Corpus-based Pronunciation Variation Rule Analysis for Singapore English

Wenda Chen[1], *Nancy F. Chen*[2] , *Boon Pang Lim*[3], *Bin Ma*[4]

Institute for Infocomm Research, A*STAR, Singapore

{chen-w, nfychen, bplim, mabin}@i2r.a-star.edu.sg

Abstract

In this paper, we evaluate a set of linguistic rules for pronunciation variations in Singapore English. We collect and annotate a speech corpus for Singapore English and label it with IPA narrow transcriptions. Data driven pronunciation rules are derived using American English (Buckeye corpus) as a reference. We compare the data driven rules with linguistic rules proposed by phoneticians, and found that some pronunciation variations observed in Singapore English are also observed in American English, but with different frequencies of occurrence. Our analysis verifies the linguistic rules previously proposed for Singapore English and demonstrates an approach to utilizing our findings to improve pronunciation feedback.

Index Terms: speech recognition, context dependent phonological rules, computer assisted language learning

1. Introduction

Singapore English (SgE) pronunciation has been analyzed by many researchers as a unique English dialect [1]. English is one of the official languages of Singapore – it is used in business, law, and education environments. The pronunciation has the roots in British English (BrE) due to the colonial history. On the other hand, every Singaporean will learn his or her own mother tongue language, Malay, Mandarin, Tamil, or another type of Indian language since childhood. In the daily life, Singaporeans typically speak their own mother tongues at the food court and with elderly people [2]. These include the official languages and additional dialects such as Hokkien and Cantonese. These languages have had strong influence in Singapore English's pronunciation, grammar and vocabulary as well. In general, Singapore English patterns are categorized into standard Singapore English and colloquial Singlish. Singlish is the informal English spoken in Singapore with unique slang and syntax. The pronunciations and vocabulary have the origins from English, Mandarin, Tamil, Malay, Hokkien, Cantonese and Teochew [3]. In the following paper, we use SgE to refer to Singapore English accent but not colloquial Singlish in the research.

In the past few decades, American English (AmE) has had increased influence on the English pronunciations of Singaporeans due to the trade interactions and cultural exchange. In 2000, the government campaign of Speak Good English Movement (SGEM) was launched to "encourage Singaporeans to speak grammatically correct English that is universally understood" [4]. Consequently, there is increasing need for Singaporeans to adapt their accent to a more standard English accent worldwide such as American accent for communication with people from Silicon Valley and New York. Thus, it has become a necessity for Singaporeans to receive appropriate feedback on how their pronunciation compares with American English -- this is now a feature required for computer assisted pronunciation training (CAPT) systems.

In 2010, Ministry of Education in Singapore funded School of Computer Engineering, Nanyang Technological University on developing a corpus and automatic evaluation software for analyzing standard Singapore English (SgE) pronunciation [6,7,8]. In that work, we proposed pronunciation variation rules based on linguistic knowledge, derived the set of context dependent data driven rules from our CALL corpus, and generated a corresponding lexicon to train Singapore English acoustic models. The learning system made use of pronunciation rules and lexicon to generate sentence level scores on a 1 to 100 point scale for intensity, speaking rate prosody and phoneme pronunciation areas.

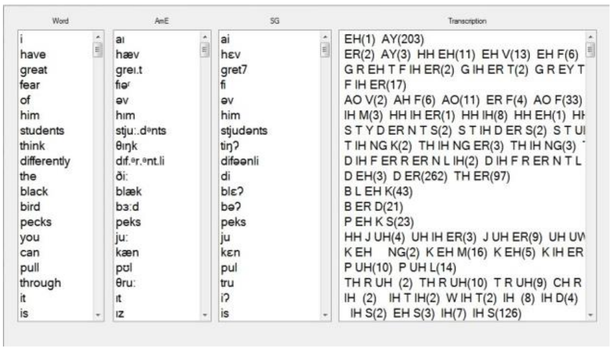
Our earlier approach can only give a single score metric per sentence or word. However, it is difficult, but definitely more useful, to give detailed pronunciation feedback at the level of phonemes. Given the set of linguistic rules defined for Singapore English, some would argue that not all Singaporeans produce the same pronunciation variations. At the same time, other dialects of English including BrE and AmE could pronounce the similar variations too. For example, /θ/ → /f/ is usually identified as a pronunciation variation rule in Singapore English at the word ending such as *perth*. But we know that southern British English dialect has the Th-fronting (/θ/) pattern as well [5].

