Sounds of Infant-Directed Vocabulary: Learned from Infants’ Speech or Part of Linguistic Knowledge?

Reiko MAZUKA***, Akiko HAYASHI*** and Tadahisa KONDO****

SUMMARY: Mothers use some segments selectively when talking to infants and young children. We examined why particular segments are favored over others by examining college-aged Japanese adults’ ratings of how good a nonsense word sounds as an item of infant-directed vocabulary (IDV). Adults’ ratings were highly consistent with mothers’ actual use of segments in IDV, as well as the predictions of Jakobson’s principle of maximal contrast. They suggest that Japanese adults possess an intuitive sense of what a good IDV should sound like, which is a part of underlying linguistic knowledge of Japanese phonology rather than learned from children.

Key words: infant-directed speech, infant-directed vocabulary, Jakobson’s principle of maximum contrast, segments, Japanese

1. Introduction

Investigation into how it is possible for human infants to acquire language in a few short years without explicit instruction could offer useful insights into the evolution of language. In particular, research on the nature of speech input is a promising topic. In humans, it is well known that adults modify their speech when they talk to infants and young children (see Soderstrom 2007, for review). This type of speech is called “motherese” or “infant-directed speech (IDS).” IDS occurs in almost every language and differs from adult-directed speech (ADS) in a variety of ways. Most notably, the modifications to the speech sound, such as an overall higher pitch, breathy (soft-sounding) voice, and exaggerated intonation, have been found to be salient to infants even when they are too young to decipher the symbolic meaning of language (Fernald 1993). Modification of speech sounds also occurs when speakers use certain segments more frequently in IDS than ADS. Although this phenomenon occurs widely across languages, it is known to occur especially prominently in languages such as Japanese and Korean (Lee, Davis and MacNeilage 2008, Tsuji, Nishikawa and Mazuka 2014, Hayashi and Mazuka 2017).

It has been proposed that there is a general relationship between the physical structures of sounds and the motivation underlying their use in animal communication, the Motivations Structure Rule Hypothesis (MSRH, Morton 1977), suggesting that investigation into the specific ways mothers modify their speech could shed light on the link between human language and animal communication. In the present paper, we will focus on the segmental aspects of IDS and examine whether there is a relationship between the use of particular segments in specialized vocabulary in IDS (infant-directed vocabulary, IDV) and native speakers’ impression of the word as a good IDV item.

During the past several decades, studies have documented the use of specialized vocabulary as particularly prevalent when Japanese mothers talk to infants and young children (Murata 1960, 1968, Murase, Ogura and Yamashita 1992, Ogura 2006, Tsuji, Nishikawa and Mazuka 2014, Hayashi and Mazuka 2017, Mazuka, Kondo and Hayashi 2017). Importantly, Japanese IDV items often bear little phonological resemblance...
mothers’ use of selective segments does not automatically mean that it is learned from children. An alternative possibility is that native speakers of Japanese have an intuitive sense of what a good IDV form should sound like, including the sense of appropriate segments, reflecting their underlying phonological knowledge of good words for young children. The theoretical support for this view may be found in a classic study by Roman Jakobson (1941/1968), who proposed a universal sequence of development in segmental contrasts. Since the segmental distributions of children’s and adults’ production are governed by the same principle, i.e., the concept of maximal contrast, a positive correlation is expected even for adults who had little chance to learn from children’s early production.

If, as the fine-tuning hypothesis predicts, mothers are producing certain segments more frequently than others because they learned them from children’s production, a mother who has extensive experience interacting with her own infant should have a better sense of which segments should be used in a good IDV than college-age adults whose experience interacting with young children is limited. If, alternatively, native speakers of Japanese have an intuitive sense of what a good IDV should sound like reflecting their underlying phonological knowledge, even college-age adults should have a good sense of what IDV should sound like. In addition, their IDV ratings should be similar to that of mothers.

Mazuka et al. (2017) tested this prediction with respect to the prominent prosodic forms in Japanese IDV; namely three-mora, two-syllable, Heavy-Light (HL) forms (e.g., maN.ma) and four-mora, two-syllable Heavy-Heavy (HH) forms (e.g., waN.waN). In a paper-and-pencil survey study, mothers of young infants were asked to rate a wide range of nonsense words in terms of how well each item sounds like an IDV on a 7-point Likert scale, from “sounds like a very good IDV word” (7) to “does not sound like an IDV word at all” (1). They found that the prominent prosodic forms mentioned above found frequently in Japanese IDV were indeed rated significantly higher as IDV forms than other forms by the mothers.

In a different study, two additional groups of college-aged adults were asked to rate a larger set of words in terms of how good they sound as IDV (The results reported in this paper were collected as part of these two larger rating experiments). Since the stimuli for the paper-and-pencil survey study in Mazuka et al. (2017) were a subset of these studies, it was possible to compare the IDV ratings of mothers and college-aged adults. Using average ratings by mothers and college-aged adults for the 529 test items common to both stud-
ies, the Pearson’s correlation coefficient was calculated ($r=87, p<.001$). This demonstrates that mothers and college-aged adults were strikingly similar in their ratings. Note that this is an exceptionally high correlation for this type of rating study.

In addition, a separate group of 147 college-aged adults (average age = 19.0, age range 17–27, 81 female) was asked to rate the same sets of words that were tested in the paper-and-pencil IDV rating in terms of how good they sound as a Japanese word on a 7-point Likert scale. The correlation between mothers’ ratings of good IDV and college-aged adults’ ratings of good Japanese words was $-0.02$ (n.s.), indicating that there is a clear distinction between words that sound like good IDV and good Japanese word (average age = 19.0, age range 17–27, 81 female).

