CHAPTER 41

PROSODY IN INFANT- AND CHILD-DIRECTED SPEECH

MELANIE SODERSTROM AND
HEATHER BORTFELD

41.1 Introduction

It has long been recognized that adults speak differently to children than they do to other adults, across cultures and languages (Casagrande 1948; Ferguson 1964; Bynon 1968). The characteristics of this 'infant-directed' or 'child-directed' speech register (henceforth collectively CDS) include a wide variety of prosodic, lexical, syntactic, and phonological changes relative to speech directed at other adults (Soderstrom 2007). Hand in hand with the prosodic characteristics described in §41.2, spoken CDS includes robust shifts in vocal timbre (Piazza et al. 2017), a specialized lexicon (Phillips 1973; Mervis and Mervis 1982), reduced utterance length (Phillips 1973), and particular phonological characteristics (Kuhl et al. 1997; though see e.g. McMurray et al. 2013). The exaggerated features of CDS extend to visual aspects of spoken language, including mouth movements (Green et al. 2010) and head motion (Smith and Strader 2014). Finally, there is evidence of gestural exaggerations in visual languages as well (Masataka 1992).

Infant preference for speech with the characteristics of CDS over those of adult-directed speech (ADS) has also been robustly documented, particularly in young infants. A meta-analysis found a Cohen's D effect size for this preference of .67, with higher effect sizes for studies using naturalistic speech samples (Dunst et al. 2012). Indeed, preference for CDS was more recently selected as the topic of a large-scale cross-laboratory replication project in no small part due to the robustness of this preference (Frank et al. 2017; ManyBabies Consortium 2020). In both theoretical and empirical works, CDS has been identified as having an important influence on language development by drawing attention to the linguistic signal, communicating positive affect, and highlighting specific linguistic elements by simplifying the overall signal.

In this chapter, we first review some of the primary prosodic characteristics of CDS (§41.2). We then discuss sources of variation across culture and context (§41.3 and §41.4).

Next, we examine the function of CDS for social and linguistic development (§41.5). We conclude with some thoughts about the future of this research (§41.6).

41.2 Primary prosodic characteristics of infant- and child-directed speech

The most salient and stable prosodic characteristic of CDS is a raised mean fundamental frequency (f0) (e.g. Fernald and Simon 1984; Fernald et al. 1989; see Figure 41.1). Another fundamental characteristic of CDS is variation in f0. This has been variously described as expanded pitch contours or increased pitch range (Fernald and Simon 1984), or increases in overall variability (Fernald et al. 1989). Differences across studies in the manner with which this pitch variation is characterized may lead to difficulties in comparisons across findings. However the variation is characterized, f0 features appear to a primary driving force behind infant preference for CDS, at least in early infancy (Fernald and Kuhl 1987).

CDS is also characterized by significant rhythmic differences that include not only shorter utterances, slower speech rate, and longer pauses (e.g. Fernald et al. 1989; Soderstrom et al. 2008) but also differing and exaggerated prosodic stress and syllable-lengthening effects compared with ADS. While there is a general consensus that lengthening effects exist, a complex picture emerges with respect to how lengthening is instantiated across the sentence and stress contexts. For example, in a task where mothers were instructed to teach their 6- and 8-month-old infants multisyllabic words, Albin and Echols (1996) found evidence of greater sentence-final lengthening in CDS than ADS, and also word-final lengthening effects in CDS, for both stressed and unstressed syllables. Bernstein Ratner (1986) similarly found increased sentence-final lengthening effects for CDS compared with ADS in a sample of pre-verbal infants, but not for mothers speaking to infants at the one-word stage or beyond (stress was not controlled or examined in this study). These studies suggest that sentence-final lengthening plays a significant role in the durational differences between CDS and ADS. Indeed, one study found that exclusion of utterance-final syllables erased