Usually, identifying pronunciation variations can be achieved using L1 dependent (we call it knowledge-based) or using L1 independent (we call it data driven) methods. The knowledge-based approach explores the existing linguistic literature and the knowledge from language teachers. It studies the cross-language transfer comparisons between the student's first language (L1) and the target language (L2) [12, 13]. Data driven approaches can obtain pronunciation variation rules through dynamic alignment of the manually transcribed L2 data with the corresponding standard canonical pronunciations [9, 10, 14, 15, 16, 17].

Given the linguistic knowledge of the speaker's native language, how can we give more specific mispronunciation feedback? This paper addresses this question. It describes the manual transcription corpus and shows a detailed comparison and analysis of differences between linguistic and data driven phonological rule patterns in Singapore English and American English. Consequently, these differences can be used to provide evaluation feedback for Singaporean leaners learning American English accent. In Section 4, we propose a framework for error detection and further develop it to provide detailed pronunciation feedback.

2. Manual Transcription Corpus in Singapore English

We constructed a manually transcribed Singapore English corpus to analyze the pronunciation variations. As far as we know, this is the only fully transcribed and time aligned corpus for Singapore English accent. In the corpus, we have transcribed 1134 utterances balanced across a pool of 44 speakers in our Singapore English accent corpus [6]. In our corpus, the students are asked to read the sentence in the oral Singapore English accent. The sentences are constructed with TIMIT scripts and designed texts and are transcribed by 5 students trained in phonetics. Each transcriber is given a set of sentences selected at random and is trained to use narrow IPA transcriptions and Praat to mark all the phonemes they heard [18]. The transcripts have the time segmentations for each phone and word stored. The detailed discussions about inter-transcriber agreement control are discussed in [6]. We developed a tool to convert the IPA narrow transcription symbols to CMU ARPAbet phoneme set [19] and collected the transcriptions for each distinct word with statistics on the transcription occurrences. ARPAbet phone set is used for representing Singapore English phone set because we have proposed to assume that the Singapore English phoneme set as a subset of American English [8]. There are a total of 560 selected distinct words transcribed.



Word	AmE	SG	Transcription
i	aɪ	ai	EH(1) AY(203)
have	hæv	hɛv	ER(2) AY(3) HH EH(11) EH V(13) EH F(6)
great	greɪ.t	gret7	G R EH T F IH ER(2) G IH ER T(2) G R EY T
fear	fiəʳ	fi	F IH ER(17)
of	əv	əv	AO V(2) AH F(6) AO(11) ER F(4) AO F(33)
him	hɪm	him	IH M(3) HH IH ER(1) HH IH(8) HH EH(1) HH
students	stju:.dᵊnts	stjudənts	S T Y D ER N T S(2) S T IH D ER S(2) S T UH
think	θɪŋk	tiŋʔ	T IH NG K(2) TH IH NG ER(3) TH IH NG(3)
differently	dɪf.ᵊr.ᵊnt.li	difəənli	D IH F ER R ER N L IH(2) D IH F R ER N T L
the	ði:	di	D EH(3) D ER(262) TH ER(97)
black	blæk	blɛʔ	B L EH K(43)
bird	bɜ:d	bəʔ	B ER D(21)
pecks	peks	peks	P EH K S(23)
you	ju:	ju	HH J UH(4) UH IH ER(3) J UH ER(9) UH UW
can	kæn	kɛn	K EH NG(2) K EH M(16) K EH(5) K IH ER
pull	pʊl	pul	P UH(10) P UH L(14)
through	θru:	tru	TH R UH (2) TH R UH(10) T R UH(9) CH R
it	ɪt	iʔ	IH (2) IH T IH(2) W IH T(2) IH (8) IH D(4)
is	ɪz	is	IH S(2) EH S(3) IH(7) IH S(126)

Figure 1: *Manual transcription results in ARPAbet with pronunciation occurrences (column 4) compared with other dictionary IPA pronunciations including the American English (column 2) and estimated canonical Singapore English (column 3) for selected words in the corpus (column 1)*

A sample list of words and their corresponding transcriptions is shown in Figure 1. The transcriptions in IPA are to be compared with the standard American English (column 2) and Singapore English pronunciations (column 3) in IPA symbols. The Singapore English dictionary was derived from the American English dictionaries using linguistic rules [7]. For the ease of representation and comparison, we will use the 39 CMU ARPAbet phoneme set to represent the manual transcriptions in column 4. But the original narrow IPA transcriptions are kept in our database for future studies. The occurrence number on the right shows the number of times the word is transcribed into the corresponding phone sequence. All the alternative pronunciations in the manual transcription corpus are listed in sequence.

