distance between each vowel token and the center of the vowel space (the latter defined as the first and second formant values (F1 and F2) averaged across /i, a, u/). Larger vowel peripherality may reflect better contrast in the articulatory space and higher intelligibility (e.g. Bradlow, Torretta, & Pisoni 1996).

They compared the average duration, average F0, and vowel peripherality in vowels across the two registers (IDS and ADS), in both stressed and
unstressed syllables, and in both utterance-final and utterance-medial positions. The results showed that vowels in both stressed and unstressed syllables exhibited higher F0 and wider vowel peripherality in IDS as compared to those in ADS. This effect was stronger in stressed syllables than it was in unstressed syllables. However, no register difference was found for measures of duration. In addition, analyses revealed that the effect of register on duration, pitch, and vowel peripherality was not dependent on utterance position or infant age. Thus, caregivers produced higher-pitched, and more extreme vowels in both stressed and unstressed syllables in IDS; and this exaggerated acoustic characteristic was found in both utterance-medial and utterance-final positions, in the speech to both 4-month-old and 11-month-old infants. This null finding for the utterance position is somewhat unexpected given that prior work has sometimes suggested that durational differences in utterance-final position should be more prominent than those in utterance-medial position (Church 2002; Kondaurova & Bergeson 2011). The different patterns may be due to the nature of the pragmatic situation in which the target words were elicited. The failure to find effects of age in this study does not necessarily provide counter-evidence to prior studies in which the age effects were found. For example, IDS-related changes as a function of age could be primarily explained by the caregivers’ awareness of their child’s phonological and lexical development. Thus, it is possible that the caregivers of both 4- and 11-month-olds in Wang, Cristia, and Seidl (2015)’s study did not differ in the perception of their child’s language development.

Taken together, these two studies provide evidence that the acoustic characteristics of IDS, such as amplitude, pitch peak, average pitch, and vowel peripherality are exaggerated in stress languages; however, the picture for average duration is still unclear given these two studies yielded different patterns of results. This may be due to different ages of the addressees in each study, but may also be due to dialect or language change (since the studies were conducted in different time periods). Finally, different statistical analyses may also serve to explain variance in results. Specifically, Wang, Cristia, and Seidl (2015) adopted a mixed model analysis which controlled for vowel quality together with three more traditional statistical methods; while Albin and Echols (1996) did not control for vowel quality.

11.2.3 Cues to Stress Contrasts in IDS Compared to ADS

This section addresses whether the contrast between stressed and unstressed syllables is as large in IDS as it is in ADS, and whether it employs the same acoustic cues in both registers. Unfortunately, a direct answer to this question is still missing from the literature. Indeed, neither of the studies reviewed earlier (or any others) has controlled for vowel quality between stressed and unstressed
syllables, which may add difficulty to the interpretation of the analyses. Future experimental studies with vowel-controlled designs are needed in order to provide better comparisons of cues to lexical stress across registers.

Furthermore, behavioral studies assessing to what extent stress-register interactions along relevant acoustic dimensions may affect infant speech processing are also needed. Although psychological evidence suggests that adults rely heavily on pitch to judge stress (Fry 1958), and that infants’ early preference for IDS is guided by pitch rather than amplitude or duration (Fernald & Kuhl 1987), the acoustic cues on which infants rely to distinguish stressed syllables from unstressed ones are still unknown.

11.3 IDS in Tone Languages

11.3.1 Definition of Lexical Tone and Tone Languages

Tone is the use of pitch to distinguish lexical or grammatical meaning. A language is a tone language if the pitch contours of syllables differentiate words (Yip 2002), i.e. a tone language uses distinctive F0 patterns to cue lexical contrasts. For example, Mandarin Chinese is often described as having four lexical tones: a high-level tone (Tone 1, H), a mid-rising tone (Tone 2, LH), a low-dipping tone (Tone 3, L), and a high-falling tone (Tone 4, HL). H refers to high tone, and L refers to low tone. Each unreduced syllable in Mandarin carries one of these tones, and the same sequences of phonemic segments can be differentiated by the tones they carry: /ti/ (LH), “carry,” /ti/ (HL), “replace.” There is general agreement that the primary acoustic correlate of tone is the F0 pattern on the tone-bearing units, with pitch height and its temporal contour contributing to adult listeners’ perception of stress in a number of languages (Khouw & Ciocca 2007; Kuo, Rosen & Faulkner 2008; Lin & Repp 1989). It should be noted that we use the term “tone languages” to refer to languages with tone; however, we do not assume that languages with tone do not also have stress.

