Large-Scale Acoustic Characterization
of Singaporean Children’s English Pronunciation

Yuling Gu*, Nancy F. Chen*

Abstract
In this work, we investigate pronunciation differences in English spoken by Singaporean children in relation to their American and British counterparts by conducting Kmeans clustering and Archetypal analysis on selected vowel pairs and approximants. Given that Singapore adopts British English as the institutional standard due to historical reasons, one might expect Singaporean children to follow British pronunciation patterns. Indeed, Singaporean and British children are more similar in their production of syllable-final rhotic approximant /ɹ/ – they do not lower their third formant (F3) nearly as much as American children do, suggesting a lack of rhoticity. Interestingly, we also observe that Singaporean children present similar patterns to American children when it comes to their fronting of vowels as demonstrated across TRAP–BATH split vowels, /æ/, /ɛ/, /ɑ/, /ɔ/, /i/ and /ɪ/. Singaporean children’s English also demonstrated characteristics that do not resemble any of the other two populations. We observe that Singaporean children’s vowel height characteristics are distinct from both that of American and British children. In tense and lax vowel pairs, /u/ and /o/, as well as /i/ and /ɪ/, we also consistently observe that the difference is less conspicuous for Singaporean children compared to the other speaker groups. Further, while American and British children demonstrate lowering of F1 and F2 formants in transitions into syllable-final /l/s, a wide gap between F2 and F3 formants, and small difference between F1 and F2 formants, all of

*Corresponding authors
Email addresses: yulinggu@uw.edu (Yuling Gu), nancychen@alum.mit.edu (Nancy F. Chen)

Yuling Gu is at University of Washington when this work is submitted to arXiv, but the majority of the work was completed while at New York University. Nancy F. Chen is currently with the Institute for Infocomm Research, A*STAR.
these are not exhibited in Singaporean children’s pronunciation. These findings point towards potential sociolinguistic implications of how Singapore English might be evolving to embody more than British pronunciation characteristics. Furthermore, these findings also suggest that Singapore English could be have been influenced by languages beyond American and British English, potentially due to Singapore’s multilingual environment.

Keywords: acoustic phonetics, pronunciation modeling, unsupervised clustering

1. Introduction

English varieties in the world can be represented in the form of three concentric circles – inner circle (e.g. US, UK), outer circle (e.g. Singapore, India), and expanding circle (e.g. China, Russia) [1]. The inner circle contains Anglo Englishes whereas the outer circle contains ‘New Englishes’ of which the spread of English to those regions occurred through the process of historical colonization. Extensive work has been done to investigate American English, including acoustic, phonetic or sociolinguistic studies [2, 3, 4, 5, 6] and work using machine learning to automatically find pronunciation patterns [7, 8, 9]. There is also much work on studying different varieties of British English in terms of phonetics and prosody, including [10, 11, 12, 13]. Further, these two inner circle English pronunciations have often been compared to each other [14, 15].

By contrast, investigations on English spoken by groups in the outer circle (e.g., Indian English, Singapore English) has received much less attention. For Singapore English, there has been literature providing analysis at length at the syntactic level (e.g. [16]) and at the semantic level (e.g. [17]). Analysis from a phonological perspective mainly focused on patterns from stress, rhythm and intonation (e.g. [18, 19, 20]), yet few have examined speech acoustic characteristics. Previous phonological analysis in this direction have either been based on anecdotal evidence (e.g. [21, 22]) or have been limited in scale due to the lack of available large-scale corpora and the limited number of speakers recruited for the experiments. For example, the National Institute of Education Corpus of Spoken Singapore English [23] consists of five-minute long interviews from 31 female and 15 male speakers; [24]’s phonological analysis was mainly based on a one-hour recording of a single female speaker. Previous work outlined some distinctive phonologi-
cal features of Singapore English \cite{23}; for instance, \(/æ/\) in British Received Pronunciation (RP) and general American pronunciation are more likely to be acoustically realized as other vowels like \(/ɛ/\) in Singapore English. \cite{24} 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 experiment to quantify these observations. Furthermore, all such work focuses on adult speech, while studies on child speech, which have important applications such as computer-assisted language learning, is limited, if any.

In this work, we present a large-scale analysis to acoustically quantify the characteristics of Singaporean children’s English pronunciations for various vowel pairs and approximants. The speaker number and utterance number in this study are at least an order of magnitude greater than past work such as \cite{23, 24}. This work extends and expands our prior work \cite{26} to include more experiments, analyses and discussions left out in the conference version. In particular, we expand our analyses to cover more vowels and also include studies of different approximants. This work can be seen as expanding and developing the directions briefly outlined in our earlier abstracts \cite{27, 28} for the analysis of vowels and \cite{29} for the analysis of approximants respectively. Our findings could have various applications in areas such as native language identification \cite{30}, automatic speech recognition \cite{31, 32}, self-correction behavior in vowel production \cite{33}, and be applied to improve pronunciation modeling in computer-assisted language learning \cite{34, 35, 36, 37}.

2. Experimental Design

2.1. Speech Corpora

Read speech was collected from American children (140 speakers, 43,406 utterances), British children (95 speakers, 32,542 utterances) and Singaporean children (193 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 \cite{38}, PF-STAR \cite{39, 40}, GMU Speech Accent Archive \cite{41} 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 \cite{12, 13, 14}. 

2.2. Acoustic Features

The Praat [45] software was used to extract acoustic features from the utterances on a per phoneme basis. We force-aligned the utterances to produce time boundaries, a small sample of which were manually inspected to ensure that they are accurate within 50ms of the actual boundaries. The acoustic features used included estimates of the first four formants, F0 and phoneme duration. These features were sampled at a 10ms time step and averaged for each phoneme to give us a representation of its acoustics characteristics. Using these features, unsupervised clustering approaches were used to explore patterns in the data, detailed acoustic analysis were used as a follow-up to gain phonological insights, and conjectures of articulatory gestures were made according to acoustic phonetics knowledge [46].

2.3. Archetypal Analysis

Most algorithms in unsupervised clustering such as k-means [47] use centroids to conduct cluster analysis. In this work, motivated by multilingual and multicultural influence of Singapore English, we adopt Archetypal analysis [48] to investigate how American and British (inner circle English) pronunciations might serve as anchoring Archetypal references to characterize Singapore English (outer circle English). All experiments were also conducted on k-means. As both approaches show similar trends, we will focus on the results for archetypal analysis.

Archetypal analysis represent each data point in a data set as a combination of characteristic “archetypes” (pure types) [48]. Given a set of multivariate data, \( \{x_i, i = 1, \ldots, n\} \), where each \( x_i \) is a vector of length \( m \), we seek vectors \( z_1, \ldots, z_p \) of length \( m \) that form the Archetypal extremes. The vectors \( z_1, \ldots, z_p \) are defined as

\[
    z_k = \sum_j \beta_{kj} x_j, k = 1, \ldots, p
\]

where \( \beta_{ki} \geq 0, \sum_i \beta_{ki} = 1 \), and we define \( \{\alpha_{ik}\}, k = 1, \ldots, p \) to minimize the following expression

\[
    \left\| x_i - \sum_{k=1}^{p} \alpha_{ik} z_k \right\|^2
\]

where \( \alpha_{ik} \geq 0, \sum_k \alpha_{ik} = 1 \). The archetypes are defined as vectors \( z_1, \ldots, z_p \) that minimize \( D \), where \( D \) is the sum of squares of distances from each data vector \( x_i \) to the convex hull formed by the \( z_1, \ldots, z_p \) vectors.
Archetypal analysis applies an alternating minimizing algorithm to a non-linear least squares problem. We experiment with this approach using an R package for Archetypal analysis that has been documented in [49] and [50]. For the following analysis we present, we experimented with various values for $p$ and chose $p = 2$ using “elbow criterion” on the residual sum of squares [49].

### 2.4. Data Processing

For the purposes of our analysis, with all the phoneme tokens extracted from our corpora, we first excluded silent tokens (labeled ‘SIL’), non-silent tokens with F1, F2 or F3 undefined, as well as non-silent tokens with duration greater than 1s. The remaining tokens are then averaged on a per speaker and per phoneme basis, giving one data point for each speaker, each phoneme. To analyze syllable-final and syllable-initial /l/ and /ɾ/, we used the CMU Pronouncing Dictionary version 0.6 augmented with syllable boundaries (syllabified CMU) [51] for syllabification.

Using the Nordstrom and Lindblom model [52], we computed scaling factors for normalization within (e.g. age, gender) and across speaker groups (population). The model was adopted for normalization across the languages based on speaker groups to account for anatomical differences in vocal tract length. This model estimate the total length of subject’s vocal tract using average of F3 in vowels with F1 greater than 600Hz. In our adaptation of the model, we determined vowels by referencing [53].

### 3. Overall formant space

We present an overview of the formant space for American, British and Singaporean children in Figures 1, 2 and 3 respectively. These vowel plots suggest some trends in how the three speaker groups may differ in their realization of the different vowels, including:

- American and Singaporean children both have /æ/ and /ɛ/ vowels that are greatly overlapping in the formant space. British children’s /æ/ and /ɛ/ vowels also show some overlap in F1 and F2 features.
All three populations demonstrate some overlapping of /ɑ/ and /ɔ/ in the formant space (albeit at varying degrees).

American and Singaporean children both have /u/ and /ʊ/ that are overlapping in the formant space, though this is not the case for British children.

Singaporean children show great overlap in their /i/ and /ɨ/ vowels. Such great overlap is not observed in the other two speaker groups for /i/ and /ɨ/.

We examine and analyze vowels produced by the three speakers in greater detail in Sections 4 to 8.

4. Trap-bath split

TRAP–BATH split is a vowel split that is well-known in UK (including RP) [54], where vowels in words such as glass, laugh, dance, can’t are pronounced as the [ɑ] phone instead of [æ] as in trap, cab, mad. Such splitting is not typically observed in American English [54]. In this section, we
Figure 2: Overall formant space for British children

Figure 3: Overall formant space for Singaporean children
examine how Singaporean, American, and British children might produce TRAP–BATH split vowels, where [a] and [ae] phones are the different realizations of such vowels. For the rest of this paper, we refer to vowels that could turn into the [a] phone when TRAP–BATH split is present as [a] vowels and those that are realized as the [ae] phone as [ae] vowels.

4.1. Unsupervised clustering

Table 1 shows the clustering results of Archetypal analysis using F1 and F2. We perform one clustering experiment per speaker group on TRAP–BATH split vowels from that group. For each phone (i.e. [ae] and [a]), we show the percentage of tokens from that category that get clustered into each of the cluster groups. For each row, the two numbers add up to 1.0 (stands for 100%). This illustrates the proportion of tokens that gets grouped into Group 1 and Group 2 respectively. Majority of Singaporean children’s [ae] and [a] vowels in the TRAP–BATH split are largely (> 60% of each of these vowels) grouped to one cluster, suggesting that Singaporean children may produce these vowels with less acoustic distinction in terms of the formant estimates. In contrast, more distinctive clusters were observed for American and British children’s TRAP–BATH split vowels. Results from Kmeans clustering in Table 2 similarly show that American and British children’s TRAP–BATH split vowels give cleaner clusters. In the next subsection, we investigate if cleaner clusters indeed reflect the split through formant space analysis.

