Acoustic Characterization of Singaporean Children’s English: Comparisons to American and British Counterparts

Yuling Gu
New York University, USA
yuling.gu@nyu.edu

Nancy F. Chen
Institute for Infocomm Research, A*STAR, Singapore
nfychen@i2r.a-star.edu.sg

Abstract

We investigate English pronunciation patterns in Singaporean children in relation to their American and British counterparts by conducting archetypal analysis (Cutler and Breiman, 1994) on selected vowel pairs. Given that Singapore adopts British English as the institutional standard, one might expect Singaporean children to follow British pronunciation patterns, but we observe that Singaporean children also present similar patterns to Americans for TRAP-BATH split vowels: (1) British and Singaporean children both produce these vowels with a relatively lowered tongue height. (2) These vowels are more fronted for American and Singaporean children ($p < 0.001$).

In addition, when comparing /æ/ and /e/ productions, British speakers show the clearest distinction between the two vowels; Singaporean and American speakers exhibit a higher and more fronted tongue position for /æ/ ($p < 0.001$), causing /æ/ to be acoustically more similar to /e/.

1 Introduction

English varieties in the world can be represented as three concentric circles – inner circle (e.g. US, UK), outer circle (e.g. Singapore, India), and expanding circle (e.g. China, Russia) (Kachru, 1982). The inner circle contains Anglo Englishes whereas the outer circle contains ‘New Englishes’ where English spread to those regions with historical colonization. Extensive work has been done to investigate American English, including acoustic, phonetic or sociolinguistic studies (Kuo, 2013; Clopper and Pisoni, 2007; Chen et al., 2009; Labov et al., 2006; Evanini, 2008) and work using machine learning to automatically find pronunciation patterns (Chen et al., 2014; Chen, 2011; Tauberer and Evanini, 2009). There is also much work on studying different varieties of British English in terms of phonetics and prosody, including (Henton, 1983; Grabe and Post, 2002; Wells, 1999). Further, these two inner circle English pronunciations have often been compared to each other (Khan and Alzobidy, 2019; Gomez, 2009).

By contrast, investigations on English spoken by groups in the outer circle (e.g., Indian English, Singapore English) has received much less attention. Focusing on the case of Singapore English, there has been literature providing analysis at length on a syntactic level (Alsagoff, 1998) and also work focusing on analyzing from a semantics level (Wong, 2004). However, analysis from a phonological perspective has either been based on anecdotal evidence (Deterding and HvItfeldt, 1994; Foley, 1988) or been limited in scale due to the lack of available large-scale corpora and the limited number of speakers recruited (Deterding and Ling, 2001). Deterding (2007)’s phonological analysis was mainly based on a single female speaker. Tan (2012) outlined some distinctive phonological features of Singapore English; for instance, /æ/ in British Received Pronunciation and general American pronunciation being realized as vowels like /e/ in Singapore English. Deterding (2007) gave a comprehensive description of the features of Singapore English by analyzing various phonemes in speech collected from one female undergraduate student. However, till date, there has been no large-scale studies to quantify these observations. Furthermore, all of such work has focused on adults, while studies on children speech is limited, if any.

In this work, we present a large-scale investigation to acoustically quantify the characteristics of Singaporean children’s English pronunciations. The speaker number and utterance number are at least an order of magnitude greater than past work such as (Deterding and Ling, 2001; Deterding, 2007).
2 Methods

2.1 Unsupervised clustering: Archetypal Analysis

Most unsupervised clustering algorithms such as k-means (MacQueen, 1967) use centroids to conduct cluster analysis. Archetypal analysis represent each data point in a data set as a combination of characteristic “archetypes” (pure types) (Cutler and Breiman, 1994). Motivated by multilingual and multicultural influence of Singapore English, we adopt archetypal analysis (Cutler and Breiman, 1994) to investigate how American and British (inner circle English) pronunciations might serve as anchoring archetypal references to characterize Singapore English (outer circle English). We set the cluster number to 2.

2.2 Acoustic Analysis

The natural resonant frequencies of the vocal tract are formant frequencies. Different tongue positions change the vocal tract configuration, resulting in different formant frequencies. The first two formant frequencies (F1 and F2) characterize acoustic characteristics of vowels the most; higher F1 corresponds to lower tongue positions while higher F2 corresponds to a more fronted tongue position (Stevens, 1998).

