9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

L2 Speech Rhythm Development in New Immigrants

Donald White, Peggy Mok

The Chinese University of Hong Kong, Hong Kong
donaldtimothywhite@gmail.com, peggymok@cuhk.edu.hk

Abstract

In a longitudinal study, English second-language (L2) speech rhythm development of five Hong Kong students is investigated during their first year after immigration to Canada, the United States, Australia, and the United Kingdom, respectively. Students were recorded reading a variety of English passages at three time points: before emigration, and then at approximately six months after, and one year after immigration. Identical utterances from the three recordings were isolated and segmented for analysis with durational rhythmic metrics in Praat. As well, the students were surveyed on the quantity and quality of interactions during their first year outside of Hong Kong. Although all five subjects displayed some degree of significant rhythmic development in the expected direction (i.e. greater stress timing), the strongest changes were evident in the student with the least amount of communication in Cantonese, and the highest amount of interaction with native English (L1) speakers while living in her new environment. These findings suggest that the effect of language experience is more robust than length of residence as a predictor for acquisition of L2 prosody.

Index Terms: speech rhythm, second language acquisition, length of residence, language experience

1. Introduction

This study investigates L2 English speech rhythm in five Hong Kong students during their first year after immigration to English-speaking countries. The students lived in Canada (2), the United States (1), Australia (1), and the United Kingdom (1). All five students attended the same secondary school in Hong Kong prior to emigration, and all five speak Cantonese as their L1. The rhythmic characteristics of Cantonese tend toward syllable timing, while those of English are traditionally considered stress timed. Previous work has demonstrated that the L2 English spoken by Cantonese L1 speakers tends to be more syllable-timed than that of English L1 speakers [1]. By measuring their L2 speech rhythm three times in the first year, this study expected to observe development in the direction of stress timing. In addition, a survey was given to the students to understand their language experience during their first year after immigration. This study also predicted that a greater amount of communication with English L1 speakers, along with general tendencies toward extroversion, would correlate with greater changes in the speech rhythm of the subjects.

In the speech of most L2-speaking immigrants, a foreign accent (FA) persists even after many years of committed language learning. The degree of FA depends mainly upon three factors: age of acquisition (AOA), the age at which one begins learning the language; length of residence (LOR), the length of time spent living in an environment where the L2 is the ambient language; and language experience (LE), the quality of interaction and input during the time spent in that environment.

Although a robust correlation has been established between AOA and FA [2], the roles of LOR and LE have been of special interest to frameworks such as Flege's Speech Learning Model [3], which emphasizes the importance of language input for the L2 learner. Some studies have demonstrated that foreign accent is significantly reduced with increased LOR [4, 5]; others have found that LOR has a negligible effect [6, 7]. What these studies all have in common is a segmental focus, often investigating the quality of L2 vowels. Very little previous research has investigated the effect that LOR and LE have on suprasegmentals. One notable exception to this is work based on the same subjects and data as [5], which, like the present study, has begun to examine the effects of LOR on speech rhythm. The inclusion of durational speech rhythm metrics in analyses may give a more objective indication of the ways in which FA is affected by increased LOR. In order to test this claim, it is first necessary to understand the battery of durational speech rhythm metrics, and what it is that they measure.

1.1. Speech Rhythm Metrics

Early research into speech rhythm proposed that perceived differences between rhythm classes stemmed from a fundamental difference in isochronous units [8, 9]; however, these units could not be empirically demonstrated. In [10] a reassessment of rhythmic typology was proposed, based on a correlation between stress timing and three phonological characteristics: stressed syllables, vowel reduction and complexity of syllable structure. Syllable timing, on the other hand, was characterized by a lower prevalence of these three factors. Rather than a categorical distinction, rhythmic typology was thenceforth framed as a continuum in which languages exhibit a greater or lesser amount of the three variables identified by [10].

