A study on the production-perception link of English vowels produced by native and non-native speakers

Byunggon Yang

Department of English
Dongeui University
bgyang@byomin.dongeui.ac.kr

Abstract
This study explored the relationship between the production of the nine English vowels and the perception of synthesized vowels by thirty-five American, Chinese, and Korean, male and female speakers. The average formant values of the ten American English speakers were employed to synthesize the nine vowels that were presented to the thirty-five speakers. The center formant values of the highest and lowest formant boundary of the same vowel quality were collected and compared to the formant values of their productions. We found that there was a strong correlation between production and perception within and across the language groups. The American group had higher coefficients than the Chinese or Korean groups. The lower coefficients were related to the non-native speakers’ deviant production of the English vowels. In the cross-linguistic comparison, the American, Chinese and Korean groups perceived the stimuli about the same. Individual comparison by regressional analyses of the formant frequency data of the produced vowels and the center formant values of the perceptual test led to a very remarkable $r^2$ value. This suggests a very lawful relationship between production and perception.

1. Introduction

Many cross-linguistic studies have been done either on the production or on the perception of vowels (Peterson and Barney 1952; Polka 1995; Best, Faber and Levitt 1996; Diehl, Lindblom, Hoemeke and Fahey 1996; Yang 1996; Ingram and Park 1997; Flege, Bohn and Jang 1997; Strange, Akahane-Yamada and Kubo 1998). Because the subject’s production varies greatly and perceptual experiment demands sophisticated procedures and long duration, the link has not been pursued frequently. Also, results of vowel perception and production appeared not so clear-cut as those of consonants.

This study aims to explore how native and non-native speakers produce and perceive English vowels and investigate the relationship between them. Specifically, the following research questions were pursued here:

1. What is the relationship between production and perception of English vowels by native and non-native speakers?
2. What is the relationship between production and perception of an individual speaker?

To answer these questions, thirty-five American, Korean and Chinese speakers produced nine English vowels in an hVd context. The first three formant values were collected to compare the speakers by gender and language group. Also, the same speakers listened to the synthesized English vowel pairs and judged whether each pair of the model vowel followed by another vowel with different formant values sounded the same or different.

2. Method

Thirty-five subjects (fourteen Americans, eight Chinese and thirteen Koreans) participated in a recording session and a perceptual test. All were graduate students at the University of Texas at Austin. We assumed that these students had a fluent command of English because they passed the high admission criteria set by the University for foreign students. None reported any speaking or hearing problems.

Nine English words (had, hawed, head, heed, hid, hoed, who’d, hood, Hudd) were chosen for production. This study adopts a dynamic time point for acoustic data collection. The same nine vowels were synthesized and randomly presented in pairs to the thirty-five listeners. The perceptual stimuli consisted of 376 synthesized vowel pairs.

Before the perceptual experiment, the thirty-five subjects produced clearly the nine English vowels using a speech microphone on a digital recorder. Audio inputs were carefully monitored by the author and digitized using a G3 notebook, at a sampling rate of 22 kHz. All acoustic measurements of vowel portions were made using Praat3.8. The formant frequencies were determined at the one-third point of the vowel duration (Yang 1996). The function Formant (sl) determined the first five formant values every 5 ms below 5.5 kHz. To avoid any extreme values, five values for each formant within a 10 ms window around the one-third point were averaged to determine the acoustic values. Pitch values were collected by the function Pitch (ac). The author carefully corrected any inappropriate formant values by simultaneously checking the formant values on each spectrogram.

The synthesis was done by SenSynPPC1.0, a Klatt88 formant synthesizer. The average formant values of the nine English vowels produced by ten American males (Yang 1996) were input to the software. The duration of each stimulus was 400 ms, synthesized and stored on the notebook at a sampling rate of 11 kHz. The general amplitude of the synthesized vowel starts from 0 dB and gradually increased to 65 dB at 100 ms and then decreases very slowly, tapering off. Its fundamental frequency contour begins from 120 Hz and reaches the peak of 132 Hz at 110 ms then gradually decreases to 90 Hz at the end.