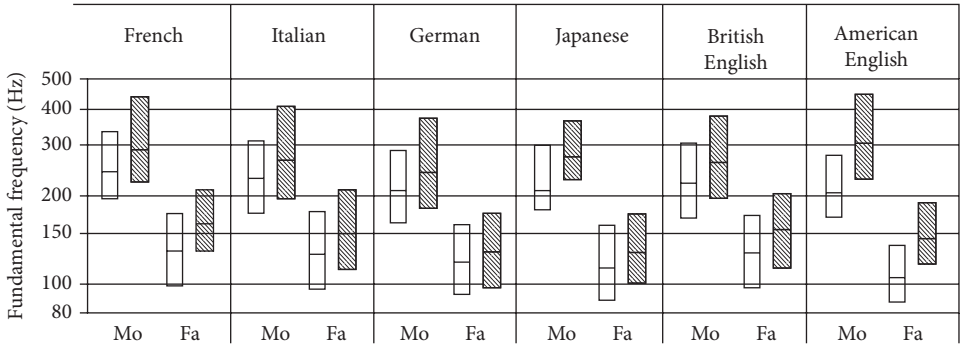

FIGURE 41.1 An example of mean f0 and f0 variability in CDS compared with ADS across six languages for both fathers (Fa) and mothers (Mo).

(Reprinted with permission from Fernald et al. 1989)

the effects of speaking rate differences (i.e. length of syllables across the sentence) between CDS and ADS (Church et al. 2005). Importantly, however, other studies have found that mothers place novel words at the ends of utterances, coincident with exaggerated pitch peaks, something not consistently done in speech to adults (Fernald and Mazzie 1991), and that mothers' use of stress across discourse to focus attention (e.g. across repeated uses of a word or to highlight a word) in speech to their word-learning infants differs from that of speech to adults (Fisher and Tokura 1995; Bortfeld and Morgan 2010). This makes it difficult to interpret whether effects of lengthening are due to position or focal stress. One recent study (Ko and Soderstrom 2013) attempted to disentangle these factors using highly controlled stimuli and systematically varying speech register (CDS vs. ADS), focus, and utterance type (question vs. declarative). In this study, lengthening effects were found across the sentence, with no proportional increase in lengthening sentence-finally. However, the highly controlled nature of these samples may not be fully reflective of naturalistic CDS, so the question remains incompletely resolved.

In addition to highlighting lexical items, CDS carries information at the paralinguistic level, such as conveying approval or comfort, arousing or soothing, or soliciting attention (Katz et al. 1996; Papoušek et al. 1991). Of note have been studies suggesting that both infants and their caregivers use specific intonational structures relevant to the dynamics of conversational interaction and turn-taking. In particular, the preponderance of questions (with their rising intonation) has been noted in a wide variety of studies (e.g. Newport et al. 1977; Soderstrom et al. 2008; but cf. van de Weijer 1997). More generally, caregivers use rising intonation to solicit behaviours from their infants (Stern et al. 1982; Ferrier 1985; Papoušek et al. 1991) or bell-shaped contours to maintain gaze (Stern et al. 1982). These different intonational structures are associated with particular grammatical structures (Stern et al. 1982) and their pragmatic meaning (e.g. approval, prohibition) is more salient in their CDS than ADS forms (Fernald 1989). Infants are sensitive to these messages, showing preferences for certain kinds of intonational patterns, such as approving contours (Papoušek et al. 1990) and questions (Soderstrom et al. 2011). Some have gone so far as to argue that this affective component is the primary distinguishing feature of CDS—comparisons of CDS and ADS have found that when emotional content is controlled, there are no differences between the prosodic characteristics of CDS and ADS or infants' responses to them (Trainor et al. 2000; Singh et al. 2002).

41.3 Cross-cultural similarities and differences

Documenting the existence of particular speech modifications across diverse cultures is important in understanding whether the use and characteristics of CDS represent a more general human communication phenomenon or are culturally specific. Indeed, evidence from a wide range of languages and cultures supports the view that there is a general drive for adults to modify their speech when speaking to children. In early work across a range of urban centres, Ferguson (1964, 1977) documented acoustic modifications of speech when adults addressed children. He also observed changes in what he referred to as 'speech register'

in adults' speech to children in 15 different languages and 23 different cultures. This latter form of CDS included a change in any linguistic feature of speech, such as vocabulary or syntax, rather than a vocal or acoustic change. It has long been argued that acoustic modification is more typical when addressing pre-linguistic infants, whereas simplified speech is more common when addressing older infants and young children (Fernald 1992), but the relative prevalence of the two forms of modification remains unclear.