4.2. Acoustic Analysis and Characterization
4.2.1. F1 F2 formant space

We present the mean and standard error for F1 and F2 estimates in Table 3 and visualize this data on a per speaker level in Figure 4: we observe some overlap between the American and Singaporean populations, which are more fronted than British pronunciations (higher F2). We observe that TRAP–BATH split vowels produced by Singaporean children generally have lower F1 values compared to American and British speakers. We then further analyzed this effect in detail in terms of F1 and F2 separately.

4.2.2. F1 formant estimates

For TRAP–BATH split vowels that become [ae] under the split, Singaporean children have the lowest F1 ($M = 766$), American children have statistically significantly higher F1 ($M = 948$); British children ($M = 937$)
Table 1: Archetypal analysis using $F1$(Hz), $F2$(Hz) estimates for Singaporean, American and British children for TRAP–BATH split $\langle ae \rangle$ and $\langle a \rangle$ vowels. Archetypal extreme points for each cluster are in the format $(F1, F2)$.

show similar trends to Americans. For vowels that could turn into $[a]$ when the split is present, Singaporean children similarly have the lowest F1 ($M = 825$), American children have slightly higher F1 ($M = 831$), and British children have the highest F1 ($M = 891$). A one-way ANOVA demonstrated that these differences are statistically significant, $F(2, 411) = 207.1, p < 0.001$ for $[ae]$ vowels under the split, and $F(2, 412) = 17.3, p < 0.001$ for vowels that could turn into $[a]$ under the split. We further investigate which pairs of these three populations are significantly different from each other. Thus we followed up our ANOVA test with a post hoc Tukey’s HSD Test: For vowels that are realized as $[ae]$ under the split, in terms of F1, American and Singaporean children differ significantly at $p < 0.001$; British and Singaporean children differ significantly at $p < 0.001$. However, American and British children are not significantly different from each other. For vowels that turn into $[a]$ under the split, British children are significantly different from the other two groups in terms of F1 ($p < 0.001$), but American and Singaporean children are not significantly different from each other. There-
| Phone | Group1 (F1, F2) | Group2 (F1, F2) |
|-------|----------------|----------------|
| æ     | 0.339          | 0.661          |
| [a]   | 0.630          | 0.370          |

**American Children**

| Phone | Group1 (F1, F2) | Group2 (F1, F2) |
|-------|----------------|----------------|
| æ     | 0.800          | 0.200          |
| [a]   | 0.350          | 0.650          |

**British Children**

| Phone | Group1 (F1, F2) | Group2 (F1, F2) |
|-------|----------------|----------------|
| æ     | 0.756          | 0.244          |
| [a]   | 0.277          | 0.723          |

Table 2: Kmeans clustering using $F1$ (Hz), $F2$ (Hz) estimates for Singaporean, American and British children for TRAP–BATH split [æ] and [a] vowels. Cluster centroids for each cluster are in the format (F1, F2).

Therefore one significant difference for TRAP–BATH [æ] vowels produced by the three populations is that those produced by Singaporean children have significantly lower F1 formant estimates compared to the other two populations. For the [a] vowels, those produced by both Singaporean and American children have significantly lower F1 formant estimates compared to British children. The articulatory implication is that Singaporean children exhibit a higher tongue height when producing both TRAP–BATH split vowels, and American children exhibit this trait when producing the [a] TRAP–BATH split vowel.

### 4.2.3. $F2$ formant estimates

For vowels that are realized as [æ] under the split, British children show the lowest F2 ($M = 1733$), American children show higher F2 ($M = 2021$), and Singaporean children show the highest F2 ($M = 2123$). A one-way ANOVA shows that these differences are statistically significant, $F(2, 411) = 258, p < 0.001$. For vowels that turn into [a] under the split, British
| Corpus | Phone | F1 mean | F1 se | F2 mean | F2 se |
|--------|-------|---------|-------|---------|-------|
| SG     | [æ]   | 766     | 5.03  | 2123    | 9.13  |
|        | [ə]   | 825     | 5.09  | 2022    | 8.89  |
| AE     | [æ]   | 948     | 7.56  | 2021    | 11.5  |
|        | [ə]   | 831     | 8.01  | 2179    | 15.9  |
| BE     | [æ]   | 937     | 13.5  | 1733    | 14.3  |
|        | [ə]   | 891     | 12.0  | 1535    | 12.4  |

Table 3: Mean and standard error (se) for each speaker group for TRAP–BATH split vowel formants.

children again show the lowest F2 \( (M = 1535) \), Singaporean children show higher F2 \( (M = 2022) \), and American children show the highest F2 \( (M = 2179) \). A one-way ANOVA shows that these differences are statistically significant, \( F(2, 412) = 518.6, p < 0.001 \). To investigate which pairs of the three speaker groups are significantly different from each other, we followed up our ANOVA test with a post hoc Tukey’s HSD Test which show that all pairwise comparisons are significantly different \( (p < 0.001) \). Such findings show that in terms of articulatory implications, Singaporean and American children’s productions of TRAP–BATH vowels are similarly more fronted, compared to British speakers.

4.3. Summary: Singaporean children’s TRAP-BATH split trends are more similar to American children and show less clear TRAP-BATH split compared to British children

Compared to British children, Singaporean and American children are more similar in fronting their TRAP–BATH vowels to produce something closer to [æ] rather than [ə]. Singaporean and American children’s production of the TRAP–BATH vowels also differ from British children by having acoustic features that reflect a higher tongue position. Singaporean children demonstrate a higher tongue position for both [æ] and [ə] TRAP–BATH vowels compared to British children while American children demonstrate the same trait for [ə] vowels. Thus, Singaporean and American children not showing as much TRAP–BATH split distinction as the two vowels are produced with similar frontness that is more characteristic of [æ] than [ə]. These articulatory features agree with our clustering results in suggesting that Singaporean children do not show TRAP–BATH split. The formant
Figure 4: Singaporean and American children are more similar in terms of $F_1$ and $F_2$ estimates for TRAP–BATH split vowels compared to British children. /ae/ phonemes that could turn into [a] in TRAP–BATH split are labeled as [a] while those that are realized as [ae] are labeled as [ae]. Smaller points: individual speaker’s mean $F_1$ and $F_2$; larger points: group means for the speaker groups.

analysis also further clarified that while American children’s TRAP–BATH vowels fall into two clusters, they are all produced with acoustic features like that of [æ]. For American children, TRAP–BATH split vowels that could be changed to the back vowel [a] when the split is present are instead articulated with even a fronter position (higher $F_2$) than [æ], and thus pronounced like [æ], reaffirming the traditional knowledge that most Americans do not exhibit TRAP–BATH split. Our formants analysis show that British children’s TRAP–BATH split exhibit more acoustic differences compared to the other two groups, and reaffirm the clustering results that they show TRAP–BATH split since vowels that could change into [a] in the split are indeed produced with lower $F_2$ like a back vowel ($M = 1535$), compared to those that could stay as the front vowel [æ] with higher $F_2$ ($M = 1733$).
5. /æ/ vs. /ɛ/

/ɛ/ is a mid-low front vowel. When compared to the /æ/ phoneme, the /ɛ/ phoneme has slightly higher F2 and lower F1 estimates. Therefore, any fronting of /æ/, resulting in a higher F2, could lead to potential confusion with /ɛ/. Having observed higher F2 of Singaporean and American children in their TRAP—BATH split vowels compared to British children (across realizations as [ʌ] and [æ] phones), we further examine how production of /æ/ and /ɛ/ phonemes might differ across the three speaker groups.

5.1. Unsupervised Clustering

We perform per-speaker-group clustering across /æ/ and /ɛ/ vowels from each group using F1 and F2 estimates. Clustering results using Archetypal analysis and Kmeans clustering are shown in Tables 4 and 5. The results from the two clustering methods agree, indicating that /æ/ and /ɛ/ vowels produced by the British children are largely (> 90 % for Archetypal analysis, > 80 % for Kmeans clustering) grouped into two clean, distinctive clusters, whereas such distinction is less clear cut for these vowels produced by Singaporean and American children.

5.2. Acoustic Analysis and Characterization

5.2.1. F1 F2 formant space

Estimates of F1 and F2 for /æ/ and /ɛ/ in the three speaker groups are summarized in Table 6. Using each speaker as a data point, we visualize this differences between speaker groups for the two vowels in Figure 5. We observe that /æ/ and /ɛ/ produced by British children are the most clearly distinguished from each other, with /ɛ/ having a higher F2 and lower F1 than that of /æ/. However, the F1 and F2 distinction between these two vowels for Singaporean and American children are less conspicuous. These observations in the formant space align with our clustering results, potentially explaining why the split for Singaporean and American children’s /æ/ and /ɛ/ are not as clear as that of British children. We then further analyzed differences across the speaker groups in terms of F1 and F2 separately.

5.2.2. F1 formant estimates

For /æ/, Singaporean children have the lowest F1 (M = 785), American children have higher F1 estimates (M = 902), and British children have the highest F1 (M = 921). Similarly, for the production of /ɛ/, Singaporean...
Figure 5: Singaporean and American children are more similar in terms of formant characteristics for /æ/ and /ɛ/ vowels and they show less distinction between these vowels in comparison to British children. Smaller points: individual speaker’s mean $F1$ and $F2$; larger points: group means for the speaker groups.
| Phoneme | Group1 (886, 2337) | Group2 (604, 1845) |
|---------|------------------|------------------|
| /æ/     | 0.651            | 0.349            |
| /ɛ/     | 0.448            | 0.552            |

| Phoneme | Group1 (1013, 2318) | Group2 (665, 1812) |
|---------|------------------|------------------|
| /æ/     | 0.693            | 0.307            |
| /ɛ/     | 0.371            | 0.629            |

| Phoneme | Group1 (1019, 1562) | Group2 (697, 1999) |
|---------|------------------|------------------|
| /æ/     | 0.928            | 0.072            |
| /ɛ/     | 0.060            | 0.940            |

Table 4: Archetypal analysis using $F_1$(Hz), $F_2$(Hz) estimates of /æ/ and /ɛ/from Singaporean, American and British children. Archetypal extreme points for each cluster are in the format (F1, F2).

children have the lowest F1 ($M = 715$), American children have higher F1 estimates ($M = 786$), which is similar to British children ($M = 779$). One-way ANOVA tests demonstrated that for both vowels, these differences are statistically significant, $F(2, 412) = 108.7$, $p < 0.001$ for /æ/ and $F(2, 413) = 43.26$, $p < 0.001$ for /ɛ/. We followed up our ANOVA tests with post hoc Tukey’s HSD Test which show that Singaporean children’s average F1 for /æ/ is significantly lower than that of American children ($p < 0.001$) and British children ($p < 0.001$). However, the difference between American and British speakers is not significant ($p = 0.244$). Similarly, Singaporean children’s average F1 for /ɛ/ is significantly lower than those of British ($p < 0.001$) children and American children ($p < 0.001$), but there is no significant difference between the latter two ($p = 0.801$). In terms of articulatory implication, this suggests that compared to American and British children, Singaporean children produce both /æ/ and /ɛ/ vowels with a higher tongue position.
| Phoneme | Singaporesan Children | American Children | British Children |
|---------|-----------------------|-------------------|-----------------|
|         | Group1 (767, 2183)    | Group1 (872, 2188)| Group1 (891, 1656) |
| /æ/     | 0.594                 | 0.550             | 0.976           |
| /ɛ/     | 0.604                 | 0.450             | 0.024           |
|         | Group2 (726, 1978)    | Group2 (816, 1957)| Group2 (795, 1972) |
|         | 0.406                 | 0.500             | 0.004           |
|         | 0.396                 | 0.564             | 0.833           |

Table 5: Kmeans clustering using $F_1$(Hz), $F_2$(Hz) estimates of /æ/ and /ɛ/from Singaporean, American and British children. Centroids for each cluster are in the format ($F_1$, $F_2$).