Then, the model formant values were modified by a step of 30 Hz for F1, 40 Hz for F2 and 50 Hz for F3 below or above the values not running into the adjacent formants. 1,109 files were synthesized and saved on the computer to
conduct a pilot study. Two American judges listened to all the pairs of a model followed by a synthesized vowel with a modified formant value and determined the range within which they heard the same vowel quality. Then, a set of 362 synthesized vowel pairs well below or above the range of the two judges were chosen out of the 1,109 files for use as perceptual stimuli.

Participants were tested individually in a sound-attenuated Phonetics Lab at the University of Texas at Austin. Each subject listened to the stimuli over a headphone on a digital player. The subjects judged whether each pair sounded the same or different. The highest and lowest formant frequency boundary, within which each subject perceived the same vowel quality, were collected from their answers. The listening session lasted for 30 minutes. The synthesized vowel pairs were played by a script of Praat3.8 and audio outputs were directly recorded on the digital player. The subject was instructed to circle the given number if the pair sounded the same in vowel quality. Otherwise they used a slash mark. The circled number was traced to find the corresponding formant value on the randomized table and marked on a page with formant values in ascending order. The highest and lowest marked values in each formant of the vowel were collected. The center formant frequency in this study refers to the middle value of the highest and lowest formant frequency boundary of the same vowel quality. The acceptable range denotes the acoustic distance between the highest and lowest formant frequency.

3. Results and Discussion

The formant values of the nine English vowels produced are summarized in Table I.

Table I. Formant values of production (Fnp) and perception (Fnc). A indicates for American, C for Chinese and K for Korean; m for male, f for female.

<table>
<thead>
<tr>
<th>Group</th>
<th>Vowel</th>
<th>F1p</th>
<th>F2p</th>
<th>F3p</th>
<th>F1c</th>
<th>F2c</th>
<th>F3c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Af</td>
<td>æ</td>
<td>653</td>
<td>1897</td>
<td>2812</td>
<td>688</td>
<td>1743</td>
<td></td>
</tr>
<tr>
<td>Af</td>
<td>å</td>
<td>634</td>
<td>1169</td>
<td>2816</td>
<td>660</td>
<td>1025</td>
<td>2538</td>
</tr>
<tr>
<td>Af</td>
<td>e</td>
<td>591</td>
<td>2085</td>
<td>2959</td>
<td>555</td>
<td>1914</td>
<td></td>
</tr>
<tr>
<td>Af</td>
<td>i</td>
<td>316</td>
<td>2761</td>
<td>3286</td>
<td>269</td>
<td>2348</td>
<td>3067</td>
</tr>
<tr>
<td>Af</td>
<td>o</td>
<td>459</td>
<td>1264</td>
<td>2742</td>
<td>509</td>
<td>1150</td>
<td></td>
</tr>
<tr>
<td>Af</td>
<td>u</td>
<td>363</td>
<td>1271</td>
<td>2749</td>
<td>329</td>
<td>1407</td>
<td>2279</td>
</tr>
<tr>
<td>Af</td>
<td>æ</td>
<td>442</td>
<td>1251</td>
<td>2776</td>
<td>450</td>
<td>1331</td>
<td></td>
</tr>
<tr>
<td>Af</td>
<td>è</td>
<td>584</td>
<td>1715</td>
<td>2872</td>
<td>394</td>
<td>1318</td>
<td></td>
</tr>
<tr>
<td>Am</td>
<td>æ</td>
<td>654</td>
<td>1597</td>
<td>2538</td>
<td>715</td>
<td>1752</td>
<td></td>
</tr>
<tr>
<td>Am</td>
<td>Å</td>
<td>633</td>
<td>1048</td>
<td>2602</td>
<td>661</td>
<td>1050</td>
<td>2527</td>
</tr>
<tr>
<td>Am</td>
<td>e</td>
<td>550</td>
<td>1823</td>
<td>2638</td>
<td>550</td>
<td>1897</td>
<td></td>
</tr>
<tr>
<td>Am</td>
<td>i</td>
<td>287</td>
<td>2301</td>
<td>2982</td>
<td>273</td>
<td>2337</td>
<td>3022</td>
</tr>
<tr>
<td>Am</td>
<td>o</td>
<td>418</td>
<td>1959</td>
<td>2688</td>
<td>405</td>
<td>2043</td>
<td></td>
</tr>
<tr>
<td>Am</td>
<td>u</td>
<td>461</td>
<td>1062</td>
<td>2461</td>
<td>522</td>
<td>1158</td>
<td></td>
</tr>
<tr>
<td>Am</td>
<td>æ</td>
<td>594</td>
<td>1290</td>
<td>2535</td>
<td>590</td>
<td>1315</td>
<td></td>
</tr>
</tbody>
</table>