Intonational characteristics of CDS have been reported in a number of languages. Fernald et al. (1989) compared acoustic aspects of CDS across six languages. These were German, British English, and American English (stress-timed languages); French and Italian (syllable-timed languages); and Japanese (a mora-timed/pitch-accented language). Acoustic analyses of mothers' speech revealed higher mean f0 and f0 range in CDS compared to ADS in all languages except Japanese, in which the f0 ranges in CDS and ADS were equivalent. In a similar vein, for the tonal language Mandarin Chinese, CDS f0 characteristics were less exaggerated than in American English (Grieser and Kuhl 1988; Papoušek and Hwang 1991). Kitamura and colleagues (2001) investigated the prosodic characteristics of CDS in a tonal (Thai) and a non-tonal (Australian English) language. Longitudinal speech samples were collected from mothers while they spoke to their children. Collection took place at three-month intervals from birth to 12 months, as well as while the mothers spoke to another adult. While the age trends across the two languages differed for each of the target measures (i.e. mean f0, f0 range, and utterance slope f0), the integrity of the tonal information in Thai was retained. Although Australian English CDS was generally more exaggerated than Thai CDS, tonal information in Thai was only slightly less identifiable in Thai CDS than in Thai ADS.

Broesch and Bryant (2015) argued that the details of such claims should be clarified with respect to which aspect of CDS was measured: acoustic modification of speech, simplified speech register, or both. While simplified speech register has been documented in both industrialized and non-industrialized cultures, data on modification of acoustic properties of speech have come entirely from large urban and industrialized societies. Broesch and Bryant (2015) set out to determine whether mothers in traditional societies likewise alter acoustic aspects of their speech when speaking to infants relative to adults. Focusing on speech from three distinct cultures (rural Fijians, Kenya's Bukusu, and middle-class Americans), their findings confirm that prosodic alteration of speech manifests similarly across quite different cultural groups (but see Ratner and Pye 1984; Ingram 1995). However, the amount of CDS heard by infants can vary quite widely (see e.g. Cristia et al. 2019), and much more work is needed to establish the extent to which CDS can be considered 'universal'. One important limitation on these kinds of comparative works to date is that researchers often use different standards of measurement (what counts as an utterance, how f0 is measured, what specific characteristics are examined, etc.) that make comparisons across studies very challenging.

Kitamura and colleagues' (2001) finding that Thai mothers preserve linguistic structure by restricting f0 movement was an important test of the parameters of acoustic changes in CDS. These and other findings (e.g. Toda et al. 1990; Bornstein et al. 1992) show that, for tonal and pitch accented languages, mothers compensate for their restricted f0 by increasing the affective content of their CDS. In other words, it is not the f0 characteristics themselves that generalize across languages and cultures, but the positive affect conveyed in the mother's voice, a finding consistent with the research described in §41.2 on the relationship

between affect and CDS. While some languages allow substantial acoustic modulation of speech to infants, others restrain prosodic production a bit more (Grieser and Kuhl 1988; Papoušek et al. 1991; Fernald et al. 1989). Further support for universality comes from the findings of 'prosodic' characteristics of CDS even in visual mode (sign) languages (e.g. Masataka 1992), and preference by both deaf and hearing infants for CDS in the visual mode (Masataka 1996, 1998). Overall, the tendency to modify speech seems to be a species-specific adaptation that facilitates effective mother–infant communication.

41.4 Other sources of variation

Cross-linguistic and cross-cultural concerns aside, CDS is often treated as a monolithic entity, a single register, whereas it is better viewed as a cluster of characteristics that vary by language and culture, but also by age, gender, context, and other factors. While CDS begins at birth (e.g. Fernald and Simon 1984) and continues through early childhood (e.g. Garnica 1977), the manner in which we speak to newborns is patently different from speech to young children. The intonational characteristics described in Fernald and Simon's work would sound very odd in communication to a 5-year-old or even a 2-year-old—the extent of the exaggeration of f0 characteristics decreases with the age of the child after the first year (e.g. Garnica 1977). Indeed, one study of Japanese parents found that prosodic modifications decreased from birth, approaching characteristics of ADS by the onset of the two-word stage (Amano et al. 2006). However, others have found age-related changes in f0 modification over the first year of life to be less linear. Kitamura and colleagues' research with Australian and Thai speakers found both linear and higher-order trends in f0 and f0 range measures across the first year of life, with speech to newborns having lower f0 than that to 6- to 9-month-olds (Kitamura et al. 2001; Kitamura and Burnham 2003). These changes are also associated with differences in the communicative intent of the speaker (Kitamura and Burnham 2003). Similarly, Stern and colleagues (1983) found the most extreme prosodic modifications at 4 months (compared with newborn, 12, or 24 months) but the longest pauses at the newborn age. Another study did find a linear trend from 4 to 16 months in changes in speaking rate, with rate of speech increasing towards the adult-to-adult rate over time (Narayan and McDermott 2016), but it did not measure speech to newborns. Overall, there are reasons to believe that some CDS characteristics are unique to the newborn period, with studies reporting particular speech styles not present at other ages, such as whispered speech (Fernald and Simon 1984) and some much longer utterances with more of a self-talk character (Phillips 1973; Snow 1977).

