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

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

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

Abstract

We investigate English pronunciation patterns in Singaporean children in relation to their American and British counterparts by conducting archetypal analysis (Cutler and Breiman, 1994) on selected vowel pairs. Given that Singapore adopts British English as the institutional standard, one might expect Singaporean children to follow British pronunciation patterns, but we observe that Singaporean children also present similar patterns to Americans for TRAP-BATH spilt vowels: (1) British and Singaporean children both produce these vowels with a relatively lowered tongue height. (2) These vowels are more fronted for American and Singaporean children ($p < 0.001$). In addition, when comparing /æ/ and /ɛ/ productions, British speakers show the clearest distinction between the two vowels; Singaporean and American speakers exhibit a higher and more fronted tongue position for /æ/ ($p < 0.001$), causing /æ/ to be acoustically more similar to /ɛ/.

1 Introduction

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

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

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

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: http://

2 Methods

2.1 Unsupervised clustering: Archetypal Analysis

Most unsupervised clustering algorithms such as k-means (MacQueen, 1967) use centroids to conduct cluster analysis. Archetypal analysis represent each data point in a data set as a combination of characteristic "archetypes" (pure types) (Cutler and Breiman, 1994). Motivated by multilingual and multicultural influence of Singapore English, we adopt archetypal analysis (Cutler and Breiman, 1994) to investigate how American and British (inner circle English) pronunciations might serve as anchoring archetypal references to characterize Singapore English (outer circle English)[1]. We set the cluster number to 2.

2.2 Acoustic Analysis

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

3 Experiments

3.1 Corpus

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

3.2 TRAP-BATH Split Vowels

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

creativecommons.org/licenses/by/4.0/.

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

Corpus	Group1	Group2
SG	**0.927**	0.073
AE	**0.614**	0.386
BE	0.317	**0.927**

Table 1: Achetypal analysis using formant features.

Corpus	F1 mean	F1 se	F2 mean	F2 se
SG	908.84	5.71	2,267.79	9.84
AE	827.55	8.17	2,186.58	15.91
BE	902.73	12.24	1,617.13	13.22

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

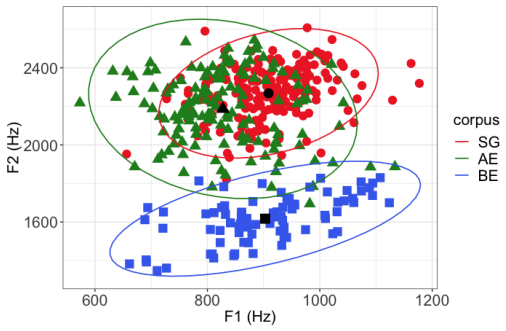

Figure 1: *F1* and *F2* estimates of TRAP-BATH split vowels across speakers. Ellipses represent 95% confidence interval for each group.Each small colored shape: speaker's mean *F1* and *F2* estimates; larger black shapes: group mean.

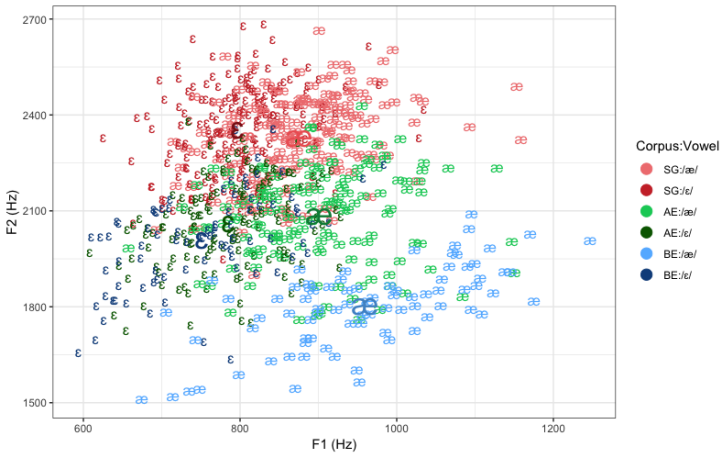

Figure 2: *F1* and *F2* estimates across speakers for /æ/ and /ɛ/. Each small colored shape: speaker's mean *F1* and *F2* estimates; larger vowel labels: group mean.

Corpus	Phone	F1 mean	F1 se	F2 mean	F2 se
SG	/æ/	875.11	5.40	2,327.39	9.77
	/ɛ/	796.35	4.64	2,353.92	10.02
AE	/æ/	901.59	7.39	2,082.68	12.36
	/ɛ/	785.44	5.87	2,059.39	11.55
BE	/æ/	959.93	13.40	1,802.39	13.89
	/ɛ/	751.79	9.80	2,009.81	16.36

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

Corpus	Phone	Group1	Group2
SG	/æ/	0.349	**0.651**
	/ɛ/	**0.552**	0.448
AE	/æ/	**0.750**	0.250
	/ɛ/	0.400	**0.600**
BE	/æ/	0.024	**0.976**
	/ɛ/	**0.963**	0.037

Table 4: Archetypal Analysis of *F1*, *F2* estimates of /æ/ and /ɛ/ from Singaporean, American and British children.

