A Longitudinal Analysis of the Spectral Peaks of Vowels for a Japanese Infant

Kentaro Ishizuka, Ryoko Mugitani, Hiroko Kato, and Shigeaki Amano

NTT Communication Science Laboratories, NTT Corporation
2-4 Hikari-dai, Keihanna Science City, Kyoto 619-0237 Japan

{ishizuka, mugitani, katohi, amano}@cslab.kecl.ntt.co.jp

Abstract

This paper describes a longitudinal analysis of the vowel development of a Japanese female infant between the ages of 4 and 60 months. The speech data were natural spontaneous speech recorded for at least an hour per month. Vowels were extracted from the data according to their phoneme labels by two transcribers. The two lower spectral peaks of the vowels were estimated by linear predictive coding method. Analyses of the mean values of the spectral peaks suggest that the values change with age according to the articulation positions of the tongue and the rapid development of the vocal tract. The values change greatly until the age of 18 months. In addition, the distribution of the two spectral peaks for individual vowels shows that the vowel space expands with age, and this also continues until the age of 18 months. Another investigation of the vowel distributions based on discriminant analysis confirms that the differentiation of the spectral peak distributions also continues until around the age 18 months in terms of a physical phenomenon. These results suggest that the infant becomes able to produce discriminable vowels corresponding to the rapid increase in the vocal tract length at around that age.

1. Introduction

Children’s speech has long been analyzed acoustically with a view to understanding the process of speech development. Eguchi and Hirsh [1] acoustically analyzed the speech of children between the ages of 3 and 13 years, and showed that the first and second formant frequencies (F1 and F2) descend with increases in age, especially between the ages of 3 and 5. Busby and Plant [2] analyzed the F1 and F2 of the vowels produced by children of 5 to 11 years old, and also showed that the F1 and F2 values descend with age. Recently, Lee et al. [3] systematically analyzed speech data obtained from a large number of children between 5 and 17 years old, and reported the acoustic changes that occurred after 5 years of age.

In contrast, there have been few acoustical analyses of the frequency spectra of the speech of infants less than 3 years old. Kent and Murray [4] analyzed infant vocalic utterances at 3, 6, and 9 months, and showed that the range of the F1 and F2 values increases with age. However, they did not introduce phonemic aspects into the analysis. Gilbert et al. [5] analyzed the F1 and F2 values of infants of 15 to 36 months from the perspective of the articulation points of the tongue. Their results indicate that the F1 and F2 values remain unchanged prior to 24 months of age, and decrease significantly from 24 to 36 months. Vocalic development in the early stages has a wide interindividual variability, and this has stimulated the need for longitudinal data collection and analysis. As regards longitudinal studies, Buhr [6] collected speech produced by an infant aged 4 and 16 months, and the analysis of formant frequency plots over time revealed the emergence of a well-developed vowel triangle. Bond et al. [7] conducted a longitudinal analysis of a child’s speech from 17 to 29 months, and showed that vowel formants shift over time towards a precisely defined vowel space.

From an anatomic point of view, Vorperian et al. [8] investigated the anatomic restructuring of the vocal tract using magnetic resonance imaging obtained between the ages of 2 weeks and 6 years 9 months. Their results indicate that there is accelerated growth of the vocal tract structures between birth and 18 months. These results suggest that the developmental acoustic changes in vocalization probably reflect such accelerated growth. Although the previous longitudinal studies [6][7] both cover ages of 4 to 29 months, there has been no single longitudinal observation of an infant for the period of the growth acceleration indicated by Vorperian et al. [8].

This study provides a longitudinal analysis of vowel development of a Japanese female infant between the ages of 4 to 60 months in order to reveal the relationship between acoustic vowel change and anatomical growth. The acoustical analyses are conducted in terms of the changes that occur in the spectral peaks of vowels with age. Section 2 describes the infant speech data properties, the acoustic analysis method used for extracting spectral peaks, and statistical analyses of the spectral peaks from perceptual and physical standpoints. In section 3 we discuss our results, and section 4 concludes this study.

