Segmental distributions and consonant-vowel association patterns in Japanese infant- and adult-directed speech*

SHO TSUJI

Radboud Universiteit Nijmegen, International Max-Planck Research School for Language Sciences, and Laboratory for Language Development, RIKEN Brain Sciences Institute

KENYA NISHIKAWA

Laboratory for Language Development, RIKEN Brain Sciences Institute

AND

REIKO MAZUKA

Laboratory for Language Development, RIKEN Brain Sciences Institute, and Duke University

ABSTRACT

Japanese infant-directed speech (IDS) and adult-directed speech (ADS) were compared on their segmental distributions and consonant-vowel association patterns. Consistent with findings in other languages, a higher ratio of segments that are generally produced early was found in IDS compared to ADS: more labial consonants and low-central vowels, but fewer fricatives. Consonant-vowel associations also favored the early produced labial-central, coronal-front, coronal-central, and dorsal-back patterns. On the other hand, clear language-specific patterns included a higher frequency of dorsals, affricates, geminates, and moraic nasals in IDS. These segments are frequent in adult Japanese, but not in the early productions or the IDS of other studied languages. In combination with previous results, the current study suggests that both fine-tuning (an increased use of early produced

segments) and highlighting (an increased use of language-specifically relevant segments) might modify IDS on the segmental level.

INTRODUCTION

Understanding the nature of infants' input is indispensable for research on language acquisition. Infants show an impressive ability to extract information from various sources of their ambient language such as the distribution of phonetic units, sequential probabilities between phonemes, and word stress. While dictionary counts or adult-directed speech (ADS) had been assumed a sufficient approximation of infants' input, evidence on differences between ADS and infant-directed speech (IDS), a speech style used by caregivers when addressing their infants, is accumulating. Speech modifications in IDS have been reported at the phonological, prosodic, syntactic, and lexical levels (for an overview, see Soderstrom, 2007), documenting both commonalities and differences between languages. At the segmental level, differences in vowel and consonant quality between IDS and ADS have been investigated (for an overview, see Cristia, 2013). Interestingly, however, only a few studies have addressed the segmental distribution characteristics of IDS.

This lack of studies contrasts with the extensive literature on the development of early segment productions (e.g. Boysson-Bardies & Vihman, 1991; Jakobson, 1941/1968) and the grouping of segments into the most basic of association patterns, the consonant-vowel (CV) sequence (e.g. MacNeilage & Davis, 2000; Vihman, 1992). Only recently, studies comparing IDS and ADS in Korean (Lee, Davis & MacNeilage, 2008) and English (Lee & Davis, 2010) have reported that segmental distribution patterns of Korean and English IDS show both commonalities and cross-linguistic differences. However, research from a wider range of languages is needed to determine the role of input on infants' acquisition of phonological segments.

The current study compared the frequency of occurrence of segments (consonants and vowels) and CV combinations in Japanese IDS with that of Japanese ADS. Patterns in Japanese IDS and ADS were also compared with those of other languages, in particular with English and Korean. English has been studied extensively in terms of segment input and production, and is typologically and historically close to many other well-studied European languages. Korean is, like Japanese, typologically and historically quite distinct from English, and as such instrumental to broadening our database. It is typologically and historically close to Japanese, but the phonologies of the two languages differ substantially, with Korean having a larger vowel and consonant inventory in addition to more complex syllable structure.

1277

The following literature review describes the ontogeny of the production of segments and CV combinations in languages other than Japanese, highlighting cross-linguistic similarities and differences. Distributional properties of segments and CV combinations in languages other than Japanese are then described and compared with those of Japanese. Finally, the structure of Japanese is outlined.

Development of early segment production

Ambient language input, in interaction with early production constraints, is considered a crucial source for learning to produce the native language segment inventory. Early claims of a rigid universal order of phoneme acquisition (Jakobson, 1941/1968) were not supported by subsequent studies that demonstrated a substantial variability in early production both within (Vihman, 1993) and across (Ingram, 1999) languages. Nonetheless, there is little doubt that motor constraints lead to a tendency of some segments to emerge earlier than others in babbling and early word production.

Based on an overview of several early production studies, Bernhardt & Stemberger (1998) reported that stops, nasals, and glides are the manners of articulation that are produced earliest across languages, while fricatives, affricates, and liquids occur comparatively late. This finding is consistent with more recent overviews of both American English (Smit, 2007) and British English (Howard, 2007), and members of other language families such as Cantonese (So, 2007), Finnish (Kunnari & Savinainen-Makkonen, 2007), Greek (Mennen & Okalidou, 2007), Spanish (Goldstein, 2007), and Thai (Lorwatanapongsa & Maroonroge, 2007). The fricative /h/ has also been reported to occur early across the languages Dutch (Fikkert, 1994), English, Swedish, and French (Vihman, 1992).

Regarding place of articulation, labials and coronals tend to be produced early compared to dorsals across languages (cf. Bernhardt & Stemberger, 1998). Languages do differ in the acquisition order of labials and coronals, but there is no overall tendency for one to be predominantly produced earlier than the other. A consistent finding in studies and overviews of American English (Boysson-Bardies & Vihman, 1991; Smit, 2007), British English (Howard, 2007), French (Boysson-Bardies & Vihman, 1991; Rose & Wauquier-Gravelines, 2007), Spanish (Goldstein, 2007), German (Fox, 2007), Jordanian Arabic (Dyson & Amayreh, 2007), Cantonese (So, 2007), and Greek (Mennen & Okalidou, 2007) is an earlier onset of labial and coronal place of articulation compared to dorsals.

For vowels, front/central mid/low vowels (i.e., vowels located in the LOWER LEFT QUADRANT of the F_1/F_2 vowel space) have been reported to be most frequent in early productions (Davis & MacNeilage, 1990). Similarly, for American English (Smit, 2007) it was reported that back vowels and

1278

the front-high vowel /i/ are rare in early productions, and that front-high /i/ and front-mid /ɛ/ remain erroneous for children between one and three years of age.

These common tendencies have mainly been explained with reference to articulatory restrictions. Stops and nasals, which are produced by a complete closure of the vocal tract, are considered relatively easy to produce compared to fricatives and affricates, which require a more complex coordination of articulatory position and airflow (cf. Kent, 1992; MacNeilage, Davis, Kinney & Matyear, 2000). Vihman and colleagues suggested that labial and coronal stops are easy to articulate as they require simple mandibular oscillations (e.g. Vihman, 1993) and because the accompanying lip closure is considered an especially salient visual cue (Boysson-Bardies & Vihman, 1991). Despite these common tendencies, cross-linguistic variation exists. This variation is generally attributed to the nature of the input, which will be reviewed in the following section.

Segmental characteristics of the input

One of the first studies on segmental properties of IDS was a qualitative description of baby talk, words modified for infants, across fifteen languages by Ferguson (1977). A later study (Vihman, Kay, Boysson-Bardies, Durand & Sundberg, 1994) quantitatively compared the characteristics of IDS in running speech, content words, word-initial segments of content words, and adult target models of children's attempted words across American English, French, and Swedish. In order to more specifically address the differences between IDS and ADS, two recent studies directly and quantitatively compared segmental properties of Korean (Lee *et al.*, 2008) and English (Lee & Davis, 2010) IDS and ADS. In both samples, the speech of ten mothers to their one-year-old infants was compared to a sample of ten women speaking to an adult experimenter.

With regard to place of articulation, Ferguson (1977) reported a high frequency of labial and coronal stops. Vihman *et al.* (1994) also found support for a higher frequency of coronals and labials compared to dorsals in running speech across languages, with coronals being most frequent. Finally, Lee *et al.* (2008) also lent support to this pattern by reporting a significantly higher frequency of labial place, and a significantly lower frequency of glottal place, in Korean IDS compared to ADS. However, they also reported coronal place to be significantly less frequent in IDS than ADS. No differences with regard to place were found in the English sample (Lee & Davis, 2010).

With regard to manner, Ferguson's (1977) sample contained a high frequency of nasals and a low frequency of liquids. This pattern was not found in the sample of Vihman *et al.* (1994), in which stops were most

1279

frequent, followed by fricatives/affricates, nasals, and glides. While Lee & Davis (2010) found a significantly higher frequency of stops and glides, and a lower frequency of fricatives, affricates, nasals, and liquids for English IDS compared to ADS, no differences were found in the Korean sample (Lee *et al.*, 2008). Thus, in sounds other than stops, the data show a rather low consistency regarding manner.

Fortis and geminate consonants were found to be more frequent in Korean IDS than ADS (Lee *et al.*, 2008). Mid-central and low-central vowels were significantly more frequent and high-central and mid-front vowels were significantly less frequent in IDS compared to ADS. We can find some consistencies between these characteristics of IDS and the early productive tendencies reviewed in the previous section: fricatives both emerged later in early productions and were less frequent in IDS than ADS across the languages studied. The early produced labial consonants and lower left quadrant vowels were relatively frequent in Korean IDS, and stops and glides were relatively frequent in English IDS. The relatively late produced affricate and liquid consonants were less frequent in English IDS compared to ADS.