Taken together, the results of Mazuka et al. (2017) show that native speakers of Japanese have a good sense of what a good IDV should sound like with respect to the prosodic forms of the words, and their ratings mirror the actual usage of these forms in IDS. Furthermore, mothers and college-aged adults were extremely similar in their ratings, indicating that this ability is not something one needs to learn from experience interacting with infants or young children. These results were more consistent with the underlying phonological knowledge account than the fine-tuning hypothesis.

However, the analyses in Mazuka et al. (2017) were limited to the prosodic forms of IDV, and the words’ segmental properties were not examined. As the main claims of Jakobson’s analyses (1941/68) relate to the acquisition of segmental contrasts, it is necessary to examine whether Jakobson’s predictions are supported by the segmental characteristics of IDV as well. In the present paper, therefore, we will focus on the segmental aspects of IDV and ask: i) whether usage of certain segments would reliably contribute to a higher (or lower) ratings of IDV by Japanese adults, and ii) whether the higher ratings for selective segments correlate with children’s early production and/or Jakobson’s predictions. In addition, whether IDV ratings are correlated with mothers’ use of the same segments in IDS will also be examined.

To address these questions, results from two rating experiments are reported in which two separate groups of college-aged adults were asked to rate a number of nonsense words in terms of how good each of them sounds as Japanese IDV. In the first rating experiment, the focus was on the segmental features. Thus, vowels and consonants were systematically controlled, while the range of prosodic form variations was limited. In the second rating experiment, the focus was on the prosodic forms of three- and four-mora words. Thus, the range of vowels and consonants was limited.

2. Rating Experiment 1

The first rating experiment was designed primarily to examine the effects of segments, both vowels and consonants. The stimuli were also varied in terms of word length in morae and syllables and presence or absence of duplicated morae or syllables.

2.1 Material

A total of 707 words were created, which varied in length in morae (1, 2, 3, or 4) and syllables (1, 2, 3, or 4). Of these, 526 were test stimuli, created from one of 25 base syllables. Base syllables were light syllables created by combining one of thirteen consonants ($p$, $t$, $k$, $b$, $s$, $f$, $m$, $n$, $tf$, $ts$, $dz$, $z$ in $za$, and $z$ in $z^2$) and one of the three corner vowels ($a$, $i$, $u$). Consonants were selected so that the effects of manner, place, voicing, and nasality could be examined.

As shown in Table 1, 22 patterns of word forms were created from each base syllable. Prosodic forms were systematically varied by adding one of the non-syllabic morae (long vowel, $H$; geminate obstruent, $Q$; or moraic nasal, $N$). The remaining 181 words were control words, which also varied in length in morae (1, 2, 3, 4) and syllables (1, 2, 3, 4). However, every syllable in a control word contained different vowels and consonants.

Using a Latin Square method, test stimuli were divided into four sets of 144 or 145 items. They were combined with 180 or 181 of the control stimuli so that each set consisted of 324 items total.

2.2 Participants and Procedure

In this experiment, 107 college-aged adults (average age = 20.7, age range 18–27, 59 female) who are native speakers of Japanese from the Tokyo area participated. They had little or no experience interacting with infants or young children. Each participant rated one of the four sets of stimuli, each of which contained 324 words. Participants were asked to rate each word on a 7-point Likert scale, from “This word sounds like very good IDV,” 7, to “This word does not sound like IDV at all,” 1. Five to seven participants were tested in a group, and they were asked to rate words that were presented visually on a computer screen. Words were presented one at a time in hiragana characters except for long vowels,
which were written “—.” The stimuli were presented visually to avoid assigning pitch-accent patterns arbitrarily to a large number of nonsense words.

2.3 Results

Results of the rating experiment were submitted to a series of one- or two-way repeated-measure ANOVAs to analyze the factors discussed below. Comparison of means are referred to as statistically significant when planned contrasts were significant at \( p < .05 \), or post-hoc comparisons of means were significant at \( p < .05 \) with Bonferroni correction.

### 2.3.1 Vowels

As vowels were controlled only among the test stimuli, results of the test stimuli were submitted to a 3 (vowels \( a, i, u \)) x 4 (word length in mora) repeated-measure ANOVA. Results revealed that the main effects of vowel \( (F(2, 212) = 7.90, p < .001) \) and word length \( (F(3, 318) = 29.32, p < .001) \) were significant, while the interaction between the two was not \( (F(6, 636) = 0.93, p < .475) \). This shows that words with /a/ as a base vowel were rated significantly higher than those with /i/ or /u/ \( (4.70, 4.55, 4.53) \). The lack of interaction between word length and vowel shows that /a/ was rated higher than /i/ or /u/ regardless of the word length. This pattern was observed when words with or without heavy syllables were analyzed separately.

### 2.3.2 Consonants

Effects of consonants were analyzed in two steps. First, ratings of all consonants \( 2 \) included in this experiment were submitted to a one-way ANOVA with 13 consonants. Results revealed that the main effect of consonant was highly significant \( (F(12, 1272) = 88.50, p < .001) \). Figure 1 shows the average IDV ratings for consonants from the highest to the lowest. In this figure, the rating of a consonant is statistically different from those that are beyond the arrow originating from the circle. In the second step, consonants were analyzed