3.3 /æ/ and /ɛ/ contrast

/ɛ/ has slightly higher F2 and lower F1 estimates (Stevens, 1998) than /æ/. Any fronting of /æ/, resulting in a higher F2, could lead to potential confusion with /ɛ/. Table 4 shows that these vowels produced by the British children are largely (> 95 %) two distinctive groups, whereas there is less acoustic distinction for Singapore and American children. From Table 3 and Figure 2, for /æ/, Singaporean children have the lowest F1 estimates (M= 875.11), American children's F1 are higher (M = 901.59), and British children's F1 are the highest (M = 959.93). The opposite trend was observed for /ɛ/, where Singaporean children have the highest F1 estimates (M = 796.35), American children's are slightly lower (M = 785.44), and British children's are the lowest (M = 751.79). Articulatorily, Singaporean and American children produce /æ/ with a higher tongue position and /ɛ/ with a lower tongue position instead. For /æ/, Singaporean children have the highest F2 (M = 2327.39), American children's are lower (M = 2082.68), and British children's are the lowest (M = 1802.39). The same trend is observed for /ɛ/, where Singaporean children also have the highest F2 estimates (M = 2353.92), American children's are lower (M = 2059.39), and British children's are the lowest (M = 2009.81). This suggests that unlike British children, Singaporean and American children exhibit fronting of both vowels such that the backness of /æ/ and /ɛ/ are similar.

4 Discussion

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

References

Lubna Alsagoff. 1998. *Society, Style and Structure in Language : The Grammar of Singapore English*. Singapore: Prentice Hall.

Anton Batliner, Mats Blomberg, Shona D'Arcy, Daniel Elenius, Diego Giuliani, Matteo Gerosa, Christian Hacker, Martin Russell, Stefan Steidl, and Michael Wong. 2005. The PF STAR Children's Speech Corpus. In *INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, September 4-8*, pages 2761–2764.

Nancy F. Chen, Wade Shen, Joseph Campbell, and Reva Schwartz. 2009. Large-Scale Analysis of Formant Frequency Estimation Variability in Conversational Telephone Speech. In *INTERSPEECH 2009, 10th Annual Conference of the International Speech Communication Association, Brighton, United Kingdom, September 6-10,*, pages 2203–2206.

Nancy F. Chen, Sharon W. Tam, Wade Shen, and Joseph P. Campbell. 2014. Characterizing Phonetic Transformations and Acoustic Differences Across English Dialects. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22:110–124.

Nancy F. Chen, Rong Tong, Darren Wee, Peixuan Lee, Bin Ma, and Haizhou Li. 2016. SingaKids-Mandarin: Speech Corpus of Singaporean Children Speaking Mandarin Chinese. In *INTERSPEECH 2016, 17th Annual Conference of the International Speech Communication Association, San Francisco, USA, September 8-12*, pages 1545–1549.

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

Cynthia G. Clopper and David B. Pisoni. 2007. Free classification of regional dialects of american english. *Journal of phonetics*, 35 3:421–438.

Adele Cutler and Leo Breiman. 1994. Archetypal analysis. *Technometrics*, 36(4):338–347, November.

David Deterding and Robert Hvitfeldt. 1994. The Features of Singapore English Pronunciation: Implications for Teachers. *Teaching and Learning*, 15:98–107.

David Deterding and Low Ee Ling. 2001. The NIE Corpus of Spoken Singapore English (NIECSSE). *SAAL Qarterly* , pages 2–5.

David Deterding. 2007. *Singapore English*. Edinburgh University Press.

Keelan Evanini. 2008. Classifying and Clustering Dialects of North American English. *Presented at North East Student Colloquium on Artificial Intelligence (NESCAI)*.

Joseph Foley. 1988. *New Englishes : the case of Singapore*. Singapore University Press, National University of Singapore, Singapore.

John S. Garofolo and et al. 1993. TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1.

Paco Gomez. 2009. British and american english pronunciation differences. *Cambridge: Mayflower Press*, pages 3–8.

Esther Grabe and Brechtje Post. 2002. Intonational Variation in the British Isles. In *Speech Prosody 2002 Aix-en-Provence, France, April 11-13*, pages 343–346.

C. G. Henton. 1983. Changes in the vowels of received pronunciation. *Journal of Phonetics*, 11:353–371.

Braj Kachru. 1982. *The Other Tongue: English Across Cultures*. Urbana: University of Illinois Press.

Afzal Khan and Soleman Awad Mthkal Alzobidy. 2019. Vowel variation between american english and british english. *International Journal of English Linguistics*, 9:350–356.

Christina L. Kuo. 2013. Formant transitions in varied utterance positions. *Folia phoniatrica et logopaedica : official organ of the International Association of Logopedics and Phoniatrics*, 65 4:178–84.

William Labov, Sharon Ash, and Charles Boberg. 2006. *The Atlas of North American English: Phonetics, Phonology, and Sound Change : a multimedia reference tool*. Mouton de Gruyter, Berlin.

Lisa Lim and Umberto Ansaldo. 2015. *Languages in Contact*. Cambridge University Press.

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

Martin Russell. 2006. The PF-STAR British English Children's Speech Corpus. *The Speech Ark Limited*, December.

Kenneth N. Stevens. 1998. *Acoustic phonetics*. MIT Press.

Peter K W Tan. 2012. English in Singapore. *International Journal of Language, Translation and Intercultural Communication*, 1:123–138, January.

Joshua Tauberer and Keelan Evanini. 2009. Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of North American English. In *INTERSPEECH 2009, 10th Annual Conference of the International Speech Communication Association, Brighton, United Kingdom, September 6-10,*, pages 2211–2214.

S. Weinburger. 2015. Speech Accent Archive.

John C. Wells. 1982. *Accents of English*. Cambridge: Cambridge University Press.

J. C. Wells. 1999. British english pronunciation preferences: A changing scene. *Journal of the International Phonetic Association*, 29:33–50.

Jock Wong. 2004. The particles of Singapore English: a semantic and cultural interpretation. *Journal of Pragmatics*, 36:739–793.