3. Linguistic and Data-driven Rules for Singapore English and American English

3.1. Linguistic Rules

In [8], we proposed and selected 17 linguistic knowledge based pronunciation variation rules according to the existing literature about Singapore English [8] using CMU phoneme set, as demonstrated above. They are categorized into 6 vowel rules and 11 consonant rules. The vowel rules can be represented as context independent phoneme substitution pairs in Table 1:

No.	Rule	Sample
V1	IY→IH (/i:/→/i/)	Sit, seat
V2	UW→UH (/u:/→/u/)	Pull, pool
V3	AA→AH(/ɑ:/→/ɑ/)	Cut, cart
V4	AE→EH (/æ/→/ɛ/)	Bat, bet
V5	EY→EH (/eɪ/→/e/)	Play, may
V6	OW→AO (/oʊ/→/ɔ/)	Go, cold

Table 1: *Context independent rewrite rules that capture pronunciation variation in Singapore English (vowels)*

The consonant rules have the substitution pairs in Table 2:

No.	Rule	Sample
C1	Z→S/ _#	Daze, dogs
C2	TH→T/ #_	Think
C3	DH→D/ #_	That
C4	TH→F/ _#	Bath, death
C5	SP→PS/ _#	Crisp, wasp
C6	R→sil/ _#	Pour, dear
C7	P→sil/ _#	tap
C8	T→sil/ _#	Last
C9	K→sil/ _#	Task
C10	L→sil/ ER_#	Pearl
C11	D→sil/ (N/M)_#	tend

Table 2: *Context independent rewrite rules that capture pronunciation variation in Singapore English (consonants)*

These rules will be shown in the following sections that they are truly effective in characterizing Singapore English and can be well used in the accent detection and language learning tasks.

3.2. Data-driven Rules for Singapore English

We generated data-driven rules by aligning the canonical phone transcriptions to manual transcriptions with a minimum edit distance algorithm. For the data driven rules generation, for each word w_i in the identical utterances, the corresponding canonical pronunciation phone sequence $p_{i,1}$, $p_{i,2}$, …, $p_{i,n}$ from CMU dictionary are aligned with the manual transcriptions $t_{j,1}$, $t_{j,2}$, …, $t_{j,n}$ using dynamic programming. The purpose is to find the optimal mapping between two transcriptions, so as to minimize the distance cost in terms of insertions, substitutions and deletions. The cost function $C(p_{i,1}, t_{j,1})$ is the reciprocal of a confusion matrix values of the two phone sets. The confusion matrix is generated from the speech recognition phone substitution experiments in standard Wall Street Journal American English dataset. A numerical number is assigned to each phoneme pair as the substitution frequency in lattice and the maximum number is normally the values on the diagonal.

The data driven Singapore English context dependent phonological rules are then derived from the dynamic alignment mappings based on this transcription corpus. Minimum edit distance is to find the minimum total cost of the substitutions in two phoneme sequences. The cost q(i,j) for phone sequences p_1, p_2, …, p_i aligned with t_1, t_2, …, t_j is:

$$q(i,j) = min(q(i-1,j)+1, q(i,j-1)+1, q(i-1,j-1)+cost) \quad (1)$$

where cost refers to substitution cost of p_i and t_j. The detailed procedures can be found in [6,7]. The rule list with number of occurrence in the data set more than 5 is shown in Table 3 in the format of *phoneme1→phoneme2 / left phone _right phone* and $\mathrm{freq} = 100 \times \frac{\text{Count of rule}}{\text{total occurrence of initial phone}}$. For the easy reading purposes, we use "sil" to represent all the sp (short pauses), insertions and deletion ("-") cases and # to represent word boundaries.