### Table 1 Stimuli items used in rating Experiment 1.

<table>
<thead>
<tr>
<th>Word length</th>
<th>N of item</th>
<th>Subtotal</th>
<th>Word form</th>
<th>Example</th>
<th>In Kana</th>
<th>Word length</th>
<th>N of item</th>
<th>Example</th>
<th>in kana</th>
</tr>
</thead>
<tbody>
<tr>
<td>In mora</td>
<td>In syll.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>In mora</td>
<td>In syll.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>25</td>
<td>cv</td>
<td>pa</td>
<td>ば</td>
<td>12</td>
<td>12</td>
<td>re</td>
<td>れ</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>25</td>
<td>cv.cv</td>
<td>papa</td>
<td>ぱぱ</td>
<td>24</td>
<td>24</td>
<td>dori</td>
<td>どり</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>50</td>
<td>cvN</td>
<td>paN</td>
<td>ぱん</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>25</td>
<td>cv.cv</td>
<td>pa, pa, pa</td>
<td>ぱぱぱ</td>
<td>24</td>
<td>24</td>
<td>pukagi</td>
<td>ぷかぎ</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>19</td>
<td>cvQ.cv</td>
<td>paQ, pa</td>
<td>ぱぱぱぱ</td>
<td>12</td>
<td>36</td>
<td>chaNki</td>
<td>ちゃんき</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cvH.cv</td>
<td>paH, pa</td>
<td>ぱぱぱぱ</td>
<td>12</td>
<td>12</td>
<td>niQte</td>
<td>にって</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>119</td>
<td>cvN, cv</td>
<td>paN, pa</td>
<td>ぱんぱん</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cvH, cv</td>
<td>paH, pa</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cvN, cv</td>
<td>pa, paN</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cvH, cv</td>
<td>pa, paH</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cv, cvN</td>
<td>pa, paN</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cv, cvH</td>
<td>pa, paH</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>25</td>
<td>cv, cv, cv</td>
<td>pa, pa, pa</td>
<td>ぱぱぱぱ</td>
<td>25</td>
<td>25</td>
<td>yumarika</td>
<td>ゆまりか</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>188</td>
<td>cvN, cv, cv</td>
<td>pa, pa, pa</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cvN, cv, cv</td>
<td>pa, pa, pa</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cvH, cv, cv</td>
<td>paH, pa</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>19</td>
<td>cvH, cv, cv</td>
<td>paH, pa</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cvH, cv, cv</td>
<td>paH, pa</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cvH, cv, cv</td>
<td>paH, pa</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cvH, cv, cv</td>
<td>paH, pa</td>
<td>ぱぱぱぱ</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>69</td>
<td>cvN, cvN</td>
<td>pa, paN</td>
<td>ぱぱぱぱ</td>
<td>12</td>
<td>60</td>
<td>moNeN</td>
<td>もんてん</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>25</td>
<td>cvH, cvH</td>
<td>paH, paH</td>
<td>ぱぱぱぱ</td>
<td>12</td>
<td>12</td>
<td>iHbuH</td>
<td>いぶー</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>19</td>
<td>cvQ, cvN</td>
<td>paQ, paN</td>
<td>ぱぱぱぱ</td>
<td>12</td>
<td>12</td>
<td>noQtoN</td>
<td>のっとん</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>0</td>
<td>cvN, cvN</td>
<td>paQ, paN</td>
<td>ぱぱぱぱ</td>
<td>12</td>
<td>12</td>
<td>yaNniH</td>
<td>やにー</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>0</td>
<td>cvQ, cvH</td>
<td>paQ, paN</td>
<td>ぱぱぱぱ</td>
<td>12</td>
<td>12</td>
<td>neQhiH</td>
<td>ねっー</td>
</tr>
</tbody>
</table>

Total | 526 | 181 |

Base syllables: pa, ba, sa, ma, na, ʧa, ta, ka, pi, bi, jì, zi, mi, ni, ʧi, ki, bu, su, dzu, mu, nu, ʧu, tsu, ku
2.3.2.1 Nasality

Two consonants /m/ and /n/ were classified as nasals, while the other consonants were classified as orals. When all test stimuli were submitted to a one-way ANOVA, nasals were rated significantly higher than orals \((F(1, 106)=16.95, p<.001)\), even though the numerical difference was slight between nasals (4.93) and orals (4.71). To isolate the effect of nasality, stimuli items that were created from the base syllables /ba/ and /ma/ were selected. When they were compared, words with nasal /ma/ (5.10) \((F(1, 106)=18.89, p<.001)\), but the numerical difference was larger. Interestingly, however, this difference was carried entirely by the words that contained heavy syllables. When words with heavy syllables were not included, the difference disappeared \((F(1, 106)=.01, p<.922)\); the ratings for /ma/ and /ba/ were 5.18 and 5.20, respectively. When only those words with heavy syllables were analyzed, the rating of /ma/ (5.77) was significantly higher than /ba/ (5.04) \((F(1, 106)=30.09, p<.001)\).

2.3.2.2 Manner

For the analysis of manner, /p, b, t, k, m, n/ were classified as stops, /s, z, j/ as fricatives, and /tʃ, ts, dz/ as affricates. When the ratings for all of the test stimuli were submitted to a one-way ANOVA with three manners (stops, fricatives, affricates), the effect of manner was highly significant \((F(2, 212)=134.12, p<.001)\). Post-hoc comparison of means showed that affricates (4.97) and stops (4.93) were rated significantly higher than fricatives (4.16). The same pattern remained when words with and without heavy syllables were analyzed separately.

To further isolate the effects of manner, ratings for minimal pairs of words were examined. For the fricative-stop contrast, only the words with /ta/ and /sa/ were compared. The ratings for /ta/ words (5.31) were significantly higher than those with /sa/ (3.83) when all words were included \((F(1, 106)=142.93, p<.001)\), only words with no heavy syllables \((F(1, 106)=35.08, p<.001)\) (4.76 for /ta/, 3.53 for /sa/), or only words with heavy syllables \((F(1, 106)=154.94, p<.001)\) (5.48 for /ta/, 3.89 for /sa/).

For the fricative-versus-affricate comparison, two pairs of words /ʃi/-/tʃi/ and /su/-/tsu/ were included. A two-way repeated-measure ANOVA with two manners (affricate, fricative) and two pairs (/ʃi/-/tʃi/, /su/-/tsu/) was performed. The main effect of manner \((F(1, 106)=52.54, p<.001)\) and pair \((F(1, 106)=73.26, p<.001)\) and the interaction between the two were all significant \((F(1, 106)=43.50, p<.001)\). Post-hoc analysis of means shows that although the affricate /tʃi/ was rated significantly higher than the fricative /ʃi/, the affricate /tsu/ (4.10) was not rated any higher than the fricative /su/.