In summary, if anything, IDS patterns show a better fit with early production patterns than with ADS patterns. One explanation offered is that early produced segments are favored, and late produced segments are avoided or substituted in IDS (cf. Ferguson, 1977; Lee *et al.*, 2008; Lee & Davis, 2010). The usage of late produced segments in IDS has also been suggested to reflect an increased use of language-specific, perceptually salient segments. Evidence from other languages is necessary to evaluate these interpretations.

Phonological segments in Japanese

Compared to the languages reviewed thus far, studies on Japanese children's segmental development reported notable differences. In a study comparing English-, French-, Japanese-, and Swedish-learning children's babbling and early speech (Boysson-Bardies & Vihman, 1991), Japanese children produced a relatively low number of labials and a high number of dorsals. Contrary to English, French, and Swedish children's decrease of fricatives and affricates in first words compared to babbling, they showed no such decrease.

Edwards & Beckman (2008) reported that substitution patterns of Japanese- and English-acquiring children's early pronunciation errors reflected differences in segmental distributions of the input: while English-learning children tended to substitute coronal [t] for dorsal /k/, Japanese children rather substituted [k] for /t/. A longitudinal study following phoneme mastery of ten Japanese children from 1;0 to 4;0

(Uno, 2007) revealed that they first mastered the labial stop /b/ and nasal /m/ (by 1;3), immediately followed by the stops and nasals /p, t, d, k, n, g/. The postalveolar affricate /tʃ/ was acquired by 1;6, relatively early in comparison to other languages (e.g. Bernhardt & Stemberger, 1998; Kent, 1992). Thus, Japanese children's early productions included a high amount of labial, stop, and nasal consonants, consistent with universal tendencies. However, they also included a comparatively high amount of affricates and dorsal consonants.

Ferguson (1977) reported that baby talk words in Japanese often included geminates and affricates. More recently, it was reported that Japanese IDS contained a higher frequency of dorsal stops (/k, g/) than coronal stops (/t, d/) (Beckman, Yoneyama & Edwards, 2003), in contrast to studies in other languages that have reported the opposite pattern. This recent study, though, was focused specifically on place of articulation in stop consonants. In order to study language-specific and language-general patterns beyond this subgroup, an overall quantitative analysis of segmental distributions of Japanese IDS and ADS is mandatory.

Early CV association patterns

Segments, especially consonants, rarely occur in isolation. The grouping of segments into CV sequences is an important milestone towards the acquisition of speech, first occurring between the ages of 0;6 and 0;8 in the stage of canonical babbling and necessarily preceding speech (Vihman, 1992). Like research on early segment production, research on early CV association patterns has considered constraints and regularities. MacNeilage *et al.* (2000) proposed that basic biomechanical constraints lead to three preferred association patterns in early production. In their Frame/Content theory, the association of labial consonants with central vowels reflects a PURE FRAME resulting from simple mandibular oscillations. By adding a tongue movement to this basic oscillation, two additional associations, coronal-front and dorsal-back, are formed. These three association patterns were observed to be more frequent than expected by chance in the babbling and early speech of fifteen English-learning infants, as well as overall in dictionary counts of the nine languages French, Swahili, Estonian, Hebrew, German, Spanish, English, Maori, and Quichua.

Another investigation of early CV association patterns (Vihman, 1992) with samples of American, French, and Swedish children found support for the labial-central association but not for the other two association patterns. Instead of a coronal-front association, the sample showed a positive association between coronal consonants and central vowels. Regarding associations of dorsal consonants with vowels, it was difficult to find any pattern due to the low frequency of dorsal segments.

1281

The relationship between IDS and early production of CV sequences has been investigated in two recent studies. A study on the relationship between IDS and the output of infants aged 0;7 to 1;6 in Mandarin Chinese (Chen & Kent, 2005) found strong correlations between a subset of infants' predominant production patterns and caregivers' speech. Infant output provided support for the coronal-front and dorsal-back frame, as well as a labial-back association pattern. IDS correlated with the labial-back and dorsal-back, but not the coronal-front association pattern.

A study of Korean compared CV association patterns of infants' babbling and first words with IDS and ADS (Lee, Davis & MacNeilage, 2007). They found support for the association patterns proposed by MacNeilage *et al*. (2000) in babbling, which were suggested to reflect early and possibly intrinsic constraints. In first words, on the other hand, the coronal-front and dorsal-back, but not the labial-central, associations were found. Both babbling and early words showed an imperfect but large overlap with IDS in the predominant association patterns, while there was no overlap with ADS. In IDS, a predominant coronal-front association pattern was observed, while ADS did not show any of the previously suggested basic patterns. These two studies show that there is a stronger relationship of early CV association patterns with IDS than with ADS, suggesting that the former resembles early production patterns more closely.

In sum, support for the labial-central association pattern comes from Swedish, English, and French early productions as well as from Korean babbling; for the coronal-front pattern from English, Chinese, and Korean early productions; and for the dorsal-back pattern from Chinese and Korean early productions. Further frequent patterns are coronal-central and labial-back, and early produced CV association patterns overlap with patterns in IDS, but not ADS.

Japanese again seems to diverge from the above-mentioned languages. The aforementioned dictionary study (MacNeilage *et al*., 2000) showed Japanese to be the only language in which the average observed-to-expected frequency ratios for the three suggested patterns did not exceed chance level. Labial-central and dorsal-back associations showed a tendency in the expected direction, but the coronal-central association led to a higher observed-to-expected ratio than the coronal-front one. Similarly, Vihman (1992) found labial-central, dorsal-back, and coronal-central associations for Japanese children. Notably, in her sample of four languages (Swedish, English, French, and Japanese), only Japanese children produced a substantial frequency of dorsal consonants and back vowels, and consequently they alone contributed a high quantity of dorsal-back associations. Japanese children's early productions thus do show a labial-central association similar to that observed in most other languages. They deviate from the coronal-front association proposed by MacNeilage

1282

et al., but agree with Vihman's coronal-central associations. Finally, they show a strong dorsal-back association, consistent with the proposal made by MacNeilage *et al.* Japanese, along with Mandarin (Chen & Kent, 2005) and Korean (Lee *et al.*, 2007), is a language in which children seem to produce many dorsal and back segments.

IDS patterns

The above review has shown that IDS is distinct from ADS at the segmental level. Overall, IDS segmental distributions parallel infants' early productions better than ADS. This is generally interpreted as a fine-tuning of caregivers' articulations to infants' capacities, favoring segments that are generally produced early while avoiding segments that are generally produced late. We will call this the FINE-TUNING pattern, following Cross (1977). This pattern predicts a universal tendency for a higher frequency of segments that are generally produced early and a lower frequency of segments that are generally produced late. However, other patterns of modification are also conceivable. Caregivers could produce language-specific segments more frequently, thus highlighting patterns that are important for the native language but are not necessarily acquired early in general. We will call this the HIGHLIGHTING pattern.

To distinguish fine-tuning and highlighting, it is necessary to examine a language in which the segmental distribution shows divergences from common patterns, as otherwise the most frequent segments will match what is easy for infants to produce. The above literature review shows that Japanese is such a language, with part of the early productions strongly reflecting language-specific characteristics. Before summarizing the aims of the current study, some relevant characteristics of Japanese phonology are described.

Japanese linguistic structure

Japanese is a mora-timed language, where one mora is a subsyllabic unit that can consist of a single vowel (V), a CV sequence, the moraic nasal /N/, or the first half of a geminate consonant /Q/. Japanese light syllables consist of either V or CV, and heavy syllables are formed by vowel lengthening or by adding /N/ or /Q/ to a CV sequence. Consequently, Japanese syllables are mostly V or CV, and the occurrence of consonant clusters is rare.

The Japanese vowel inventory consists of the five mono-moraic short vowels /a, i, u, e, o/, and their long bi-moraic counterparts /a:, i:, u:, e:, o:/. There are no quality differences between short and long vowels (Saito, 1997). As we are going to compare our findings to previous findings in Korean and English later on, we are referring to characteristics of these languages where adequate. In terms of monophthong vowels, the Japanese

inventory of five is smaller than the inventories of both Korean and English. Korean consists of eight, and English of twelve, monophthong and an additional three diphthong vowels (Lee *et al.*, 2008; Lee & Davis, 2010). Taking into account that Japanese distinguishes long and short vowels, the total number of vowel categories is ten, putting it in between Korean and English.

Japanese has twenty-three consonants, and all consonants except the moraic /N/ necessarily precede a vowel in a CV sequence. Additionally, Japanese has a geminate segment /Q/, which forms a geminate or long consonant combined with a singleton plosive or fricative consonant combined, and is in phonemic contrast to singleton consonants. As the consonantal status of the geminate segment is controversial (cf. Vance, 1987), we did not include it in our analysis of place and manner, but its frequency of occurrence in IDS and ADS separately. Based on the higher frequencies of geminate consonants in both Japanese baby talk words (Ferguson, 1977) and Korean IDS (Lee *et al.*, 2008), a difference might be expected. The number of phonemic consonants was reported to be twenty-four in English (Lee & Davis, 2010), and nineteen in Korean (Lee *et al.*, 2008). Like English, Japanese distinguishes voiced and unvoiced stops, while Korean makes a three-way distinction between lenis, fortis, and aspirated. English contains nine fricatives while Japanese has eight (two of which are extremely rare) and Korean only three.

