Cognition 166 (2017) 84–93

Contents lists available at ScienceDirect

Cognition

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

Original Articles

# Vowels in infant-directed speech: More breathy and more variable, but not clearer

Kouki Miyazawa [a,b,*], Takahito Shinya [c], Andrew Martin [a,d], Hideaki Kikuchi [e], Reiko Mazuka [a,f]

[a] Laboratory for Language Development, RIKEN BSI, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan
[b] Fairy Devices Inc, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[c] Otsuma Women's University, Faculty of Language and Literature, 12 Sanban-cho, Chiyoda-ku, Tokyo 102-8357, Japan
[d] Konan University, Department of English Literature and Language, 8-9-1 Okamoto, Higashi Nada-ku, Kobe 658-0072, Japan
[e] Waseda University, Faculty of Human Sciences, 2-579-15 Mikajima, Tokorozawa, Saitama 359-1192, Japan
[f] Duke University, Department of Psychology & Neuroscience, 417 Chapel Drive, Durham, NC 27708-0086, United States

ARTICLE INFO

*Article history:*
Received 6 April 2016
Revised 17 April 2017
Accepted 1 May 2017
Available online 26 May 2017

*Keywords:*
Speech production
Vowel variation
Breathiness
Language acquisition
Acoustic analysis
Infant-directed speech

ABSTRACT

Infant-directed speech (IDS) is known to differ from adult-directed speech (ADS) in a number of ways, and it has often been argued that some of these IDS properties facilitate infants' acquisition of language. An influential study in support of this view is Kuhl et al. (1997), which found that vowels in IDS are produced with expanded first and second formants (F1/F2) on average, indicating that the vowels are acoustically further apart in IDS than in ADS. These results have been interpreted to mean that the way vowels are produced in IDS makes infants' task of learning vowel categories easier. The present paper revisits this interpretation by means of a thorough analysis of IDS vowels using a large-scale corpus of Japanese natural utterances. We will show that the expansion of F1/F2 values does occur in spontaneous IDS even when the vowels' prosodic position, lexical pitch accent, and lexical bias are accounted for. When IDS vowels are compared to carefully read speech (CS) by the same mothers, however, larger variability among IDS vowel tokens means that the acoustic distances among vowels are farther apart only in CS, but not in IDS when compared to ADS. Finally, we will show that IDS vowels are significantly more breathy than ADS or CS vowels. Taken together, our results demonstrate that even though expansion of formant values occurs in spontaneous IDS, this expansion cannot be interpreted as an indication that the acoustic distances among vowels are farther apart, as is the case in CS. Instead, we found that IDS vowels are characterized by breathy voice, which has been associated with the communication of emotional affect.



## 1. Introduction

How human infants acquire highly complex language systems in a short time and without explicit instruction is a perplexing issue that has been the focus of much research in psychology. In order to understand how this can be accomplished, one must first understand the nature of the input to learning. This is particularly critical in the case of infant language acquisition due to the fact that adults typically modify their speech when talking to infants—the resulting speech style is known as infant-directed speech (IDS), as opposed to adult-directed speech (ADS). It has often been argued that the modifications speakers make when addressing infants are in some sense specifically designed to aid language acquisition, not only for the obvious reason that IDS constitutes the primary input from which infants learn language, but also because infants have been shown to exhibit a specific preference for IDS (Cooper & Aslin, 1990; Fernald, 1985; Hayashi, Tamekawa, & Kiritani, 2001; Kitamura & Lam, 2009; Pegg, Werker, & McLeod, 1992).

The focus of the present study is the nature of vowels in IDS, one of the most extensively investigated aspects of this speech style. We will comprehensively examine how IDS vowels differ from those in ADS, and what function, if any, they may serve in infants' language development. Cumulative research has revealed that infants' perception of vowels and consonants becomes attuned to the phonological system of their ambient language during the first year of life. The specific ways in which segments are modified in IDS have been proposed to play an important role in this process (Kuhl et al., 1997). In an influential paper, Kuhl and colleagues measured the frequencies of the first and second formants (F1

* Corresponding author at: Laboratory for Language Development, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan.
*E-mail address:* kouki.miyazawa@riken.jp (K. Miyazawa).

http://dx.doi.org/10.1016/j.cognition.2017.05.003
0010-0277/

and F2) of vowels in ADS and IDS in American English, Russian, and Swedish, and found that the triangular area that contains the average F1 and F2 of the vowels /i/, /a/, /u/ (in other words, the Vowel Space) is significantly larger for IDS than for ADS across these languages (Kuhl et al., 1997). Previous research with adult speech has shown that a larger vowel space is associated with an increased acoustic distance among vowels, and that these vowels are perceived as being more clearly articulated (Bradlow, Torretta, & Pisoni, 1996). Thus, an increased vowel space in IDS has been interpreted as an indication that vowels in IDS are pronounced more clearly than those in ADS, a process sometimes called "hyperartic ulation," which should make the infant's task of learning vowel categories easier. In the present paper, we will call this view the *hyperarticulation hypothesis*, following McMurray, Kovack-Lesh, Goodwin, and McEchron (2013) and Martin, Utsugi, and Mazuka (2014).

Subsequent studies found expanded vowel spaces in the IDS of other languages, such as French, Chinese, and Japanese (Dodane & Al-Tamimi, 2007; Liu, Kuhl, & Tsao, 2003). Furthermore, a number of studies, taking this hypothesis as given, have investigated the computational and neural processes underlying the perception of hyperarticulated vowels, as well as the consequences of hearing hyperarticulated vowels on infants' ability to discriminate them (Bernsteinn, 1984; Malsheen, 1980; Sebastián-Gallés & Bosch, 2009; Whiteside & Hodgson, 2000; Zhang, Miller, Tess, & Carney, 2010). Through these studies, the hyperarticulation hypothesis has become a widely accepted view with regard to the nature of IDS segments.

