Characterization of Singaporean Children's English: Comparisons to American and British Counterparts using Archetypal Analysis

Yuling Gu[1], Nancy F. Chen[2]

[1]New York University, USA
[2]Institute for Infocomm Research, A*STAR, Singapore
`yuling.gu@nyu.edu, nfychen@i2r.a-star.edu.sg`

Abstract

In this work, we investigate pronunciation differences in English spoken by Singaporean children in relation to their American and British counterparts by conducting archetypal clustering and formant space analysis on selected vowel pairs. Given that Singapore adopts British English as the institutional standard due to historical reasons, one might expect Singaporean children to follow British pronunciation patterns, but interestingly we observe that Singaporean children present similar patterns to American children when it comes to TRAP−BATH split vowels and /æ/ vs. /ɛ/ productions: Singaporean and American speakers both exhibit more fronted characteristics ($p < 0.001$) for vowels in these vowel pairs, resulting in less contrast compared to British speakers. In addition, when producing these vowels, the first formant frequency estimates of Singaporean children is consistently lower, suggesting a higher tongue position, distinguishing them from American and British speakers ($p < 0.05$).

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 phonological features of Singapore English [25]; 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. [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 selected vowel pairs[1]. The speaker number and utterance number in this study are at least an order of magnitude greater than past work such as [23, 24].

2. Experimental Design

2.1. Speech Corpora[2]

Read speech was collected from American children (140 speakers, 43,406 utterances, ∼21.48 hours), British children (82 speakers, 32,542 utterances, ∼16.12 hours) and Singaporean children (192 speakers, 34,457 utterances, ∼20.41 hours). The best attempt was made to recruit general American English speakers and speakers of Received Pronunciation for respective American and British populations. For Singaporean English speakers, the best attempt was made to recruit students who are growing in a household where parents also grew up in Singapore. 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 comprises sentences from TIMIT [26], PF-STAR [27, 28], GMU Speech Accent Archive [29] 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 [30, 31].

2.2. Acoustic Features

The speech utterances were forced-aligned to obtain phonetic time boundaries; a subset was manually inspected to ensure time boundary offsets are within 50ms. For each utterance, pitch estimates (F0) and formant frequencies (F1, F2 in Hz) were extracted with a time step of 10ms using Praat [32]. Us-

---

[1]A more comprehensive study covering other vowels and approximants is being prepared for journal submission.

[2]D. Wee and R. Tong helped design the speech corpora.

ing the Nordstroem and Lindblom model [33], 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.

2.3. Archetypal Analysis

Most algorithms in unsupervised clustering such as k-means [34] use centroids to conduct cluster analysis. In this work, motivated by multilingual and multicultural influence of Singapore English, we adopt archetypal analysis [35] to investigate how American and British pronunciations might serve as anchoring archetypal references to characterize Singapore English [3].

Archetypal analysis represent each data point as a combination of "archetypes" (pure types) [35]. Given a set of multivariate data, $\{\mathbf{x_i}, i = 1, ..., n\}$, where each $\mathbf{x_i}$ is a vector of length m, we seek vectors $\mathbf{z_1}, ..., \mathbf{z_p}$ of length m that form the archetypal extremes. The vectors $\mathbf{z_1}, ..., \mathbf{z_p}$ are defined as

$$\mathbf{z_k} = \sum_j \beta_{kj} \mathbf{x_j}, k = 1, ..., p \quad (1)$$

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

$$\left\| \mathbf{x_i} - \sum_{k=1}^{p} \alpha_{ik} \mathbf{z_k} \right\|^2 \quad (2)$$

where $\alpha_{ik} \geq 0, \sum_k \alpha_{ik} = 1$. The archetypes are defined as vectors $\mathbf{z_1}, ..., \mathbf{z_p}$ that minimize D, the sum of squares of distances from each data vector $\mathbf{x_i}$ to the convex hull formed by the $\mathbf{z_1}, ..., \mathbf{z_p}$ vectors:

$$D = \sum_i \left\| \mathbf{x_i} - \sum_{k=1}^{p} \alpha_{ik} \mathbf{z_k} \right\|^2 \quad (3)$$

Archetypal analysis applies an alternating minimizing algorithm to a nonlinear least squares problem. We used the 'archetypes' R package as documented in [36, 37] and set $p = 2$ using the elbow criterion [36].