Another important source of variation is gender, both that of the caregiver and that of the infant. Findings related to gender have been varied and complex, and appear to involve interactions between the gender of the caregiver and child, age of the child, language, and context. For example, Kitamura and colleagues (2001) found changing patterns of mean f0 raising by mothers over their children's development that differed between speech to their male versus female infants—but this developmental pattern was different for Thai speakers and speakers of Australian English. On the other hand, both Thai and Australian English mothers produced greater overall f0 range for girls than boys.

Comparisons between fathers and mothers yield a similarly complex story. Overall, it is clear that the prosodic characteristics of CDS are found in both male and female speech (e.g. Fernald et al. 1989). However, some studies have found no difference in f0 measures of male and female adults' production of CDS (e.g. Jacobson et al. 1983) while others have shown a complex pattern of differentiation across child ages that varies between mothers and fathers (e.g. Warren-Leubecker and Bohannon 1984). While one study found greater increases in f0 range in fathers' than mothers' speech to 2-year-olds (Warren-Leubecker and Bohannon 1984), another study found less increase in f0 range in speech to 1-year-olds by fathers than mothers (Fernald et al. 1989). Importantly, baseline measures of ADS also vary across genders, not just for mean f0 but for f0 variability as well (Warren-Leubecker and Bohannon 1984), adding to the complexity in interpreting these gender findings. Furthermore, gender differences may be highly influenced by cultural expectations, which may vary not only by language or culture but also by cultural shifts over time.

Beyond systematic variation, there are individual differences in the implementation of CDS that often go unrecognized and may confound analyses with small numbers of individuals. To our knowledge, only one study has attempted to examine this directly. Bergeson and Trehub (2007) found systematic individual differences in the implementation of a particular prosodic contour (rising) by particular mothers, which they referred to as 'signature tunes'.

One of the most important concerns is variation due to the context in which the speech sample is collected, as this touches on the ecological validity of the characteristics of CDS. This is particularly important given that many experimental studies on the effects of CDS for infant language acquisition rely on laboratory-recorded 'CDS' in the absence of an infant target. A few studies have attempted to directly examine the impact of recording context on CDS characteristics. Some studies have shown impacts of the knowledge of observation on the quantity and quality of CDS (e.g. Graves and Glick 1978; Field and Ignatoff 1981; Shneidman and Goldin-Meadow 2012). Specific to prosody, Fernald and Simon (1984) found that speech produced by a mother to her newborn without the infant actually present ('simulated' CDS) contained some of the same expanded contours of CDS, but these characteristics were reduced compared to CDS produced while holding the infant. In another study, simulated CDS from female students and trained female actors was compared. The trained actors produced higher f0 in their CDS than the students (Knoll et al. 2009). However, both groups showed similar increases in ratings of positive affect in their CDS compared with ADS. By contrast, Schaeffler and colleagues (2006) showed no difference in the impact of presence or absence of a child on f0 of CDS. However, in that study, the child was present in the room during the recording of the ADS, which may have served to reduce any measurable CDS effect in the child-present condition.

41.5 Function of prosodic characteristics

Numerous researchers over the decades have pointed to the functional benefits of CDS (see references in Snow and Ferguson 1977, particularly Sachs 1977 for an early articulation), which can be classified in one of three ways. First, CDS engages and maintains attention.

Second, it communicates affect and facilitates social interaction. Third, it facilitates language. Each of these three functional benefits will be reviewed in turn.