11.3.2 The Acoustic Properties of IDS in Tone Languages in Comparison with Non-tone Languages

In this section we provide a review of literature on acoustic differences between IDS and ADS in tone languages a comparison with non-tonal languages. Some findings suggest that IDS differs from ADS along similar dimensions in stress and tone languages, including F0 and F0 range (Burnham, Kitamura & Vollmer-Conna 2002; Grieser & Kuhl 1988). For example, Grieser and Kuhl (1988) analyzed the prosodic features of eight mothers addressing their 2-month-old Mandarin-learning infants, and their speech to an adult. They
found higher F0 and a wider F0 range (calculated over the entire sample) in IDS as opposed to in ADS. Furthermore, an experimental study conducted by Liu, Tsao, and Kuhl (2007) provides evidence that IDS exhibits higher pitch and wider pitch range in the four Mandarin tones in IDS compared to ADS.

Although the modification of pitch height and pitch range seems to be present in tone languages as well as in non-tone languages, it has been suggested that the degree of these modifications varies across languages. For example, Kitamura, Thanavishuth, Burnham, and Luksaneeyanawin (2002) investigated the prosodic characteristics of IDS in Thai and Australian English. While both groups of mothers used higher average F0 in IDS, the IDS pitch was higher for Australian mothers than for Thai mothers. This is consistent with findings showing less exaggerated pitch features of IDS in Mandarin than in American English (Papoušek, Papoušek & Symmes 1991). Aside from differences in the degree of modification across languages with different prosodic profiles, variations have also been observed within tonal languages. For example, Mandarin-speaking mothers have been reported to use higher pitch in IDS than Thai-speaking mothers (Khaoonoo 1996). In fact, such differences are not unheard of in non-tone languages, with American mothers being renowned for their marked pitch ranges (Fernald et al. 1989). Thus, it is not clear whether such differences are cultural or may be attributed to more subtle typological differences related to the prosody of the language in which the IDS is manifest, as we will discuss in the final section.

Similar to some of the work cited above for stress languages, IDS-ADS differences are modulated by infant age in tone languages. For example, Liu, Tsao, and Kuhl (2009) recorded mothers addressing an adult and their children at two different stages of development: once at 7–12 months and again at 5 years of age. In this study, mothers raised their pitch and expanded their pitch range in both IDS and child-directed speech (CDS) compared to ADS, but to a greater extent in speech to the infants in the younger age group.

In sum, this literature suggests that pitch is modified in IDS in tone languages, although maybe to a lesser degree than in stress languages. This could be related to the fact that caregivers of tone languages unconsciously restrict the use of pitch so as not to disrupt the integrity of the lexical tones in the language (Luksaneeyanawin 1998). However, pitch range can still be expanded by increasing the polarity between high tones and low tones in tonal languages (note that this occurs in Mandarin to mark modality and focus). The next section focuses on the implementation of the contrast among lexical tones in each register.

Before moving on, it should be mentioned that most work seems to suggest that temporal and segmental differences between IDS and ADS are also present in tone languages, with the occasional modulation for age of addressee. For example, the duration of vowels in IDS spoken to the youngest group were
significantly greater than that of CDS vowels to 5-year-olds, which were, in turn, longer than ADS ones (Liu, Tsao & Kuhl 2009). Other features may exhibit cross-linguistic variation. Indeed, Mandarin IDS to infants between 6 and 12 months, and CDS to 5-year-olds, exhibited overall larger vowel spaces as compared to ADS (Liu, Kuhl & Tsao 2003), with no difference between the age groups (Liu, Tsao & Kuhl 2009). However, no difference in vowel spaces was observed in Cantonese IDS versus ADS in a longitudinal study by Xu Rattanasone, Burnham, and Reilly (2013), which will be discussed in detail subsequently. This diversity of results is also found in non-tone languages. Although English and Dutch are prosodically similar, IDS in these two languages exhibits distinct properties in terms of vowel space, with a significant expansion in English (e.g. Kuhl et al. 1997), and a significant reduction in Dutch (Benders 2013). Contradictory results could be caused by several factors, which we will discuss in Section 11.5: Implications and Future Directions.