Rules	Count, freq	Rules	Count, freq
AE → EH / DH _ T	6, 50	D → sil / N _ #	14, 48.3
AH → ER / SH _ N	10, 15.4	D → JH / # _ R	8, 27.6
AH → IH / B _ F	9, 13.8	DH → D / # _ EH	10, 62.5
AH → AO / K _ N	7, 10.8	R → sil / AO _ #	12, 27.3
EH → ER / N _ R	8, 100	R → ER / IH _ #	9, 42.9
IY → IH / L _ #	17, 44.7	R → AH / EH _ #	6, 28.6
OW → AO / # _ R	8, 57.1	T → sil / N _ #	22, 52.4
OW → UH / S _ #	6, 42.9	T → CH / # _ R	11, 26.2
UW → UH / T _ #	10, 58.8	Z → S / AH _ #	14, 20.9

Table 3: *Selected Data Driven Pronunciation Variation Rules for Singapore English (#means empty and sil means silence in phone deletion, in this corpus, the maximum number of occurrence of a single rule is 22, ER refers to /ə/)*

3.3. Spontaneous American English Rules

We used similar procedures to produce an analogous set of American English rules using the Buckeye corpus [11]. Manual transcriptions and canonical pronunciations of all words in the corpus are provided. We selected the 10 minute recording data from each of the first five speakers. Alignment between the lexical and manual transcriptions was performed using a minimum edit distance algorithm. The rules are considered to be representative for spontaneous American English pronunciation variations. We then counted the number of instances for each substitution under various original contexts, and present the most frequently occurring ones in Table 4. Only instances which occur more than five times are considered as candidate rules for American English.

Rules	Count, freq	Rules	Count, freq
AE → IH / # _ N	23, 19.3	DH → N / # _ EH	18, 21.4
AE → EH / # _ T	14, 11.8	DH → TH / # _ IY	16, 19.0
AH → sil / F _ R	15, 23.0	DH → D / # _ EH	10, 11.9
EH → IH / G _ T	14, 100	R → sil / OW _ #	36, 100
IY → sil / DH _ #	12, 10.4	T → Tq / IH _ #	17, 5.0
IY → AH / DH _ #	85, 73.9	T → D / AE _ #	20, 5.8
IY → IH / DH _ #	18, 15.7	T → sil / N _ #	47, 13.5
OW → ER / # _ R	13, 24.5	T → CH / # _ R	25, 7.2
OW → AO / P _ R	15, 28.3	Z → S / AH _ #	16, 53.3
OW → AH / S _ #	11, 20.8	UW → AH / T _ #	59, 50
D → sil / N _ #	127,81.4	UW → IH / Y _ #	12, 10.2
D → EN / N _ #	29, 18.6	UW → sil / T _ #	14, 11.9

Table 4: *Selected Data Driven Pronunciation Variation Rules for American English (original phonemes occur in Singapore English rules, Tq: T with glottal stop)*

3.4. Comparison between Singapore English and American English

We can observe from Table 3 and Table 4 that for the same original phonemes, the substitutions in American English are similar to Singapore English in the consonant deletion cases such as T, Z, D and R deletions while the vowel substitutions can be very different. These consonant deletion cases could be the typical coarticulation effects in spontaneous speech which are expected to occur in both American English and Singapore English. Specifically, the linguistic rules AE → EH (word initial: *apple*), IY → IH, OW → AO, D → sil, DH → D, R → sil, T → sil, Z → S all occur in both Singapore English and American English data but UW → UH (*good*) and AE → EH (not word initial: *bad*) only occurs in Singapore English data. This shows that UW → UH and AE → EH (not word initial) are unique Singapore English rules.

The pattern of AH→ER (/ʌ/→/ə/) in SgE rarely exists in AmE could be due to the fact that CMU dictionary also uses AH to represent short /ə/ (ER) in SgE. For example, AH is used to represent /ə/ for the context of /ʃən/ and ACCOMMODATION is transcribed as AH K AA M AH D EY SH AH N in CMU dictionary. Therefore although the AH → ER mapping rule never occurred in the American English data but happens in Singapore English data for the major number of times (more than 75%), we won't propose it to be a new rule.

We can identify unique Singapore English rules by comparing the Singapore English and American English rules. The frequently occurring rules in American English are deleted from the Singapore English if they exist at both sides. The remaining rules are evaluated with the linguistic knowledge based rules in section 3.5.

The common rules between the American English and Singapore English are in Table 5:

T → sil / N _ #	D → sil / N _ #
T → CH / # _ R	Z → S / AH _ #
T → sil / S _ #	

Table 5: *Rules in Common between SgE and AmE*

The common rules are mainly for consonant substitutions and deletions. These consonant rules typically appear in the coarticulation cases. They show that there are common patterns in both American English and Singapore English for the spontaneous speech. The vowel substitutions are really different and comprise the majority of the unique Singapore English rule patterns.