---

Figure 1 The average ratings for consonants in Experiment 1, presented from the highest to the lowest. Ratings of a consonant is statistically different from those that are beyond the arrow originating from the circle.
2.3.2.3 Place

For the analysis of place, /p, b, m/ were classified as labials, /t, s, z, dz, n/ as alveolar, /tʃ, ʃ/ and /z/ in /zi/ as palatal, and /k/ as velar. When ratings of all test stimuli were analyzed in a one-way ANOVA with four places (labial, alveolar, palatal, velar), the effect of place was highly significant ($F(3, 318)=164.63, p<.001$). Post-hoc comparison of means revealed that labials (5.15) and palatals (5.11) were rated significantly higher than alveolars (4.31) or velars (4.28).

To further isolate the effect of place, minimal triplets of /pa/, /ta/, and /ka/ were compared directly. In this analysis, a different pattern from the overall place effect emerged. Although the main effect of the place remained significant ($F(2, 212)=76.03, p<.001$), both labial /pa/ (5.39) and alveolar /ta/ (5.31) were rated significantly higher than velar /ka/ (4.14), but there was no significant difference between /pa/ and /ta/. Among words without heavy syllables, labials (5.35) were rated significantly higher than alveolars (4.76), and alveolars in turn were rated higher than velars (3.73). Among words with heavy syllables, alveolars (5.48) and labials (5.40) were rated higher than velars (4.28). Among words with nasal stops /ma/ and /na/, words with labial /ma/ (5.15) were rated significantly higher than those with alveolar /na/ (4.75).

To examine whether the higher rating of labials over velars would change depending on the vowel, minimal-pair words /pa/, /pi/ and /ka/, /ki/ were compared. A two-way ANOVA with two places (labial, velar) and two vowels (a, i) was performed. The results revealed that although the main effects of place ($F(1, 106)=161.17, p<.001$) and vowel ($F(1, 106)=10.63, p<.001$) were significant, the interaction between the two was modest ($F(1, 106)=6.17, p<.05$). Labial /pa/ (5.39) was rated higher than /pi/ (5.02), and /pi/ in turn was rated higher than both of the velars /ka/ (4.14) and /ki/ (4.11). Velars /ka/ and /ki/ did not differ from each other.

2.3.2.4 Voicing

For this analysis, /p, t, k, s, tʃ, ʃ, ts, ts/ were classified as voiceless, and /b, dz, z, m, n/ and /z/ in /zi/ were classified as voiced. Overall, although the main effect of voicing was significant ($F(1, 106)=7.16, p<.01$), the numerical difference was small: voiced (4.80), voiceless (4.70). When minimal pairs /pa–ba/, /sa–za/, /ʃi–zi/, and /tsu–dzu/ were compared, the effect of voicing disappeared ($F(1, 106)=.02, p<.90$), indicating that there is no overall difference in rating between voiced and voiceless consonants.

2.3.3 Effects of Heavy Syllables

Only the test stimuli were analyzed for the effect of heavy syllables. Since the type and position of heavy syllables differ depending on the length of words, analyses were carried out separately for two-, three-, and four-mora words. For two-mora words, a one-way ANOVA with three types of heavy syllables (0, N, H) was performed. The effect of syllable type was significant ($F(1, 106)=304.32, p<.001$). Two-mora monosyllabic words with a long vowel (H) were rated higher (5.61) than two-mora words with two light syllables (4.71). They were in turn rated higher than mono-syllabic words with N (3.83). Among three-mora words, a two-way repeated measure ANOVA with 4 heavy-syllable types (0, N, H, Q) x 3 heavy-syllable positions (0, 2, 3) was performed. The main effect of syllable type was significant ($F(2, 212)=158.72, p<.001$). Words with H (5.58) were rated higher than all other types. Words with Q (5.07) were in turn rated higher than those with N (4.51) or none (4.69). The main effect of position was also significant ($F(1, 106)=71.49, p<.001$). Those in H(eavy)–L(ight) form (5.23) were rated higher than LH (4.77) or LLL (4.69). However, the interaction between syllable type and position was not significant ($F(1, 106)=1.47, p=.224$).

For four-mora words, a two-way repeated measure ANOVA was performed with 7 heavy syllable types (0, N, N–N, Q, Q–N, H, H–H) x 5 positions (0, 2, 3, 4, 2–4). The results revealed that the main effect of syllable type ($F(4, 424)=147.39, p<.001$) and position ($F(2, 212)=6.98, p<.01$) were significant while the interaction between the two was not ($F(3, 318)=.38, p=.769$). Words with two heavy syllables H–H (5.98) were rated higher than all other four-mora words, followed by two-heavy-syllable words N–N (4.99), and three-syllable words with one heavy syllable with H (5.00). Two-syllable words with Q and N (4.07) were rated lower than the remaining three types: four-mora, four-syllable words without a heavy syllable (4.52), one heavy syllable with N (4.31), and one heavy syllable with Q (4.31).

2.3.4 Duplication and Word Length

Reduplication is a prominent property of infants’ canonical babbling (Oller 1980), and it has been noted to occur frequently in IDS cross-linguistically (Ota and Skarabera 2016). Mazuka et al. (2008) found that 65% of IDV items that Japanese mothers reported using in a survey contained some form of repetition. Thus, we tested whether test stimuli that contained duplicated syllables were rated higher than control items.

