THE ACOUSTICS OF VOWELS IN JAPANESE WOMENS’ SPEECH TO INFANTS AND ADULTS

Jean E. Andruski*, Patricia K. Kuhl†, Akiko Hayashi‡

*Wayne State University, Detroit MI, †University of Washington, Seattle WA, ‡Tokyo Gakugei University, Tokyo, Japan

ABSTRACT

The point vowels ([i], [u] and [a]) reach more extreme acoustic targets in infant-directed (ID) speech than in adult-directed (AD) speech. This increases the distance between vowel categories along the F1-F2 dimensions and may assist young infants in separating ambient vowel sounds into native language categories. In AD speech, citation form vowels also reach more extreme acoustic targets than vowels from conversation. Are ID vowels equivalent in structure to AD citation form vowels? This study examines the acoustic structure of point vowels in the speech of 14 Japanese women, and compares AD citation form vowels to ID conversation form vowels. ID vowels reach more extreme acoustic targets than AD citation form vowels, and vowel triangle size increases significantly in ID speech over AD citation form speech. When vowel structure is compared across 4 languages, however, ID vowels retain their language-specific character.

1. INTRODUCTION

Cross-linguistically, the point vowels [i], [u] and [a] reach more extreme acoustic targets in mothers’ speech to infants than in speech to another adult [2]. This results in greater acoustic separation between vowel tokens belonging to different categories in infant-directed (ID) speech than in adult-directed (AD) speech. Fernald describes ID speech as a speech register intended to conform to the infant’s level of linguistic competence [1]. Lindblom’s H&H Theory also suggests that whenever adults are tacitly aware that a listener lacks information that could assist with signal interpretation, they tend to hyperarticulate, i.e. produce acoustically very well-defined tokens [3]. Clearly, pre-linguistic infants are listeners who lack information that could assist with signal interpretation, and exaggeration of the acoustic properties by which vowels are recognized in ID speech may facilitate the infant’s task of dividing ambient vowel sounds into native-language categories.

In AD speech, conversational tokens of vowels often undershoot the expected acoustic target. Lindblom refers to this as hypoarticulation [3]. Even in AD speech, however, more extreme F1 – F2 values tend to be reached when words are produced in isolation or in careful speech. These citation form F1 – F2 values are, in fact, generally viewed as the expected or intended acoustic targets for vowels. How does the acoustic structure of vowels in ID speech compare with the acoustic structure AD citation forms?

2. METHOD

2.1. Subjects and Target Words

Subjects for this study were 14 native-speaking Japanese mothers with an infant of either 5 ½ months or 8 ½ months of age. AD citation form tokens were recorded in interactions with another female native speaker of Japanese. To obtain AD citation form tokens, each mother was asked to identify 3 toys whose names contained one of the point vowels [i], [u] or [a] (see Table 1). All mothers produced vowel tokens either in isolation or in the context “[Target] desu” (“It’s a / They’re [target].”) To obtain ID tokens of the vowels, each mother was then asked to show the 3 toys to her infant and play with the infant for several minutes using the toys.

All 3 target words were content words that consisted of two syllables and contained a point vowel in the first, stressed syllable. All point vowels were in b[V][alveolar] context.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Target Word</th>
<th>English Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>[i]</td>
<td>bi:zu beads</td>
<td></td>
</tr>
<tr>
<td>[u]</td>
<td>bu:tsu boots</td>
<td></td>
</tr>
<tr>
<td>[a]</td>
<td>batto bat</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Target words containing the point vowels [i], [u] and [a].

2.2. Equipment

Recordings were made in a sound treated room using a Sony TCD-D8 Digital Audio Tape recorder and a Sony Electret Condenser Microphone ECM-MS907. Target words were sampled for acoustic analysis on a Gateway Computer using Kay’s CSL (Computerized Speech Laboratory) system and a sampling rate of 10,000 Hz.

2.3. Acoustic Analysis

All target word tokens produced by each mother were included in the analysis unless the vowel in the first syllable overlapped with noise, infant vocalizations, or speech from another adult. Tokens whose amplitude was too low were also excluded from the analysis. For each point vowel token included in the study, the first and second formants (F1 and F2) were measured at the durational center of the vowel. Vowel duration was measured from formant onset to formant offset.

Formant measurements were made using narrowband spectrograms supplemented by wideband spectrograms plus FFT and LPC analyses. Narrowband spectrograms were used in preference to wideband spectrograms because in many cases the fundamental frequency (F0) was too high to allow accurate
formant measurement using wideband analyses. I.e. if bandwidth was sufficiently wide to include more than one harmonic, it was often too wide to allow accurate measurement of formant locations. Instead, formant locations were estimated from the relative amplitude of harmonics, and by observing changes in harmonic amplitude as they entered or exited the frequency region of a vocal tract resonance.