3.5. Comparison of the Data Driven Rule Set with Linguistic Rules on Two Data Sets

The rules identified in the data driven approach can be traced back to the linguistic knowledge. We can test whether the linguistic rules agree with the data driven rules. At the phoneme level, we can evaluate the effectiveness of the linguistic rules by assessing the frequency of occurrence in the actual Singapore English data. By comparing the frequency of occurrence of the linguistic rules for Singapore English in the American English dataset, we can observe the threshold of the rules for accent detection.

To compute the rule coverage in percentage, we normalize the words with word distribution to be equal. The common words include the, and, you, are, etc and the total distinct words are 1401 (SgE data) and 1517 (AmE data) respectively. After normalizing the occurrence frequency of each word in the 560 distinct words, we compute the percentage of the rule application as

$$\mathit{Rule\ Application}\ (A \rightarrow B) = \frac{\text{number of phone mappings A→B}}{\text{total number of phone occurrences of A}}. \quad (2)$$

Table 6 shows the percentage of the rules that were applied in the data when all the occurrences of the original phonemes are collected. It shows that the linguistic rules can be applied to the majority of the Singapore English speech. On the other hand, there are also significant other possible phoneme substitutions spoken by Singaporeans. These other possible pronunciation variations can be captured only through data driven methods.

From Table 7 we can see that the percentage of the occurrences of the Singapore English rules is significantly reduced in general while the standard pronunciation percentage increases significantly. However there are certain cases such as Z→S (e.g. in the word *Jazz*) show that Americans also frequently pronounces in this way. The "other possible variations" section also shares similar phonemes with Singapore English. Therefore in order to characterize Singapore English pronunciation variations and detect the accent accurately, we could only rely on the combination of the rule set with the average percentage difference between rule application and standard pronunciations to be at least 20%.

Singapore English data	Rule coverage	Standard pronunciation percentage	Other pronunciation variations of the original phoneme
V1	89.6%	0.0%	EH, sil
V2	57.0%	14.9%	IH, ER, OW, sil
V3	32.7%	0.0%	AO, sil, ER, OW
V4	72.7%	6.7%	AH
V5	60.5%	36.0%	sil
V6	80.9%	17.0%	OW, UH
C1	79.6%	12.6%	Sil
C2	60.7%	32.1%	F
C3	45.0%	14.0%	TH, F, T
C4	78.6%	14.3%	T
C5	87.5%	12.5%	∅
C6	60.4%	37.5%	ER, AH
C7	50.0%	50.0%	∅
C8	73.9%	13.6%	TH, CH, IH, N, AH, D, AO, ER
C9	65.0%	35.0%	∅
C10	80.0%	20.0%	∅
C11	14.8%	72.2%	∅

Table 6**:** *Occurrences of Pronunciation Rules in Singapore English. (The rule set is the same as explained in section 3.1. Rule application means the corresponding Singapore English rule is applied and standard pronunciation means the phoneme is pronounced as the original phone)*

American English data	Rule coverage	Standard pronunciation percentage	Other pronunciation variations of the original phoneme
V1	16.5%	53.8%	EH, AH, sil
V2	2.0%	18.6%	IH, IY, S H, AH, sil
V3	14.5%	62.3%	AO, ER
V4	31.6%	46.9%	AH, IH, sil
V5	15.9%	42.9%	sil, AH, IH
V6	19.5%	43.7%	AH, ER, sil
C1	47.6%	50.2%	Sil
C2	0.0%	84.3%	DH
C3	0.0%	60.3%	TH, N, AH
C4	0.0%	84.5%	DH
C5	0.0%	100.0%	∅
C6	21.5%	75.9%	ER
C7	0.0%	100.0%	∅
C8	18.1%	51.5%	EH, CH, IH, OW, N, AH, D, AO, ER
C9	2.0%	98.0%	∅
C10	10.2%	89.8%	∅
C11	32.7%	57.7%	∅

Table 7**:** *Occurrences of Pronunciation Rules in American English*

Specifically, we see that UW→UH (*good*), K→sil (*Jack*), L→sil/ ER_# (/l/ deletion: *girl*) have the biggest difference in rule application coverage between Singapore English and American English. Considering the observations in section 3.3, we would propose that UW → UH, AE → EH (not word initial), K→sil, ERL→ER (/l/ deletion) are the four truly unique Singapore English pronunciation rules. These four rules are successfully observed from comparison of our corpus, linguistic rules and Buckeye corpus.