This hypothesis has led to a number of durational measurements that quantify the acoustic correlates of speech rhythm. First, [11] devised three metrics to distinguish between rhythm types. Percent V measures the overall percentage of vocalic (vowel) intervals in an utterance, which is higher in syllable-timed than stress-timed languages. The other two metrics, ΔV and ΔC, measure the standard deviations of vocalic and intervocalic (consonant) duration in an utterance respectively, which gauge the durational variability of these two intervals. Stress-timed languages tend to have higher scores in both of these measurements. Although %V was found to be a stable predictor of rhythm types, ΔV and ΔC were not as reliable because they varied inversely with speech rate. To remedy this problem, [12] developed a variation coefficient (VarcoC) to normalize ΔC across varying speech rates. This metric, which divides ΔC by the mean

10.21437/SpeechProsody.2018-169

duration of intervocalic intervals in a given utterance, was found to classify rhythm types more accurately. Subsequently, the same variation coefficient was applied to Δ measures of other intervals (vowels and syllables), resulting in Varco V [13] and Varco S [1].

Another approach to rhythmic measurements was proposed by [14]. Instead of overall variability of an utterance, it measured the difference in duration between successive pairs of intervals. These metrics, called Pairwise Variability Indices (PVI), compare adjacent vocalic intervals (nPVI-V), and adjacent intervocalic (consonantal) intervals (nPVI-C). These measurements were thought to reflect two features of stress timing more accurately: alternating stressed syllables and larger variation in consonant clusters. For both of these PVI measurements, a higher result indicated greater variability, and therefore more stress timing. Later, [15] used the PVI measurements to gauge the variability of syllables in the same way with a new metric called nPVI-S. Like the Δ metrics, the PVI measurements were normalized for speech rate, because variations in speech rate tended to distort the results. To normalize the results, the durational difference between the two intervals was divided by the mean duration of the two intervals.

The present study will use 8 of these metrics Percent V; normalized PVI metrics for intervocalic, vocalic, and syllabic durations; and Varco metrics for intervocalic, vocalic, and syllabic durations. In addition, the speech rate (syllables/second) will be reported for each of the subjects.

For every metric except for Percent V, it was expected that scores would increase with more time spent in an English-speaking environment. The Varco and PVI metrics measure durational variability, which is generally higher in stress timed languages. An increase in speech rate was also expected to increase because a longer time living abroad would tend to increase L2 fluency.

2. Methodology and Data

2.1. Subjects

All five subjects were attending the same Hong Kong secondary school prior to emigration, which occurred between 2011 and 2015. Table 1 details their ages at the time of emigration, and their destination. In every case, the primary reason for leaving Hong Kong was to continue their studies at secondary schools or universities overseas.

Table 1: Ages of emigration and Destination of the Subjects

Subject	Age (years; months)	Destination
CanGirl 1	17;9	Markham, Canada
CanGirl 2	17;10	Toronto, Canada
USAGirl	16;7	Wausau, WI, USA
AusBoy	20;5	Sydney, Australia
UKBoy	16;11	Cambridge, UK

2.2. Recordings

Students were recorded before emigration and then at approximately six months, and one year after immigration. (One exception is AusBoy, who was recorded at 3 months and 9 months after moving to Australia. Because of the differences in Hong Kong's and Australia's school schedules, AusBoy returned to Hong Kong for his December (summer) holiday three months after his departure.) Recordings were sometimes conducted in person and sometimes remotely over Skype. In every case, the subjects were recorded using a Zoom H2 recorder, with digital sampling at 44.1 Hz.

The subjects were recorded reading three passages from which the data for the present study are taken: The North Wind and the Sun, The Rainbow, and fourteen sentences composed by the authors. Several criteria determined which utterances were suitable for analysis. First, the utterances had to be longer than five syllables in length and spoken within the same breath group. Second, any utterance with a pause or false start was rejected. Finally, the utterance had to meet these criteria in all three recordings.