2. Analysis

2.1. Infant speech data

We used the speech data of an infant (C) in the NTT Japanese infant speech database [9]. C was born in Japan and grew up in the Tokyo area. The utterances of the infant and parents were recorded in a room in their house with a digital audio recorder (SONY TCD-D10) and a stereo microphone (SONY ECM-959) with 16-bit quantization and at a sampling rate of 48 kHz. The microphone was held by a parent or placed in a microphone stand during recording. The infant and parents were not required to undertake any particular task for the recording so that their utterances would occur in natural situations in daily life. The speech data were recorded for at least an hour per month. The recording period was from 1 to 60 months of age, and the total recording time was 68 hours.
In this study, we only used the speech data obtained at 4, 6, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 55, and 60 months. The speech data were down-sampled to a sampling rate of 16 kHz, and every 1 ms segment was phoneme labeled by a Japanese transcriber. The labels were then checked and corrected by another Japanese trained transcriber. Based on these phoneme labels, we automatically extracted the five Japanese vowels /a/, /e/, /i/, /o/, and /u/ with temporal lengths of 50 to 300 ms. These vowel data were divided into 10 groups each representing a 6 month age period. Table 1 shows the number of vowels for each of these groups.

2.2. Spectral peak extraction

The vowel waveforms extracted as described above were wholly analyzed by 12-order linear predictive coding (LPC) analysis regardless of the temporal lengths of the waveforms. Frequency spectra were estimated based on the LPC coefficients, and the two lower spectral peaks were extracted automatically from each frequency spectrum. Henceforth in this paper, the lower spectral peak between the two spectral peaks is denoted as "\( f_1 \)", and the other is denoted as "\( f_2 \)". Both spectral peak values in Hz are transformed into values on the Bark scale using the approximated formula given below [10];

\[
B = 13 \tan^{-1}(0.76f) + 3.5 \tan^{-1}(f/7.5)
\]

where \( f \) is the frequency value in Hz, \( B \) is the corresponding value in Bark.

2.3. Mean values of the spectral peaks

The upper part of Fig. 1 shows the mean \( f_1 \) values as a function of the month age groups. Linear regression analyses with the month ages as explanatory variables reveal significant changes in the value as described below. (1) The \( f_1 \) values for vowel /a/ increase until 35 months of age (\( p < .0001 \)), and subsequently decrease (\( p < .0001 \)). (2) The values for vowel /e/ remain unchanged until 35 months of age, and then decrease (\( p < .0001 \)). (3) The values for vowel /i/ decrease until 18 months of age (\( p < .0005 \)), then remain unchanged from 20 to 35 months, and finally decrease after 40 months (\( p < .0005 \)). (4) The values for vowel /o/ increase from 8 to 20 months (\( p < .0001 \)), whereas they decrease after that age (\( p < .0001 \)). (5) The values for vowel /u/ first decrease until 14 months (\( p < .0001 \)), then increase from 16 to 30 months (\( p < .0001 \)) before reaching a constant value.

The lower part of Fig. 1 shows the mean values of \( f_2 \) as a function of the month age groups. Again linear regression analyses with the month ages as explanatory variables reveal significant changes in the value as described below. (1) The \( f_2 \) values for vowel /a/ remain unchanged until the age of 16 months, and then decrease (\( p < .0001 \)). (2) The values for vowel /e/ increase until the age of 30 months (\( p < .0001 \)), and then decrease (\( p < .0001 \)). (3) The values for vowel /i/ increase until the age of 40 months (\( p < .0001 \)), and then remain the same. (4) The values for vowel /o/ decrease throughout the entire period (\( p < .0001 \)). (5) The values for vowel /u/ decrease until the age of 16 months (\( p < .0001 \)), and then remain unchanged.

2.4. Phonemic distribution of the spectral peaks

Figure 2 depicts the \( f_1-f_2 \) planes for each month age group. Each dot indicates one of the vowels, and 50 % probability density ellipses for each vowel are plotted. Although there are large overlaps among the density ellipses before the age of 12 months, the overlaps become smaller with age due to the spread of centroids and a reduction in the distributions of the density ellipses. In particular, the overlaps between the density ellipses for vowels /a/ and /i/, and vowels /i/ and /o/ have almost disappeared after the age of 12 months. These plots clearly show the change in the range of the vowel space with age.