3. TRAP−BATH Split Vowels

TRAP−BATH split is a vowel split that is well-known in UK (including RP) [38], where vowels in words such as *glass, laugh, dance, can't* are pronounced as the [ɑ] phone instead of [ae] as in *trap, cab, mad*. Such splitting is not typically observed in American English [38]. In this section, we examine how Singaporean, American, and British children might produce TRAP−BATH split vowels, where [ɑ] 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 [ɑ] phone when TRAP−BATH split is present as [ɑ] vowels and those that are realized as the [ae] phone as [ae] vowels.

3.1. Archetypal Analysis Clustering

Table 1 shows the clustering results of archetypal analysis using F1 and F2[4]. We perform one clustering experiment per speaker group on TRAP−BATH split vowels from that group. For each

---

[3]Experiments conducted on k-means showed similar trends. Only results for achetypal analysis are shown due to space constraints

[4]Mel frequency cepstral coefficient (MFCC) features reveal similar trends, thus not shown due to space constraints

Singaporean Children		
Phone	Group1 (915, 2290)	Group2 (596, 1715)
[æ]	**0.667**	0.333
[ɑ]	**0.688**	0.312

American Children		
Phone	Group1 (696, 2422)	Group2 (1087, 1772)
[æ]	0.179	**0.821**
[ɑ]	**0.786**	0.214

British Children		
Phone	Group1 (1173, 1951)	Group2 (651, 1347)
[æ]	**0.707**	0.293
[ɑ]	0.317	**0.683**

Table 1: *Archetypal Analysis using F1(Hz), F2(Hz) estimates for Singaporean, American and British children for TRAP−BATH split [æ] and [ɑ] vowels. Archetypal extreme points for each cluster are in the format (F1, F2).*

Corpus	Phone	F1 mean	F1 se	F2 mean	F2 se
SG	[æ]	756	4.91	2087	9.15
	[ɑ]	803	5.02	2003	8.62
AE	[æ]	949	7.85	2017	11.8
	[ɑ]	828	8.17	2187	15.9
BE	[æ]	948	13.9	1724	14.1
	[ɑ]	885	12.1	1586	13.1

Table 2: *Mean and standard error (se) for each speaker group for TRAP−BATH split vowel formants.*

phone (i.e. [ae] and [ɑ]), we show the percentage of tokens from that category that get clustered into each of the cluster groups. Majority of Singaporean children's [ae] and [ɑ] vowels in the TRAP−BATH split are largely ($> 65\%$ 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. In the next subsection, we investigate if cleaner clusters indeed reflect the split through formant space analysis.

3.2. Acoustic Analysis and Characterization

F1 F2 formant space: We present the mean and standard error for F1 and F2 estimates in Table 2 and visualize this data on a per speaker level in Figure 1: 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.

F1 formant estimates: For TRAP−BATH split vowels that become [ae] under the split, Singaporean children have the lowest F1 ($M = 756$), American children have statistically significantly higher F1 ($M = 949$); British children ($M = 948$) show similar trends to Americans. For vowels that could turn into [ɑ] when the split is present, Singaporean children similarly have the lowest F1 ($M = 803$), American children have statistically significantly higher F1 ($M = 828$), and British children have the highest F1 ($M = 885$). A one-way ANOVA demonstrated that
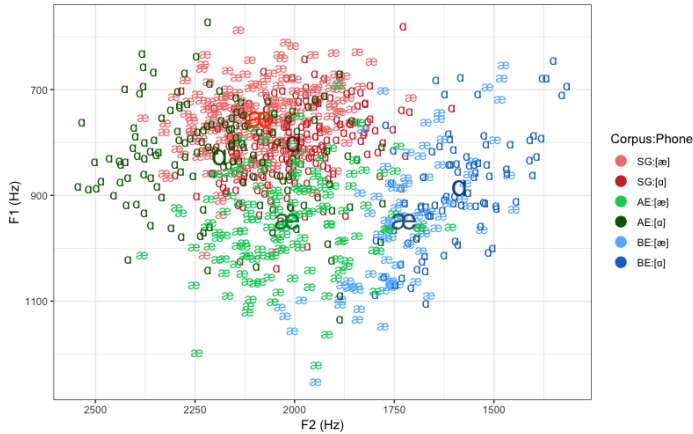
Figure 1: *Comparing F1 and F2 estimates of TRAP−BATH split vowels across Singapore, American and British children. /æ/ phonemes that could turn into [ɑ] in TRAP−BATH split are labeled as [ɑ] while those that are realized as [ae] are labeled as [ae]. Smaller points: individual speaker's mean F1 and F2; larger points: group means for the speaker groups.*