Japanese IDS contains a high amount of specialized vocabulary that is often phonologically unrelated to the adult form. This is known as infant-directed vocabulary (IDV). A survey of mothers of infants aged 0;8 to 1;0 reported 237 distinct infant-directed word-types (Mazuka, Kondo & Hayashi, 2008). For example, *kuruma* 'car' in ADS becomes *buHbu* in IDS, and *gohan* 'meal' becomes *maNma*. Many of the expressions in IDV have their roots in onomatopoeia, and occur most frequently in heavy-light or heavy-heavy disyllabic forms (79% of word forms reported in the survey). Since heavy syllables necessarily contain either a geminate, a moraic nasal, or a long vowel, it is of interest whether the frequency of occurrence of these three segment types differs between IDS and ADS. The moraic nasal is pooled with non-moraic nasals in consonant analysis, but in order to capture its exceptional status a separate analysis compares the frequency of occurrence of moraic and non-moraic nasals in IDS and ADS.

A further language-specific input factor is the high frequency of YOUON consonants. As the definition of youon depends on orthographic characteristics, it cannot directly be related to phonological categories. Orthographically, youon are formed by adding a small kana symbol that represents glides to a normal-sized one, for example in ギャ [gʲa] or チャ [tʃa]. One group of youon consists of consonants that are palatalized before a vowel, for example in [gʲa] or [kʷo]. The other group of youon includes

fricative or affricate consonants that precede the vowels /a, u, e, o/, for example in [tʃa] or [ʃo]. Note that consonants preceding /i/ are palatalized in most cases but are not classed as youon as they do not have the orthographic distinction described above. Youon are often associated with a familiar/casual style of speech and with speech directed to young children and infants. Examples include [tʃitt **t**ʃai] instead of [tʃi: **s**ai] 'small' and diminutive suffixes for people's names or kinship terms such as [onii-**t**ʃaN] instead of [onii-**sa**N] 'older brother'. They are also used frequently in onomatopoeic expressions (e.g. [tʃokitʃoki], to describe the action of cutting something with scissors), which are in turn used often in IDV, as discussed above. Therefore, a difference in the frequency of youon between IDS and ADS might be expected.

Word boundaries in Japanese can be determined by either referring to SHORT-UNIT or LONG-UNIT words. Short-unit words roughly correspond to dictionary entries and are monomorphemic or at most bimorphemic. Long-unit words are combinations of words that may correspond to compound words. For instance, *baikiNmaN* (a Japanese cartoon character, 'germ-man') may be analyzed as two short-word units, *baikiN* 'germ' and *maN* 'man', or as one long unit. To our knowledge, there exists neither a strict agreement concerning when to use short and long units nor any reference discussing this topic. As long units are often perceived as the more natural boundaries, we chose those for analysis.

Aims of the current study

The above review shows that, if anything, IDS patterns fit early productions better than ADS patterns. These differences could be due to caregivers' fine-tuning, accounting for an increased use of segments that are generally produced early, and/or due to highlighting of language-specific patterns. The available data are too sparse and varied to establish the above tendencies, and data from additional languages are necessary to evaluate systematic differences between IDS and ADS at the segmental level. Analyzing segmental frequencies in a relatively large corpus of Japanese, a language in which early production patterns show both language-general and language-specific patterns, will provide a further step towards answering this question.

The current study will thus evaluate differences and similarities between Japanese IDS and ADS in light of the fine-tuning and highlighting accounts. If caregivers are fine-tuning their speech to infants' production capacities, we expect IDS to have higher frequencies for segment groups that are generally produced early. This would be labial and coronal place of articulation; stop, nasal, and glide manner; lower left quadrant vowels; and labial-central, coronal-central, coronal-front, and dorsal-back

consonant-vowel associations. If on the other hand, caregivers are highlighting language-specific patterns, we expect this to show in those segment groups that are both frequent in Japanese and acquired rather late in general. Among these patterns are dorsals, affricates, and dorsal-back consonant-vowel associations. Additionally, segment types occurring in Japanese infant-directed vocabulary (geminates, moraic nasals, youon) are expected to occur frequently in IDS.

We will compare segment frequencies of IDS and ADS for place of articulation, manner of articulation, and vowels in that order, as well as evaluate the occurrence of consonant-vowel sequences in IDS and ADS.

The methods of segment comparison will closely follow those in Lee *et al.* (2008) and Lee & Davis (2010) in order to facilitate comparison (a more detailed explanation is provided in the 'Methods' section). However, running speech, which was used in these studies, might not be the most representative measure of what matters to infants. There is evidence that children orient to initial consonants in word selection (Boysson-Bardies & Vihman, 1991) and that content words are especially salient to infants (Shi & Werker, 2001). Moreover, segmental frequency of word-initial content words has been found to better reflect children's early productions than running speech (Vihman *et al.*, 1994). Therefore, both running speech and word-initial content words were examined.

METHODS

Corpus

The corpus used in this study contains the speech of twenty-two Japanese mothers from the Tokyo area and their children aged 1;6 to 2;0 (Mazuka, Igarashi & Nishikawa, 2006). Children of this age are at the early stage of their production and comprehend some of what their mothers say to them. Recordings of each mother–child dyad took place in a sound-attenuated room. The mother's utterances were recorded by a head-mounted dynamic microphone, and a condenser microphone placed on a table recorded the child's utterances. Additionally, dyads were video-recorded by means of a ceiling camera and microphone. Audio recordings were made with DAT tapes, and video recordings with mini DV tapes. For IDS samples, two separate recordings were made. During the first 15 minutes, the mother was asked to play with the child using a number of picture books. Mothers could choose from seven books depicting a variety of animals, toys, and actions and contained very little text. For the remaining 15 minutes the books were replaced by a set of silent toys such as animals, soft blocks, and finger puppets. Mothers were free to use any of the materials but were not specifically instructed to do so. Some mothers in fact played with their child without using any of the materials that were provided.

1286

TABLE 1. *Japanese consonant inventory*

	Labial	Coronal	Dorsal	Unclassified
Stop	p b	t d	k g	
Fricative	ɸ v	s z ʃ ʒ ç	h	
Affricate		ts ʧ dʒ		
Nasal	m	n ɲ	ŋ n	
Glide		j		w
Liquid		ɾ		

TABLE 2. *Japanese vowel inventory*

	Front	Central	Back
High	i		u
Mid	e		o
Low		a	

For ADS samples, a female experimenter subsequently entered the room and talked with the mother for ten minutes, in the child's presence, about topics related to child-raising. A total of approximately 45 minutes of recording per dyad was obtained.

Data coding

The IDS recordings totaled about 11 hours of speech and 50,000 words; the ADS recordings 3 hours and 30,000 words. Annotations were based on the schemes developed for the Corpus of Spontaneous Japanese (Maekawa, 2003). The phonetic transcriptions were performed by three highly trained phoneticians. In cases of disagreement or uncertainty, they examined the original sound files together in an effort to resolve the issue. When no agreement could be reached regarding some section, it was marked and excluded from the analysis. The entire corpus was double-checked for its accuracy by a single phonetician.

Phonemes were transcribed according to the Japanese consonant and vowel inventory (cf. Tables 1 and 2). Additionally, the geminate segment /Q/ was coded. Transcribed consonants were classified for place of articulation as labials [p, b, ɸ, v, m], coronals [t, d, s, z, ʒ, ʃ, ç, ts, tʃ, dʒ, n, ɲ, j, ɾ], and dorsals [k, g, h, ŋ]. The place of articulation of the moraic nasal /N/ depends on the place of articulation of the following consonant such that it is realized as [m] preceding labial consonants, [n] preceding coronal consonants, and [ŋ] preceding dorsal consonants. We classified it post-hoc according to these rules. However, if /N/ was followed by a

1287

vowel or a pause, it could not be classified and was thus excluded from the analysis of place of articulation. About 34% of moraic nasals were excluded for that reason. Additionally, the glide [w] and the palatalized stops [kʷ] and [gʷ] were not classifiable for place of articulation post-hoc and therefore excluded from this analysis. For manner analysis, [p, b, t, d, k, g] were classified as plosives, [ɸ, v, s, z, ʃ, ʒ, ç, h] as fricatives, [ts, ʧ, dʒ] as affricates, [m, n, ɲ, ŋ, ɴ] as nasals, [j, w] as glides, and [ɾ] as a liquid. The segments that could not be classified for place of articulation were included in the manner analysis. Adjectives, adjectival nouns, adnominals, adverbs, nouns, and verbs were considered content words.

Data analysis

The two types of IDS sample (book reading and toy playing) were collapsed after an initial comparison of the two data types did not show systematic differences. As total sample sizes of IDS and ADS differed, segment frequency ratios rather than absolute frequencies were used for analysis. Ratios were calculated separately for IDS and ADS, for vowels and consonants, and for the analysis considering running speech and the analysis considering only word-initial content words. For example, the ratio of the labial stop [b] in infant-directed running speech was calculated by dividing the total number of [b] occurrences by the number of all consonants in infant-directed running speech. The obtained ratios were then subjected to an arcsine transformation, a common transformation recommended for stabilizing variances in proportional variables (Cohen, Cohen, West & Aiken, 2003).