### Table 1: Number of analyzed vowel data by month age group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Data</th>
<th>/a/</th>
<th>/e/</th>
<th>/i/</th>
<th>/o/</th>
<th>/u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-6</td>
<td>4, 6</td>
<td>221</td>
<td>261</td>
<td>26</td>
<td>30</td>
<td>354</td>
</tr>
<tr>
<td>7-12</td>
<td>6, 10, 12</td>
<td>589</td>
<td>382</td>
<td>198</td>
<td>65</td>
<td>448</td>
</tr>
<tr>
<td>13-18</td>
<td>14, 16, 18</td>
<td>913</td>
<td>372</td>
<td>469</td>
<td>398</td>
<td>528</td>
</tr>
<tr>
<td>19-24</td>
<td>20</td>
<td>497</td>
<td>124</td>
<td>265</td>
<td>277</td>
<td>180</td>
</tr>
<tr>
<td>24-30</td>
<td>25, 30</td>
<td>665</td>
<td>250</td>
<td>402</td>
<td>243</td>
<td>233</td>
</tr>
<tr>
<td>31-36</td>
<td>35</td>
<td>222</td>
<td>101</td>
<td>89</td>
<td>153</td>
<td>80</td>
</tr>
<tr>
<td>37-42</td>
<td>40</td>
<td>727</td>
<td>394</td>
<td>303</td>
<td>512</td>
<td>213</td>
</tr>
<tr>
<td>43-48</td>
<td>45</td>
<td>1,628</td>
<td>725</td>
<td>626</td>
<td>1,075</td>
<td>483</td>
</tr>
<tr>
<td>49-54</td>
<td>50</td>
<td>277</td>
<td>128</td>
<td>95</td>
<td>222</td>
<td>71</td>
</tr>
<tr>
<td>55-60</td>
<td>55, 60</td>
<td>531</td>
<td>248</td>
<td>216</td>
<td>383</td>
<td>111</td>
</tr>
</tbody>
</table>

**Figure 1:** Mean values of \( f_1 \) and \( f_2 \) by month of age.

---

1 The \( f_1 \) and \( f_2 \) basically correspond to the first and second formant (F1 and F2) of the vowels. However, automatic estimation may result in subtle differences from the precise formant peaks, especially when for vowels with high fundamental frequencies such as infant speech. Therefore, we use the terms \( f_1 / f_2 \) rather than F1 / F2 in this paper.
2.5. Physical distribution of the spectral peaks

To investigate changes in the vowel distribution based on the physical, transcription-independent characteristics of \( f_1 \) and \( f_2 \), we employed an analysis method motivated by discriminant analysis. First, we fixed three reference points, namely the centroids of the adult Japanese female’s (C’s mother’s) “corner” vowels of the vowel space i.e. /a/, /i/, and /o/. Henceforth, these centroid vectors are called \( c_i \) (\( i = 1, 2, 3 \)) which have \{\( f_1, f_2 \)\} values of \{6.85, 12.78\}, \{5.71, 14.85\}, and \{6.18, 12.20\} in Bark, respectively. The positions of \( c_i \) on the \( f_1-f_2 \) plane are plotted in Fig. 3. Using these vectors as a basis, we then investigated the problem of classifying the data \( d_j \) indicating each vowel (\( j = 1 \ldots J \); \( J \) is the total number of vowels in a month age group) into one of three categories:

\[ \Pi_1: \text{the category whose centroid is } c_1, \text{ (vowel } /a/) \]

\[ \Pi_2: \text{the category whose centroid is } c_2, \text{ (vowel } /i/) \]

\[ \Pi_3: \text{the category whose centroid is } c_3, \text{ (vowel } /o/) \]

Then, classification was performed by the divergence \( |d_j - c_i| \).