3 Experiments

3.1 Corpus

Read speech was collected from American children (140 speakers, 43,406 utterances), British children (82 speakers, 32,542 utterances) and Singaporean children (192 speakers, 34,457 utterances). The age range is 6-13 years old and the gender ratio is balanced. The reading material were customized for each of the three populations, and consists of sentences from TIMIT (Garofolo and et al., 1993), PF-STAR (Russell, 2006; Batliner et al., 2005), GMU Speech Accent Archive (Weinburger, 2015) and carefully designed sentences containing minimal pairs and words that elicit possible acoustic and pronunciation differences across speakers and speaker populations. All three corpora were designed to be phonetically balanced, and in part designed according to the considerations laid out in (Chen et al., 2016).

3.2 TRAP-BATH Split Vowels

TRAP–BATH split occurs in mainstream English in UK (Wells, 1982), where words such as bath are pronounced with /æ/ instead of /æÆ/ as in trap. Such splitting is not observed in general American English. Using archetypal analysis, we represent each data point as a combination of characteristic pure types (Cutler and Breiman, 1994); results in Table 1 show that Singaporean children are more similar to American children in producing these vowels. From Table 2 and Figure 1, in terms of F1 estimates, American children have the lowest F1 ($M = 827.55$), British children have significantly higher F1 ($M = 902.73$), and Singaporean children have the highest F1 ($M = 908.84$). This suggests that Singaporean and British children produce TRAP-BATH vowels with a lower tongue position compared to American children. In terms of F2, British children show the lowest F2 ($M = 1617.13$), American children show higher F2 ($M = 2186.58$), and Singaporean children show the highest F2 ($M = 2267.79$), indicating that TRAP-BATH vowels are more fronted for both Singaporean and American children, resulting in a vowel closer to /æ/ rather than /Æ/; i.e., Singaporean children do not show much TRAP-BATH split distinction.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>F1 mean</th>
<th>F1 se</th>
<th>F2 mean</th>
<th>F2 se</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td>908.84</td>
<td>5.71</td>
<td>2,267.79</td>
<td>9.84</td>
</tr>
<tr>
<td>AE</td>
<td>827.55</td>
<td>8.17</td>
<td>2,186.58</td>
<td>15.91</td>
</tr>
<tr>
<td>BE</td>
<td>902.73</td>
<td>12.24</td>
<td>1,617.13</td>
<td>13.22</td>
</tr>
</tbody>
</table>

Table 2: Mean and standard error (se) for each speaker group for TRAP-BATH split vowels.

All experiments were also conducted using k-means clustering. As both approaches show similar trends, we only show results for archetypal analysis due to space constraints.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Group1</th>
<th>Group2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td>0.927</td>
<td>0.073</td>
</tr>
<tr>
<td>AE</td>
<td>0.614</td>
<td>0.386</td>
</tr>
<tr>
<td>BE</td>
<td>0.317</td>
<td>0.927</td>
</tr>
</tbody>
</table>

Table 1: Archetypal analysis using formant features.
Figure 1: $F_1$ and $F_2$ estimates of TRAP-BATH split vowels across speakers. Ellipses represent 95% confidence interval for each group. Each small colored shape: speaker’s mean $F_1$ and $F_2$ estimates; larger black shapes: group mean.

Table 3: Mean and standard error (se) for each speaker group for /æ/ and /e/ formant estimates.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Phone</th>
<th>$F_1$ mean</th>
<th>$F_1$ se</th>
<th>$F_2$ mean</th>
<th>$F_2$ se</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td>/æ/</td>
<td>875.11</td>
<td>5.40</td>
<td>2,327.39</td>
<td>9.77</td>
</tr>
<tr>
<td></td>
<td>/e/</td>
<td>796.35</td>
<td>4.64</td>
<td>2,353.92</td>
<td>10.02</td>
</tr>
<tr>
<td>AE</td>
<td>/æ/</td>
<td>901.59</td>
<td>7.39</td>
<td>2,082.68</td>
<td>12.36</td>
</tr>
<tr>
<td></td>
<td>/e/</td>
<td>785.44</td>
<td>5.87</td>
<td>2,059.39</td>
<td>11.55</td>
</tr>
<tr>
<td>BE</td>
<td>/æ/</td>
<td>959.93</td>
<td>13.40</td>
<td>1,802.39</td>
<td>13.89</td>
</tr>
<tr>
<td></td>
<td>/e/</td>
<td>751.79</td>
<td>9.80</td>
<td>2,009.81</td>
<td>16.36</td>
</tr>
</tbody>
</table>

4 Discussion

We presented a large-scale study (at least an order of magnitude more speakers and utterances than previous work), showing that Singaporean children are more similar to American children in pronunciation patterns of the /æ/ vs. /æ/ and /æ/ vs. /e/. This alludes to sociolinguistic perspectives of how Singapore English has been changing beyond the British influence during historical colonization (Lim and Ansaldo, 2015) and could be increasingly moving towards embodying American pronunciation characteristics.
References