For each subject, the total numbers of acceptable utterances per recording were as follows: CanGirl1 – 17, Cangirl2 – 20, USAGirl – 20, AusBoy – 28, UKBoy – 16.

After the suitable utterances were identified and isolated, they were segmented in Praat [16] on two tiers. The first demarcated vocalic (vowel) and consonantal boundaries; syllable boundaries were established on the second tier. Segmenting syllables is a somewhat controversial process because it requires commitment to a hypothesis regarding the exact composition of a syllable. In this process, the authors adhered to the Maximum Onset Principle [17] to determine syllable boundaries; however, this did not preclude a number of choices that were essentially judgment calls based on careful listening and observation of spectrograms. These cases were, for the most part, instances when the final coda syllable was resyllabified across a word boundary.

After segmentation was complete, and the rhythmic scores were tabulated, a paired-sample comparison of means was carried out between each set of data. The results of these comparisons are shown in the following section.

3. Results

Tables 2-6 show the individual results for each of the five subjects. T1 is the recording before emigration, T2 is the half-year recording, and T3 is the one-year recording (except in the case of Table 5 (AusBoy)—see section 2.2) In each column, the numbers on the left are the mean scores for each metric. An asterisk denotes a significant difference between that mean and the mean in the column to its right; an asterisk in the T3 column denotes a significant difference between that mean and the mean in T1. The numbers on the right, in parentheses, are the standard deviations. Beneath each table, the results of the paired sample t tests are listed for the comparisons that were significantly different.

Table 2*: Results for CanGirl1*

Metric	T1	T2	T3
Speech Rate	4.2* (.72)	4.8 (.78)	4.6 (.7)
VarcoS	33.1 (11.9)	33.2 (10.8)	34.5 (9.5)
VarcoV	40.7 (20.4)	39.7 (10.9)	42.5 (12.6)
VarcoC	42.4 (11.2)	47.5 (18.4)	50.3* (14.5)
PercentV	52.5 (8.1)	49.9 (5.4)	53.3 (6.9)
nPVI-V	43.6 (14.9)	47.5 (16.2)	46.8 (12.8)
nPVI-C	48.7 (15.6)	51.7 (19.3)	53.5 (17.4)
nPVI-S	40.3 (17.6)	38 (16.6)	40.5 (13.9)

(Speech Rate: T1/T2- t(17) = -7.439, p = < .001) (VarcoC: T1/T3- t(17) = -2.917, p = .01)

Table 3*: Results for CanGirl2*

Metric	T1	T2	T3
Speech Rate	4.6* (.65)	4.9 (.66)	4.9* (.87)
VarcoS	46.1 (11.3)	47.2* (11.9)	51.4* (12.1)
VarcoV	44.7 (17.6)	43.4* (18.5)	47.9 (20.1)
VarcoC	48.9 (12.5)	50.4 (15.1)	55.7 (16.6)
PercentV	43.67 (6)	43.51* (5.8)	41.25 (5.8)
nPVI-V	48.44 (18.3)	48.8 (21.9)	53.78 (22.8)
nPVI-C	57.9 (16.7)	61.3 (18.3)	65.2 (23)
nPVI-S	63.8 (14.4)	60.7 (17.6)	63.3 (18.8)

(Speech Rate: T1/T2- t(20) = -5.743, p < .001; T1/T3- t(20) = -4.036, p = .001) (VarcoS: T2/T3- t(20) = -2.803, p = .011; T1/T3- t(20) = -2.703, p = .014) (VarcoV: T2/T3- t(20) = -2.151, p = .044) (PercentV T2/T3- t(20) = 2.711, p = .013)