5.2.3. $F_2$ formant estimates

For /æ/, Singaporean children ($M = 2089$) and American children ($M = 2083$) have similarly high $F_2$, while British children have a much lower $F_2$ ($M = 1665$). The same trend is observed for /ɛ/, where Singaporean children also have the highest $F_2$ ($M = 2113$), American children have slightly lower $F_2$ ($M = 2060$), and British children have the lowest $F_2$ ($M = 1918$). One-way ANOVA showed that for both vowels, these differences are statistically significant, $F(2, 412) = 353.8, p < 0.001$ for /æ/ and $F(2, 413) = 55.79, p < 0.001$ for /ɛ/. We followed up our ANOVA tests with post hoc Tukey’s HSD Test, which show that for both /æ/ and /ɛ/, Singaporean and American children’s $F_2$ are significantly higher ($p < 0.001$) than those of British children. Within Singaporean and American children, they are not significantly different ($p = 0.916$) in terms of $F_2$ for /æ/. For /ɛ/, Singaporean children demonstrate higher $F_2$ compared to American children ($p < 0.01$). In terms of articulatory implications, Singaporean and American children consistently exhibit fronting of both /æ/ and /ɛ/ such that they are produced
with a much fronter tongue position compared to those of British children. Within Singaporean and American children, their /æ/ vowels are produced with similar frontness while for /ɛ/, Singaporean children demonstrate even more fronting than American children.

5.3. Summary: Singaporean and American children similarly demonstrate fronting of /æ/ and /ɛ/ vowels and show less clear distinction between these vowels compared to British children

Inferring from the acoustic characterizations of /æ/ and /ɛ/, the corresponding articulatory implication is that Singaporean children, compared to American and British children, are producing these vowels with a higher tongue position. Interestingly, this tendency for producing vowels with a higher tongue position, is not only observed for /æ/ and /ɛ/ but also for the TRAP−BATH split [æ] and [ɑ] vowels we examined earlier in subsection 4.2. Compared to British children, Singaporean and American children are more similar in terms of their fronting of both /æ/ and /ɛ/. This fronting behavior also bears resemblance to the trend we observed for TRAP−BATH split [æ] and [ɑ] vowels in subsection 4.2 With the fronting of both /æ/ and /ɛ/, Singaporean and American children do not show as much distinction between these two vowels compared to British children as the two vowels are produced with similar frontness that is more characteristic of /ɛ/ than /æ/.

These trends observed from formant analysis agree with and provide further explanations for the results from our clustering experiments. Particularly, the fronting of /æ/ and /ɛ/ by Singaporean and American children such that these two vowels are produced with similar frontness (of the tongue position) likely explains why /æ/ and /ɛ/ from these two speaker popula-
tions do not fall into clean distinct clusters like those of British speakers in our unsupervised clustering experiments.

6. /ɑ/ vs. /ɔ/

/ɑ/ and /ɔ/ are both back vowels, with /ɔ/ having slightly lower F1 and F2 estimates (difference in F2 estimates even less prominent) compared to the /ɑ/ phoneme [46]. Peter and Keith [55] noted that many Midwestern and Californian speakers of American English do not distinguish [ɑ] and [ɔ] as in cot and caught and usually have a vowel intermediate, whereas many speakers of British English have an additional vowel in this area and distinguish between the vowels [ɑ], [ɒ], [ɔ] as balm, bomb and bought.

Given the acoustic characteristics of the two phonemes, producing /ɑ/ with a higher tongue position, resulting in lower F1, could lead to potential confusion with /ɔ/. Having observed that Singaporean children exhibit lower F1 in their TRAP–BATH split vowels (across realizations as [æ] and [a] phones) in subsection 4.2 as well as in /æ/ and /ɛ/ vowels in subsection 5.2, we further examine if they show similar traits for /ɑ/ and /ɔ/ vowels. We also investigate how production of /ɑ/ and /ɔ/ phonemes might differ across the three speaker groups.

6.1. Unsupervised clustering

We perform per-speaker-group clustering across /ɑ/ and /ɔ/ vowels from each group using F1 and F2 estimates. Clustering results using Archetypal analysis and Kmeans clustering are shown in Tables 7 and 8. The results from the two clustering methods both show that /ɑ/ and /ɔ/ vowels produced by each of the speaker groups are largely (about at least 80%) grouped into two clean, distinctive clusters. This distinction is only less clear for British children’s /ɑ/ vowels when using Archetypal analysis (67.1 % in majority group).

6.2. Acoustic Analysis and Characterization

6.2.1. F1 F2 formant space

We present the mean and standard error for F1 and F2 estimates in Table 9 and visualize this data on a per speaker level in Figure 6. We observe that for all three speaker groups, the /ɑ/ and /ɔ/ vowels are largely distinguished from each other, with /ɔ/ vowels being produced with lower F1 and F2 estimates. The /ɑ/ vowels produced by Singaporean children share similar
Table 7: Archetypal analysis using $F_1$(Hz), $F_2$(Hz) estimates of /$\alpha$/ and /$\beta$/ from Singaporean, American and British children. Archetypal extreme points for each cluster are in the format ($F_1$, $F_2$).

F1 and F2 characteristics as those produced by American children. This similarity is also observed for the /$\beta$/ vowels produced by the two speaker groups. Compared to the /$\alpha$/ and /$\beta$/ vowels produced by Singaporean and American children, the two vowels seem closer in the formant space and show slight overlapping for British speakers. We further analyzed these formant characteristics in detail in terms of F1 and F2 separately.

6.2.2. F1 formant estimates

For /$\alpha$/-, British children have the lowest F1 ($M = 699$), Singaporean children have higher F1 estimates ($M = 896$), and American children show slightly higher F1 ($M = 916$) compared to that of Singaporean children. Similarly, for the production of /$\beta$/-, British children have the lowest F1 ($M = 569$), and the F1 estimates for Singaporean children ($M = 751$) and American children ($M = 760$) are close to each other and higher than that of British children. One-way ANOVA tests demonstrated that for both vowels, these differences are statistically significant, $F(2, 414) = 286.5$, $p < 0.001$.
Figure 6: Singaporean and American children are more similar in terms of $F1$ and $F2$ estimates for /a/ and /ɔ/ vowels compared to British children. Smaller points: individual speaker’s mean $F1$ and $F2$; larger points: group means for the speaker groups.
Table 8: Kmeans clustering using $F1$(Hz), $F2$(Hz) estimates of /a/ and /ɔ/ from Singaporean, American and British children. Centroids for each cluster are in the format (F1, F2).

for /a/ and $F(2, 412) = 311.9$, $p < 0.001$ for /ɔ/. We followed up our ANOVA tests with post hoc Tukey’s HSD Test which show that British children’s average F1 for /a/ and /ɔ/ vowels are significantly lower than those of Singaporean and American children ($p < 0.001$). However, the difference between Singaporean and American children is not significant ($p = 0.0284$ for /a/ and $p = 0.364$ for /ɔ/). In terms of articulatory implication, this suggests that compared to Singaporean and American children, British children produce both /a/ and /ɔ/ vowels with a higher tongue position.

6.2.3. $F2$ formant estimates

For /a/, Singaporean children ($M = 1535$) and American children ($M = 1433$) both have higher F2 compared to that of British children ($M = 1255$). Whereas for /ɔ/, the F2 estimates across the three speaker groups are more similar, with American children having the highest F2 ($M = 1234$), Singaporean children having just slightly lower F2 ($M = 1195$), and British children having the lowest F2 ($M = 1153$). One-way ANOVA showed that for
both vowels, these differences are statistically significant, $F(2, 414) = 239.7, p < 0.001$ for /a/ and $F(2, 412) = 15.28, p < 0.001$ for /o/. We followed up our ANOVA tests with post hoc Tukey’s HSD Test, which show that for both /a/ and /o/, Singaporean and American children’s F2 estimates are significantly higher ($p < 0.01$) than those of British children. This suggests that in terms of articulation, Singaporean and American children exhibit fronting of both /a/ and /o/ such that they are produced with a more fronted tongue position compared to those of British children. Between Singaporean and American children, Singaporean children’s F2 estimates are higher for /a/ ($p < 0.001$) and American children’s are higher for /o/ ($p = 0.00299$).

### 6.3. Summary: Singaporean and American children similarly demonstrate lower but more fronted tongue positions for /a/ and /o/ vowels compared to British children and show clear distinction between the two vowels

Compared to British children, Singaporean and American children are more similar in producing their /a/ and /o/ vowels with a lower tongue position and in fronting both vowels. The observation that British children are producing these vowels with a higher tongue position (lower F1) such that their /a/ and /o/ vowels are produced with more similar heights possibly explains the unsupervised clustering results that distinction between the two vowels is only less clear for British children (when using Archetypal analysis).

F1 characteristics of /a/ and /o/ revealed an interesting trend for this vowel pair where British children articulate these vowels with higher vowel heights, in contrast to the trends we observed for TRAP−BATH split, /æ/, and /ɛ/ vowels (subsections 4.2, 5.2) where it is usually Singaporean and sometimes also American children who produce the vowels with a higher

| Corpus | Phone | F1 mean | F1 se | F2 mean | F2 se |
|--------|-------|---------|-------|---------|-------|
| SG     | [a]   | 896     | 4.43  | 1535    | 7.40  |
|        | [o]   | 751     | 4.15  | 1195    | 9.13  |
| AE     | [a]   | 916     | 6.98  | 1433    | 8.45  |
|        | [o]   | 760     | 5.83  | 1234    | 6.59  |
| BE     | [a]   | 699     | 7.63  | 1255    | 9.32  |
|        | [o]   | 569     | 5.79  | 1153    | 10.4  |

Table 9: Mean and standard error (se) for each speaker group for /a/ and /o/ formant estimates.
tongue position. In terms of the frontness of the vowels, Singaporean and American children’s fronting of /ɑ/ and /ɔ/ is similar to the trends we observed for TRAP–BATH split [æ] and [a] vowels in subsection 4.2 as well as for /æ/ and /ɛ/ in subsection 5.2.

