Cognition 156 (2016) 52–59

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Utterances in infant-directed speech are shorter, not slower

Andrew Martin[a,b,*], Yosuke Igarashi[c], Nobuyuki Jincho[d], Reiko Mazuka[a,e]

[a] Laboratory for Language Development, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan
[b] Konan University, Department of English Literature and Language, 8-9-1 Okamoto, Higashi Nada-ku, Kobe 658-0072, Japan
[c] Hitotsubashi University, Graduate School of Social Sciences, 2-1 Naka, Kunitachi-shi, Tokyo 186-8601, Japan
[d] Waseda University, School of Human Sciences, 2-579-15 Mikajima, Tokorozawa-shi, Saitama 359-1192, Japan
[e] Duke University, Department of Psychology and Neuroscience, Box 90086, 417 Chapel Drive, Duke University, Durham, NC 27708-0086, United States

ARTICLE INFO

Article history:
Received 1 April 2014
Revised 16 July 2016
Accepted 29 July 2016

Keywords:
Infant-directed speech
Speech rate
Final lengthening

ABSTRACT

It has become a truism in the literature on infant-directed speech (IDS) that IDS is pronounced more slowly than adult-directed speech (ADS). Using recordings of 22 Japanese mothers speaking to their infant and to an adult, we show that although IDS has an overall lower mean speech rate than ADS, this is not the result of an across-the-board slowing in which every vowel is expanded equally. Instead, the speech rate difference is entirely due to the effects of phrase-final lengthening, which disproportionally affects IDS because of its shorter utterances. These results demonstrate that taking utterance-internal prosodic characteristics into account is crucial to studies of speech rate.

1. Introduction

Infants learning a spoken language are faced with the formidable task of extracting information at multiple levels—phonemic, lexical, syntactic, semantic, pragmatic—from a densely encoded speech signal, all without explicit instruction. One approach to the problem of understanding how infants accomplish this has focused on the nature of the speech input that forms the basis of language acquisition. Speech directed to infants differs in a number of ways from that directed to adults, and this approach asks if and how these differences might facilitate the acquisition process.

Research on how infant-directed speech (IDS) differs from adult-directed speech (ADS) has identified a range of possible functions for this speech style, including the communication of affect and maintaining the infant's attention. There is also a substantial literature dedicated to testing the hypothesis that IDS serves an additional, pedagogical function. On this view, changes in acoustic dimensions, such as vowel space area (Burnham, Kitamura, & Vollmer-Conna, 2002; Kuhl, Andruski, Chistovich, & Chistovich, 1997), pitch range (Fernald & Simon, 1984), and pitch height (Ferguson, 1964; Fernald & Simon, 1984), represent speakers' attempts to emphasize certain features of the linguistic system in order to make that system easier to learn.

One such property is the focus of the present paper—speech rate. Speech rate has long interested linguists because of the profound effects it can have on the phonetic dimensions of the speech signal. Linguistic units such as phonemes and syllables are not only shorter in faster speech, but are often altered or deleted entirely. An extensive literature on fast speech phonomena in English has documented increased reduction and deletion of both vowels (Davidson, 2006; Hooper, 1978) and consonants (Fosler-Lussier & Morgan, 1999; Raymond, Dautricourt, & Hume, 2006) as speech rate increases. These changes are a consequence of the physical realities of speech production—speaking involves moving a number of articulators, such as the jaw, the tongue, and the teeth, according to a precisely coordinated plan. The less time available for each individual movement, the more simplification of the speech plan becomes necessary.

How is speech rate measured? The literature on automatic speech recognition has traditionally made a distinction between *local speech rate*, the speed at which individual units such as phonemes or syllables are pronounced, and *global speech rate*, which represents an average rate computed over a larger speech unit, such as an entire utterance (e.g., Hermansky & Morgan, 1994; Ohno & Fujisaki, 1995; Wang & Narayanan, 2007). Studies of speech rate in IDS have typically relied on measures of global speech rate, on the assumption that it approximates the local speech rate. Our goal in this paper is to demonstrate that utilizing a more fine-grained measure of speech rate reveals a more nuanced picture of the changes speakers are implementing in IDS.

Measured at the utterance level, speech directed to infants has been shown in a number of languages to be pronounced with a

* Corresponding author at: Konan University, Department of English Literature and Language, 8-9-1 Okamoto, Higashi Nada-ku, Kobe 658-0072, Japan.
E-mail address: amartin@center.konan-u.ac.jp (A. Martin).

http://dx.doi.org/10.1016/j.cognition.2016.07.015
0010-0277/

slower tempo than speech directed to adults, regardless of the age of the infant. Fernald and Simon (1984), for example, found that German mothers spoke more slowly to their newborn infants. Speech rate was measured in syllables per second over an entire speech sample, with pauses removed. Thanavisuth and Luksaneeyanawin (1998) performed a longitudinal study of Thai-speaking mothers addressing their infants at birth, 3 months, 6 months, 9 months, and 12 months. They found that the average syllable length (across the entire utterance) was longer in IDS than ADS for all ages, but found no effect of the age of the infant. Shute and Wheldall (1995, 2001) recorded British English-speaking mothers (Shute & Wheldall, 1995) and grandmothers (Shute & Wheldall, 2001) reading a story to their children (or grandchildren), who were between one and three years of age, and reading the same story to an adult. Both studies found that the duration of the story was longer in IDS than ADS for both groups of speakers. Finally, in a study whose main focus was sibilant contrasts, Cristià (2010) recorded speech from American English-speaking mothers to their infants, who were taken from two age ranges: 4–6 months and 12–14 months. She reported speech rate in syllables per second, calculated over entire utterances, was lower in IDS than in ADS, with no difference between the two age groups.