Table 4*: Results for USAGirl*

Metric	T1	T2	T3
Speech Rate	3.8* (.5)	4.6 (.74)	4.8* (.79)
VarcoS	37.2 (11)	38.2 (13.2)	38.3 (14.5)
VarcoV	36.7* (13)	42.5 (14.5)	46.6* (13.9)
VarcoC	49.9 (9.5)	56.7 (13.1)	57.7* (13.7)
PercentV	51.7* (4.6)	54.5 (5.8)	54.4* (5.9)
nPVI-V	40.7* (14)	49.6 (18.6)	54.7* (16.7)
nPVI-C	57.9* (14.6)	68.6 (19.1)	69.1* (16.7)
nPVI-S	49.8 (18.9)	47.3 (22)	52.5 (19.9)

(Speech Rate: T1/T2- t(20) = -7.066, p < .001; T1/T3- t(20) = -9.682, p < .001;) (VarcoV: T1/T2- t(20) = -2.671, p = .015; T1/T3- t(20) = -3.152, p = .005) (VarcoC: T1/T3- t(20) = -2.634, p =.028) (PercentV: T1/T2- t(20) = -3.407, p = .003; T1/T3- t(20) = -2.756, p = .012) (nPVI-V: T1/T2- t(20) = -3.956, p = .001; T1/T3- t(20) = -4.68, p < .001) (nPVI-C: t(20) = -2.451, p = .024; T1/T3- t(20) = -2.838, p = .01)

Table 5*: Results for AusBoy*

Metric	T1	T2	T3
Speech Rate	4.7 (.7)	4.8* (.81)	5.6* (.88)
VarcoS	45 (11.3)	45.5 (11.1)	48.5 (11.8)
VarcoV	47.5 (13.9)	49.8 (17.7)	53* (19.3)
VarcoC	49 (9.9)	48 (10.9)	47.6 (12.6)
PercentV	48.2* (7)	45.62 (7.3)	45.7* (6.7)
nPVI-V	53.7 (19.2)	54.1 (17.9)	55.9 (17.5)
nPVI-C	59.4 (18)	58.5 (15.5)	57.8 (22.9)
nPVI-S	57 (17.6)	55.7 (17)	60.5 (17.4)

(Speech Rate: T2/T3- t(28) = -8.598, p < .001; T1/T3- t(28) = -8.224, p < .001) (VarcoV: T1/T3- t(28) = -2.405, p = .023) (PercentV: T1/T2- t(28) = 3.283, p = .003; T1/T3- t(28) = 2.227, p = .034)

Table 6*: Results for UKBoy*

Metric	T1	T2	T3
Speech Rate	4.7 (.71)	4.5 (.48)	4.7 (.55)
VarcoS	38.1 (10.8)	40.4 (11.6)	43.4* (8.4)
VarcoV	37.9 (13.4)	37.1 (14)	39 (13.9)
VarcoC	50 (12.8)	55 (15.6)	52.8 (11.1)
PercentV	48.5* (2.8)	46.1* (3.6)	48.1 (3.3)
nPVI-V	42.1 (17.6)	39.6 (16.3)	40.6 (12.5)
nPVI-C	59.7 (12.4)	59.6 (13.4)	61.7 (15.6)
nPVI-S	45.9 (15.4)	47.1 (14.6)	51 (13.4)

(VarcoS: T1/T3- t(16) = -2.867, p = .011) (PercentV: T1/T2- t(16) = 2.246, p = .039; T2/T3: t(16) = -2.127, p = .49)

Tables 7 and 8 show some of the data from the Language Experience survey that was given to the students. In Table 7, the students estimated the average days per week and hours per day that they spent speaking English to L1 English speakers during their first year after immigration. In Table 8, the students estimated the average days per week and hours per day that they spent speaking Cantonese to L1 Cantonese speakers during their first year after immigration.