7. Tense and lax vowels - /u/ vs. /ʊ/

We can characterize the differences between tense and lax vowels by considering them in pairs. In each tense and lax vowel pair, the lax vowel is produced with a shorter duration, lower tongue position and is slightly more centralized in the formant space compared to the corresponding tense vowel \[55\]. One such pair is [u, ʊ] as in boot, foot, where /u/ is a tense vowel and /ʊ/ is a lax vowel. In terms of acoustic characteristics, this means that the /ʊ/ vowel would have higher F1 and F2 estimates compared to the tense vowel /u/. Further, the duration of /ʊ/ would also be shorter than that of /u/.

Given these characteristics, if /u/ is produced with slightly higher F1 and F2 estimates, /ʊ/ is produced with slightly lower F1 and F2 estimates, or the duration difference is less conspicuous, the two vowels can be easily confused. We observed in sections 4, 5 and 6 that some speaker groups tend to produce some vowels with a higher tongue position, or show fronting of some vowels. In this section, we investigate if the speaker groups show similar trends for the /u/ and /ʊ/ vowel pair, and how that might affect the distinction between the tense and lax vowels in this pair.

7.1. Unsupervised clustering

We perform per-speaker-group clustering across /u/ and /ʊ/ vowels from each speaker group, using F1, F2 and duration estimates as features. The results from unsupervised clustering experiments using Archetypal analysis and Kmeans clustering are summarized in Tables 10 and 11. The Archetypal analysis results indicate that /u/ and /ʊ/ vowels produced by Singaporean children are similar as they largely fall into the same group, whereas /u/ and /ʊ/ vowels produced by American and British children largely (≥ 90 %) fall into two clean, distinctive groups. For this vowel pair, Kmeans clustering generally give less clean groups, and a clear distinction (≥ 90 %) between the two vowels is only observed for British children.

\[55\] also points that there can be variations amongst American and British speakers...
Table 10: Archetypal analysis using $F_1$(Hz), $F_2$(Hz), duration(seconds) estimates of /u/ and /U/ from Singaporean, American and British children. Archetypal extreme points for each cluster are in the format (F1, F2, duration).

### 7.2. Acoustic and Duration Analysis

#### 7.2.1. $F_1$ $F_2$ formant space

Estimates of $F_1$ and $F_2$ for /u/ and /U/ in the three speaker groups are summarized in Table 10. Using each speaker as a data point, we visualize the differences between speaker groups for these two vowels in Figure 7. We observe that /u/ and /u/ produced by British children are the most clearly distinguished from each other, with /u/ having higher $F_1$ and lower $F_2$ estimates than that of /u/. However, the $F_1$ and $F_2$ distinction between these two vowels for Singaporean and American children are less conspicuous – /u/ and /u/ share relatively more similar $F_2$ estimates for American children, and the two vowel groups are generally close to each other in formant space for Singaporean children. We further analyzed this tense and lax vowel pair produced by the speaker groups in detail, looking into $F_1$ and $F_2$ characteristics separately, and also in terms of vowel duration.
Figure 7: Singaporean children are more similar to American children in terms of $F1$ and $F2$ for /u/ and /u/ vowels. Smaller points: individual speaker’s mean $F1$ and $F2$; larger points: group means for the speaker groups.
Table 11: Kmeans clustering using $F_1$(Hz), $F_2$(Hz), duration(seconds) estimates of /u/ and /U/ from Singaporean, American and British children. Centroids for each cluster are in the format (F1, F2, duration).

| Phoneme | Singaporean Children | American Children | British Children |
|---------|----------------------|-------------------|------------------|
| /u/     | Group1: (545, 1223, 0.146) | Group1: (546, 1562, 0.134) | Group1: (522, 1726, 0.0840) |
|         | 0.458                | 0.493             | 0.098            |
| /u/     | 0.724                | 0.664             | 0.976            |
|         | 0.542                | 0.507             | 0.902            |
|         | 0.276                | 0.336             | 0.024            |

7.2.2. $F_1$ formant estimates

For /u/, British children have the lowest F1 ($M = 426$), America children have higher F1 estimates ($M = 477$), and Singaporean children show the highest F1 ($M = 520$) amongst the three speaker groups. For the production of /U/, it is similar in that British children also have the lowest F1 ($M = 533$), and the F1 estimates for Singaporean children ($M = 574$) and American children ($M = 607$) are higher than that of British children. One-way ANOVA tests demonstrated that for both vowels, these differences are statistically significant, $F(2, 411) = 105.4$, $p < 0.001$ for /u/ and $F(2, 411) = 32.89$, $p < 0.001$ for /o/. We followed up our ANOVA tests with post hoc Tukey’s HSD Test which show that British children’s average F1 for /u/ and /U/ vowels are significantly lower than those of Singaporean and American children ($p < 0.001$). In terms of articulatory implication, this suggests that compared to Singaporean and American children, British children produce both /u/ and /U/ vowels with a higher tongue position. Between Singaporean and American children, Singaporean children’s tongue positions are
Table 12: Mean and standard error (se) for each speaker group for /u/ and /o/ formant estimates.

| Corpus | Phoneme | F1 mean | F1 se | F2 mean | F2 se |
|--------|---------|---------|-------|---------|-------|
| SG     | /u/     | 520     | 3.86  | 1393    | 11.9  |
|        | /o/     | 574     | 4.54  | 1279    | 12.7  |
| AE     | /u/     | 477     | 4.28  | 1680    | 12.8  |
|        | /o/     | 607     | 6.26  | 1641    | 11.7  |
| BE     | /u/     | 426     | 4.15  | 2083    | 15.4  |
|        | /o/     | 533     | 6.27  | 1722    | 13.7  |

lower for /u/ ($p < 0.001$) and American children’s are lower for /o/ ($p < 0.001$).

7.2.3. F2 formant estimates

For /u/, Singaporean children have the lowest F2 estimates ($M = 1393$), American children have higher F2 estimates ($M = 1680$), and British children have the highest F2 ($M = 2083$). The same trend is observed for /o/, where Singaporean children also have the lowest F2 ($M = 1279$), American children have higher F2 ($M = 1641$) and British children demonstrate the highest F2 ($M = 1722$). One-way ANOVA showed that for both vowels, these differences are statistically significant, $F(2, 411) = 578.3, p < 0.001$ for /u/ and $F(2, 411) = 336.8, p < 0.001$ for /o/. We followed up our ANOVA tests with post hoc Tukey’s HSD Test, which show that for both /u/ and /o/, Singaporean and American children’s F2 estimates are significantly lower ($p < 0.001$) than those of British children. This suggests that in terms of articulation, Singaporean and American children’s /u/ and /o/ are produced with a less fronted tongue position compared to those of British children. Between Singaporean and American children, it is consistent across the two vowels that Singaporean children demonstrate even less fronted tongue position compared to those of American children ($p < 0.001$).

7.2.4. Duration

We compare the differences in vowel duration between the tense and lax vowels for /u/ and /o/. T-tests for correlated samples revealed that all three speaker groups produce /u/ vowels that are of significantly longer duration compared to their /o/ vowels (Singaporean children: $t(191) = 33.12, p < 0.001$ (two-tailed); American children: $t(139) = 36.09, p < 0.001$ (two-tailed);
Figure 8: /u/ and /u/ duration ratio comparison across speaker groups. All groups show clear duration distinction between the two vowels, with /u/ being produced with a vowel duration least 1.6 times that of /u/. Colored points mark the mean duration ratio for each speaker group, whereas the error bars indicate standard error.

British children: $t(81) = 23.64, p < 0.001$ (two-tailed)). We then investigate if this duration distinction is even more prominent for some speaker group(s). Figure 8 shows the trend for duration ratio across speaker groups where for each speaker, the /u/ and /u/ duration ratio is calculated by

$$\text{duration ratio} = \frac{\text{average duration of } /u/ \text{ from that speaker}}{\text{average duration of } /u/ \text{ from that speaker}}.$$ 

One-way ANOVA showed that the differences in duration ratio across speaker groups are statistically significant, $F(2, 411) = 43.71, p < 0.001$. Post hoc Tukey’s HSD Test show that the /u/ and /u/ duration ratio for Singaporean ($M = 1.65$) and British children ($M = 1.61$) are significantly smaller ($p < 0.001$) compared to American children ($M = 1.93$). However, the /u/ and /u/ duration ratio for Singaporean and British children are not significantly different ($p = 0.535$).
7.3. Summary: Singaporean and American children similarly demonstrate lower and less fronted tongue positions for /u/ and /u/ vowels compared to British children. All groups show clear duration distinction between the two vowels.

Singaporean children’s /u/ and /u/ vowels are more similar to American children in that they are produced with a lower tongue position (higher F1 estimates) compared to British children. This trend is similarly observed when the speakers produce other back vowels like /a/ and /o/ (subsection 6.2). Whereas, for TRAP–BATH split, /æ/, and /ɛ/ vowels (subsections 4.2, 5.2), it was Singaporean and sometimes also American children who usually produce the vowels with a higher tongue position than that of British children.

Singaporean children’s /u/ and /u/ are also similar to American children in that they are produced with a less fronted tongue position (lower F2 estimates) compared to those of British children. Interestingly, this characteristic of Singaporean children’s /u/ and /u/ tense lax vowel pair is different from when they produce other vowels like the TRAP–BATH split [æ] and [a] vowels, /æ/, /ɛ/, /a/ and /ɔ/ (subsections 4.2, 5.2, 6.2) where they consistently demonstrated fronting of the vowels.

In terms of F1 and F2 characteristics, as seen in Figure 7, British children show the greatest distinction with /u/ being produced with a higher tongue height and more fronted tongue position than /u/. This possibly explains the results from kmeans clustering where clear distinction (≥90 %) between the two vowels is only observed for British children. Although all speakers show significant duration difference between /u/ and /u/, the difference is the most prominent for American children. Compared to British and American children, Singaporean children neither mostly clearly distinguishes /u/ and /u/ through formant feature differences nor duration ratio differences. This potentially accounts for why the /u/ and /u/ tokens produced by Singaporean children do not fall into two distinct groups for our unsupervised clustering experiments, a trend that is consistently observed across Archetypal analysis and Kmeans clustering.

Finally, it is interesting that across children in all three speaker groups, /u/ is articulated with a lower F2 (if not, similar) compared to /u/, which differs from the usual canonical characteristics of /u/ and /u/. 55’s chapter on “English Vowels” also illustrated diagrams showing how there could be such variations amongst American and British speakers. From our work,
such variation may also be an interesting characteristic in children speech and worth further investigation in future work.