Repeated-measures analyses of variance (ANOVA) were conducted on place, manner, and vowel contrasts, followed by Bonferroni-adjusted pairwise comparisons where appropriate. Greenhouse–Geisser corrected values were reported where the sphericity assumption was violated. Separate paired *t*-tests compared the frequencies of youon and of geminates, and a separate ANOVA compared the frequency of moraic and non-moraic nasals in IDS and ADS. Recently, the use of parametric statistical tests like ANOVA for analyzing segment frequencies in corpus data has been criticized due to their distributional properties, and a non-parametric alternative was proposed (Daland, 2012). Results from an analysis following this method were comparable to those reported below, and are omitted due to space constraints.

For analysis of CV sequences, we calculated observed-to-expected ratios for each of the nine possible consonant-vowel association patterns for labial, coronal, and dorsal consonants with front, central, and back vowels, adopting the procedure introduced in Lee *et al.* (2007). Expected frequencies were obtained by multiplying the number of consonants in the respective place

of articulation with the number of vowels in the respective position, and dividing this number by the total number of CV association patterns. For instance, the expected frequency of labial-front associations was obtained by multiplying the number of labial consonants with the number of front vowels, and dividing the result by the total number of CV associations. The observed-to-expected ratio was then calculated by dividing the observed frequency of each CV association pattern by its expected frequency. Chi-square tests were conducted to indicate if observed frequencies overall differed significantly from expected frequencies. If so, to determine which of the CV association patterns contributed to this result, the standardized residuals for every association pattern were obtained, where a category with a standardized residual value above 2 is considered to be a major contributor to significance. As analyses of early CV association patterns in production mainly concentrate on the early acquired groups of stops and nasals (cf. MacNeilage & Davis, 2000), we report results on this subgroup of segments in addition to results including all segments.

RESULTS AND DISCUSSION

Overall, there were a total of 75,199 consonants in IDS and of 34,973 consonants in ADS. Vowel numbers totaled 78,583 in IDS and 37,154 in ADS. Assuming that the number of vowels roughly corresponds to the number of syllables in a corpus, our data is approximately eleven times the size of previous Korean and English studies.

Consonant place

Running speech. Coronal place of articulation was most frequent in both IDS (59%) and ADS (66%), followed by dorsal (22% for IDS and 19% for ADS) and labial (13% for IDS and 10% for ADS) places of articulation. The percentages do not reach 100% because of the moraic nasals that were not classifiable for place of articulation, the glide [w] and the palatalized stops [k^w] and [g^w]. A two-way repeated measures ANOVA with the factors speech style (2) × place of articulation (3) revealed significant main effects for both speech style [$F(1,21) = 5\cdot66$, $p=\cdot027$, $\eta_p^2=\cdot212$] and place of articulation [$F(2,42) = 3680\cdot65$, $p<\cdot001$, $\eta_p^2=\cdot994$], as well as a significant interaction between the two [$F(2,42) = 39\cdot80$, $p<\cdot001$, $\eta_p^2=\cdot655$]. Post-hoc pairwise comparisons between IDS and ADS for labial, coronal, and dorsal place of articulation showed significant differences for all three places. As shown in Table 3, labial and dorsal place were more frequent in IDS, but coronal place more frequent in ADS (Figure 1A).

Word-initial content words. For word-initial segments of content words, coronal place of articulation was again most frequent in both IDS (52%) and ADS (60%), followed by dorsals in IDS (24%) and ADS (23%)

TABLE 3. *Pairwise comparisons of consonant proportions in IDS and ADS. IDS and ADS ratio values are untransformed ratios; difference values, standard errors, and p values are based on arcsine-transformed ratios*

|	Ratio IDS	Ratio ADS	Mean difference (IDS$_{arc}$-ADS$_{arc}$)	Standard error$_{arc}$	p value
			Running speech		
Consonant place					
Labial	·127	·103	·074	·015	<·001
Coronal	·586	·661	−·156	·019	<·001
Dorsal	·218	·192	·066	·016	·001
Consonant manner					
Stop	·386	·375	·024	·014	·104
Fricative	·142	·162	−·056	·020	·011
Affricate	·051	·032	·098	·011	<·001
Nasal	·282	·300	−·040	·013	·006
Glide	·056	·059	−·014	·013	·297
Liquid	·082	·072	·034	·015	·031
			Word-initial content words		
Consonant place					
Labial	·228	·158	·181	·024	<·001
Coronal	·517	·599	−·166	·035	<·001
Dorsal	·237	·230	·016	·027	·558
Consonant manner					
Stop	·459	·297	·340	·038	<·001
Fricative	·216	·321	−·242	·036	<·001
Affricate	·046	·048	−·009	·031	·787
Nasal	·192	·209	−·047	·027	·097
Glide	·069	·115	−·168	·028	<·001
Liquid	·019	·010	·077	·033	·030
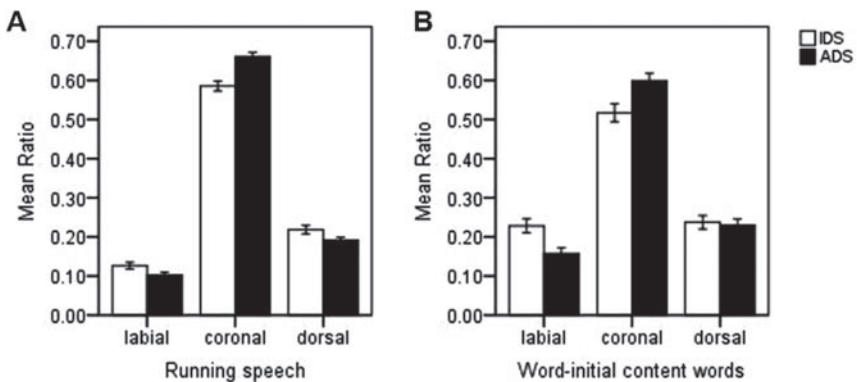
Fig. 1. Percentage of each consonant place of articulation. A: running speech; B: word-initial content words. Error bars represent ±2 standard errors.
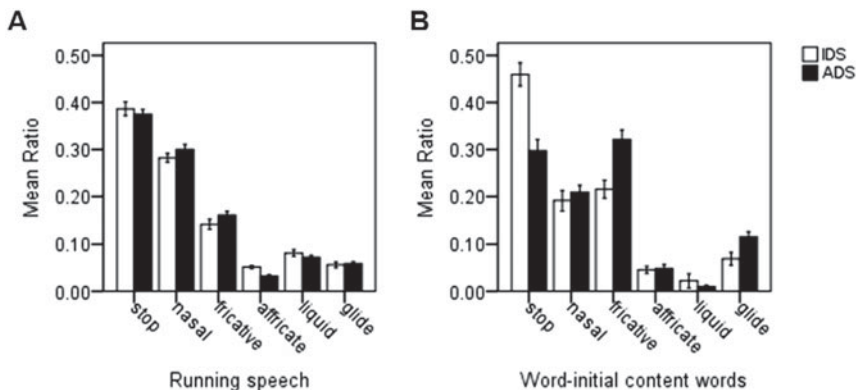
1290

Fig. 2. Percentage of each consonant manner of articulation. A: running speech; B: word-initial content words. Error bars represent ±2 standard errors.

and labials in IDS (23%) and ADS (16%). A (2) speech style×(3) place of articulation repeated measures ANOVA showed significant main effects for both speech style [$F(1,21)=8\cdot04$, $p=\cdot01$, $\eta_p^2=\cdot277$] and place of articulation [$F(2,42)=573\cdot97$, $p<\cdot001$, $\eta_p^2=\cdot965$], and a significant interaction effect between the two factors [$F(1\cdot31,27\cdot44)=24\cdot16$, $p<\cdot001$, $\eta_p^2=\cdot535$]. Post-hoc paired comparisons between speech styles for each place of articulation showed significant differences for labial and coronal place of articulation (Table 3), with labials being more frequent in IDS, and coronals in ADS (Figure 1B). The findings for word-initial content words are generally parallel to those of running speech, except that the difference between IDS and ADS for dorsal segments is not significant here.

Consonant manner

Running speech. For both IDS and ADS, stops were the most frequent manner category with 39% in IDS and 38% in ADS. The second most frequent category was nasals with 28% in IDS and 30% in ADS, followed by fricatives (14% for IDS and 16% for ADS), liquids (8% for IDS and 7% for ADS), glides (6% for both IDS and ADS), and affricates (5% for IDS and 3% for ADS). A two-way repeated measures ANOVA: speech style (2)×manner of articulation (6) revealed significant main effects for speech style [$F(1,21)=11\cdot65$, $p=\cdot003$, $\eta_p^2=\cdot357$] and manner [$F(3\cdot00,63\cdot09)=1576\cdot50$, $p<\cdot001$, $\eta_p^2=\cdot987$], and a significant interaction between the two [$F(3\cdot24,67\cdot94)=12\cdot79$, $p<0\cdot001$, $\eta_p^2=\cdot378$]. Post-hoc paired comparisons showed significant differences in fricative, affricate, nasal, and liquid manner between IDS and ADS (Table 3), affricates and liquids being more frequent in IDS, and fricatives and nasals in ADS (Figure 2A).