Note, however, that the expansion of vowels' F1/F2 values refers to the acoustic measurements of vowel formants, while the hyperarticulation of vowels refers to the process of vowel articulation. In particular, hyperarticulation refers to clear articulation, which speakers may or may not be necessarily consciously aware of. This distinction is crucial, since, as we will demonstrate in the present paper, the expansion of F1/F2 is not always an indication that the vowels are hyperarticulated.

To date, many studies have already challenged the claims of the hyperarticulation hypothesis. Empirically, not all studies that have examined the vowel spaces of IDS and ADS found the IDS vowel space to be significantly larger than that of ADS. For example, no increase was found in spontaneous IDS in Norwegian (Englund & Behne, 2005) or Jamaican English (Wassink, Wright, & Franklin, 2007). In addition, in contrast to studies which are limited to measuring peripheral point vowels such as /a/, /i/, and /u/, F1 and/or F2 values of more central vowels do not seem to be expanded in a way that is consistent with the hyperarticulation hypothesis (Cristia & Seidl, 2014). It has also been argued that other changes in IDS, such as the slower speech rate, lexical stress, and the prosodic context in which the target vowels tend to appear in IDS, may be responsible for changes in F1/F2 values, rather than hyperarticulation per se (McMurray et al., 2013; Wang, Seidl, & Cristia, 2015).

Methodologically, problems with using vowel space as a measure of hyperarticulation have also been raised. As mentioned above, the vowel space is typically calculated by computing the average F1 and F2 values of the point vowels for each speaker, a process in which within-speaker variability across vowel tokens is eliminated. IDS is known to be highly variable in almost every aspect that has been examined (Soderstrom, 2007), and vowels in IDS are likely to be more variable than those of ADS. If variability is high and the overlap between vowels is large, a larger vowel space, as measured by average formant values for each vowel, would not necessarily make vowels more distant from each other (e.g., Sumner, 2011; Wassink et al., 2007). Computationally, Kirchhoff and Schimmel (2005) demonstrated this using a Hidden Markov Model (Rabiner, 1989). Experimentally, McMurray and colleagues (McMurray et al., 2013) measured the variability among vowel tokens in an elicited production task, and found that IDS vowels were in fact more variable and had larger overlap among them than those in ADS.

In sum, even though the hyperarticulation hypothesis remains an influential view of IDS segments, a number of studies have challenged the hypothesis both on empirical and methodological grounds, casting serious doubts on its wide-reaching implications. Nonetheless, a number of important issues need to be addressed before we can reconcile the mixed results from previous studies. The goal of the present study is to address these issues systematically such that we can come to a comprehensive perspective on the nature of IDS vowels and the role they may play in infants' language acquisition.

First of all, it must be determined whether an expansion of F1/F2 values for IDS vowels occurs reliably in spontaneous IDS. (Note that we will refer to F1/F2 expansion, rather than to the "vowel space," since our analysis will include other vowels than the point vowels /a/, /i/, and /u/.) Thus far, data on expanded F1/F2 values has tended to come from elicited production tasks (Kuhl et al., 1997; Liu et al., 2003; McMurray et al., 2013), while those studies which use spontaneous speech have produced inconsistent results (Englund & Behne, 2005). If vowel expansion is a phenomenon that occurs only when mothers recite prepared scripts, and not in spontaneous IDS, the basic assumption of the hyperarticulation hypothesis is undermined. Furthermore, as McMurray and colleagues have argued (McMurray et al., 2013), even if a numerical expansion is found on average, it could be an "unintended by-product" (p. 36) of other changes in IDS, such as biases in lexical selection and the prosodic positions in which vowels appear. Until now, it has not been possible to examine these factors in spontaneous IDS, since no existing corpus of spontaneous IDS has been coded with prosodic level information. In the present study, we will address both of these issues by analyzing a corpus of spontaneous IDS and ADS in Japanese that is fully coded with both segmental and prosodic level information (Mazuka, Igarashi, & Nishikawa, 2006).

Second, even if expanded F1/F2 values are found in spontaneous IDS, and the effect remains after the other factors discussed above are taken into account, the question would still remain – are expanded F1/F2 values a valid measure of vowel hyperarticulation? We will approach this question in two steps. As in previous studies, we will first measure the variability of F1 and F2 values in spontaneous IDS and ADS vowels. Using these values, we will examine the discriminability of vowels computationally, taking the magnitude of variability into account by computing Mahalanobis distance (MD) (Mahalanobis, 1936, see Section 4 for details). The usefulness of MD in measuring acoustic distance was demonstrated by Nakamura, Iwano, and Furui (2008), who compared Japanese adult speech across three registers; dialogue, academic presentation, and read speech. They measured the acoustic distance between phoneme pairs using MD, and found that the MDs among the three registers differed significantly from each other. They also computed the clarity of speech by means of a speech-recognition algorithm, and found that the MD correlated positively with the clarity of speech.

Next, in order to test whether expanded F1/F2 values in IDS can be interpreted as an indication of hyperarticulation, we add another corpus to our analysis, i.e., careful, read speech (CS) spoken by the same mothers that produced our IDS and ADS data (Martin et al., 2014). As discussed above, the link between an expanded vowel space and the hyperarticulation of vowels was originally based on a comparison between carefully spoken adult speech and spontaneous adult speech (Bradlow et al., 1996). If the expanded F1/F2 is in fact an indication that vowels are hyperarticulated, CS vowels should differ from spontaneous ADS in similar ways as IDS. Namely, vowels in both IDS and CS should have expanded F1/F2 values compared to those in ADS, and they should