11.3.3 Cues to Tonal Contrasts in IDS Compared to ADS

The modification of lexical tones was primarily explored in studies conducted by Liu and colleagues (Liu, Kuhl & Tsao 2003; Liu, Tsao & Kuhl 2007, 2009). They recorded 32 Mandarin-speaking mothers (16 with infants aged 6–8 months and 16 with infants aged 10–12 months) describing toys and pictures to either their infants or an adult experimenter. The target words were twenty-one Mandarin bisyllabic words with a (C)V.CV(C) structure. All the four tones were represented in the first syllable, whereas the second syllable always had a high level tone (Tone 1). Mean pitch, pitch range, lexical tone duration, and turning point were measured in the first syllable of each target word. Results suggested that tone-bearing syllables were overall longer and the pitch range was overall larger in IDS than in ADS. Additionally, the turning point, which distinguishes two tones with similar pitch contours, was similar in IDS to infants between 6 and 12 months, CDS to 5-year-olds, and ADS (Liu, Tsao & Kuhl 2009).

Other work suggests that tones are sometimes hyperarticulated in IDS, but perhaps only at certain ages. Xu Rattanasone, Burnham, and Reilly (2013) investigated IDS in Cantonese, which has an inventory of six tones (high, mid, and low level tones; as well as high-rising, mid-rising, and falling). Among these six tones, Barry and Blamey (2004) have identified the high level, high-rising, and falling tones as the three most disperse categories in terms of their pitch onset and offset level, thus defining a “tonal triangle”. Xu Rattanasone, Burnham, and Reilly (2013) measured tonal triangles in ADS and IDS collected from twenty-two native Cantonese-speaking mother-infant dyads at four different developmental stages: 3 months, 6 months, 9 months, and 12 months. Careful comparisons revealed that tonal triangles were larger in IDS than in ADS at 3, 6, and 9 months, but not at 12 months. We will return to possible
reasons behind the presence or absence of tone register-dependent modification across languages and ages after discussing pitch-accented languages.

11.4 IDS in Pitch-accented Languages

11.4.1 Definition of pitch accent and Pitch-accented Languages

The definition of pitch accent has long been vague and controversial (e.g. Hyman 2009; Hyman 2014; Van der Hulst 2014). Traditionally, pitch-accented languages are defined as those which use pitch as the primary correlate of prominence. In these languages, the pitch patterns for words are often restricted (e.g. Bybee, Chakraborti, Jung & Scheibman 1998; Ding 2006). However, Hyman (2009) has argued that there is no pitch accent prototype. “The alleged ‘pitch-accent’ systems freely pick-and-choose properties from the tone and stress prototypes, producing mixed, ambiguous, and sometimes analytically intermediate systems which appear to be ‘intermediate’” (Hyman 2009:213). Thus, Hyman has proposed that the pitch accent system is actually a sub-type of a tonal system. Given these controversies, this review does not attempt to give an exact definition for pitch-accented languages; instead, we describe the languages exhibiting related properties of pitch accent, which do not fall into the category of either a strict stress or tone language. The most cited example of a pitch-accented language is Japanese. In Japanese, the location of the pitch accent distinctively characterizes words. For example, when the high pitch is on the first mora, the word /e-ga/ (HL) means “picture-NOM”; in contrast, when the high pitch is located on the second mora, /e-ga/ (LH) means “handle-NOM.” In addition, Norwegian falls into this category as it has two distinct pitch patterns on stressed syllables, Tone 1 and Tone 2, which normally occur on the first syllable of a word in the native vocabulary to distinguish lexical meaning (Kristoffersen 2007).

11.4.2 The Acoustic Properties of IDS in Pitch-accented Languages in Comparison with Non-pitch-accented Languages

The main focus of this section will be on studies using more global measures, typically gathered from prominent syllables, with the goal of comparing pitch-accented languages in terms of the extent to and the manner in which IDS and ADS differ.