4. Decision Threshold and Proposed Pronunciation Feedback Algorithm

4.1. Finding Decision Threshold

Given the rule coverages in our current data, we would like to predict the best threshold of each rule in the final language learning system's testing data to distinguish whether the speaker is using Singapore English accent or American English accent for the corresponding phoneme cases. This will help us provide the pronunciation variation feedback. To find the threshold value and define the threshold region for each rule, we can model each occurrence of the left hand side phone in the lexical transcription as a random event in which either the rule is applied or it is not. This follows a Bernoulli distribution. If we assume that rule application is independent across instances, we can show that the rule coverage, essentially a sum of i.i.d Bernoulli Random Variables, follows a Binomial distribution of the form:

$$X \sim B(n, p) \tag{3}$$

And the probability mass function of binomial distribution is used to model the coverage of the rules on Singapore English data and American English data:

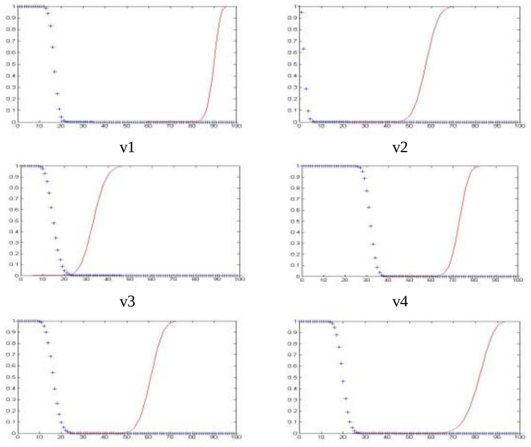
$$f(k; n, p) = Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{4}$$

The final optimal threshold for deciding if an utterance is generated from a model consistent with American English or with Singapore English rule based on the rule coverage threshold in the training data is formulated under a Maximum A Priori (MAP) framework [24]. The threshold is computed based on comparing normalized binomial distribution probability density functions and finding the equalized values as shown in Table 8.

Singapore English rules	Sg data	Am data	Threshold based on binomial distribution
V1	0.8960	0.1650	0.5518
V2	0.5700	0.200	0.1974
V3	0.3270	0.1450	0.2274
V4	0.7270	0.3160	0.5243
V5	0.6050	0.1590	0.3612
V6	0.8090	0.1950	0.5028
C1	0.7960	0.4760	0.6472
C2	0.6070	0.10	0.1271
C3	0.4500	0.10	0.890
C4	0.7860	0.10	0.1877
C5	0.8750	0.10	0.2348
C6	0.6040	0.2150	0.3985
C7	0.5000	0.010	0.1002
C8	0.7390	0.1810	0.4484
C9	0.6500	0.200	0.2283
C10	0.8000	0.1020	0.4220

Table 8**:** *Singapore English Rules with thresholds*

A refinement to this framework considers confidence-intervals based on the same probabilistic models. The threshold region is determined by finding the interval of rule coverage percentage that will cause the p values of SgE rules on AmE data to be less than 0.05 while the 1-p values for SgE rules on SgE data to be less than 0.05. In Figure 2, we can find the confidence intervals on both sides for the 6 vowel rules as samples. Hence we can observe the 0.05 significance level lower and upper boundaries as shown in Figure 3. The p value curves on the right hand side show the Singapore English rules applied on Singapore data and the left hand side shows the rules applied on American English data. It shows that the rules' applications are widely separated between Singapore English and American English hence can be used to detect the accent and variations of the given pronunciations. For rule v1, the lower, upper and threshold boundaries in Table 8 are shown in Figure 3.



v1 v2

v3 v4

v5 v6

Figure 2: *P value (y axis) vs Probability of Binomial event (x axis) of Singapore English rules (v1-c2) on SgE data (solid) and AmE data (dashed)*

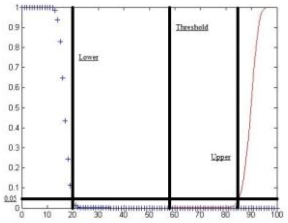


Figure 3: *Lower, Upper and Threshold values for V1*

4.2. Pronunciation Feedback Algorithm

The purpose of the section is to describe our model of pronunciation error detection and phonetic feedback for Singapore English accent to be compared with American English accent. The procedure for giving pronunciation feedback of a given speaker is as following. Users of the speech evaluation system are evaluated by first reading designed phonetically balanced text such as *The North Wind and the Sun* for several times [20] or the 80 designed sentences in our SgE corpus. Next we compare the forced alignment with phone recognition results with time segmentation and extract the rules according to the steps in section 2. Then the system will compute the rule coverage percentage compared with the thresholds to decide whether the speaker follows SgE or AmE patterns. Finally we will provide the corresponding pronunciation variation feedback. The feedback takes two parts. Firstly, it will detect all the erroneous phonemes spoken by the user compared with the reference phonemes. Then it will provide the suggested phonemes that have been identified to be the actual variations to help the user realize his or her mispronunciations.