1291

A separate (2) nasal type × (2) speech style repeated-measures ANOVA was conducted to separate the moraic and non-moraic nasal. This analysis was only conducted for running speech, as the moraic nasal rarely occurs word-initially. A significant interaction between nasal type and speech style was found $[F(1,21) = 157 \cdot 94,\ p < \cdot 001,\ \eta_p^2 = \cdot 883]$, with a higher frequency of the non-moraic nasal for ADS ($M = 0 \cdot 14$) than IDS ($M = 0 \cdot 10$), and a higher frequency of the moraic nasal for IDS ($M = 0 \cdot 13$) than ADS ($M = 0 \cdot 10$).

Word-initial content words. Word-initially, stops were again the most frequent category for IDS (46%), and fricatives for ADS (32%). These were followed by fricatives (22%), nasals (19%), glides (7%), affricates (5%), and liquids (2%) in IDS, and by stops (30%), nasals (21%), glides (12%), affricates (5%), and liquids (1%) in ADS. A two-way repeated-measures ANOVA with the factors speech style (2) × manner of articulation (6) revealed no main effects for speech style $[F(1,21) = 2 \cdot 845,\ p = \cdot 106,\ \eta_p^2 = \cdot 119]$, a significant effect of manner $[F(5,105) = 533 \cdot 62,\ p < \cdot 001,\ \eta_p^2 = \cdot 962]$, and a significant interaction between the two $[F(5,105) = 33 \cdot 97,\ p < \cdot 001,\ \eta_p^2 = \cdot 618]$. Post-hoc paired comparisons showed that stops and liquids were significantly more frequent in IDS, while fricatives and glides were more frequent in ADS (Table 3; Figure 2B).

Youon

The ratio of occurrence for all youon in IDS and ADS was compared. Youon were significantly more frequent in IDS than in ADS both for running speech (IDS: $M = 0 \cdot 075$; ADS: $M = 0 \cdot 031$; $t(21) = 13 \cdot 32,\ p < \cdot 001,\ d = 2 \cdot 561$), and for word-initial content words (IDS: $M = 0 \cdot 083$; ADS: $M = 0 \cdot 047$; $t(21) = 4 \cdot 69,\ p < \cdot 001,\ d = 1 \cdot 013$).

Geminate stops and fricatives

Ratios for geminates were calculated by dividing the number of geminates by the number of consonants plus geminates. As geminates rarely occur word-initially, only running speech was considered. A paired *t*-test revealed significant differences with a higher geminate ratio in IDS ($M = 0 \cdot 062$) than in ADS ($M = 0 \cdot 050$) $[t(21) = 2 \cdot 94,\ p = \cdot 008,\ d = 0 \cdot 607]$.

Vowels

Running speech. As can be seen in Figure 3, the majority of vowels are short. A (2) speech style × (2) vowel length × (5) vowel quality repeated measures ANOVA was conducted. The results revealed a significant main effect of vowel quality $[F(4,84) = 295 \cdot 45,\ p < \cdot 001,\ \eta_p^2 = \cdot 934]$ and vowel length $[F(1,21) = 85299 \cdot 35,\ p < \cdot 001,\ \eta_p^2 = 1 \cdot 00]$, a significant interaction between speech style and vowel quality $[F(4,84) = 27 \cdot 54,\ p < \cdot 001,\ \eta_p^2 = \cdot 567]$, between
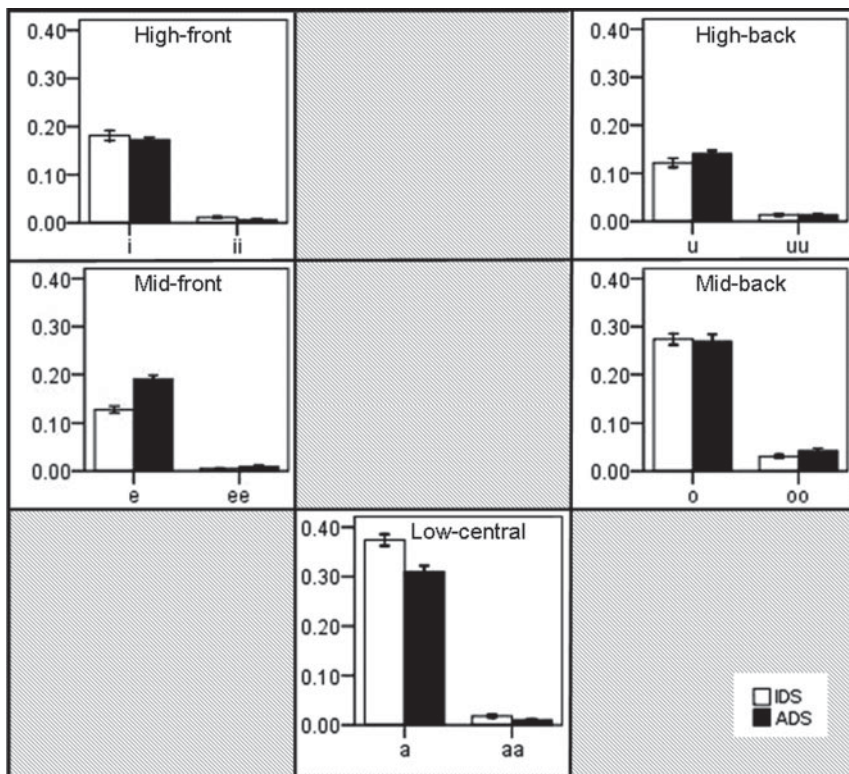
Fig. 3. Percentage of each vowel place of articulation, running speech. Error bars represent ±2 standard errors.

speech style and vowel length [$F(1,21) = 5\cdot53$, $p=\cdot029$, $\eta_p^2=\cdot208$], between vowel length and vowel quality [$F(2\cdot62,55\cdot09) = 433\cdot28$, $p<\cdot001$, $\eta_p^2=\cdot954$] and between speech style, vowel quality, and vowel length [$F(4,84) = 39\cdot02$, $p<\cdot001$, $\eta_p^2=\cdot650$]. Post-hoc paired comparisons showed significant vowel quality differences such that long high-front /ii/, and short and long low-central vowels /a, aa/ were more frequent in IDS, and short high-back /u/, short mid-front /e/, and long mid-back vowels /oo/ were more frequent in ADS (Table 4; Figure 3).

Word-initial content words. A (2) speech style × (2) vowel length × (5) vowel quality repeated measures ANOVA showed a significant main effect of speech style [$F(1,21) = 12\cdot71$, $p=\cdot003$, $\eta_p^2=\cdot377$], of vowel quality [$F(4,84) = 171\cdot74$, $p<\cdot001$, $\eta_p^2=\cdot891$], and of vowel length [$F(4,84) = 2330\cdot65$, $p<\cdot001$, $\eta_p^2=\cdot991$], a significant interaction between speech style and vowel quality [$F(4,84) = 128\cdot32$, $p<\cdot001$, $\eta_p^2=\cdot338$] and between vowel length and

1293

TABLE 4. *Pairwise comparisons of vowel proportions in IDS and ADS. IDS and ADS ratio values are untransformed ratios; difference values, standard errors, and p values are based on arcsine-transformed ratios*

	Ratio IDS	Ratio ADS	Mean difference ($IDS_{arc}-ADS_{arc}$)	Standard $error_{arc}$	p value
			Running speech		
Vowels					
a	·371	·310	·134	·014	<·001
i	·181	·172	·022	·016	·182
u	·122	·140	−·056	·016	·002
e	·128	·190	−·170	·014	<·001
o	·274	·269	·011	·020	·595
aa	·018	·010	·069	·016	<·001
ii	·011	·007	·050	·011	<·001
uu	·014	·014	·000	·016	·954
ee	·005	·009	−·033	·019	·102
oo	·031	·043	−·064	·017	·001
			Word-initial content words		
Vowels					
a	·062	·048	·056	·027	·049
i	·103	·064	·142	·022	<·001
u	·015	·019	−·035	·024	·165
e	·007	·006	·041	·023	·084
o	·066	·071	−·021	·023	·384
aa	·004	·003	·002	·017	·905
ii	·019	·007	·107	·018	<·001
uu	·020	·009	·020	·014	·166
ee	·003	·006	·006	·016	·687
oo	·003	·085	·032	·022	·166

vowel quality $[F(4,84) = 10·71, p < ·001, \eta_p^2 = ·859]$, a marginally significant interaction between speech style and vowel length $[F(1,21) = 3·67, p = ·068, \eta_p^2 = ·150]$, and a three-way-interaction between speech style, vowel quality, and vowel length $[F(4,84) = 3·27, p = ·015, \eta_p^2 = ·135]$. Post-hoc paired comparisons revealed a significantly higher frequency of short low-central vowel /a/ and of short and long high-front vowels /i, ii/ for IDS (Table 4).