Table 7: *Communication with L1 English Speakers*

Subject	Mean Days/Week	Mean Hours/Day
CanGirl1	6	>5
CanGirl2	2	2
USAGirl	7	>5
AusBoy	6	4
UKBoy	7	1

Table 8: *Communication with L1 Cantonese Speakers*

Subject	Mean Days/Week	Mean Hours/Day
CanGirl1	2	4
CanGirl2	7	5
USAGirl	1	2
AusBoy	4	2
UKBoy	2	1

4. Discussion

Although all subjects seem to undergo at least some rhythmic modifications in the expected direction, the most notable results are those of USAGirl (Table 4), who had significant differences in 11 out of 27 comparisons. According to her survey results, USAGirl was also the subject who spoke English the most and Cantonese the least during her first year after immigration. Surely one of the biggest factors contributing to these communication patterns was Wausau, Wisconsin, which was the small city of 40,000 in which USAGirl lived. While living there, she did not know any other Cantonese speaker in the city. Furthermore, she was boarding with an English-speaking family in Wausau. In other words, the opportunities to communicate in her L2 were abundant, but in her L1 were quite limited. The Cantonese she did speak was to her family in Hong Kong over the internet.

All of USAGirl's significant differences are in the expected direction except for one surprising exception: Percent V increases significantly during the first six months and remained at a similar level after one year. This is a bit unusual because her vocalic variability increases significantly, as reflected in VarcoV and nPVI-V. When these numbers increase, the usual interpretation is that vowel reduction has increased, and therefore it would be expected that the overall vocalic content of the utterances should decrease. This suggests, perhaps, that her vowel reduction may have actually been achieved by a lengthening of the non-reduced vowels rather than a shortening of the reduced ones. In addition, the pattern of USAGirl's significant differences suggest that the changes to her speech rhythm occurred mainly within the first six months. In almost every case, there is a significant difference between T1 and T2, as well as T1 and T3, but not between T2 and T3. In other words, the rhythmic changes occurred early and seem to have been maintained in the second half of the year.

AusBoy did not undergo as many rhythmic modifications as USAGirl, but his patterns of English communication, as shown in Table 8, were quite similar. During his first month away, Ausboy's living situation was similar to that of USAGirl in that he boarded with an English-speaking family. Subsequently, he moved into an apartment of his own, but he was very socially active with a number of L1 English speakers. In contrast to USAGirl, Ausboy's Cantonese communication was quite frequent, which was attributable to a fair number of Cantonese speakers at the university in which he studied. Also in contrast to USAGirl were the PercentV scores of Ausboy, which were significantly different in the expected direction. This change also seems to have occurred early (within the first three months after immigration), and were maintained up until the nine-month mark (T3).

CanGirl1 and CanGirl2 were both living in Cantonese-speaking households after immigration. CanGirl1 boarded with her aunt, with whom she did not speak very often. It is important to point out, however, that CanGirl1 was living in Markham, Ontario, in which nearly 16% of the citizens are L1 Cantonese speakers, and 45% are ethnically Chinese [17]. Even if her household communication was actually as low as she reports, she was not isolated from her L1 in the same way that USAGirl was. CanGirl1 also reports frequent interaction with L1 English speakers during the time period observed. Nevertheless, this apparent extroversion did not translate into a high number of significantly different rhythmic scores. Her speech rate increased significantly in the first six months, and her T3 VarcoV score was significantly higher than that of T1.

Similar to CanGirl1, CanGirl2's speech rate increased significantly within the first six months. All of her other rhythmic changes, however, occurred in the second half of the year. These metrics included VarcoV, VarcoS, and PercentV. CanGirl2 had, by her own estimation, the highest exposure to L1 Cantonese, and a relatively low level of interaction with L1 English speakers. Nevertheless, she seems to have made some rhythmic modifications to her L2 English. In considering these facts, it is also worth noting that, according to the first author's judgement, CanGirl 2 had the lowest amount of FA in the T1 recording when compared with the other subjects. Perhaps her competence in speaking English allowed her to strengthen and further develop her rhythmic patterns while she was living in Canada, notwithstanding her exposure to Cantonese.