In terms of pitch height, most work documents higher pitch in IDS than ADS in Japanese, similar to non-pitch-accented languages (Fernald et al. 1989). Vowel duration is in general found to be greater in IDS than ADS, as exemplified in the Norwegian data gathered by Englund and Behne (2006). This dataset will be discussed in more detail since it exemplifies an area of divergence in pitch-accented languages.
Englund and Behne (2006) collected spontaneous speech from 6 Norwegian mothers producing both IDS and ADS between birth and 6 months of age. Each mother was recorded in ten sessions during diaper changes at home. Acoustic measurements included F1, F2, and duration for vowels, mainly in stressed syllables (94 percent in primary stressed syllables, and 6 percent in secondary stressed syllables). Six vowels were studied, including 3 long vowels /a:, i:, u:/, and their 3 short counterparts /a, i, u/. Only vowels in focal sentence position were included. A smaller and more front vowel space was observed in IDS. Using a small and diverse corpus with various infant age, Dodane and Al-Tamimi (2007) documented that Japanese may also have a slightly reduced and downward shifted vowel space. However, other languages with pitch-accented properties can exhibit vowel space expansion in IDS (Swedish, Kuhl et al. 1997). While most work documents vowel space expansion regardless of prosodic typology (e.g. Kuhl, et al. 1997; Liu, Kuhl & Tsao 2003), shifts and/or reduction of vowel space have also sometimes been observed in both stress (Dutch, Benders 2013) and tone languages (Cantonese, Xu Rattanasone, Burnham & Reilly 2013).

An additional dimension that has been discussed previously is pitch range, which early evidence suggested to be similar across registers (or possibly reduced) in Japanese (Fernald et al. 1989). Recent evidence suggests that such global analyses, which ignore language-specific phonological patterns, may be particularly damaging for the description of pitch-accented languages. For example, Igarashi (2014) analyzed IDS and ADS in the RIKEN Japanese Mother-Infant Conversation Corpus. This corpus included both samples collected during interactions of twenty-two mothers with their 18- to 24-month-old infants by showing them picture books, as well as free conversation with an experimenter. The authors first replicated Fernald et al.’s (1989) results when pitch range was calculated over whole sentences. In addition, they demonstrated that this analysis was inappropriate because in Japanese there is a process of downstep, which affects the range of each pitch accent by the presence of preceding words in the same sentence. As a result, ADS seemed to have an expanded range because sentences were longer and contained more pitch-accented (and therefore more downstepped) syllables. Once this difference was controlled for, larger pitch ranges were observed within the appropriate domains (tone boundary units) in IDS as compared with ADS.

11.4.3 The Acoustic Properties of Lexical Pitch in IDS Compared to ADS

The only evidence directly concerning the realization of pitch accent across registers is from Mazuka, Igarashi, and Nishikawa (2006), who analyzed the same RIKEN corpus. It should be noted that pitch accent is cued by a pitch rise,
followed by an abrupt drop. Mazuka, Igarashi, and Nishikawa (2006) reported that pitch-accented units had a higher maximum pitch, minimum pitch, and mean pitch, as well as a wider pitch range in IDS than ADS. Additionally, the temporal structure of pitch accent may be modified in IDS. Indeed, the abrupt pitch fall for ADS does not have to occur within the pitch-accented unit, but can occur later, sometimes in the following syllable. While this delayed fall is not altered by speech rate or speaking style in ADS (Ishihara 2006), the peak pitch and the whole contour were shifted rightward in IDS as compared with ADS. Additional work with Japanese is necessary to understand this shift. Moreover, it would be particularly interesting to extend this analysis to other pitch-accented languages.

11.5 Implications and Future Directions

This chapter has reviewed acoustic properties of IDS to answer two key questions. First, do IDS and ADS differ along the same acoustic dimensions, and to a similar extent, in typologically different languages? The answer depends on the actual dimension under consideration. IDS and ADS differ in pitch height and average syllable duration in many stress, tone, and pitch-accented languages; moreover, vowel space expansion has often, but not always, been observed. There are, nonetheless, some salient cross-linguistic differences that are equally informative.

On the one hand, some cross-linguistic differences underline the usefulness of a phonologically sound approach, which is illustrated with two examples discussed earlier. The first one relates to the reanalysis of pitch range differences between stress-timed English and pitch-accented Japanese (Fernald 1989; Igarashi 2014). The other one concerns pitch height in English, Mandarin, and Cantonese (Albin & Echols 1996; Liu, Tsao & Kuhl 2007; Wang, Cristia & Seidl 2015; Xu Rattanasone, Burnham & Reilly 2013). Indeed, it was argued that pitch height is modified to the greatest extent in English, which does not have lexical tone, less in Mandarin, which has a four-tone system, and Cantonese, which has a six-tone system.