CV association patterns

Before turning to the actual analysis of CV association patterns, we compared the ratio of consonants to vowels in this corpus to the other corpora where this information was available. The consonant/vowel ratio in this corpus was ·49/·51 for both IDS and ADS. In English (Lee *et al.*, 2008), 14,450 (IDS) and 14,990 (ADS) consonants per 10,000 vowels were reported, resulting in a consonant/vowel ratio of ·59/·41 for IDS and ·60/·40 for ADS. In Korean (Lee *et al.*, 2008), 11,800 (IDS) and 12,500 (ADS) per

10,000 vowels were reported, resulting in a consonant/vowel ratio of ·52/·48 in IDS and ·56/·44 in ADS. Thus, in comparison, Japanese speech contains the highest rate of vowels, followed by Korean and English. This is consistent with Ramus, Nespor & Mehler (2000), who found that syllables in stress-timed languages tend to have more complex syllables than syllable-timed languages. In their sample the reported consonant/vowel ratio for English was ·60/·40, while Japanese, a mora-timed language, was reported to have the least complex syllables with a ratio of ·47/·53. Interestingly, both English and Korean IDS contain fewer consonants than ADS, suggesting that syllables with fewer consonants are favored in IDS.

Observed-to-expected ratios of serial consonant-vowel organization patterns in IDS and ADS were analyzed. As analyses of early CV association patterns mainly concentrate on stops and nasals, we report separate results on the subgroup of stops and nasals and on all segments for running speech and for word-initial content words.

Running speech. For IDS in stops and nasals there was an overall significant difference between observed and expected frequencies [χ^2 (4, $N=40,182)=2978\cdot93$, $p<\cdot001$, Cramer's V=·1593]. The four associations significantly contributing to the result were dorsal-back, coronal-front, coronal-central, and labial-central. In ADS, there was also an overall significant difference [χ^2 (4, $N=20,189)=1279\cdot11$, $p<\cdot001$, Cramer's V=·193]. The four significantly contributing patterns were dorsal-central, labial-back, coronal-front, and labial-central (cf. Table 5). Considering all tokens in all segments, both IDS [χ^2 (4, $N=58,411)=2426\cdot29$, $p<\cdot001$, Cramer's V=·144] and ADS [χ^2 (4, $N=29,876)=1449\cdot76$, $p<\cdot001$, Cramer's V=·156] showed overall significant differences in observed and expected frequencies. The observed-to-expected ratios that significantly contributed to the difference in IDS were dorsal-back, coronal-front, and labial-central. In ADS, the patterns were dorsal-central, labial-back, coronal-front, and labial-central (Table 5).

Word-initial content words. For stops and nasals of word-initial content words, both IDS [χ^2 (4, $N=8,987)=337\cdot79$, $p<\cdot001$, Cramer's V=·137] and ADS [χ^2 (4, $N=3,470)=475\cdot25$, $p<\cdot001$, Cramer's V=·262] observed and expected ratios were overall significantly different. Dorsal-back, labial-front, coronal-central, and labial-central associations significantly contributed to significance in IDS, whereas in ADS the significantly contributing patterns were in that order coronal-central, dorsal-back, labial-front, and labial-back (cf. Table 5). For all segments, again both association patterns in IDS [χ^2 (4, $N=12,735)=288\cdot46$, $p<\cdot001$, Cramer's V=·106] and ADS [χ^2 (4, $N=6,457)=92\cdot34$, $p<\cdot001$, Cramer's V=·085] showed overall significant differences. The CV associations contributing to significant differences were dorsal-back, labial-central, labial-front, and

1295

TABLE 5. *Observed-to-expected ratios in IDS and ADS. Ratio: observed-to-expected ratio of segment group. SR: standardized residual of each ratio. Ratios contributing to a siginificant Chi-square effect are in bold*

Segments analyzed	Speech style (number of CV sequences analyzed)	Consonant place	Vowel position		
			Front ratio (SR)	Central ratio (SR)	Back ratio (SR)
			Running speech		
All segments	**IDS** (58,411)	Labial	0·78 (−10·5)	**1·22 (10·4)**	0·95 (−0·9)
		Coronal	**1·23 (22·7)**	0·97 (−3·4)	0·86 (−16·1)
		Dorsal	0·56 (−27·4)	0·96 (−2·5)	**1·37 (25·8)**
	ADS (29,876)	Labial	0·62 (−12·4)	**1·10 (2·8)**	**1·27 (9·3)**
		Coronal	**1·17 (14·4)**	0·82 (−13·6)	0·99 (−1·7)
		Dorsal	0·64 (−16·9)	**1·54 (22·5)**	0·90 (−3·8)
Plosives and nasals	**IDS** (40,182)	Labial	0·84 (−6·9)	**1·11 (5·3)**	0·97 (0·4)
		Coronal	**1·27 (19·0)**	**1·12 (10·0)**	0·67 (−26·9)
		Dorsal	0·68 (−18·2)	0·74 (−16·7)	**1·54 (33·3)**
	ADS (20,189)	Labial	0·61 (−13·0)	0·99 (−0·80)	**1·48 (14·8)**
		Coronal	**1·27 (17·6)**	0·89 (−6·9)	0·81 (−11·7)
		Dorsal	0·65 (−15·7)	**1·26 (10·8)**	**1·13 (5·6)**
			Word-initial content words		
All segments	**IDS** (12,735)	Labial	**1·16 (5·2)**	**1·17 (4·8)**	0·76 (−8·3)
		Coronal	**1·09 (3·3)**	0·91 (−3·5)	1·02 (0·6)
		Dorsal	0·71 (−9·6)	1·00 (0·1)	**1·19 (7·3)**
	ADS (6,457)	Labial	**1·29 (4·0)**	0·70 (−5·5)	1·04 (1·5)
		Coronal	1·01 (0·7)	**1·08 (3·0)**	0·95 (−2·7)
		Dorsal	0·80 (−4·4)	1·04 (0·0)	**1·07 (3·0)**
Plosives and nasals	**IDS** (8,987)	Labial	**1·21 (7·1)**	**1·10 (2·7)**	0·75 (−8·3)
		Coronal	1·02 (−0·8)	**1·11 (3·1)**	0·93 (−2·4)
		Dorsal	0·83 (−6·5)	0·81 (−5·9)	**1·33 (10·9)**
	ADS (3,470)	Labial	**1·29 (4·2)**	0·62 (−7·4)	**1·15 (3·6)**
		Coronal	0·89 (−1·9)	**1·63 (13·0)**	0·55 (−10·3)
		Dorsal	0·93 (−1·8)	0·65 (−7·0)	**1·35 (7·7)**

coronal-front in IDS. For ADS, significantly contributing association patterns included labial-front, coronal-central, and dorsal-back (cf. Table 5).

GENERAL DISCUSSION

The current study analyzed differences between IDS and ADS in Japanese segments and segment association patterns in order to identify possible modifications of Japanese IDS. We will first discuss the results separately for consonants, vowels, and CV association patterns, comparing them to results from previous studies in English and Korean. Our discussion will focus primarily on running speech in order to make our results directly

comparable to previous findings. In a separate section, we will discuss the results of word-initial content words where they diverge from running speech. The final section discusses the relevance of IDS modifications on the segmental level.

Consonants

In IDS, a significantly higher frequency of labials, dorsals, affricates, and liquids, and a lower frequency of coronals, fricatives, and nasals were found compared to ADS. The findings for labials and fricatives are consistent with previous reports in Korean and English (Lee *et al.*, 2008; Lee & Davis, 2010), and fit the fine-tuning pattern: an increased use of segments that are generally produced early, and a decreased use of segments that are generally produced late. The higher frequency of dorsals and affricates in the present study is inconsistent with previous findings on both general tendencies in infant production and IDS. These findings do, however, match early production tendencies in Japanese infants, who do produce dorsal and affricate segments from relatively early on. This in turn parallels the high prevalence of both dorsals and affricates (Beckman *et al.*, 2003) in Japanese adult language. Dorsals and affricates in Japanese IDS thus follow the highlighting pattern, a higher frequency of language-specific, not generally early produced segments.

The higher frequency of liquid manner and lower frequency of coronal place and nasal manner in Japanese IDS fit neither fine-tuning nor highlighting. The findings regarding coronals are consistent with those found in Korean (Lee *et al.*, 2008), but not English (Lee & Davis, 2010). Nasals were also less frequent in IDS than ADS in English (Lee & Davis, 2010), although they are among the early produced segments (Bernhardt & Stemberger, 1998).

When looking separately at moraic and non-moraic nasals, however, moraic nasals were more frequent in Japanese IDS than ADS. Geminate segments were also more frequent in IDS than ADS, and these patterns mirror the pattern of geminate and non-geminate nasals in Korean (Lee *et al.*, 2008). Japanese IDV predominantly consists of words with heavy-light and heavy-heavy syllables, which include moraic nasals, geminates, or long vowels (Mazuka *et al.*, 2008). Both the moraic nasal and the geminate segment are exceptional because they are the only mora types that consist of a single consonant. Therefore, they have a distinct, perceptually salient rhythm, which conceivably helps initial segmentation (Vihman, 1993) and therefore might occur frequently in IDV.

There were significantly more youon in IDS than in ADS. Youon are also frequently used in Japanese IDV, where words are often realized with a substitution of affricates for other segments or a palatalized form of adult

1297

words. The sound symbolism literature proposes palatalized sounds to be associated with 'childishness and immaturity' (Hamano, 1998). Since many youon are palatalized, their use may be a way of fine-tuning. Interestingly, increased palatalization after dentals has also been reported for English child-directed speech (Ratner, 1996). Future studies are necessary to investigate if there is an auditory or productive preference for youon-like sounds in infants.