Taken together, these results present a remarkably consistent picture—across languages and ages, speech directed to infants is slower than that addressed to adults. Within the literature on IDS, this claim has attained the status of an established fact, regularly cited in research articles and introductory textbooks (e.g., Coulmas, 2013; Harley, 2008; Lust, 2006) as being one of the defining hallmarks of the infant-directed speech style.

In addition to the widespread consensus that IDS is slower than ADS, there appears to be support for the idea that this phenomenon is helpful to infants. Having input presented at slower speeds confers clear benefits for recognizing phonemes and words—the performance of automatic speech recognition systems, for example, suffers at higher speech rates (Siegler & Stern, 1995). There is also experimental evidence that a lower speech rate helps infants recognize words. Zangl, Klarman, Thal, Fernald, and Bates (2005) found that infants recognized tokens of words pronounced at normal speed more accurately than the same tokens played at twice the speed. Song, Demuth, and Morgan (2010) showed that target words embedded in IDS sentences were recognized more quickly and accurately by infants when the sentences were played at natural speed, compared to a condition in which they were played at twice normal speed.

Given the evidence from multiple studies that IDS is slower than ADS, the results from these two experimental studies would seem to be evidence that infants do in fact have a word recognition advantage when listening to IDS over ADS. This conclusion, however, only follows if the speeded-up conditions in these experiments actually resemble ADS. Both Song et al. (2010) and Zangl et al. (2005) employ stimuli which are uniformly compressed throughout their duration, meaning that their results would only be relevant to infant acquisition if in fact real infant-directed speech is a uniformly slower version of adult-directed speech, with every syllable being expanded equally.

In fact, however, syllables or words at the ends of prosodic phrases are known to be pronounced with increased duration, a phenomenon known as *phrase-final lengthening* (Downing, 1970). Because utterances in IDS are on average shorter than in ADS, final lengthening has a disproportionate effect on the mean speech rate in IDS. The slower speech rates found in previous corpus studies of IDS could thus have simply been the result of this register's greater proportion of finally lengthened syllables, rather than evidence of an overall slowing on the part of speakers. If by speech rate we mean local speech rate, utterances in two registers cannot be considered to have different speech rates unless the durations of syllables in corresponding phrasal positions (in non-final positions and in phrase-final positions) between the two registers are different. Therefore, in order to determine precisely how speakers adjust their production to accommodate infants, speech rate must be calculated independently for phrase-final and non-final positions.

Previous research on speech rate in IDS has primarily relied on global speech rate averages computed over entire utterances. In their influential study of 24 German-speaking mothers of newborns, Fernald and Simon (1984) took the number of syllables occurring in a two-minute speech sample from each mother, divided by the duration of the utterance in seconds, minus pauses, as a measure of speech rate. In some studies, the calculation of speech rate was done without removing the pauses. For example, Shute and Wheldall (2001) recorded 16 English-speaking grandmothers reading stories both to their grandchild and to an adult. Their measure of speech rate was simply the duration of the entire story in each condition.

It is not the case that there are no studies of speech rate in IDS which take into account the effects of final lengthening. Two studies have approached this problem by using a small, controlled set of sentences read in both registers. Swanson, Leonard, and Gandour (1992), for example, recorded fifteen mothers reading a set of stories both to their infant (18–28 months old) and to an adult. They measured four vowels (/æ/, /ɪ/, /ʌ/, /ɛ/), which were identified as being either phrase-final or phrase-non-final. They found that although there was evidence of final lengthening, vowels in content words were longer in IDS than in ADS even in non-final position. More recently, Ko and Soderstrom (2013) also looked at durational differences in IDS versus ADS as produced by five speakers reading a list of prepared sentences, although they used simulated IDS rather than interactions with real infants. Like Swanson et al. (1992), they found an overall lengthening effect across the utterance in IDS.

Although both of these studies are carefully done, from their results we can safely conclude only that read IDS is slower than read ADS. It is known that read and spontaneous speech differ both prosodically and in terms of speech rate (Howell & Kadi-Hanifi, 1991), making it difficult to conclude from these studies how spontaneous ADS and IDS differ. If our goal is to understand the role IDS plays in language acquisition, the crucial data will be spontaneous speech, which makes up the bulk of an infant's input.

There is to date only one study which has analyzed the speech rate of spontaneous speech data, separating utterance final syllables from non-final syllables. Church, Bernhardt, Pichora-Fuller, and Shi (2005) compared speech rates in spontaneous IDS and ADS produced by two English-speaking mothers (the infants were 8.5 and 11 months old). They found an IDS > ADS duration difference in utterance-final syllables, but no difference in non-final syllables. These results suggest that the overall lengthening found in studies of read speech may not generalize to spontaneous IDS. The study we describe in this paper is intended to build on the findings of Church and colleagues by expanding the scope of their basic methodology. We do this in three main ways: first, we will collect data from a much larger number of speakers. Second, while Church et al. distinguished only between utterance final syllables and non-final syllables, we consider lengthening at the end of a number of different types of prosodic phrase. Finally, we will test their hypothesis on a non-English language, namely, Japanese.

Our approach involves recording a large number of mothers actually interacting with their infants, and transcribing and annotating these recordings with segmental, word, and syllable boundaries, as well as the positions of multiple types of prosodic boundaries. Our data consists of recordings of spontaneous speech from 22 Japanese mothers. We will show that when the local speech rate is considered, there is no evidence that IDS is produced more slowly than ADS, demonstrating that apparent speech rate