Attention is required for learning, and different forms of speech have more or less success in attracting infants' attention. Since 'happy talk' draws infants' attention in a positive way, caregivers (and doting others) are more inclined to manipulate their vocal acoustics to elicit this response. Indeed, and perhaps unsurprisingly, adults rate infants' facial responses to CDS as more 'attractive' than their facial responses to ADS (Werker and McLeod 1989). While only one study has measured the association of prosody with infant attention longitudinally (Roberts et al. 2013), the findings confirm the long-term effects of the attention-gating of such speech, showing that CDS at 6 months predicts infant joint attention skills at 12 months. In other words, the intonational modifications inherent to CDS increase the salience of the input, probably by increasing its variability relative to ADS and by reflecting positive emotions.

Not surprisingly, positive speech greatly affects infants' social and linguistic development. There is growing evidence that an abundance of negative or neutral speech can have a detrimental effect on early development. For example, Weinberg and Tronick (1998) found that infants as young as 3 months are sensitive to their mothers' depression. In turn, infants of depressed mothers show impairment in social, emotional, and cognitive functions (Weinberg and Tronick 1998), as well as in associative learning, something necessary for language development (Kaplan et al. 2002). While affect (or lack thereof) may account for such findings, another important factor is infants' own active elicitation of responses from caregivers. This communicative give-and-take creates an environment rich in linguistic structure, which is fundamental for language development to take place, and depressed mothers may not partake in it as readily as non-depressed mothers.

Perhaps most critically, the exaggerated intonational characteristics of CDS highlight linguistic structure and how different components of language are strung together. These properties affect infants' organization of, and memory for, speech. Moreover, young learners can use a variety of distributional strategies to pull individual words out of the components. The simplest example of this is that a priori knowledge of certain high-frequency words (e.g. the infant's own name) (Bortfeld et al. 2005) combined with CDS can help to further delineate where other words begin and end, even at the age of 6 months. In other words, while the structure of CDS provides initial edges in otherwise continuous speech, continued exposure to the regular patterns within the smaller 'chunks' of speech that those edges create allows infants to break them down further. Consistent with this, a wealth of recent evidence has highlighted different forms of structural information in the speech signal. For example, 9-month-olds prefer to hear artificial pauses at grammatical boundaries over pause insertions at non-boundaries, but only for CDS and not ADS (Kemler Nelson et al. 1989); infants can segment artificial speech with CDS characteristics, but not when the stimuli are produced in ADS (Thiessen et al. 2005); and CDS has been found to contribute to lexical learning (Ma et al. 2011b). Other linguistic properties of CDS (e.g. an abundance of questions) may also serve to highlight these chunks and syntactic regularities in the language (Soderstrom et al. 2008).

41.6 Conclusion and future directions

Communication is inherently social. At the earliest stages of development, infants are being influenced by the sounds around them, particularly the speech directed at them. Fortunately, caregivers' biases to communicate in particular ways help infants to focus their attention specifically on speech sounds. The structure of the speech signal together with the contingent structure of the infant–caregiver interaction (mothers responding dynamically to infants and vice versa) serve to highlight regularities in speech and communicate affect; infants respond to this. These maternal responses correlate positively with language development. Through these responses, infants appear to learn the association between the production of certain sounds and their outcomes. Finally, caregivers' input during social interactions and early 'conversations' scaffold language learning by providing information about activities and objects that are the focus of infants' attention in the first place.

While research on CDS dates back to the 1970s and even earlier and continues to flourish, recent methodological advances in two domains warrant particular attention in the coming years. First, much of the data available on CDS are behavioural in nature. Although there are a growing number of studies using neurophysiological methods to examine infants' processing of different forms of CDS (e.g. Bortfeld et al. 2005, 2007; Saito et al. 2007; Naoi et al. 2012; Fava et al. 2014a, 2014b), an increase in such research will broaden our understanding of what role, if any, this form of speech plays in early brain development. Second, recent advances in the use of full-day recordings to examine infants' and young children's language experiences (Vandam et al. 2016) present an important opportunity to expand our understanding of how CDS manifests in the real world, particularly across different cultures. Recent collaborative approaches increasingly allow for data sharing (e.g. Vandam et al. 2016) and focus on the development of standardized annotation approaches (e.g. the DARCLE Annotation Scheme: https://osf.io/4532e), which will greatly improve comparative analyses.

These and other approaches will contribute to our understanding of the interplay between language development and important competencies, both social and emotional. Nonetheless, it is clear that this special form of speech is fundamental to infants' initial vocal development and lays the foundation for subsequent advances in language learning.