The correlation coefficient between the formant values of the American male speakers and the synthesis model amounted to 0.988 for F1; 0.981 for F2; 0.992 for F3. This result corroborates that the current American male data are comparable to those of Peterson and Barney (1952) and Yang (1996). The formant ratios of the American female to male determined from the nine English vowels are 0.91 for the Americans; 0.89 for the Chinese; 0.93 for the Koreans. The results are comparable to the Swedish and Dutch data. The vocal tract ratios were 0.89 for Swedish (Fant 1975), 0.89 for Dutch (van Nierop, Pols and Plomp 1973; Pols, Tromp and Plomp 1973) and 0.86 for English (Peterson and Barney 1952).

The averages of male and female listeners are about the same. It appears that the three groups perceived the synthesized vowel pairs almost the same even though the acceptable ranges were wide and their formant values of
production between the male and female groups were quite different.

To explore the link between production and perception of English vowels by native and non-native speakers, we conducted correlational analyses between the formant values of production and perception within each language group. Table II lists the correlation coefficients.

Table II. Correlation coefficients of F1-F3 of the English vowels produced and the center formant frequency of the nine English vowels.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Slope</th>
<th>Intercept</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Af1</td>
<td>0.855</td>
<td>91</td>
<td>0.990</td>
</tr>
<tr>
<td>Af2</td>
<td>0.825</td>
<td>176</td>
<td>0.968</td>
</tr>
<tr>
<td>Af3</td>
<td>0.784</td>
<td>171</td>
<td>0.959</td>
</tr>
<tr>
<td>Af4</td>
<td>0.838</td>
<td>66</td>
<td>0.983</td>
</tr>
<tr>
<td>Af5</td>
<td>0.825</td>
<td>173</td>
<td>0.937</td>
</tr>
<tr>
<td>Af6</td>
<td>0.917</td>
<td>-34</td>
<td>0.949</td>
</tr>
<tr>
<td>Af7</td>
<td>0.885</td>
<td>12</td>
<td>0.981</td>
</tr>
<tr>
<td>Am8</td>
<td>1.017</td>
<td>44</td>
<td>0.974</td>
</tr>
<tr>
<td>Am9</td>
<td>0.932</td>
<td>123</td>
<td>0.968</td>
</tr>
<tr>
<td>Am10</td>
<td>0.985</td>
<td>64</td>
<td>0.980</td>
</tr>
<tr>
<td>Am11</td>
<td>0.922</td>
<td>129</td>
<td>0.976</td>
</tr>
<tr>
<td>Am12</td>
<td>0.907</td>
<td>101</td>
<td>0.956</td>
</tr>
<tr>
<td>Am13</td>
<td>1.113</td>
<td>-68</td>
<td>0.986</td>
</tr>
<tr>
<td>Am14</td>
<td>1.005</td>
<td>-12</td>
<td>0.989</td>
</tr>
<tr>
<td>Cm15</td>
<td>0.769</td>
<td>174</td>
<td>0.939</td>
</tr>
<tr>
<td>Cm16</td>
<td>0.717</td>
<td>194</td>
<td>0.980</td>
</tr>
<tr>
<td>Cm17</td>
<td>0.817</td>
<td>153</td>
<td>0.960</td>
</tr>
<tr>
<td>Cm18</td>
<td>0.794</td>
<td>219</td>
<td>0.907</td>
</tr>
<tr>
<td>Cm19</td>
<td>0.969</td>
<td>113</td>
<td>0.964</td>
</tr>
<tr>
<td>Cm20</td>
<td>0.801</td>
<td>255</td>
<td>0.919</td>
</tr>
<tr>
<td>Cm21</td>
<td>0.825</td>
<td>159</td>
<td>0.939</td>
</tr>
<tr>
<td>Cm22</td>
<td>0.990</td>
<td>72</td>
<td>0.959</td>
</tr>
<tr>
<td>Cm23</td>
<td>0.765</td>
<td>238</td>
<td>0.929</td>
</tr>
<tr>
<td>Cm24</td>
<td>0.764</td>
<td>279</td>
<td>0.881</td>
</tr>
<tr>
<td>Cm25</td>
<td>0.837</td>
<td>282</td>
<td>0.834</td>
</tr>
<tr>
<td>Cm26</td>
<td>0.697</td>
<td>255</td>
<td>0.895</td>
</tr>
<tr>
<td>Cm27</td>
<td>0.794</td>
<td>155</td>
<td>0.944</td>
</tr>
<tr>
<td>Kf28</td>
<td>0.799</td>
<td>223</td>
<td>0.945</td>
</tr>
<tr>
<td>Kf29</td>
<td>0.704</td>
<td>296</td>
<td>0.903</td>
</tr>
<tr>
<td>Kf30</td>
<td>1.065</td>
<td>13</td>
<td>0.979</td>
</tr>
<tr>
<td>Km31</td>
<td>0.948</td>
<td>132</td>
<td>0.964</td>
</tr>
<tr>
<td>Km32</td>
<td>0.913</td>
<td>136</td>
<td>0.972</td>
</tr>
<tr>
<td>Km33</td>
<td>1.066</td>
<td>51</td>
<td>0.949</td>
</tr>
<tr>
<td>Km34</td>
<td>0.931</td>
<td>137</td>
<td>0.959</td>
</tr>
<tr>
<td>Km35</td>
<td>0.863</td>
<td>158</td>
<td>0.934</td>
</tr>
</tbody>
</table>