Finally, UKBoy was in a largely English environment at home, but he was rather introverted in his behaviour. He spoke more English than Cantonese, but overall he did not speak very much of either language. In his interviews with the first author, UKBoy has repeatedly suggested that he led a very quiet life after immigration that was very focused on his studies. Perhaps this is the reason that his rhythmic differences were marginal.

The results suggest that speech rhythm metrics may correlate to some extent with LOR and LE. To corroborate these findings, the next phase of this ongoing study will enlist native English speakers to make judgements about the FA and intelligibility of the subjects. In this way, a better understanding of the ways in which L2 speech rhythm effects L1 English perception. The present study suggests that greater stress timing may increase with time spent in an L1 English environment. The question is whether these speakers are also making themselves more intelligible when they adjust their rhythm in a more stress-timed direction.

5. Acknowledgements

The authors would like to acknowledge the staff and participating students from F.D.B.W.A. Szeto Ho Secondary School in Lam Tin, Kowloon, Hong Kong S.A.R.

6. References

[1] P. Mok and V. Dellwo, "Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English," *4th Conference on Speech Prosody,* Campinas, Brazil. 423-426, 2008.

[2] R. DeKeyser and J. Larson-Hall, "What does the critical period mean?" In J. Kroll, and A. DeGroot, (Eds.) *Handbook of Bilingualism, Psycholinguistic Approaches.* Oxford: Oxford University Press, pp. 88-108, 2005.

[3] J. E. Flege, "Second language speech learning: theory, findings, and problems." In W. Strange (Ed), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Timonium, Maryland: York Press, pp. 233-269, 1995.

[4] E. T. Purcell and R. W. Suter, "Predictors of pronunciation accuracy: a reexamination," *Language Learning*, vol. 30, 271-287, 1980.

[5] M. Munro, and T. Derwing, "Segmental acquisition in adult ESL learners: a longitudinal study of vowel production," *Language Learning*, vol. 58, no. 3, 479-502, 2008.

[6] S. Tahta, M. Wood, and K. Lowenthal, "FA: Factors relating to transfer of accent from the first to the second language," *Language and Speech*, vol. 24, 265–272, 1981.

[7] K. Tsukada, D. Birdsong, E. Bialystok, M. Mack, H. Sung, and J. Flege, "A developmental study of English vowel production and perception by native English adults and children," *Journal of Phonetics*, vol. 33, 263-290, 2005.

[8] K. Pike, *The intonation of American English*. Ann Arbor: University of Michigan Press, 1945.

[9] D. Abercrombie, *Elements of general phonetics*. Chicago: Aldine Publishing Company, 1967.

[10] R. M. Dauer, "Stress-timing and syllable-timing reanalyzed," *Journal of Phonetics,* vol. 11, no. 1, pp. 51-62, 1983.

[11] F. Ramus, M. Nespor and J. Mehler, "Correlates of linguistic rhythm in the speech signal," *Cognition,* vol. 73, no. 3, 265-292, 1999.

[12] V. Dellwo, "Rhythm and speech rate: A variation coefficient for ΔC," In P. Karnowski and I. Szigeti (Eds.), *Language and*

language processing. Frankfurt am Main: Peter Lang, pp. 231-241, 2006.
[13] L. White and S. L. Mattys, "Calibrating rhythm: First language and second language studies," *Journal of Phonetics,* vol. 35, no. 4, 501-522, 2007.
[14] E. Grabe, and E. Low, "Durational variability in speech and the rhythm class hypothesis," *Papers in Laboratory Phonology 7,* 515-546, 2002.
[15] D. Deterding, "The measurement of rhythm: A comparison of Singapore and British English," *Journal of Phonetics,* vol. 29, no. 2, 217-230, 2001.
[16] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, 341-345, 2001.
[17] D. Kahn, *Syllable-based generalizations in English phonology*. Doctoral dissertation, MIT, 1976.
[18] Statistics Canada, *2011 Census*, 2012.