Results were submitted to a 2 (repeated vs. variable) by 4 (word length, 1, 2, 3, 4 in mora) repeated-measure ANOVA. The main effects of
Sounds of Infant-Directed Vocabulary

Repetition ($F(1, 106)$=403, $p<.001$) and of Length ($F(3, 318)$=6.78, $p<.001$) and the interaction between Repetition and Length ($F(3, 318)$=64.52, $p<.001$) were all significant. Post-hoc comparison of means revealed that words with repeated syllables were rated significantly higher (4.75) than variable ones (3.52).

Among test stimuli that were created by repeated syllables, three- and four-mora words (4.98 and 4.70, respectively) were rated higher than one- or two-mora words (4.01 and 3.43, respectively). When syllables were variable, there were no significant differences in word length (3.71, 3.43, 3.49, and 3.54 for one-, two-, three-, and four-mora words, respectively).

3. Rating Experiment 2

The second rating experiment was designed primarily to examine the prosodic forms of IDV. Accordingly, the stimuli were designed to examine the effects of various repetition patterns in three- and four-mora words. Some of the results for the prosodic forms were reported in Mazuka et al. (2017). In order to accommodate a large number of repetition patterns, it was not possible to include enough variation in vowels and consonants. Nonetheless, segments excluded from Experiment 1 were included: vowels /o, e/, glides /ʃ, w/ and liquids /i, e/ among consonants, and diphthongs /AI/. In addition, some segments were included to examine the effect of kana spelling: i.e., yo-on (e.g., が sa vs. さ ja), sei-on (a syllable written without any superscript, e.g., じ f), daku-on (kana with two dots as a superscript, で ba, generally indicating that the consonant is voiced), and han-daku-on (syllable with a circle as a superscript, は pa).

3.1 Material

In this experiment, different sets of vowels and consonants were included for two-mora words, while three- and four-mora words used the same set of base syllables. For two-mora words, 13 consonants ($p, t, k, b, s, f, z, tf, ts, m, n, dz, zi$) and 3 vowels ($a, i, u$) were used. For three- and four-mora words, 3 vowels ($a, u, o$), and 11 consonants ($p, b, d, s, f, tf, m, n, j, w, r$) were used. Other segments were also used in order to create various repetition patterns, although they are not included in the analyses. As mentioned above, it was not possible to include a full combination of consonants and vowels in this experiment. Thus, 26 base syllables for two-mora words and 19 base syllables each for three- and four-mora words were tested. In addition, not all of the base syllables were used in all variation patterns. This made it difficult to analyze some of the interactions between vowels and consonants as some of the cells contained zero observations.

A total of 1452 nonsense words were created. Of those, 1362 were test items, while 90 were control items. Four-mora words were created by: i) systematically varying the consonants of a base syllable while keeping the vowel constant, ii) varying the vowels of a base syllable while keeping the consonant constant, or iii) varying both consonants and vowels of a base syllable in AAAA, ABAA, AABA, AABAB, AAB, ABBA, ABAB, ABAC, etc. Words with heavy syllables were similarly created by systematically varying the position of non-syllabic morae (special mora, S) in ASAA, AASA, AAAS, ASAS etc. Three-mora words were created by varying just consonants, just vowels, or both vowels and consonants in AAA, ABA, AAB, etc. (See Mazuka et al. (2017) for the details of the stimuli variation patterns). As shown in Table 2, four-mora, four-syllable words contained four light (L) syllables LLLL, and no heavy (H) syllable. Four-mora, three-syllable words had either an HLL, LHL, or LH pattern, in which a heavy syllable contained a long vowel H, geminate obstruent Q, moraic nasal N, or diphthong D. Four-mora, two-syllable words contained two heavy syllables, HH. Similarly, three-mora words were created by systematically varying consonants, vowels, or consonants and vowels. Using a Latin Square method, stimuli were divided into four sets of 363 words each.

3.2 Participants and Procedure

In this experiment, 53 college-aged adults (average age=21.0, age range 18–27, 22 female), none of whom participated in Experiment 1, rated one of the four sets of words, each containing 363 words. The procedures were the same as Experiment 1.

3.3 Results

As in Experiment 1, results of the rating experiment were analyzed in a series of one- or two-way repeated-measures ANOVAs. Results are referred to as statistically significant when planned contrasts were significant at least at the $p<.05$ level, or post-hoc comparisons of means were significant at $p<.05$ with Bonferroni correction. As mentioned above, not every combination of vowels and consonants was used to create base syllables, nor was every base syllable tested in every variation pattern. Consequently, not all interaction was estimable in two-way ANOVAs.

3.3.1 Vowels

Since test stimuli in this experiment included words
that varied in consonants, vowels, or both from the base syllables, we selected a subset of the stimuli that were also tested in Experiment 1, i.e., no consonant or vowel variation within a word, except for the addition of non-syllabic morae for heavy syllables. Note that in addition to the three corner vowels, /o/ was included in this experiment. A repeated-measures ANOVA with 4 (vowels) × 6 (consonant manners) was performed. The main effect of vowel was significant (F(3,156) = 10.65, p < .001). Post-hoc comparison of means showed that words with /a/ (4.73) and /o/ were rated significantly higher than those with /i/ (4.39). Ratings of words with /u/ (4.61) did not differ significantly from the other three. Thus, in both rating experiments, low vowels (a, o) were rated higher than high vowels (i, u).

### 3.3.2 Consonants

As in the vowel analysis, only the words created from a single base syllable were included in the analyses. Figure 2 shows the average ratings for each base syllable for two-mora words without heavy syllables. Figure 3 shows the average ratings for each base syllable for three- and four-mora words without heavy syllables.

For statistical analyses, consonants were grouped together primarily by their manners, so that the interactions between consonants and vowels are estimable. Consonants /p, b, d/ for three- and four-mora words and /p, b, k, t/ in two-mora words were classified as stops.
Affricates included /tʃ/ in three- and four-mora words and /tʃ, ts, dz/ in two-mora words. Fricatives included /s, ʃ/ for three- and four-mora words and /s, ʃ, z/ in two-mora words. Glides /j, w/ and liquid /ɾ/ were included only in three- and four-mora words. For the purpose of this analysis, /m, n/ were classified as nasals. Words with base syllables /pe/ and /pu/ were eliminated, as they appeared only in limited patterns.