Vowels

A higher ratio of short and long low-central as well as of long high-front vowels and a lower ratio of short high-back, long mid-back, and short mid-front vowels were found in IDS compared to ADS. Previous studies have reported a higher occurrence of the lower left quadrant vowels in early productions (e.g. Davis & MacNeilage, 1990) and IDS (e.g. Lee *et al.*, 2008). Of these, the Japanese vowel inventory only possesses the low-central vowels. These were indeed more frequent in IDS compared to ADS, thus supporting fine-tuning, the increased use of segments that are generally produced early. The findings on long high-front, short high-back, long mid-back, and short mid-front vowels do not fit previous results, however.

Concerning vowel length, a higher ratio of long vowels was found for ADS, showing that prosodic vowel lengthening does not affect phonological vowel length. Consistent with this, Japanese mothers maintained two distinct phonological vowel length categories despite their use of non-lexical vowel lengthening in IDS (Werker *et al.*, 2007).

CV association patterns

Preferred early association patterns previously reported were labial-central, coronal-frontal, and dorsal-back in MacNeilage *et al.* (2000), and labial-central, coronal-central, and dorsal-back in Vihman (1992). The patterns found in the present study in IDS very closely resembled all of these articulatory patterns. The analysis of stops and nasals perfectly mirrored all four suggested patterns in running speech plus the patterns suggested by Vihman (1992) in word-initial content words, while the analysis of all segments mirrored the three patterns of MacNeilage *et al.* (2000) in both running speech and word-initial content-words.

In contrast, ADS showed a correspondence to the suggested patterns only in parts, and not consistently across analyses. The rankings of observed-to-expected ratios were highly variable in ADS, while they were fairly consistent across analyses in IDS. Although IDS and ADS were not compared directly, these data show that the pattern of observed-to-expected ratios in IDS matches the suggested basic production patterns more closely

1298

than ADS. Thus, Japanese mothers are producing CV association patterns that correspond to the suggested basic production patterns of children in IDS but not ADS, providing support for the fine-tuning account.

Labial-central and dorsal-back patterns were constantly the most frequent association patterns in IDS. This is in line with both Vihman's (1992) and MacNeilage *et al.*'s (2000) suggestion of labial-central as the most basic of productive association patterns. Both authors had also reported the dorsal-back pattern as a preferred grouping. In the languages studied previously, however, dorsal consonants were infrequent, and Vihman (1992) suggested that a preference for this pattern may show later in development with more frequent use of back consonants and vowels. As studies both in Mandarin Chinese (Chen & Kent, 2005) and Korean (Lee *et al.*, 2007) found a relationship between IDS and early production of CV association patterns, future studies of Japanese should investigate how far the dorsal-back association pattern is preferred in early productions given a comparatively high amount of dorsals.

Differences between running speech and word-initial content words

In addition to running speech, the present study reported segmental distributions for the subgroup of word-initial content words, because these are known to be especially salient to infants (Boysson-Bardies & Vihman, 1991; Shi & Werker, 2001) and to differ in their distributional properties (Vihman *et al.*, 1994). With a few exceptions, the direction of results was the same for these two type samples. Statistical analyses sometimes showed significant differences for one, but not the other type of analysis, though.

For consonants of word-initial content words, labials, stops, and liquids were more frequent in IDS, and coronals, fricatives, and glides were more frequent in ADS. The difference in stop frequencies between IDS and ADS did not reach significance in running speech, but does also fit into the picture of IDS containing articulatory simple segments and is consistent with the results in Korean and English (Lee *et al.*, 2008; Lee & Davis, 2010). The occurrence of word-initial vowels was generally low, which is due to the moraic structure of Japanese, where vowels mostly follow a consonant. Among these, low-central /a/, as well as high-front /i/ and /ii/, had a significantly higher ratio in IDS. CV association patterns of IDS plosives and nasals mirrored the patterns suggested by Vihman (1992), and association patterns of all segments mirrored the patterns suggested by MacNeilage *et al.* (2000) in word-initial content words.

The role of IDS on the segmental level

The fine-tuning account describes a pattern predicting language-general emphasis of segments that are acquired early in general. It argues that

1299

caregivers match their speech to their infants' production capacities, which was originally suggested based on correlational analyses of mothers' and children's speech in English (Cross, 1977). In Japanese, Murase, Ogura & Yamashita (1992) reported an increase of caregivers' use of adult forms and decrease of baby-talk forms between the ages of 1;10 and 2;2, corresponding to the age where Japanese children start producing adult forms. Matching speech to infants' productions makes sense according to Vihman's (1993) articulatory filter model, which suggests that infants perceive input matching their own productions as especially salient, picking up those patterns for which they already have a motor representation. The prevalence of patterns that match early production tendencies in IDS in the current study and other languages studied so far suggests fine-tuning as one way in which IDS differs from ADS on the segmental level.

At the same time, language-specific differences in the distribution of segments in IDS compared to ADS clearly show that fine-tuning is not the only way segmental distributions in IDS are modified. One source of these differences could be mothers' highlighting of segments that are prevalent in the native language but are not produced early in general. For Japanese, we specifically predicted dorsals and affricates to be more frequent in IDS than ADS, and we indeed found this to be the case. Based on these results, we suggest highlighting as a further way in which IDS could be modified on the segmental level, but further studies are necessary to strengthen this account.

Other differences in patterns across languages cannot easily be explained by either fine-tuning or highlighting. These differences could be due to some systematic language-specific factors at the phonological or lexical level. As for the phonological level, language-specific phoneme inventories could make important contributions. For example, Korean mothers in Lee *et al.* (2008) did not highlight dorsal segments in IDS even though they are highly frequent. The simple phonotactics of Japanese, in particular the low frequency of consonant clusters, may contribute to a higher amount of highlighting in Japanese IDS by inducing less pressure to avoid certain segments: the acquisition of consonants in clusters is late compared to singletons (McLeod, Doorn & Reed, 2001) and infants pay more attention to consonants in syllable onsets than to those in codas (Vihman *et al.*, 1994). In contrast to Japanese, English and Korean both allow CVC syllables, and consonant clusters occur frequently. The observation that the consonant-to-vowel ratio of Japanese IDS does not differ from ADS, while it decreases in both Korean and English, speaks to this interpretation.

On the lexical level, Lee & Davis (2010) assign part of the differences between Korean and English to the different usage of IDS: Lee & Nakayama (2000) found that Korean and Japanese mothers frequently use specific infant-directed vocabulary like nonsense words and onomatopoeia,

1300

which American mothers (Fernald & Morikawa, 1993) do less frequently. Our findings of an increased use of geminates, moraic nasals, and youon in IDS are likely to reflect Japanese mothers' frequent usage of such lexical items. Lee and Nakayama, based on reports of differences in Korean, Japanese, and American mothers' speech, additionally proposed a role of cultural differences in lexical choice: Korean mothers frequently use verbs to teach actions, Japanese mothers use words related to social actions to teach social skills, and American mothers use nouns to teach object names.

This latter proposal touches upon an important point: the modifications in IDS segment distributions may not be an end of their own, but rather a by-product of other modifications. For example, Trainor, Austin, and Desjardins (2000) found that both emotional adult speech and IDS contain more exaggerated vowel contours than unemotional adult speech, which suggests that they are rather a by-function of emotional expression. Similarly, segment distributions in IDS might be a by-product of lexical choice (cf. Daland, 2012). For example, the higher frequency of the long high-front vowel in IDS is likely related to lexical factors, because in Japanese the word *ii* means 'good'. In the current study, the word *ii* comprised 24% of all long high-front vowels in ADS, while it comprised 42% in IDS.

Another factor to be considered is that IDS may change during development, with caregivers' input adapting to the needs of the infant in a certain stage (Cross, 1977). The age range 1;6 to 2;0 in the current study differs from the one-year-olds in the previous studies, which may impair comparison. Following studies on developmental changes in IDS on the acoustic (e.g. Kitamura, Thanavishuth, Burnham & Luksaneeyanawin, 2001) and semantic (e.g. Snow, 1977) aspects of IDS, future studies should track such changes on the segmental level.

Lastly, IDS is not the only input for infants (Soderstrom, 2007): ADS, as well as siblings' speech, occurs frequently in an infant's environment. It is still not clear to what extent the speech not directly addressed to the infant influences language acquisition. The current study considers both IDS and ADS, providing a starting point for comparing the impact of these speech styles. Further studies in the tradition of Vihman *et al.* (1994) and Chen & Kent (2005) are necessary to investigate the exact relationship between IDS, other speech styles, and segment acquisition.

CONCLUSION AND OUTLOOK

Overall, and consistent with previous studies (Lee *et al.*, 2008; Lee & Davis, 2010), we found evidence for an increased use of segments and association patterns that occur early in production in IDS compared to ADS (fine-tuning). We also found evidence for an increased use of segments

1301

and association patterns that are acquired rather late overall but that are very frequent and acquired early in Japanese (highlighting). Concerning the latter, it is not clear how far this pattern is specific to Japanese or whether it can be generalized across languages. Moreover, some of the other differences between IDS and ADS cannot be explained by fine-tuning or highlighting. These differences could be due to language-specific phonological or lexical factors, or a by-product of other factors. Further research in additional languages and corpora is necessary to assess these alternatives.