8. Tense and lax vowels - /i/ vs. /ɪ/

Another tense and lax vowel pair is [i, ɪ] as in *beat*, *bit*. The lax vowel /ɪ/ is shorter, lower, and more centralized compared to the corresponding tense vowel /i/. In terms of acoustic characteristics, this means that the /ɪ/ vowel would have higher F1 and lower F2 estimates compared to the tense vowel /i/. Further, the duration of /ɪ/ would also be shorter than that of /i/.

Given these characteristics, if /i/ is produced with slightly higher F1 and lower F2 estimates, /ɪ/ is produced with slightly lower F1 and higher F2 estimates, or the duration difference is less conspicuous, the two vowels can be easily confused. It is consistent across the previous sections (sections 4, 5, 6, 7) that some speaker groups tend to produce particular vowels with a higher tongue position, or show fronting of particular vowels. We also observed that the effect may differ based on the vowel concerned. In this section, we investigate the trends for the /i/ and /ɪ/ vowel pair, and make comparisons across different speaker groups.

8.1. Unsupervised clustering

We perform per-speaker-group clustering across /i/ and /ɪ/ vowels from each speaker group using F1, F2 and vowel duration estimates. Clustering results using Archetypal analysis and Kmeans clustering are shown in Tables 13 and 14. The Archetypal analysis results indicate that /i/ and /ɪ/ vowels produced by American and British children largely (> 90 %) fall into two clean, distinctive groups, whereas the distinction is less clear for the /i/ and /ɪ/ produced by Singaporean children. Kmeans clustering generally give less clean groups, but the results similarly show that the distinction between the two vowels for Singaporean children is less clear compared to the other two speaker groups.

8.2. Acoustic and Duration Analysis

8.2.1. F1 F2 formant space

Estimates of F1 and F2 for /i/ and /ɪ/ in the three speaker groups are summarized in Table 15. Using each speaker as a data point, we visualize the differences between speaker groups for the two vowels in Figure 9.
Table 13: Archetypal analysis using F1(Hz), F2(Hz), duration(seconds) estimates of /i/ and /I/ from Singaporean, American and British children. Archetypal extreme points for each cluster are in the format (F1, F2, duration).

| Phoneme | Group1          | Group2          |
|---------|-----------------|-----------------|
|         | (507, 2241, 0.0902) | (526, 2899, 0.279) |
| /i/     | 0.344           | 0.656           |
| /I/     | 0.844           | 0.156           |

| Phoneme | Group1          | Group2          |
|---------|-----------------|-----------------|
|         | (607, 2142, 0.0801) | (363, 2898, 0.242) |
| /i/     | 0.086           | 0.914           |
| /I/     | 0.993           | 0.007           |

| Phoneme | Group1          | Group2          |
|---------|-----------------|-----------------|
|         | (564, 2040, 0.0622) | (403, 2779, 0.232) |
| /i/     | 0.071           | 0.929           |
| /I/     | 1.000           | 0.000           |

observe that the /i/ and /I/ vowels produced by Singaporean children are close to each other in the formant space. This observation is different from both that for American and British children whose /i/ and /I/ vowels form more distinctively separate groups in the formant space, with /I/ having a higher F1 and lower F2 than that of /i/.

8.2.2. F1 formant estimates

Singaporean children’s F1 estimates for /i/ (M = 515) is much higher compared to that of American (M = 432) and British children (M = 437). Whereas, for the production of /I/, Singaporean children have the lowest F1 (M = 516), the F1 estimates for British children is higher (M = 547) and American children (M = 578) show the highest F1 estimates amongst the three groups. One-way ANOVA tests demonstrated that for both vowels, these differences are statistically significant, $F(2, 413) = 112.8$, $p < 0.001$ for /i/ and $F(2, 413) = 81.24$, $p < 0.001$ for /I/. We followed up our ANOVA tests with post hoc Tukey’s HSD Test which show that Singaporean chil-
Figure 9: Compared to American and British children, Singaporean children’s \( /i/ \) and \( /I/ \) are less distinguished in terms of \( F1 \) and \( F2 \) estimates. Smaller points: individual speaker’s mean \( F1 \) and \( F2 \); larger points: group means for the speaker groups.
| Phoneme | Group1 (F1, F2, Duration) | Group2 (F1, F2, Duration) |
|---------|---------------------------|---------------------------|
| /i/     | 0.266 (513, 2370, 0.148)  | 0.734 (517, 2658, 0.193)  |
| /i/     | 0.583 (515, 2179, 0.107)  | 0.417 (460, 2598, 0.169)  |

**Table 14**: Kmeans clustering using $F_1$(Hz), $F_2$(Hz), duration(seconds) estimates of /i/ and /i/ from Singaporean, American and British children. Centroids for each cluster are in the format (F1, F2, duration).

dren’s F1 estimates for /i/ are significantly higher ($p < 0.001$) than those of American and British children, while American and British children are not significantly different from each other ($p = 0.832$). For /i/, American children’s F1 estimates are significantly higher than that of the other two groups ($p < 0.001$). Between Singaporean and British children, Singaporean children’s F1 estimates for /i/ are even lower than British children. In terms of articulatory implication, this suggests that amongst the three speaker groups, Singaporean children produce /i/ with the lower tongue position and /i/ with a higher tongue position compared to the other two speaker groups. American children produce /i/ with the lowest tongue position.

**8.2.3. F2 formant estimates**

For /i/, British children have the lowest F2 estimates ($M = 2511$), Singaporean ($M = 2609$) and American children ($M = 2633$) have higher F2 estimates which are more similar to each other. For /i/, British children also have the lowest F2 ($M = 2201$), American children have higher F2 ($M =$
2282) and Singapore children have the highest F2 ($M = 2463$). One-way ANOVA showed that for both vowels, these differences are statistically significant, $F(2, 413) = 10.2, p < 0.001$ for /i/ and $F(2, 413) = 96.38, p < 0.001$ for /I/. We followed up our ANOVA tests with post hoc Tukey’s HSD Test, which show that Singaporean and American children’s F2 estimates for /i/ are significantly higher ($p < 0.001$) than those of British children, whereas the difference within Singaporean and American children is not significant ($p = 0.535$). For /I/, all pairwise differences are significant ($p < 0.001$). Singaporean and American children’s F2 estimates for /I/ are also significantly higher ($p < 0.001$) than those of British children. Between Singaporean and American children, Singaporean children’s F2 estimates for /I/ are even higher ($p < 0.001$) than American children, indicating that they demonstrate more fronting. These suggest that in terms of articulation, Singaporean and American children’s /i/ and /I/ are produced with a more fronted tongue position compared to those of British children.

### 8.2.4. Duration

We compare the differences in vowel duration between the tense and lax vowel pair for /i/ and /I/. T-tests for correlated samples revealed that all three speaker groups the produce /i/ vowels that are of significantly longer duration compared to their /I/ vowels (Singaporean children: $t(191) = 39.28, p < 0.001$ (two-tailed); American children: $t(139) = 47.12, p < 0.001$ (two-tailed); British children: $t(83) = 21.83, p < 0.001$ (two-tailed)). We then investigate if this duration distinction is more prominent for some speaker group(s). Figure[10] shows the trend for duration ratio across speaker groups.

| Corpus | Phoneme | F1 mean | F1 se | F2 mean | F2 se |
|--------|---------|---------|-------|---------|-------|
| SG     | /i/     | 515     | 4.99  | 2609    | 10.5  |
|        | /I/     | 516     | 2.96  | 2463    | 12.7  |
| AE     | /i/     | 432     | 3.09  | 2633    | 19.1  |
|        | /I/     | 578     | 3.78  | 2282    | 11.7  |
| BE     | /i/     | 437     | 4.55  | 2511    | 28.3  |
|        | /I/     | 547     | 5.27  | 2201    | 17.0  |

Table 15: Mean and standard error (se) for each speaker group for /i/ and /I/ formant estimates.
Figure 10: /i/ and /I/ duration comparison across speaker groups. All groups show clear duration distinction between the two vowels, with /i/ being produced with a vowel duration least 1.5 times that of /I/. Colored points mark the mean duration ratio for each speaker group, whereas the error bars indicate standard error.

where for each speaker, the /i/ and /I/ duration ratio is calculated by

\[
\text{duration ratio} = \frac{\text{average duration of } /i/ \text{ from that speaker}}{\text{average duration of } /I/ \text{ from that speaker}}.
\]

One-way ANOVA showed that the differences in duration ratio across speaker groups are statistically significant, \( F(2, 413) = 137.2, p < 0.001 \). Post hoc Tukey’s HSD Test show that the /i/ and /I/ duration ratio for Singaporean children (\( M = 1.51 \)) is significantly smaller (\( p < 0.001 \)) compared to those of American (\( M = 1.78 \)) and British children (\( M = 1.86 \)). Between American and British children, the duration ratio is even greater for British children (\( p = 0.00902 \)). Therefore, the characteristic of /i/ being produced with a longer vowel duration compared to /I/ is significantly more prominent in American and British children compared to Singaporean children, and is most prominent in British children.
8.3. Summary: Singaporean children produce /i/ with a lower tongue position and /I/ with a higher tongue position compared to the other two speaker groups. They show the least distinction between this vowel pair.

Amongst the three groups, Singaporean children produce the tense vowel /i/ with the lowest tongue position and American children produce the lax vowel /I/ with the lowest tongue position (highest F1 estimates). This trend bears resemblance to what we have observed for the tense lax vowel pair /u/ and /u̯/ (subsection 7.2), where Singaporean children also produce the tense vowel /u/ with the lowest tongue position and American children produce the lax vowel /u̯/ with the lowest tongue position amongst the speaker groups.

Singaporean children’s /i/ and /I/ are more similar to those of American children in that they are produced with a more fronted tongue position compared to those of British children. This aligns with the trends we observed when these speakers produce other vowels like the TRAP−BATH split [æ] and [a] vowels, /æ/, /ɛ/, /ɑ/, and /ə/ (sections 4.2, 5.2, 6.2) where they consistently demonstrated fronting of the vowels compared to British children.

In terms of F1 and F2 characteristics, as seen in Figure 9, American and British children show greater distinction between this tense lax vowel pair with /i/ being produced with a higher tongue height and more fronted tongue position than /I/. Singaporean children tend to produce /i/ with a lower tongue position and /I/ with a higher tongue position compared to the other groups such that their /i/ and /I/ are not significantly different in terms of vowel height, $t(191) = 0.311, p = 0.756$ (two-tailed). Further, Singaporean children also consistently exhibit fronting of the vowels such that the vowels are closer in terms of the frontness. For duration differences, although all speaker groups show significant duration differences between /i/ and /I/, the difference is the more prominent for British and American children. These trends support why /i/ and /I/ for Singaporean children give less clean clusters in the unsupervised clustering experiments. Therefore, exploratory analysis using supervised learning, formant space analysis and vowel duration analysis all provide evidence that point towards Singaporean children distinguishing less clearly between /i/ and /I/ in their articulation compared to American and British children.