The rule set to be used to give the pronunciation feedback for Singapore learners of American English are V1-V6 and C1-C10. We will drop the rule C11 as it is not a representative rule for Singapore English due to the less rule coverage compared with American English. The detailed framework of pronunciation feedback can be viewed as two parts: detection of pronunciation variations and feedback of suggested errors. The detection and feedback process are both based on our rule patterns with the decision thresholds.

The detailed procedure is as following. At frame level, let φ be the forced alignment results and θ be the decoding results. We extract the phonological rules and compute the rule coverage s for each rule phone A → phone B with the confidence interval value C at significance level 0.05 ($C_{0.05}$=upper value in Figure 3).

Then we identify the number of occurrences g(x) of the rules as following and compare with the total occurrences of the original phonemes to get the coverage s:

$$g(x) = g(\varphi, \theta) = \forall k\ s.t.\ \varphi_k = A, \theta_k = B \qquad (5)$$

Finally the feedback decision f(s) is:

$$f(s) = \begin{cases} if : s > C_{0.05}, feedback : A \rightarrow B \\ if : C_{0.05} > s > threshold, feedback : error(A) \\ if : s < threshold, normal \end{cases} \qquad (6)$$

The detected phonemes A in the first part and identified phonemes B in the second part are recorded and tested to be more accurate than solely applying the phoneme confidence scores and one best decoding such as Goodness of Pronunciation (GOP) [6] for each phoneme.

For an example of our feedback framework with rule V1, we will collect all the 10 occurrences of phoneme IY in the testing script. The threshold is 0.5518 and the upper bound value is 0.8340 from the statistical test as shown in Figure 3. If the speaker is recognized to have pronounced IY to be IH for 9 times or more, we would report that there is an erroneous pronunciation IH for IY for the speaker. If it is recognized that the speaker has pronounced IY as IH for 6 to 8 times, we would report that IY is detected as an erroneous phoneme but we wouldn’t suggest the errors feedback of IH as it is not confirmed from the data. If the number of times IY is pronounced as IH is less than 6, we will not report any error.

For comparison, the GOP approach is to compute the log likelihood of the phonemes (p) based on the acoustic scores (O) [16]:

$$GOP(p) \equiv P(p|O) = \frac{P(O|p)P(p)}{\sum_{q \in Q} P(O|q)P(q)} \qquad (7)$$

It will report error in part one if GOP likelihood value of a certain phoneme is less than certain threshold and can give suggestion of the one best substitution phonemes in part 2 based on the acoustic scores of the substitution phonemes in the lattice.

To evaluate the effectiveness of the two approaches, we compare average F1 scores of them compared with manual transcribed sentences. For the 90 test sentences, we compute precision and recall scores of the erroneous phonemes detected in part one and suggested in part two by the two approaches and compare with the manual phonetic transcriptions. The average F1 $(= \frac{2 \times precision \times recall}{precision + recall})$ scores for the rule based approach and GOP approach (threshold of error detection fined tuned to be the optimum) and the GOP one best phoneme substitution compared with rule based feedback are shown in Table 9. In addition, the erroneous phonemes identified by our rules can cover more than 85% of the phone errors identified by the manual transcribers. The F1 scores show that our rule patterns are really effective in providing the suggestions of error detection and feedback. This can be further applied in the applications of Singapore English accent detection and tone recognition by transforming the linguistic knowledge into the detection process [22, 23].

F1 score	Rule based	GOP
Detected (part one)	0.67	0.51
Suggested (part two)	0.54	0.33

Table 9**:** *F1 scores for the detected and suggested phonemes based on Rule based and GOP approaches*

5. Conclusions and Future Work

This paper presents analyses to identify the pronunciation feedback rules and determine the rule thresholds for Singaporeans to learn American English accent and presents an improvement on error detection framework. We have collected a manual transcription corpus and derived the context dependent phonemic rules in Singapore English and American English. The result shows that the pronunciation errors generated from data agree with the linguistic knowledge and the differences between the accents are significant. It reflects the effectiveness of the set of linguistic rules with thresholds in identifying the unique accent and providing mispronunciation feedback in Singapore English. In the following work, we will find the optimum detection thresholds in the region for each rule considering the multinomial distribution and generate our extended recognition and decoding network based on the rule frequencies. The rules generated from read speech data and spontaneous data will play a part in the phoneme decoding process.