As the two-mora words and three- and four-mora words contained different segments, analyses were carried out for each word length separately. For two-mora words, a two-way ANOVA with 3 vowels (a, i, u) x 4 consonants (stop, fricative, affricate, nasal) revealed that the main effect of vowel was not significant ($F(2, 104)=.74, p=.48$), while the main effect of consonant ($F(3, 156)=17.45, p<.001$) and the interaction between the two were significant ($F(8, 97)=3.33, p<.01$). Post-hoc comparisons showed that affricates (5.38) were rated higher than stops (4.46), nasals (4.34), or fricatives (3.88). Fricatives were rated lower than all other consonants, while the difference between stops and nasals was not significant.

For three-mora and four-mora words, a two-way ANOVA with 3 vowels (a, o, u) x 6 consonants (stop, fricative, affricate, nasal, glide, liquid) was estimated. Results for the three-mora words revealed that although the main effect of vowel was significant ($F(2, 104)=3.39, p<.05$), the numerical differences were very small (a=4.89, o=4.85, u=4.81), and individual comparisons among means did not reach significance. The main effect of consonant ($F(5, 260)=21.16, p<.001$) and the interaction between the two were significant ($F(4, 167)=14.53, p<.01$). For four-mora words, the main effect of vowel was not significant.
(F(2, 104)=.02, p=.82), while the main effect of con-
sonant (F(5, 260)=30.62, p<.001) and the interaction
between the vowel and consonant (F(4, 168)=27.45,
p<.001) was. Since we found little differences among
vowels for three- and four-mora words, post-hoc com-
parisons among means were carried out only for conso-
nants. Results revealed that affricates, stops, and nasals
were rated higher than fricatives and liquids. Glides
were rated significantly lower than affricates, but
higher than liquids and fricatives. Figure 4 summarizes
the average ratings for consonant manners by vowels.

3.3.3 Effects of Heavy Syllables

For three-mora words, a one-way ANOVA with three
prosodic forms (HL, LH, LLL) revealed that the effect
of prosodic form was significant (F(2, 104)=38.35,
p<.001). HL words (4.65) and LH words (4.69) were
rated significantly higher than LLL words (4.07),
showing that the presence of heavy syllables resulted
in higher ratings among three-mora words. To examine
the effect of heavy syllable type, a two-way ANOVA
with eight variation patterns and four heavy syllable
types (0, N, Q, H) was performed. Words with a long
vowel (H, 5.26) were rated higher than all others.
Geminates (Q, 4.54) and nasals (N, 4.04) were rated
higher than diphthongs (D, 3.99) or words with no
heavy syllables (4.07).

For four-mora words, a one-way ANOVA with five
prosodic forms (HH, HLL, LHL, LLH, LLLL) re-
vealed that the effect of prosodic form was significant
(F(4, 208)=26.04, p<.001). HH words were rated
higher than LHL (4.27) or HLL words (3.95). LLLL
words (4.41) and LLH words (4.45) were rated higher
than HLL words. HLL words were rated lower than all
others. These results indicated that among four-mora
words, simply containing heavy syllable did not mean
words were rated higher.

To further examine the effect of heavy syllables,
four-mora words with two heavy syllables were se-
lected and their ratings were submitted to a one-way
ANOVA with seven heavy-syllable types (DD, NN,
NH, QN, QH, HN, HH). The results revealed that the
heavy-syllable type was significant (F(6, 312)=25.22,
p<.001). HH words (5.04) were rated the highest, and
it was significantly higher than those with NH (4.55),
NN (4.37), QN (4.25), or DD (3.63). HN (4.81) words
and QH words (4.75) were rated higher than NN, QN,
or DD words. NH words were rated higher than DD
words.

3.3.4 Duplication and Word Length

In Experiment 2, the number of variable stimuli
was only 1/20 of the total stimuli and the test stimuli
and control stimuli were not directly compared via
ANOVA. However, numerical comparison clearly indi-
cates that variable items were rated much lower (2.97)
than test items which included some form of repetition
(4.34). This replicates the result of Experiment 1.

The results of a one-way ANOVA with four levels of
length (1, 2, 3, 4) revealed that the length was signifi-
bears out the results of the two rating experiments. We will discuss whether these predictions are consistent with children's early production, Jakobson's prediction, where the nasal stops were rated highest, the higher rating of two-mora words over four-mora words was reversed from Experiment 1.

3.3.5 Effects of Kana Orthography
Stimulus words were classified into four categories of kana orthography based on the base syllable. Yo-on included /ʃ, tʃ/ with a vowel /a, u, o/, syllables with /p/ were classified as han-daku-on, /b, z, d/ as daku-on, and the remaining syllables were classified as sei-on. A one-way ANOVA with four types of kana orthography showed the main effect of kana was significant ($F(3, 156) = 66.96$, $p < .001$). Words with han-daku-on (5.37) and yo-on (5.15) were rated higher than daku-on (4.71), which in turn was rated higher than sei-on (4.25).

4. Discussion
In the present paper, we investigate why certain segments are favored over others when mothers talk to infants and young children. Using the results from two rating experiments by college-aged native speakers of Japanese, we examined the validity of two alternative accounts: mothers' use of the selected segments is learned from their experience of interacting with young children (i.e., the fine-tuning hypothesis) or it is based on the linguistic knowledge of Japanese native speakers.