The average r² values were 0.97 for Americans, 0.93 for Chinese and Koreans. These r² values are remarkably high, which suggests a very lawful relationship between individual production and perception. Here again, the American group shows higher values than the other two groups, which could be related to the non-native speakers’ different vowel system. One can notice that the males have slopes higher than one,
while all the females have slopes lower than one. Some subjects show deviations from this. Some males have lower formant values than the synthesized vowels so that they must have tuned highly to the synthesized ones. Since the regression analyses were applied with different intercepts, the negative intercept or lower slope comes from over-compensating. So, these slopes with negative intercept may be lower than the statistical analyses involved here.

4. Conclusion

This study has explored the relationship between the production of the nine English vowels and the perception of synthesized vowels by thirty-five American, Chinese, and Korean, male and female speakers. We found that there was a strong correlation between production and perception within and across the language groups. The American group had higher coefficients than those of the Chinese or Korean groups. The lower coefficients were related to the non-native’s deviant production of the English vowel. In the cross-linguistic comparison, the American, Chinese and Korean groups perceived the stimuli almost the same. Individual comparison by regressive analyses on the formant frequency data of the produced vowels and the center formant values of the perceptual test led to a very remarkable $r^2$ value (0.97 for the Americans, 0.93 for the Chinese and Koreans on average), which suggests a very lawful relationship between production and perception. The present data proved that human listeners possess a broad but accurate perceptual mechanism to distinguish between complex acoustic stimuli.

It should be noted that we have looked at the cross-linguistic comparison of native and non-native speakers only. It may clarify the production and perception link better with higher correlation coefficients if we use stimuli in the subject’s native language. Also, the perceptual experiment was conducted simply to judge whether the two sound pairs are the same or different. Further rigorous discriminatory study will be necessary to capture the subtle auditory perception.

Acknowledgment. This work was supported by a grant from the Interdisciplinary Research Program (Contract No. 1999-2-302-106-5) of the KOSEF.

5. References