these differences are statistically significant, $F(2, 411) = 234$, $p < 0.001$ for [ae] vowels under the split, and $F(2, 411) = 25.3$, $p < 0.001$ for vowels that could turn into [ɑ] 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 [ɑ] under the split, all three groups are significantly different in terms of F1 ($p < 0.05$). Therefore one significant difference for TRAP−BATH 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. The articulatory implication is that, compared to British and American children, Singaporean children exhibit a higher tongue height when producing TRAP−BATH split vowels.

F2 formant estimates: For vowels that are realized as [ae] under the split, British children show the lowest F2 ($M = 1724$), American children show higher F2 ($M = 2017$), and Singaporean children show the highest F2 ($M = 2087$). A one-way ANOVA shows that these differences are statistically significant, $F(2, 411) = 223$, $p < 0.001$. For vowels that turn into [ɑ] under the split, British children again show the lowest F2 ($M = 1586$), Singaporean children show higher F2 ($M = 2003$), and American children show the highest F2 ($M = 2187$). A one-way ANOVA shows that these differences are statistically significant, $F(2, 411) = 440$, $p < 0.001$. 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 ($p < 0.001$). This shows that in terms of articulatory implications, Singaporean and American children's productions of TRAP−BATH vowels are similarly more fronted, compared to British speakers.

3.3. Overall comparisons

The TRAP−BATH vowels are more fronted for both Singaporean and American children, suggesting that these two groups

Singaporean Children		
Phoneme	Group1 (853, 2291)	Group2 [579, 1814)
/æ/	**0.609**	0.391
/ɛ/	0.453	**0.547**

American Children		
Phoneme	Group1 (1033, 2345)	Group2 (648, 1791)
/æ/	**0.650**	0.350
/ɛ/	0.343	**0.657**

British Children		
Phoneme	Group1 (998, 1574)	Group2 (677, 2034.)
/æ/	**0.976**	0.024
/ɛ/	0.049	**0.951**

Table 3: *Archetypal Analysis using F1(Hz), F2(Hz) estimates of /æ/ and /ɛ/from Singaporean, American and British children. Archetypal extreme points for each cluster are in the format (F1, F2).*

of speakers are more similar in producing a vowel closer to [æ] rather than [ɑ], and thus not showing much TRAP−BATH split distinction. These articulatory features agree with our clustering results which also suggest that Singaporean children do not show TRAP−BATH split. However, Singaporean children's production of the TRAP−BATH vowels differ from both of the other groups by having acoustic features that reflect a higher tongue position. The formant analysis 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 [ɑ] when the split is present are instead articulated with even a fronter position (higher F2) 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 archetypal clustering results that they show TRAP−BATH split since vowels that could change into [ɑ] in the split are produced with lower F2 like a back vowel ($M = 1586$), compared to those that could stay as the front vowel [æ] with higher F2 ($M = 1724$).

4. /æ/ and /ɛ/ contrast

/ɛ/ is a mid-low front vowel. When compared to the /æ/ phoneme, the /ɛ/ phoneme has slightly higher F2 and lower F1 estimates [39]. 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 (across realizations as [æ] and [ɑ] phones), we further examine how production of /æ/ and /ɛ/ phonemes might differ across the three speaker groups.

4.1. Archetypal Analysis clustering

Similar to clustering experiments for TRAP−BATH split vowels, we perform per-speaker-group clustering across /æ/ and /ɛ/ vowels from each group. Clustering results are shown in Table 3, indicating that /æ/ and /ɛ/ vowels produced by the British children are largely (> 95 %) grouped into two clean, distinctive clusters, whereas such distinction is less clear cut for these vowels produced by Singaporean and American children.