6. Acknowledgements

We would like to thank Professor Chng Eng Siong, Professor Tan Ying Ying from Nanyang Technological University and Professor Li Haizhou from I2R for valuable discussions and advising Wenda's research during Master degree project.

7. References

[1] R. K. Tongue, The English of Singapore and Malaysia. 2nd edition. Singapore: Eastern Universities Press. 1979

[2] J. Platt, and H. Weber., English in Singapore and Malaysia. Kuala Lumpur, New York: Oxford University Press. 1980

[3] M. W. J. Tay, The phonology of educated Singapore English. English World Wide. 3:2. pp. 135-145. 1982

[4] "Singapore to launch Speak-good-English campaign", Agence France Presse in Singapore, 30 August 1999.

[5] Wells, John C. Accents of English 2, Cambridge University Press. pp. 96–97, 328–30, 498, 500, 553, 557–58, 635. ISBN 0-521-24224-X, 1982

[6] Wenda Chen, Computer Assisted Pronunciation Learning for English learners in Singapore, Master thesis, School of Computer Engineering, Nanyang Technological University, 2014

[7] W. Chen, Y. Y. Tan, E. S. Chng, and H. Li, Computer Assisted Language Learning in Singapore - Modeling Singapore English for Pronunciation Variation Detection, The 15th International CALL Research Conference, Taiwan, 24-27 May 2012

[8] W. Chen, Y. Y. Tan, E. S. Chng, and H. Li, The Development of A Singapore English CALL Resource, Oriental COCOSDA 2010

[9] Witt S. M., "Automatic error detection in pronunciation training: Where we are and where we need to go", Proc. IS ADEPT, 6 June, 2012

[10] Ann Lee, James R. Glass, Context-dependent pronunciation error pattern discovery with limited annotations. INTERSPEECH 2014: 2877-2881

[11] Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. (2007) Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

[12] Meng, H., Lo, Y. Y., Wang, L. and Lau, W. Y., "Deriving salient learners' mispronunciations from cross-language phonological comparisons", Proc. ASRU, 2007

[13] Lo, W. K., Zhang, S., and Meng, H., "Automatic derivation of phonological rules for mispronunciation detection in a computer assisted pronunciation training system", Proc. Interspeech, 2010

[14] Cucchiarini, C., Van den Heuvel, H., Sanders, E., and Strik, H., "Error selection for ASR-based English pronunciation training in "My Pronunciation Coach"", Proc. Interspeech, 2011

[15] Hong, H., Kim, S., and Chung, M., "A corpus-based analysis of Korean segments produced by Japanese learners", Proc. SLaTE, 2013

[16] Silke M. Witt. "Use of Speech recognition in computer-assisted language learning", unpublished thesis, Cambrige Uni. Eng. Dept, 1999.

[17] S. Witt and S. J. Young, Language learning based on non-native speech recognition, 5th European Conference on Speech Communication and Technology (Eurospeech 1997), 22-25 September 1997, Rhodes, Greece

[18] Boersma, Paul and Weenink, David (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved 24 March 2015 from http://www.praat.org/

[19] Weide R. The CMU pronunciation dictionary, release 0.6[J]. 1998.

[20] David Deterding and Low Ee Ling, The NIE Corpus of Spoken Singapore English (NIECSSE), SAAL Quarterly No 56, Nov 2001, pp.2-5.

[21] Cambridge English Pronunciation Dictionary, http://dictionary.cambridge.org/dictionary/british/pronunciation

[22] Nancy F. Chen, Sharon Tam, Wade Shen, Joseph P. Campbell, "Characterizing Phonetic Transformations and Acoustic Differences Across English Dialects", IEEE Transactions on Audio, Speech, and Language Processing, 2014.

[23] Rong Tong, Nancy F. Chen, Bin Ma, Haizhou Li, "Goodness of Tone (GOT) for Non-native Mandarin Tone Recognition", Interspeech 2015.

[24] Murphy, Kevin P. (2012). Machine learning : a probabilistic perspective. Cambridge, MA: MIT Press. pp. 151–152.