On the other hand, some intriguing differences cannot be captured by such descriptions and require further investigations. One salient example is vowel space, which has been variably reported to be expanded, shifted, or reduced in languages of all three prosodic types. This may indicate that vowel space is orthogonal to pitch variation, and that stress, tone, and pitch-accented languages are likely to differ to the greatest extent along pitch but not in the segmental dimension. This may be because more or less peripheral vowels, such as tense/lax vowels, occur in many languages, thus hyperarticulation along the segmental dimension may blur the distinction between these vowels. This is in
line with the Functional Load Hypothesis (FLH) which suggests that acoustic properties that are used contrastively in a language will tend to be avoided, or at least be less prominent cues (Vogel et al. Chapter 5, this volume).

There are, however, other interpretations. One of them is that the diversity is related to methodological differences between studies. Indeed, one obvious difference between most studies reporting expanded vowel space size contain speech to infants elicited with a few objects with minimally different names (e.g. sheep, shoes, shop), whereas most studies reporting no such expansion contain IDS elicited with many more toys or with no props at all (hence, not a clear contrast set). This is because when provided with fewer toys or objects, parents may hyperarticulate words referring to those objects, since they are aware that this small contrastive set is the focus of the experiment. Thus, the methodological differences may account for some of the differences that we see in the aforementioned studies.

Another explanation relates to whether acoustic changes in IDS are learning-oriented. Indeed, one could attempt to explain the divergences above (e.g. reduced vowel space) by postulating that parents were implementing a sort of trade-off in their speech, enhancing dimensions that are more informative over others that are less so in the infant’s native language. For example, it has been proposed that Cantonese mothers do not hyperarticulate vowels because they pay more attention to, and exaggerate, tonal information, and thus vowel information (somehow) suffers as a consequence. This line of thought would greatly benefit from a consideration of how exactly this trade-off occurs, whether at the physiological level (i.e. vowels are necessarily less distinct when tones are more so) or at a higher cognitive level (i.e. they stem from talkers’ strategic behavior, but they are not necessary). In any case, this discussion goes beyond the scope of the present chapter.

In our second research question, we discussed whether the relevant prosodic dimensions (stress, tones, pitch accent) are equally accurately cued in both registers; in other words, do the units with stress, tone, or pitch accent have more prominent acoustic cues than the units without stress, tone or pitch accent? The tentative answer appears to be positive. Stress, tone, and lexical pitch contrasts in IDS seemed just as pronounced as in ADS. Further work with more diverse languages is nonetheless desirable because this question has only been investigated in a few papers.

This review further highlights how limited the extant empirical coverage is. In fact, register-stress interactions have only been described in one language, English, which is a rather special case among stress languages. While stress is contrastive, the system is heavily biased toward a trochaic pattern. Moreover, stress can be predicted to a certain extent from syllable complexity and vowel length, and thus stress is overall redundant with syllable weight. An entirely different picture may emerge by studying very different stress languages, such
as Spanish, which is less biased toward trochees, and Finnish, which has a fixed stress pattern. Interestingly, more diverse systems have been studied among tone languages (e.g. Mandarin and Cantonese have a different number and composition of tones), although in this case one can criticize the lack of coverage in terms of language families. In this sense, pitch-accented IDS research is more diverse, with representatives of both Sino-Tibetan and North Germanic languages. The general focus is clearly centered in Eurasia, with little research on other rich areas in Africa and the Americas, not to mention Oceania. We hope that future work will gain in breadth, as well as in depth, by more carefully controlling the phonological and extralinguistic factors that may modulate register effects.

One limitation of the present review is that it has not taken into account other aspects of the phonological system. Most importantly, neither rhythm (see Payne, Post, Astruc & Prieto 2009 for relevant empirical evidence), nor contrastive vowel length has been discussed. These two salient dimensions may also compete with register along durational cues. Therefore, it would be desirable for empirical studies to move on from the description of single dimensions, to the incorporation of a multidimensional, integrated perspective, in which modifications along segmental and suprasegmental dimensions are described together.

In summary, the findings discussed in this review point to a tendency for IDS and ADS to be distinct even across a diverse set of languages. Furthermore, the registers differ along similar acoustic dimensions, specifically, along the dimension of pitch. Nonetheless, differences between IDS and ADS appear to be affected by other factors such as linguistic structure, context, as well as cultural background and infant age or linguistic development. However, there is a great deal of work that needs to be done in order to provide a better understanding of the acoustic characteristics of IDS and how they interact with the prosodic typology of languages. Well-controlled studies comparing the cue usage in different languages, settings, as well as cultures could better tap into the potential effects of these other factors.

References