Finally, we want to address the potential relevance of such segmental differences between IDS and ADS for language acquisition. Daland (2012) points out that, even if IDS and ADS segment distributions differ, such small differences are unlikely to affect phoneme category learning. To date, there is indeed no study that looks at the effect of small frequency differences in the input. Thus, caregivers' fine-tuning might just be a way caregivers adjust their speech to infants' production capabilities without necessarily impacting phoneme category learning in a significant way.

REFERENCES

Beckman, M. E., Yoneyama, K. & Edwards, J. (2003). Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. *Journal of the Phonetic Society of Japan* **7**, 18–28.

Bernhardt, B. H. & Stemberger, J. P. (1998). *Handbook of phonological development*. San Diego, CA: Academic Press.

Boysson-Bardies, B. & Vihman, M. M. (1991). Adaptation to language: evidence from babbling and first words in four languages. *Language* **67**, 297–319.

Chen, L.-M. & Kent, R. D. (2005). Consonant-vowel co-occurrence patterns in Mandarin-learning infants. *Journal of Child Language* **32**, 507–34.

Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Earlbaum Associates.

Cristia, A. (2013). Input to language: the phonetics and perception of infant-directed speech. *Language and Linguistics Compass* **7**(3), 157–70.

Cross, T. G. (1977). Mothers' speech adjustments: the contributions of selected child listener variables. In K. Snow & C. A. Ferguson (eds), *Talking to children: language input and acquisition*, 151–88. Cambridge: Cambridge University Press.

Daland, R. (2012). Variation in the input: a case study of manner class frequencies. *Journal of Child Language*. Available at: doi:10.1017/S0305000912000372.

Davis, B. L. & MacNeilage, P. F. (1990). Acquisition of correct vowel production. *Journal of Speech and Hearing Research* **33**, 16–27.

Dyson, A. T. & Amayreh, M. M. (2007). Jordanian Arabic speech acquisition. In S. McLeod (ed.), *The international guide to speech acquisition*, 288–99. Clifton Park, NY: Thomson Delmar Learning.

Edwards, J. & Beckman, M. E. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in the acquisition of consonant phonemes. *Language Learning & Development* **4**(2), 122–56.

Ferguson, C. A. (1977). Baby talk as a simplified register. In K. Snow & C. A. Ferguson (eds), *Talking to children: language input and acquisition*, 209–35. Cambridge: Cambridge University Press.

Fernald, A. & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development* **64**(3), 637–56.

Fikkert, P. (1994). On the acquisition of rhyme structure in Dutch. In R. Bok-Bennema & C. Cremers (eds), *Linguistics in the Netherlands 1994*, 37–48. Amsterdam: John Benjamins.

Fox, A. V. (2007). German speech acquisition. In S. McLeod (ed.), *The international guide to speech acquisition*, 386–97. Clifton Park, NY: Thomson Delmar Learning.

Goldstein, B. A. (2007). Spanish speech acquisition. In S. McLeod (ed.), *The international guide to speech acquisition*, 539–53. Clifton Park, NY: Thomson Delmar Learning.

Hamano, S. (1998). *The sound-symbolic system of Japanese* (Studies in Japanese Linguistics, Vol. 10). Stanford, CA: CSLI Publications; Tokyo: Kuroshio Publishing Company.

Howard, S. (2007). English speech acquisition. In S. McLeod (ed.), *The International guide to speech acquisition*, 188–203. Clifton Park, NY: Thomson Delmar Learning.

Ingram, D. (1999). Phonological acquisition. In M. Barrett (ed.), *The development of language*, 73–97. Hove: Psychology Press.

Jakobson, R. (1941/1968). *Child language, aphasia and phonological universals*. The Hague: Mouton.

Kent, R. D. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (eds), *Phonological development: models, research, implications*, 65–90. Parkton, MD: York Press.

Kitamura, C., Thanavishuth, C., Burnham, D. & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development* **24**(4), 372–92.

Kunnari, S. & Savinainen-Makkonen, T. (2007). Finnish speech acquisition. In S. McLeod (ed.), *The international guide to speech acquisition*, 351–63. Clifton Park, NY: Thomson Delmar Learning.

Lee, S. A. & Davis, B. (2010). Segmental distribution patterns of English infant- and adult-directed speech. *Journal of Child Language* **37**(4), 767–91.

Lee, S. A., Davis, B. & MacNeilage, P. (2007). 'Frame dominance' and the serial organization of babbling, and first words in Korean-learning infants. *Phonetica* **64**, 217–36.

Lee, S. A., Davis, B. & MacNeilage, P. (2008). Segmental properties of input to infants: a study of Korean. *Journal of Child Language* **35**(3), 591–617.

Lee, S. A. & Nakayama, M. (2000). Characteristics of maternal speech in Korean: Do Korean and Japanese maternal speech show similar characteristics? In S. C. Howell, S. A. Fish & T. Keith-Lucas (eds), *Proceedings of the Annual Boston University Conference on Language Development* **24**, 486–97. Boston: Cascadilla Press.

Lorwatanapongsa, P. & Maroonroge, S. (2007). Thai speech acquisition. In S. McLeod (ed.), *The international guide to speech acquisition*, 554–65. Clifton Park, NY: Thomson Delmar Learning.

MacNeilage, P. & Davis, B. (2000). On the origin of internal structure of word forms. *Science* **288**(5465), 527–31.

MacNeilage, P., Davis, B., Kinney, A. & Matyear, C. L. (2000). The motor core of speech: a comparison of serial organization patterns in infants and languages. *Child Development* **71**(1), 153–63.

Maekawa, K. (2003). Corpus of Spontaneous Japanese: Its design and evaluation. *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)*, 7–12. Tokyo.

Mazuka, R., Igarashi, Y. & Nishikawa, K. (2006). Input for learning Japanese: RIKEN Japanese mother-infant conversation corpus. *Institute of Electronics, Information and Communication Engineers Technical Report* **16**, 11–15.

Mazuka, R., Kondo, T. & Hayashi, A. (2008). Japanese mothers' use of specialized vocabulary in infant-directed speech: infant-directed vocabulary in Japanese. In N. Masataka (ed.), *The origins of language*, 39–58. Tokyo: Springer.

1303

McLeod, S., Doorn, J. V. & Reed, V. A. (2001). Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology* **10**(2), 99–110.

Mennen, I. & Okalidou, A. (2007). Greek speech acquisition. In S. McLeod (ed.), *The international guide to speech acquisition*, 398–411. Clifton Park, NY: Thomson Delmar Learning.

Murase, T., Ogura, T. & Yamashita, Y. (1992). Ikujigo no kenkyuu (1). Doobutsu meishoo ni kansuru hahaoya no shiyoogo: Ko no geturei ni yoru chigai. [Study of child-rearing vocabulary (1). Mothers' use of animal terms: effects of children's age.] *Annual Bulletin of Shimane University, Faculty of Education* **17**, 37–54.

Ramus, F., Nespor, M. & Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition* **75**, AD3–AD30.

Ratner, N. B. (1996). From signal to syntax – But what is the nature of the signal? In J. Morgan & K. Demuth (eds), *From signal to syntax: bootstrapping from speech to grammar in early acquisition*, 35–150. Hillsdale, NJ: Erlbaum.

Rose, Y. & Wauquier-Gravelines, S. (2007). French speech acquisition. In S. McLeod (ed.), *The international guide* to speech acquisition, 364–85. Clifton Park, NY: Thomson Delmar Learning.

Saito, Y. (1997). *Nihongo no onseigaku nyumon.* [Introduction to Japanese phonetics.] Tokyo: Sanseido.

Shi, R. & Werker, J. F. (2001). Six-month-old infants' preference for lexical words. *Psychological Science* **12**(1), 70–75.

Smit, A. B. (2007). General American English speech acquisition. In S. McLeod (ed.), *The international guide to speech acquisition*, 128–47. Clifton Park, NY: Thomson Delmar Learning.

Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language* **4**, 1–22.

So, L. K. H. (2007). Cantonese speech acquisition. In S. McLeod (ed.), *The international guide to speech acquisition*, 313–26. Clifton Park, NY: Thomson Delmar Learning.

Soderstrom, M. (2007). Beyond babytalk: re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review* **27**(4), 501–32.

Trainor, L. J., Austin, C. M. & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science* **11**(3), 188–95.

Uno, A. (2007). *Kotoba to kokoro no hattatsu to shougai.* [Development and disorders of language and mind.] Tokyo: Nagaishoten.

Vance, T. J. (1987). *An introduction to Japanese phonology.* Albany: State University of New York Press.

Vihman, M. M. (1992). Early syllables and the construction of phonology. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (eds), *Phonological development: models, research, implications*, 393–422. Timonium, MD: York Press.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics* **21**, 61–82.

Vihman, M. M., Kay, E., Boysson-Bardies, B., Durand, C. & Sundberg, U. (1994). External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to one-year-old children. *Developmental Psychology* **30**(5), 651–62.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L. & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition* **103**(1), 147–62.