If the latter alternative is correct and Japanese native speakers have an intuitive sense of what a good IDV should sound like, reflecting their phonological knowledge, college-aged adults' ratings of IDV should correlate positively with children's early production as well as the patterns predicted by Jakobson (1941/1968), even when they have had little opportunity to learn from young children's production. Given that mothers' production of segments in IDS is known to correlate with children's early production, adults' ratings should also correlate positively with mothers' production. Alternatively, if the first alternative is correct, as the fine-tuning hypothesis predicts, no systematic correlations are expected between college-aged adults' ratings and children's early production, Jakobson's prediction, or mothers' use of segments in IDS. In the following sections, we will discuss whether these predictions are borne out by the results of the two rating experiments for each of the phonological properties in turn.

4.1 Vowels
Studies examining children's early production have found that /a/ is a vowel that is produced from the earliest age, followed by /i, u, then /e, o/ appear last. Children's production of /a/ becomes clearly recognizable at the earliest age, followed by /i, u/ then /e, o/ (Kirikae and Sawashima, 1978). It is also the vowel produced most frequently by young children (Ichishima, 2009). In IDS, Tsuji et al. (2014) also reported that /a/ was produced significantly more frequently in IDS than ADS. Thus, the high ratings of /a/ in the rating study are consistent with children's production as well as mothers' production in IDS.

In contrast, /o/ was rated as high as /a/ in Experiment 2. This is not consistent with children's late production or the IDS data, in which /o/ does not occur more frequently in IDS than ADS.

4.2 Consonants
According to Nagasawa (1987), children's earliest production includes oral and nasal labial stops /m/ and /b/, followed by /n/ and other stops /p, t, k, d/. Glides /j, w/, affricates /f, dʒ/, and the liquid /ɾ/ appear next. Fricatives /ʃ, s/ and affricates /dz, ts/ are among the consonants that appear last. Ichishima (2009) also reported that /b, m, n, p/ were the consonants that are produced earliest by children, while /k/ and /tʃ/ were produced most frequently by children who are at the 30–60-word stage.

As can be seen in Figures 1 to 4, the results of the rating experiments are generally consistent with the order of children's production. The labials and stops were rated higher while fricatives and affricates tended to be rated lower. An exception to these general patterns was the affricate /tʃ/, which was rated among the highest even though it is not a consonant that appears at the earliest stage of children’s production. We will return to /tʃ/ below.

Regarding nasality, IDV containing nasal stops was rated higher than those without nasals. This is consistent with children’s production, where nasal stops are among the earliest to be produced by children. It is also consistent with Jakobson’s prediction, where the nasal/ oral distinction is predicted to be the earliest to emerge. Interestingly, however, nasal segments appeared more frequently in ADS than IDS.

Among the manners, stops, nasals, and affricates were rated higher than others, glides were in the middle, and fricatives and liquids were rated lower than the
others. The relative status of stops versus fricatives was clearly supported by all the data we obtained. When the minimal pair of a stop /ta/ and a fricative /sa/ was compared, /ta/ was rated higher than /sa/. This pattern is consistent with children’s production and the prediction of Jakobson’s principle of maximum contrast. It is also consistent with mothers’ production in IDS. The relative standings of glides and liquids, which were rated at the medium level, were generally consistent with children’s production. They appear later than nasals and stops, but earlier than fricatives.

The results for affricates were mixed. Although the overall rating of affricates was high, this was due primarily to the very high ratings for /tʃ/. This was not predicted by children’s production or by Jakobson’s theory. The other affricates /ts, dz/ were rated among the lowest, consistent with children’s production data. According to Ichishima (2009), although /tʃ/ was not among the consonants that appeared earliest, it was one of the consonants most frequently produced by children at the 30–60-word stage, when other affricates are rarely produced. This indicates that /tʃ/ is an easier affricate for children to produce compared to other affricates. Thus, the high rating of /tʃ/ relative to other affricates is consistent with children’s production patterns. Since Tsuji et al. (2014) did not report the frequency of individual affricates, we are unable to estimate the frequency of /tʃ/ versus /ts, dz/ in IDS and ADS.

Among four places of articulation, labials and palatal consonants were rated higher than fricatives or velars. Higher ratings of labials over velars are consistent with Jakobson’s predictions. In contrast, palatal consonants were rated as high as labials while alveolars were rated as low as velars. These patterns are not consistent with Jakobson’s prediction that palatal consonants appear later than coronals or dorsals, nor are they consistent with children’s production, where alveolar consonants tend to appear early, while palatal consonants /tʃ, dʒ, tʃ/ appear later.

There is an interesting discrepancy between the IDV ratings and actual occurrence of velar stops by children or mothers. In the literature, the coronal place of articulation has been argued to be the default position among labial, coronal, and dorsal positions (Paradis and Prunet 1991). Japanese is an exception to this pattern, as the velar stop /k/ occurs more frequently than labial /p/ or coronal /t/ both in IDS and ADS (Tsuji, Nishikawa and Mazuka 2010, Tsuji, Mazuka, Cristia and Fikkert 2015). Consistent with this pattern, /k/ emerges as early as labial and alveolar stops /p, t, d/ in children’s production. Nonetheless, Japanese adults’ rating of /k/ was lower than /p/ or /t/, which was more consistent with the universal pattern and Jakobson’s prediction than the language-specific pattern of Japanese.

4.3 Heavy Syllables
Among three-mora words, a heavy-light word form was rated highest, while among four-mora words, two-syllable heavy-heavy words were rated highest, as reported in Mazuka et al. (2017). This is consistent with the actual occurrence of these word types in Japanese IDV as well as children’s early production (Mazuka et al. 2008) and mothers’ use of these forms in IDS, even though Jakobson’s theory is not applicable to the prosodic forms of IDV.

4.4 Duplication and Word Length
In both experiments, words in which the same syllables were duplicated were rated higher than those with variable syllables. This is consistent with the fact that reduplication is a prominent pattern in infants’ babbling, as well as the high frequency of duplication in IDS in Japanese (Mazuka et al. 2008) and other languages (Ota and Skarabera 2016).