9. Analysis of lateral approximant /l/

The lateral approximant /l/ has traditionally been observed to have two variants, namely, dark /l/ which appears in syllable rimes (syllable-final /l/)
and clear /l/ in syllable onsets (syllable initial /l/) [56, 57]. Previous studies such as Hansen [58] and He [59] have shown that speakers of Mandarin experienced difficulties in producing syllable-final /l/, and tend to modify syllable-final /l/ in their production. Given that many Singaporeans are bilingual speakers of both English and Mandarin, and the Chinese languages spoken in Singapore, we devote this section towards investigating if this sociolinguistic background affects Singaporean children’s production of syllable-final /l/ like the way Mandarin speakers’ production are affected [58, 59].

Acoustic characteristics of syllable-final /l/ include 1) lowering of F1 and F2 formant features when transitioning from a preceding vowel [61], 2) a relatively wide gap between F2 and F3 [62, 46], and 3) F1 and F2 are expected to have values close to each other when the phoneme is present [63]. We focus on these formant features and compare how the production of syllable-final /l/ vary across Singaporean, American and British children.

9.1. Unsupervised clustering

Archetypal analysis and Kmeans clustering on syllable-final /l/ tokens across the three speakers groups were performed using the first three formants as features. The results for Archetypal analysis are summarized in Table 16 whereas the results obtained using Kmeans clustering are summarized in Table 17. For each row, the two numbers add up to 1.0 (stands for 100%). This illustrates the proportion of each speaker group’s tokens that gets grouped into Group 1 and Group 2 respectively. Both sets of results show that syllable-final /l/ tokens produced by Singaporean children are mostly grouped into a different group from those produced by British children speakers, whereas those produced by American children is somewhere between those two groups with more tokens being grouped into the British dominant group. To quantify this difference, detailed acoustic analysis using the characteristic formant features of syllable-final /l/ tokens was carried out.

9.2. Acoustic Analysis

Our acoustic analysis serves to compare how well the syllable-final /l/ tokens produced by each speaker group demonstrates the standard acoustic features of the phoneme as reported in literature, in terms of the first three formants.
9.2.1. Falling F1

When transitioning into the syllable-final /l/ phoneme, we anticipate a lowering of F1 as compared to that in the preceding vowel [61]. That is, if we calculate the F1 difference between the syllable-final /l/ and the preceding vowel using Equation 4, we expect a negative value when a characteristic syllable-final /l/ has been produced, or in some cases, due to the formant features of the preceding vowel, a very small value.

\[ \text{F1 difference} = \frac{\text{Average F1 in syllable-final /l/} - \text{Average F1 in preceding vowel}}{} \]  

From the visualization of F1 differences during transition as shown in Figure 11, syllable-final /l/ tokens produced by American (\(M = -64.7\)) and British children (\(M = -90.8\)) show a decrease in F1 estimates whereas Singaporean children demonstrated a very slight increase (\(M = 4.77\)). One-way ANOVA demonstrated that these differences are statistically significant across speaker groups, \(F(2, 825) = 99.72, p < 0.001\). We then followed up our ANOVA test with a post hoc Tukey’s HSD Test which showed that,

| Corpus | Group1 \((830, 1870, 3450)\) | Group2 \((419, 1040, 3001)\) |
|--------|-------------------------------|-------------------------------|
| SG     | \(0.651\)                    | 0.349                        |
| AE     | 0.414                        | \(0.586\)                    |
| BE     | 0.012                        | \(0.988\)                    |

Table 16: Archetypal analysis using \(F1(\text{Hz})\), \(F2(\text{Hz})\), and \(F3(\text{Hz})\) estimates of syllable-final /l/ from Singaporean, American and British children. Archetypal extreme points for each cluster are in the format (F1, F2, F3).

| Corpus | Group1 \((661, 1558, 3298)\) | Group2 \((554, 1278, 3114)\) |
|--------|-------------------------------|-------------------------------|
| SG     | \(0.755\)                    | 0.245                        |
| AE     | 0.471                        | \(0.529\)                    |
| BE     | 0.061                        | \(0.939\)                    |

Table 17: Kmeans clustering using \(F1(\text{Hz})\), \(F2(\text{Hz})\), and \(F3(\text{Hz})\) estimates of syllable-final /l/ from Singaporean, American and British children. Centroids for each cluster are in the format (F1, F2, F3).
Figure 11: Comparison of F1 transition into syllable-final /l/. Singaporean children do not show conspicuous lowering of F1, compared to the other two speaker groups. The F1 difference plotted is calculated using the formula in Equation 4 on a per speaker basis. The error bars indicate standard error.

significant difference was observed between Singaporean and American children speakers, as well as Singaporean and British children speakers at $p < 0.001$ but not between American and British children speakers ($p = 0.00377$). This shows that, where a syllable-final /l/ phoneme is involved, Singaporean children speakers maybe producing something that is acoustically much less characteristic of a typical syllable-final /l/ compared to other speakers in terms of F1 characteristics.

In Figure 12, we perform a more detailed analysis by breaking down the syllable-final /l/ tokens according to their different preceding vowels and observe that this trend for F1 transition is generally consistent across most preceding vowels.

9.2.2. Falling F2

In addition to falling F1, when transitioning into the syllable-final /l/ phoneme, we also anticipate a lowering of F2 as compared to that in the preceding vowel [61]. Similar to our analysis for falling F1, we plot the
Figure 12: Syllable final /l/ F1 transition comparison by preceding phoneme across speaker groups. Compared to the other two speaker groups, Singaporean children generally do not demonstrate lowering of F1 as clearly in this transition. The error bars indicate standard error. (Syllable final /l/ preceded by /aʊ/ is not present in the Singapore corpus)

corresponding differences for F1 as in Figure 13, where the differences are calculated using Equation 5.

\[
F2 \text{ difference} = \frac{\text{Average F2 in syllable-final } /l/}{-}\frac{\text{Average F2 in preceding vowel}}
\]  

(5)

Similar to the trend observed for F1 transition, syllable-final /l/ tokens produced by American (\(M = -86.5\)) and British children (\(M = -191\)) show a decrease in F1 estimates compared to its preceding token whereas Singaporean children demonstrated an increase (\(M = 301\)). One-way ANOVA tests demonstrated that these differences are statistically significant across speaker groups, \(F(2, 825) = 350.9, p < 0.001\). We then followed up our ANOVA test with a post hoc Tukey’s HSD Test which showed that, the differences between Singaporean and American children speakers, as well as Singaporean and British children speakers are significant at \(p < 0.001\). This
Figure 13: Comparison of F2 transition into syllable-final /l/. Singaporean children do not show conspicuous lowering of F2, compared to the other two speaker groups. The F2 difference plotted is calculated using the formula in Equation 5 on a per speaker basis. The error bars indicate standard error.

analysis on the F2 values further affirmed our conclusion based on F1 values, showing that Singaporean children maybe producing something that is acoustically much less characteristic of a typical syllable-final /l/ compared to other speakers in terms of F2 characteristics as well.

We perform a more detailed analysis by breaking down the syllable-final /l/ tokens according their different preceding vowels in Figure 14 and show that this trend for F2 transition (where Singaporean children do not exhibit falling F2 as clearly) is consistent across most preceding vowels.

9.2.3. Gap between F2 and F3

In addition, if a phoneme is indeed articulated as a characteristic syllable-final /l/, we expect a relatively wide gap between F2 and F3 [62, 46]. We compare the difference between F2 and F3 values across different speaker groups in Figure 15 which shows that this difference in formant values is the least in Singaporean children (M = 1563) compared to American (M = 2094) and British children (M = 1904) who demonstrated more conspicuous
Figure 14: Syllable final /l/ F2 transition comparison by preceding phoneme across speaker groups. Compared to the other two speaker groups, Singaporean children generally do not demonstrate lowering of F2 as clearly in this transition. The error bars indicate standard error. (Syllable final /l/ preceded by /au/ is not present in the Singapore corpus)

Gaps as should be prominent in syllable-final /l/. One-way ANOVA tests demonstrated that these differences are statistically significant across speaker groups, $F(2, 825) = 383, p < 0.001$. We then followed up our ANOVA test with a post hoc Tukey’s HSD Test which showed that the differences in F3 - F2 gap is significant across all pairwise comparisons ($p < 0.001$). Therefore, although should be a conspicuous gap between F2 and F3 if a characteristic syllable-final /l/ was articulated, but this characteristic is significantly less clearly demonstrated by Singaporean children compared to the other two speaker groups.

In Figure 16, we show that across the different preceding vowel conditions, this trend of Singaporean children showing a less conspicuous gap between F2 and F3 for syllable-final /l/ is consistent.

This analysis on differences between F2 and F3 formant values further support what we have concluded using F1, F2 formants alone, that Singaporean children do not produce a characteristic syllable-final /l/ at where it
Figure 15: Comparison of F3 - F2 gap in syllable-final /l/. Singaporean children show a less conspicuous gap between F2 and F3, compared to the other two speaker groups. The F3 - F2 difference plotted is calculated using the formula in Equation 6 per token. The error bars indicate standard error.

\[
F3 - F2 \text{ difference } = \frac{\text{Average F3 in syllable-final } /l/}{-} \frac{\text{Average F2 in syllable-final } /l/}{(6)}
\]

9.2.4. Difference between F1 and F2

In syllable-final /l/, we also expect the difference between F1 and F2 values to be small. The differences between F1 and F2 values across the speaker groups are visualized in Figure 17, showing that such difference is the greatest in Singaporean children (\(M = 950\)) where F1 and F2 should be close together if the syllable-final /l/ phoneme was indeed present, compared to American (\(M = 640\)) and British children (\(M = 713\)) who demonstrated smaller differences between F1 and F2 values. One-way ANOVA tests demonstrated that these differences are statistically significant across
speaker groups, $F(2, 825) = 348.1$, $p < 0.001$. A post hoc Tukey’s HSD Test showed that across all pairwise comparisons, $F2 - F1$ differences are significantly different across the speaker groups ($p < 0.001$). Therefore, although the difference between $F1$ and $F2$ should be small if a characteristic syllable-final /l/ was articulated, Singaporean children instead demonstrated significantly greater difference between the two formants compared to the other two speaker groups.

In Figure 18, we compare differences between $F1$ and $F2$ values for the three speaker groups across the various preceding vowel conditions. We observe that this trend of Singaporean children showing a greater difference between $F1$ and $F2$ for syllable-final /l/ is consistent across almost all preceding vowel conditions, when the difference is instead expected to be small for a characteristic syllable-final /l/.

This analysis on differences between $F1$ and $F2$ formant values yields results that are consistent with what we have found using $F1$, $F2$ formants...
as well as analyzing the F2 and F3 gap. All four analyses agree and together, provide evidence for the observation that Singaporean children are not producing a characteristic syllable-final /l/ at where it occurs.

\[
\text{F2 - F1 difference} = \frac{\text{Average F2 in syllable-final } /l/ - \text{Average F1 in syllable-final } /l/}{7}
\]

9.3. **Summary:** Singaporean children’s acoustic realization of syllable-final /l/ differs from British and American children, and from the canonical syllable-final /l/.