Corpus	Phoneme	F1 mean	F1 se	F2 mean	F2 se
SG	/æ/	753	5.01	2046	8.85
	/ɛ/	687	4.20	2075	8.95
AE	/æ/	900	7.47	2075	12.7
	/ɛ/	777	5.98	2056	11.8
BE	/æ/	920	13.1	1662	12.7
	/ɛ/	768	11.1	1927	17.5

Table 4: *Mean and standard error (se) for each speaker group for /æ/ and /ɛ/ formant estimates.*
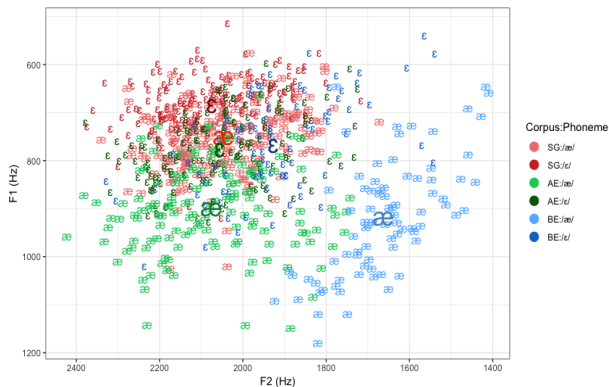

Figure 2: *Comparing F1 and F2 estimates across Singapore, American and British children for /æ/ and /ɛ/. Smaller points: individual speaker's mean F1 and F2; larger points: group means for the speaker groups.*

4.2. Acoustic Analysis

F1 F2 formant space : Estimates of F1 and F2 for /æ/ and /ɛ/ in the three speaker groups are summarized in Table 4. Using each speaker as a data point, we visualize this differences between speaker groups for the two vowels in Figure 2. 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 F2 and F1 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.

F1 formant estimates: For /æ/, Singaporean children have the lowest F1 ($M = 753$), American children have higher F1 estimates ($M = 900$), and British children have the highest F1 ($M = 920$). Similarly, for the production of /ɛ/, Singaporean children have the lowest F1 ($M = 687$), American children have higher F1 estimates ($M = 777$), which is similar to British children ($M = 768$). A one-way ANOVA demonstrated that for both vowels, these differences are statistically significant, $F(2, 411) = 161, p < 0.001$ for /æ/ and $F(2, 411) = 74.9, p < 0.001$ 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 ($p = 0.234$). 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 ($p = 0.665$). 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.

F2 formant estimates: For /æ/, Singaporean children ($M = 2046$) and American children ($M = 2075$) have similarly high F2, while British children have a much lower F2 ($M = 1662$). The same trend is observed for /ɛ/, where Singaporean children also have the highest F2 ($M = 2075$), American children have just slightly lower F2 ($M = 2056$), and British children have the lowest F2 ($M = 1927$). A one-way ANOVA showed that for both vowels, these differences are statistically significant, $F(2, 411) = 301, p < 0.001$ for /æ/ and $F(2, 411) = 35.2, p < 0.001$ for /ɛ/. We followed up our ANOVA tests with post hoc Tukey's HSD Test, which show that Singaporean and American children's F2 for both /æ/ and /ɛ/ are significantly higher ($p < 0.001$) than those of British children. This suggests that in terms of articulation, Singaporean and American children exhibit fronting of both /æ/ and /ɛ/ such that they are produced with a much fronter tongue position compared to those of British children.

4.3. Overall comparisons

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 (lower F1). Interestingly, this tendency for producing vowels with a lower F1 which suggests higher tongue position, is not only observed for /æ/ and /ɛ/ but also in the TRAP−BATH split [æ] and [ɑ] vowels we examined earlier in subsection 3.2. American and Singaporean children's fronting of both /æ/ and /ɛ/ is also similar to the trend we observed for TRAP−BATH split [æ] and [ɑ] vowels in subsection 3.2. The fronting of /æ/ and /ɛ/ by American and Singaporean children such that the two vowels are produced with a similar frontness of the tongue position (similar F2) likely explains why /æ/ and /ɛ/ from these two speaker populations do not fall into clean distinct clusters like those of British speakers in our clustering experiment.