In both experiments, three-mora words were rated highest, while four-mora words were rated as high as three-mora words in Experiment 1 and two-mora words were rated as high as three-mora words in Experiment 2. In Japanese IDS, three-mora words occur most frequently among nouns, while two-mora and four-mora words occur relatively more frequently in ADS and as nouns in newspaper articles (Mazuka et al., 2008). Again, Japanese adults’ intuitive ratings of IDV in terms of length are consistent with the actual occurrence in IDS. Although these are characteristics that were not included in Jakobson’s theory, college-aged adults’ ratings were generally consistent with children’s production.

4.5 Yo-on
Words that are written in han-daku-on (including only /p/) and yo-on were rated higher than the others. Although yo-on is defined by the orthography, some of them represent palatalized consonants or affricates. They are often associated with casual, friendly speech and speech directed to young children and infants. Tsuji et al. (2014) found that yo-on occurred significantly more frequently in IDS than ADS. As shown in Figures 2 and 3, all yo-on syllables (/tʃa, tʃu, tʃo, ʃa, ʃu, ʃo/) were rated quite highly. It appears as though yo-on gives rise to an impression among Japanese adults that the sounds are child-friendly. Given that many of the yo-on syllables are palatals and affricates, the
palatal affricate \(/tʃ/\) might have been judged as child-friendly even though it is not written in \(yo-on\). Yo-on is a language-specific aspect of Japanese, and generally high ratings of \(yo-on\) items are not necessarily consistent with actual children’s production. This may be one aspect of Japanese adult native speakers’ ratings that is learned through the experience of learning Japanese orthography.

4.6 Is It Learned?

Although there were notable exceptions, college-aged adults’ ratings of IDV by their segmental properties were generally consistent with children’s early production patterns, as well as the predictions of Jakobson’s principle of maximum contrast. The ratings were also consistent with mothers’ use of these segments in IDS. These results demonstrate that college-aged adults, who have had little experience interacting with young children, were highly sensitive to the segmental properties of a word that contribute to a good-sounding IDV. As mentioned above, the fine-tuning hypothesis assumes that the correlation between children’s early production and mothers’ production in IDS is the result of mothers’ fine-tuning their production to children’s production. Yet, it is difficult to imagine how the college-aged adults in our study with little experience with young children could have tuned their judgments to children’s production patterns. Recall that college-aged adults’ ratings of IDV items were strikingly similar to those of mothers (\(r = .87\), Mazuka et al. 2017). Taken together, the results indicate that adults’ judgments, children’s development of segmental production, and mothers’ production of these segments in IDS arise from the same source, rather than adults’ judgment and mothers’ production being learned from children’s production. Jakobson’s principle of maximum contrast argues that children’s early production is guided by the same principle that guides adults’ production. Our results are more consistent with this view than the fine-tuning hypothesis.

At the same time, some of our results deviated from his prediction. For example, the palatal affricate \(/tʃ/\) was one of the highest-rated segments. This may be related to \(yo-on\) orthography, which is a language-specific feature of Japanese. Another example is the case of the velar stop \(/k/\). Unlike the majority of languages of the world, velar stop \(/k/\) occurs more frequently than alveolar \(/t/\) or labial \(/p/\) in Japanese. Nonetheless, Japanese adults’ ratings were consistent with the universal pattern of higher ratings for \(/p/\) and \(/t/\) than \(/k/\), as predicted by Jakobson. Further research is needed to account for why some of the language specific characteristics are strong enough to override the universal pattern or vice versa.

5. Conclusion

In the present paper, we examined why mothers would use selected segments in IDS through two rating experiments with Japanese adults. The results reveal that the college-aged adults’ ratings were highly consistent with young children’s early production patterns and the prediction of Jakobson’s principle of maximal contrast. They are also consistent with mothers’ use of these segments in IDS as well as mothers’ own ratings of these segments. These results suggest that Japanese adults (mothers and non-mothers) possess an intuitive sense of what a good IDV should sound like, and this sense is a part of the underlying linguistic knowledge of Japanese phonology.

As mentioned in the Introduction, Morton’s MSRH proposed that there is a general relationship between the physical structures of sounds that is common across different species. If this applies to human vocal communication, mothers’ motivation to favor certain segments in their speech to infants may share this tendency. It will be an interesting topic to further investigate whether any of the phonological properties in human IDS share the characteristics of animal vocal communication.

Acknowledgements

Research reported in this paper was funded in part by Grant-in-Aid for Scientific Research S (16H06139), and on Innovative Areas #4903 (Co-creative Language Evolution), 17H06382 to the first author, and by the NINJAL collaborative research project “Cross-linguistic Studies of Japanese Prosody and Grammar.” We thank Haruo Kubozono for his suggestion to examine Jakobson’s predictions in this study.

Notes

1) Although the survey was targeted for mothers of young infants, we received responses from five fathers in addition to 146 mothers, and one without gender information. None of the responses were excluded, but the group was referred to as “mothers” as the survey was originally designed.

2) In the statistical analyses of consonants, \(/z/\) that appeared in a \(/zi/\) syllable was treated as a separate category.
from /z/ in /za/ syllables because /z/ in /zi/ syllables can be realized either as a fricative [z] or an affricate [dz].

3) The vowel /e/ was used by mistake only in /pe/ syllables. The results for /pe/ words were included in Figure 3, but they were removed from the statistical analyses.

4) Three-mora words were varied in eight patterns. See Mazuka et al. (2017) for details.

References


Tsui, S., R. Mazuka A. Cristia and P. Fikkert (2015) “Even at 4 months, a labial is a good enough coronal, but not vice versa.” Cognition 134, 252–256.


(Received May 2, 2017, Accepted Jun. 15, 2017)