From our acoustic analysis, when the phoneme /l/ is in the syllable-final position, Singaporean children tend to pronounce it differently compared to the other two speaker groups, giving it formant features differing from other speakers and from the canonical syllable-final /l/. According to the four
Figure 18: Syllable final /l/ F2 - F1 gap comparison by preceding phoneme across speaker groups. F1 and F2 should be close together for the syllable-final /l/ phoneme, but Singaporean children generally show the greatest F2 - F1 gap. The error bars indicate standard error. (Syllable final /l/ preceded by /aʊ/ is not present in the Singapore corpus)

acoustic characteristics that we have examined, the phonological interpretation would be that Singaporean children speakers’ production of syllable-final /l/ tokens undergo extreme articulatory change differing from an characteristic syllable-final /l/ as reported in literature. That is, Singaporean children’s syllable-final /l/ tokens, unlike the characteristic syllable-final /l/, are less prominent in showing characteristics of falling F1 and F2, conspicuous gap between F2 and F3, as well as small F2 - F1 difference. From these observations, we conclude that Singaporean children’s syllable-final /l/ tokens could well be altered into a very different sound or considered to be deleted. This potentially explains the results from our unsupervised clustering experiments which show that syllable-final /l/ tokens produced by Singaporean children are mostly grouped into a different group than the majority group for American and British children.

Given that studies like Hansen [58] and He [59] have shown that speakers of Mandarin experienced difficulties in producing syllable-final /l/, the
observation that Singaporean children’s syllable-final /l/s are altered into a very different sound in terms of formant features could possibly be attributed to influence from the various Chinese languages spoken in Singapore.

10. Analysis of rhotic approximant /ɹ/

The most characteristic acoustic feature of the rhotic approximant in English is that the third formant has very low frequency. Similar to our analysis on the lateral approximant /l/, we focus our analysis here on /ɹ/ tokens that occur in syllable-final positions since Mandarin speakers tend to modify English syllable codas. For /ɹ/, it is known that many Mandarin speakers drop the final /ɹ/ in Mandarin, which could then potentially influence their pronunciation of the English syllable-final /ɹ/. Given the various Chinese languages spoken in Singapore, we investigate if Singaporean children’s pronunciation of the phoneme is similarly affected. For other two speaker groups, Standard BBC English is not rhotic whereas rhotic accents are the norm in most parts of North America.

10.1. Unsupervised clustering

Archetypal analysis and Kmeans clustering using the characteristic F3 formant feature were carried out to explore potential differences across speaker groups. Clustering results using the two unsupervised clustering methods are summarized in Tables 18 and 19 respectively. The two numbers in each row add up to 1.0 (stands for 100%), illustrating the proportion of each speaker group’s tokens that gets grouped into Group 1 and Group 2 respectively. Both sets of results suggest that when a syllable-final /ɹ/ is present, Singaporean and British children are producing tokens with F3 characteristics differing from that of American children. We follow up on this result with a detailed acoustics analysis using the F3 formant estimates of syllable-final /ɹ/ tokens.

10.2. Acoustic Analysis

We compared the mean F3 formant values across the three speaker groups where a syllable-final /ɹ/ was supposed to be present. From the visualization

---

2We also tried the unsupervised clustering experiments using $F1$(Hz), $F2$(Hz), and $F3$(Hz) estimates together. Those experiments gave results very similar to that of using just $F3$(Hz) alone, therefore we only present the results from using just $F3$(Hz) since that is the most characteristic formant feature for our analysis in this section.
of F3 formant values across speakers in Figure 19, we observe that syllable-final /\textipa{1}/ tokens produced by Singaporean \((M = 2931)\) and British children \((M = 3019)\) have much higher F3 formant values compared to that of American children \((M = 2361)\). An one-way ANOVA test demonstrated that these differences are statistically significant across speaker groups, \(F(2, 411) = 404.9, p < 0.001\). A post hoc Tukey’s HSD Test further showed that the F3 estimates for American children are significantly lower than that of both Singaporean and British children \((p < 0.001)\), but the difference between the latter two groups is not as significant \((p = 0.00328)\). This aligns with our unsupervised clustering results, which also show that Singaporean and British children’s syllable-final /\textipa{1}/ have F3 characteristics differing from that of American children.

**10.3. Summary: Singaporean and British children exhibit non-rhotic characteristics at syllable-final /\textipa{1}/ in contrast to American children**

Our unsupervised clustering results show that for syllable-final /\textipa{1}/, Singaporean and British children have F3 characteristics differing from that of American children. The follow up acoustic analysis further elucidates this difference, showing that Singaporean and British children speakers do not
Figure 19: Syllable final /ı/ F3 comparison across speaker groups. Singaporean and British children exhibit non-rhotic characteristics at syllable-final /ı/ in contrast to American children. Colored points mark the mean F3 for each speaker group, whereas the error bars indicate standard error.

lower their F3 formant values nearly as much as American children speakers do for syllable-final /ı/. In terms of articulatory implications, this suggests that Singaporean and British children speakers lack rhoticity in their production of syllable-final /ı/ tokens, in contrast to American children.

11. Discussion

The vowel height and frontness characteristics that different speaker groups exhibit are summarized in Tables 20 and 21. We observe that the height and frontness of vowels produced by Singaporean children often exhibit characteristics that are opposite from that of British children. For instance, Singaporean children produce TRAP–BATH split vowels, /æ/ and /ɛ/ all with a higher tongue position, while British children demonstrate lower tongue height for these vowels and higher tongue height for the rest of the vowels examined. British children also demonstrate fronting of vowels for the only vowels that Singaporean children do not show fronting for, namely, /u/ and
Corpus

Vowels and tongue height

| Corpus | TRAP–BATH |
|--------|-----------|
|        | /æ/ | /e/ | /a/ | /o/ | /u/ | /i/ | /I/ |
| SG     | [æ] | [a] |     |     |     |     |     |
| AE     |     |     | [α] |     |     |     |     |
| BE     |     |     |     | [α] | [a] |     |     |

Table 20: Summary of tongue height characteristics exhibited by speaker groups for different vowels. Darker boxes indicate that the speaker group exhibits higher tongue positions for the particular vowel compared to other speaker groups.

Vowels and tongue frontness

| Corpus | TRAP–BATH |
|--------|-----------|
|        | /æ/ | /e/ | /a/ | /o/ | /u/ | /i/ | /I/ |
| SG     | [æ] | [a] |     |     |     |     |     |
| AE     |     |     | [α] |     |     |     |     |
| BE     |     |     |     | [α] | [a] |     |     |

Table 21: Summary of tongue frontness characteristics exhibited by speaker groups for different vowels. Darker boxes indicate that the speaker group exhibits more fronted tongue positions for the particular vowel compared to other speaker groups.

/ɑ/. In terms of fronting of the vowels, Singaporean children’s behavior appears more similar compared to American children. However, vowel height characteristics demonstrated by Singaporean children does not resemble any of the other two populations.

In tense and lax vowel pairs, /u/ and /ʊ/, as well as /i/ and /ɪ/, although all speakers show significant duration difference between the tense and lax vowels, we also consistently observe that the difference is less conspicuous for Singaporean children compared to the other speaker groups. For rhotic approximant /r/, Singaporean children are similar to British children in their lack of rhoticity. In the case of lateral approximant /l/, Singaporean children’s syllable-final /l/ exhibits characteristics unlike either American or British children, and differs from a canonical syllable-final /l/.

We show illustrative examples of syllable-final /l/ using instances of the word “cool” (/kul/) pronounced by Singaporean, American and British children in Figures 20a, 20b and 20c respectively. Comparing the formants in these spectrograms, American and British children’s F1 and F2 for syllable-
final /l/ are low and close together, but this is less so for Singaporean children whose /l/ tokens have F1 and F2 values being further apart. Further, one characteristic feature of syllable-final /l/ is that there is a huge gap between F2 and F3, but for Singaporean children F3 is closer towards being in the middle of F2 and F4, making the F2 and F3 gap less conspicuous. Specifically, in the spectrogram examples, F3 values for American and British children are around 3800Hz and 3900Hz whereas the F3 for the Singaporean speaker example is much lower (around 3300Hz) and therefore closer to the F2 formants.

We also performed preliminary analysis looking into other aspects such as prosody, as well as analyzing differences between spontaneous versus elicit speech across speakers. Such analysis revealed some interesting trends such as Singaporean children showing greater variation in their /æ/ vs. /ɛ/ vowels produced in spontaneous speech when compared to elicit speech. However, with these analysis, no consistent trends were observed across speaker groups nor for each vowel or approximant type. Therefore, in this paper, we focus on presenting our findings on formant and duration characteristics which revealed more consistent trends. Nonetheless, looking further into other aspects to analyze Singaporean children’s pronunciation patterns would definitely be an interesting direction for future work.

From our analysis in this work, we also found several pronunciation characteristics exhibited by Singaporean children that are not present for either American or British children. This work alludes to sociolinguistic perspectives of how Singapore English could have evolved beyond the British influence during historical colonization [65], moving towards also embodying American pronunciation characteristics, and even traits beyond American and British pronunciation patterns. It would be interesting to compare Singaporean children’s English to speech from other speaker groups in future work to study the different sources of influence. Investigating how Singapore English is also influenced by Malay [66] and the range of Chinese languages spoken [60] in Singapore can paint a more comprehensive picture of the complexities of Singapore English; this is a line of on-going research endeavors.

12. Conclusion

We presented a large-scale study to characterize Singaporean children’s English pronunciation patterns. We looked into different vowels (TRAP−BATH split vowels, /æ/ vs. /ɛ/, /ɑ/ vs. /ɔ/, /u/ vs. /u/, and /i/ vs. /ɪ/) and
(a) Spectrogram example for syllable-final /l/ token produced by a Singaporean child as in the word “cool” (/kul/).

(b) Spectrogram example for syllable-final /l/ token produced by an American child as in the word “cool” (/kul/).

(c) Spectrogram example for syllable-final /l/ token produced by a British child as in the word “cool” (/kul/).

Figure 20: Comparing spectrogram examples for syllable-final /l/ token produced by Singaporean, American and British children.
approximants (/l/ and /ɹ/) by first using unsupervised clustering to explore potential trends across the different speaker groups, followed by detailed acoustic analysis with linguistics insights. Our analysis showed that Singaporean children speakers are generally more similar to American children speakers in their pronunciation patterns of the vowels compared to British speakers, in exhibiting fronting. Whereas, when it comes to the pronunciation of syllable-final /ɹ/, Singaporean children speakers, like British children speakers, demonstrated a lack of rhoticity. It is also interesting that Singaporean children speakers appeared to have a different pronunciation for the approximant /l/ distinct from the other speaker groups.