5. Discussion

We presented a large-scale study (at least an order of magnitude more speakers and utterances than what was done in the past) to characterize Singaporean children's English pronunciation patterns. We examined different vowel pairs (TRAP−BATH split [æ] vs. [ɑ], /æ/ vs. /ɛ/) by first using archetypal clustering analysis to explore potential trends across the different speaker groups, followed by detailed acoustic analysis with linguistics insights. Our analysis showed that Singaporean children are similar to American children in terms of fronting (higher F2) in their pronunciation of TRAP−BATH split vowels as well as the /æ/ vs. /ɛ/ vowel pair. This finding alludes to how Singapore English has been evolving beyond the British influence during historical colonization [40], potentially moving towards embodying more American pronunciation characteristics. Compared to both American and British children, Singaporean children also consistently articulate these vowels with a higher tongue positions (lower F1). This suggests that English spoken by Singaporean children could be shaped by factors beyond the characteristics of American and British English. Investigating how Singapore English is potentially influenced by Malay [41] and the range of Chinese languages [42] spoken in Singapore can paint a more comprehensive picture of the complexities of Singapore English. There is also on-going work in applying our findings to improve pronunciation modeling in computer-assisted language learning [43, 44].

6. 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*, vol. 65 4, pp. 178–84, 2013.

[3] C. G. Clopper and D. B. Pisoni, "Free classification of regional dialects of american english," *Journal of phonetics*, vol. 35 3, pp. 421–438, 2007.

[4] N. F. Chen, W. Shen, J. Campbell, and 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, and 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 (NESCAI)*, 2008.

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

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

[9] J. Tauberer and 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*, vol. 11, pp. 353–371, 1983.

[11] E. Grabe and 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*, vol. 29, pp. 33–50, 1999.

[13] E. L. Low and E. Grabe, "A contrastive study of lexical stress placement in singapore english and british english," in *ICSLP*, 1998.

[14] A. Khan and S. A. M. Alzobidy, "Vowel variation between american english and british english," *International Journal of English Linguistics*, vol. 9, pp. 350–356, 2019.

[15] P. Gomez, "British and american english pronunciation differences," *Cambridge: Mayflower Press*, pp. 3–8, 2009.

[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*, vol. 36, pp. 739–793, 2004.

[18] L. E. Ling, E. Grabe, and F. Nolan, "Quantitative characterizations of speech rhythm: Syllable-timing in singapore english," *Language and Speech*, vol. 43, no. 4, pp. 377–401, 2000.

[19] E. Grabe, B. S. Rosner, J. E. García-Albea, and X. Zhou, "Perception of english intonation by english, spanish, and chinese listeners," *Language and Speech*, vol. 46, no. 4, pp. 375–401, 2003.

[20] Y. K. LIANG, "Rhythm in singaporean english : An instrumental study," 1986.

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

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

[23] D. Deterding and L. E. Ling, "The NIE Corpus of Spoken Singapore English (NIECSSE)," *SAAL Qarterly* , pp. 2–5, 2001. [Online]. Available: 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*, vol. 1, pp. 123–138, Jan. 2012.

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

[27] M. Russell, "The PF-STAR British English Children's Speech Corpus," *The Speech Ark Limited*, December 2006.

[28] A. Batliner, M. Blomberg, S. D'Arcy, D. Elenius, D. Giuliani, M. Gerosa, C. Hacker, M. Russell, S. Steidl, and 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.

[29] S. Weinburger, "Speech Accent Archive," 2015. [Online]. Available: http://accent.gmu.edu

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

[31] N. F. Chen, D. Wee, R. Tong, B. Ma, and H. Li, "Large-scale characterization of non-native mandarin chinese spoken by speakers of european origin: Analysis on icall," *Speech Communication*, vol. 84, pp. 46 – 56, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639315300753

[32] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

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

[34] 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.* Berkeley, Calif.: University of California Press, 1967, pp. 281–297.

[35] A. Cutler and L. Breiman, "Archetypal analysis," *Technometrics*, vol. 36, no. 4, pp. 338–347, Nov. 1994.

[36] M. J. A. Eugster and F. Leisch, "From Spider-man to Hero – Archetypal analysis in R," *Journal of Statistical Software*, vol. 30, pp. 1–23, 2009. [Online]. Available: https://www.jstatsoft.org/article/view/v030i08/v30i08.pdf

[37] M. J. A. Eugster, "archetypes: Archetypal analysis," 2010. [Online]. Available: https://cran.r-project.org/web/packages/archetypes/archetypes.pdf

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

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

[40] L. Lim and U. Ansaldo, *Languages in Contact.* Cambridge University Press, 2015.

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

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

[43] N. F. Chen and 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.

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