3. RESULTS

Central F1/F2 measurements for each of the point vowel tokens are shown in Figure 1 below. In the figure, black symbols represent AD citation form tokens and white symbols represent ID tokens. Tokens of the vowel [i] are represented by squares, [u] by triangles, and [a] by circles. Mean F1-F2 values for each of the ID and AD vowel categories are represented by gray squares. Squares representing AD means have an X in the center, and squares for ID means have a plus sign in the center.

![Figure 1: Central F1-F2 measurements for AD (black) and ID (white) [i], [u] and [a]. See above text for further explanation.](image)

It can be seen from Figure 1 that, on average, F1-F2 values for the ID vowels are acoustically more extreme than values for AD citation form vowels. In other words, the mean acoustic distance between tokens belonging to different vowel categories is greater in ID speech than between citation form vowels in AD speech.

To more directly compare the acoustic distance between the three point vowel categories in AD citation form speech and ID speech, formant measurements were first converted to Mels. The acoustic space encompassed by the 3 point vowel categories was then compared by calculating the area of the vowel triangle for each mother in AD and ID speech. Vowel triangle area was calculated as:

\[ \frac{1}{2} [X1(Y2-Y3) + X2(Y3-Y1) + X3(Y1-Y2)] \]

where X and Y are the mean F1 and F2 values, and 1, 2, and 3 are the point vowels, [i], [u], and [a].

Figure 2 shows vowel triangle areas for each of the 14 mothers for AD citation forms and ID speech. In the figure, black bars represent AD citation forms, and white bars represent ID speech. For all but two mothers (803 and 809), vowel triangle area increased from AD citation forms to ID speech. The mean ratio of ID vowel triangle area to AD vowel triangle area was 1.6:1. The significance of differences in vowel triangle areas was tested using a one-way ANOVA with Addressee (AD, ID) as within-subjects factor. Vowel triangle area increased significantly from AD citation forms to ID speech (F = 6.611, p < 0.02).

![Figure 2: Mean vowel triangle areas for AD citation forms (black bars) and ID speech (white bars). See above text for further explanation.](image)

4. DISCUSSION

In AD speech, formant values for citation form vowels are generally considered to represent the "intended" values for vowels, even though speakers often fail to reach these values during conversation because of factors such as speaking rate and coarticulation. In comparison with AD citation form values, Japanese mothers produce still more extreme acoustic values for point vowels in ID speech. Since acoustic differences must originate from articulatory differences, these results suggest that Japanese mothers hyperarticulate when speaking to their infants, even relative to citation forms. Hyperarticulation may have the dual advantage for infants of increasing the acoustic distance between vowel categories and illustrating more extreme articulatory postures, thus providing exaggerated auditory and visual cues to vowel category differences.

The ratio of ID to AD vowel triangle areas in this study was 1.6:1. This is smaller than the difference in vowel triangle areas found for American, Russian and Swedish (ARS) mothers [2]. When AD and ID point vowels were compared in these languages, the ratio of ID to AD vowel triangle area was 1.91:1, 1.94:1, and 1.90:1 for ARS mothers respectively. Since AD vowels in the ARS study were conversational, rather than citation forms, the smaller ratio of ID to AD areas in the current study suggests that Japanese mothers did increase vowel triangle area for citation form vowels relative to vowels from conversation.

In the languages examined to date, American, Russian, Swedish and Japanese mothers all show similar kinds of changes in the acoustic structure of point vowels from AD to ID speech. Are language-specific differences in vowel structure obscured in ID speech by the similarity of changes to the
acoustic structure of point vowel categories? Figures 3 and 4 show the average AD and ID vowel triangles for each of these languages. Vowel triangles for American and Swedish mothers are shown in Figure 3, and vowel triangles for Russian and Japanese mothers in Figure 4. In all cases, AD point vowels are represented by black symbols and ID point vowels by white symbols.

As the figures show, the mean F1-F2 values for [i], [u] and [a] vary with language. Although mothers in all languages produce more extreme formant values for the point vowels in ID speech than in AD speech, ID vowels appear to be no more similar across languages than AD vowels. The language-specific acoustic quality of point vowels appears to be preserved in ID speech, as is the unique geometric relationship among point vowels in each of these languages.

4. CONCLUSIONS

The acoustic structure of vowels in Japanese mothers’ speech to infants differs significantly from that of citation form vowels. Relative to citation form point vowels, the average acoustic distance between vowel tokens belonging to different point categories increases in ID speech. Although mothers in each of the languages examined to date produce similar changes in point vowel acoustic structure during ID speech, vowels nevertheless appear to consistently retain their language-specific acoustic qualities.

ACKNOWLEDGMENTS

This work was supported by NIH grants HD 35465 and DC 00520 to P.K. Kuhl.

REFERENCES