Our work is the first of its kind to characterize Singapore English pronunciation patterns by quantifying pronunciation differences across different English speaker groups on a large-scale basis, and to present a comparative study from the perspective of children speech. Given the multicultural and multilingual environment in Singapore, Singapore English could potentially have been shaped by British influence during historical colonization, American influence through media sources, as well as bear other characteristics, for instance, originating from different languages spoken in Singapore such as Malay and Chinese. Future work analyzing Singaporean children’s pronunciation patterns from other aspects and investigation into effects of the various influences will help to paint a more comprehensive picture of the complexities of Singapore English.
References

[1] B. Kachru, The Other Tongue: English Across Cultures, Urbana: University of Illinois Press, 1982.

[2] C. L. Kuo, Formant transitions in varied utterance positions., Folia phoniatrica et logopaedica : official organ of the International Association of Logopedics and Phoniatrics 65 4 (2013) 178–84.

[3] C. G. Clopper, D. B. Pisoni, Free classification of regional dialects of American English, Journal of phonetics 35 3 (2007) 421–438.

[4] N. F. Chen, W. Shen, J. Campbell, R. Schwartz, Large-Scale Analysis of Formant Frequency Estimation Variability in Conversational Telephone Speech, in: INTERSPEECH 2009, 2009, pp. 2203–2206.

[5] W. Labov, S. Ash, C. Boberg, The Atlas of North American English: Phonetics, Phonology, and Sound Change : a multimedia reference tool, Mouton de Gruyter, Berlin, 2006.

[6] K. Evanini, Classifying and Clustering Dialects of North American English, Presented at North East Student Colloquium on Artificial Intelligence (NESCAl) (2008).

[7] N. F. Chen, S. W. Tam, W. Shen, J. P. Campbell, Characterizing Phonetic Transformations and Acoustic Differences Across English Dialects, IEEE/ACM Transactions on Audio, Speech, and Language Processing 22 (2014) 110–124.

[8] N. F. Chen, Characterizing phonetic transformations and fine-grained acoustic differences across dialects, Ph.D. thesis, Harvard-MIT Division of Health Sciences and Technology, 2011.

[9] J. Tauberer, K. Evanini, Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of North American English, in: INTERSPEECH, 2009, pp. 2211–2214.

[10] C. G. Henton, Changes in the vowels of received pronunciation, Journal of Phonetics 11 (1983) 353–371.

[11] E. Grabe, B. Post, Intonational Variation in the British Isles, in: Speech Prosody 2002 Aix-en-Provence, France, April 11-13, 2002, pp. 343–346.
[12] J. C. Wells, British English pronunciation preferences: A changing scene, Journal of the International Phonetic Association 29 (1999) 33–50.

[13] E. L. Low, E. Grabe, A contrastive study of lexical stress placement in Singapore English and British English, in: ICSLP, 1998.

[14] A. Khan, S. A. M. Alzobidy, Vowel Variation Between American English and British English, International Journal of English Linguistics 9 (2019) 350–356.

[15] P. Gomez, British and American English Pronunciation Differences, Cambridge: Mayflower Press (2009) 3–8.

[16] L. Alsagoff, Society, Style and Structure in Language : The Grammar of Singapore English, Singapore: Prentice Hall, 1998.

[17] J. Wong, The particles of Singapore English: a semantic and cultural interpretation, Journal of Pragmatics 36 (2004) 739–793.

[18] L. E. Ling, E. Grabe, F. Nolan, Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English, Language and Speech 43 (2000) 377–401.

[19] E. Grabe, B. S. Rosner, J. E. García-Albea, X. Zhou, Perception of English Intonation by English, Spanish, and Chinese Listeners, Language and Speech 46 (2003) 375–401.

[20] Y. K. Liang, Rhythm in Singaporean English : an instrumental study (1986).

[21] D. Deterding, R. Hvitfeldt, The Features of Singapore English Pronunciation: Implications for Teachers, Teaching and Learning 15 (1994) 98–107.

[22] J. Foley, New Englishes : the case of Singapore, Singapore University Press, National University of Singapore, Singapore, 1988.

[23] D. Deterding, L. E. Ling, The NIE Corpus of Spoken Singapore English (NIECSSE), SAAL Quartery (2001) 2–5. URL: http://videoweb.nie.edu.sg/phonetic/niecsse/index.htm
[24] D. Deterding, Singapore English, Edinburgh University Press, 2007.
[25] P. K. W. Tan, English in Singapore, International Journal of Language, Translation and Intercultural Communication 1 (2012) 123–138.
[26] Y. Gu, N. F. Chen, Characterization of Singaporean Children’s English: Comparisons to American and British Counterparts Using Archetypal Analysis, in: INTERSPEECH, 2020.
[27] Y. Gu, N. Chen, Acoustic Characterization of Singaporean Children’s English: Comparisons to American and British Counterparts, in: Extended abstract. WiNLP@ACL, 2019.
[28] Y. Gu, N. F. Chen, Large-scale acoustic characterization of mid-low vowels across American, British, and Singaporean children, Abstract at the Acoustical Society of America (2019).
[29] Y. Gu, N. F. Chen, Acoustic characterization of Singaporean children’s English with American and British counterparts: A case study on approximants, Abstract at the Acoustical Society of America (2019).
[30] S. Malmasi, K. Evanini, A. Cahill, J. Tetreault, R. Pugh, C. Hamill, D. Napolitano, Y. Qian, A Report on the 2017 Native Language Identification Shared Task, in: Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications, Association for Computational Linguistics, 201North American English7, pp. 62–75.
[31] K. Livescu, Analysis and Modeling of Non-Native Speech for Automatic Speech Recognition, Ph.D. thesis, MIT Department of Electrical Engineering and Computer Science, 1999.
[32] K. Livescu, J. R. Glass, Lexical modeling of non-native speech for automatic speech recognition, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2000, pp. 1683–1686.
[33] S. Bakst, C. A. Niziolek, SELF-CORRECTION IN L1 AND L2 VOWEL PRODUCTION, in: Proceedings of ICPhS 2019, 2019.
[34] R. Tong, B. P. Lim, N. F. Chen, B. Ma, H. Li, Subspace Gaussian mixture model for computer-assisted language learning, in: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2014, pp. 5347–5351.
[35] A. Lee, N. F. Chen, J. Glass, Personalized mispronunciation detection and diagnosis based on unsupervised error pattern discovery, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2016, pp. 6145–6149.

[36] N. F. Chen, H. Li, Computer-assisted pronunciation training: From pronunciation scoring towards spoken language learning, in: 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), IEEE., 2016.

[37] K. Shi, K. M. Tan, S. U. M. Salleh, N. F. A. B. Suhaimi, R. s/o Vellu, T. N. T. H. Helen, N. F. Chen, Computer-Assisted Language Learning System: Automatic Speech Evaluation for Children Learning Malay and Tamil, in: Interspeech, 2020.

[38] J. S. Garofolo, et al., TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1, 1993.

[39] M. Russell, The PF-STAR British English Children's Speech Corpus, The Speech Ark Limited (2006).

[40] A. Batliner, M. Blomberg, S. D’Arcy, D. Elenius, D. Giuliani, M. Gerosa, C. Hacker, M. Russell, S. Steidl, M. Wong, The PF STAR Children's Speech Corpus, in: INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, September 4-8, 2005, pp. 2761–2764.

[41] S. Weinburger, Speech Accent Archive (2015). URL: http://accent.gmu.edu

[42] N. F. Chen, R. Tong, D. Wee, P. Lee, B. Ma, H. Li, SingaKids-Mandarin: Speech Corpus of Singaporean Children Speaking Mandarin Chinese, in: INTERSPEECH, 2016, pp. 1545–1549.

[43] N. F. Chen, R. Tong, D. Wee, P. Lee, B. Ma, H. Li, iCALL corpus: Mandarin Chinese spoken by non-native speakers of European descent, in: Sixteenth Annual Conference of the International Speech Communication Association, 2015.
[44] N. F. Chen, D. Wee, R. Tong, B. Ma, H. Li, Large-scale characterization of non-native Mandarin Chinese spoken by speakers of European origin: Analysis on iCALL, Speech Communication 84 (2016) 46 – 56. URL: http://www.sciencedirect.com/science/article/pii/S0167639315300753

[45] P. Boersma, Praat, a system for doing phonetics by computer, Glot International 5 (2001) 341–345.

[46] N. S. Kenneth, Acoustic phonetics, MIT Press, 1998.

[47] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, University of California Press, Berkeley, Calif., 1967, pp. 281–297.

[48] A. Cutler, L. Breiman, Archetypal analysis, Technometrics 36 (1994) 338–347. doi:10.2307/1269949.

[49] M. J. A. Eugster, F. Leisch, From Spider-man to Hero – Archetypal analysis in R, Journal of Statistical Software 30 (2009) 1–23. URL: https://www.jstatsoft.org/article/view/v030i08/v30i08.pdf

[50] M. J. A. Eugster, archetypes: Archetypal analysis (2010). URL: https://cran.r-project.org/web/packages/archetypes/archetypes.pdf

[51] S. Bartlett, G. Kondrak, C. Cherry, On the syllabification of phonemes, in: NAACL, 2009.

[52] P.-E. Nordström, B. Lindblom, A normalization procedure for vowel formant data, in: Proceedings of the 8th International Congress of Phonetic Sciences, 1975.

[53] D. Jurafsky, J. H. Martin, Speech and Language Processing (3rd ed. draft), 2021.

[54] J. C. Wells, Accents of English, Cambridge: Cambridge University Press, 1982.

[55] L. Peter, J. Keith, A Course in Phonetics, Sixth Edition, Cengage Learning, Inc, 2010.
[56] H. Sweet, The Sounds of English: An Introduction to Phonetics, Clarendon Press, 1923.

[57] D. Jones, An Outline of English Phonetics, Cambridge: W. Heffer and Sons, 1947.

[58] J. G. Hansen, Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese, Applied Linguistics 22 (2001) 338–365.

[59] Y. He, Production of English Syllable Final /l/ by Mandarin Chinese Speakers, Journal of Language Teaching and Research 5 (2014) 742–750.

[60] D. Deterding, The Vowels of the Different Ethnic Groups in Singapore, Newcastle, UK: Cambridge Scholars Press, 2007, pp. 2–29.

[61] J. Keith, S. V. Clayton, S. Stephanie G., Acoustic and auditory phonetics (Third edition), Chichester : Wiley-Blackwell, Hoboken: John Wiley & Sons, Incorporated, 2012.

[62] L. Philip, Speech physiology, speech perception, and acoustic phonetics, Cambridge University Press, 1988.

[63] B. Robert, C. Richard, L. Ceil, The Oxford Handbook of Sociolinguistics, Oxford University Press, 2015.

[64] C.-C. Cheng, A Synchronic Phonology of Mandarin Chinese, De Gruyter Mouton, 1973.

[65] L. Lim, U. Ansaldo, Languages in Contact, Cambridge University Press, 2015.

[66] M. D. Duraidah, Malay + English → A Malay variety of English vowels and accent, Oxford University Press, 1980.