We introduced the discriminant function:

\[ D_i^k = |d_j - c_i| - |d_j - c_k| \quad k = \text{mod}(i,3)+1 \]

where \( \text{mod}(i,3) \) is the modulus operator. The actual classification procedure is; if \( D_i^k \) (where \( k = 2 \)) > 0, then we choose category \( \Pi_k \). We performed the classification for \( \Pi_1 \) and \( \Pi_2 \), \( \Pi_2 \) and \( \Pi_3 \), and \( \Pi_3 \) and \( \Pi_1 \). We calculated all the \( D_i^k \) values for each month age group, and generated histograms of the values for each \( i \). Figure 4 shows the histograms. The histograms reveal that the \( D_i^k \) values are broadly distributed until 12 months of age, they are then biased to the two ends of the histograms. This means that the vowels are clustered around either of the two reference points. This analysis confirms that the vowels form vowel space with age solely through the use of physical acoustic features.

3. Discussion

3.1. Mean values of the spectral peaks

The \( f_1 \) values for high vowels i.e. /i/ and /u/ decrease until 14 months of age. On the other hand, the \( f_2 \) values for middle and low vowels i.e. /a/, /e/, and /o/ increase or remain the same until 20 months. \( f_1 / f_2 \) basically correspond to F1 / F2. The F1 value is subject to the vocal tract length and the height of the articulation position of the tongue, and so high vowels have a lower F1 value. Thus, the developmental decrement in the \( f_1 \) values of vowels /i/ and /u/ possibly reflects the vocal tract length and vertical tongue elevation, which leads to the distinct articulation of high vowels. The vocal tract length increment and the tongue elevation are assumed to be achieved by the rapid growth of the pharyngeal cavity and the rapid descent of the larynx and hyoid until 12 months of age [8]. The F1 value also reflects mouth opening, and \( f_1 \) increments in the middle and low vowels possibly result from the expansion of the mouth opening range caused by the rapid development of the mandible [8]. The \( f_1 \) decrements after the age of 35 months correspond to the F1 decrements reported in previous studies [1] [2].

The \( f_2 \) values for the front vowels i.e. /e/ and /i/ increase until 30 months of age. On the other hand, until 16 months, the \( f_2 \) values for the central vowel /a/ remain unchanged, and the values for the back vowels i.e. /a/ and /u/ decrease. F2 corresponds to the horizontal tongue advancement to form an articulation position. The tongue lengthens rapidly until 16 months [8], and this age coincides with the \( f_2 \) changes in the central and back vowels. Therefore, these acoustic changes may relate to the differentiation of vowel articulation by the tongue along the horizontal axis. Interestingly, the \( f_2 \) values for vowel
18 months, suggesting that the infant rapidly learned to produce discriminable vowels until around that age. Although the phoneme labels are subjectively provided by the transcribers, the discriminant analysis of the vowel distributions based on the physical acoustic characteristics of the vowels suggested that the vowels clustered in the vowel space until around 18 months. These results confirmed that the vowel space expansion occurs until around 18 months in terms of the listener’s perception and also the physical acoustic phenomenon. These results also suggest that the expansion reflects the rapid development of the vocal tract until around the age of 12 to 18 months [8]. Although this result also coincides with the vowel space depicted by Buhr [6] or Bond et al. [7], our study further revealed that the vowel space expansion continues after that age.

### 4. Conclusion

We conducted a longitudinal study of vowel development for a Japanese female infant aged between the ages of 4 and 60 months in terms of changes in the spectral peaks of vowels. The results revealed that there are longitudinal trends of vowel development that correspond to the articulation positions of the tongue and the vocal tract length. The results also suggest that the early acoustic development of the vowels corresponds to the rapid anatomic development of the vocal tract until around the age of 12 to 18 months [8], and that discriminable vowel space continues to be formed until around that age.

**Acknowledgements:** The authors thank Mr. Ken-ichi Sakakibara for comments on speech production, and Prof. Kazuhiro Kakehi for valuable suggestions about the physical analysis of the vowel acoustic features.

### References


