

L2 Speech Rhythm and Language Experience in New Immigrants

Donald White & Peggy Mok

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

ABSTRACT

This study is a longitudinal investigation of second language (L2) speech rhythm in Cantonese-first-language (L1) immigrants. Seven Hong Kong students were recorded five times throughout a two-year period while they were living abroad in English-speaking countries. The speech rhythm of the read utterances in these recordings was then measured using several durational variability metrics. In addition, the participants were surveyed on their use of L1 and L2 speech during their time abroad. The results suggest that significant increases in durational variability and speech rate occurred during the first year abroad. Additionally, there seems to be inverse correlation between the use of L1 Cantonese and rhythmic changes in the expected direction. These findings were further supported by ratings of accentedness, comprehensibility and intelligibility of their speech production by a group of native English speakers.

Keywords: L2 Speech Rhythm, Language Experience, Length of Residence

1. INTRODUCTION

In this ongoing longitudinal investigation of L2 speech rhythm, the speech of seven L1 Cantonese-L2 English-speaking immigrants is examined. These participants emigrated from Hong Kong to the following countries: Canada (3), USA (2), United Kingdom (1), and Australia (1). Previous longitudinal studies of this kind, such as [15], have provided an in-depth understanding of the ways in which L2 speech patterns change when immigrants enter a new linguistic environment. The present study takes the same approach, but with a focus on prosody. The durational rhythmic patterns of the participants L2 speech is measured and compared over a two-year observation period after immigration to their respective countries.

2. L2 SPEECH AMONG NEW IMMIGRANTS

In studies of new immigrants, previous research has demonstrated correlations between L2 pronunciation and three factors: Length of Residence (LOR) in an L2-ambient environment; Language Experience (LE), which is the quantity and quality of L2 interaction among new immigrants; and “Age of

Arrival”, the age at which they immigrated to the L2-ambient environment.

In L2 speakers with an AOA greater than 12, the segmental effects of LOR seem limited to a short initial improvement during the first year after immigration [6], [7], [15], [18]. While it is unclear whether a similar phenomenon manifests in suprasegmentals, the crucial role of speech rhythm on L2 speech has been demonstrated convincingly [2], [13], [20]. In addition, more recent research has observed prosodic changes among new immigrants [11], [16], [19]. In order to quantify rhythmic changes of this kind, researchers have relied on several metrics developed over the past two decades.

2.1. Speech Rhythm Metrics

Since Dauer [3] debunked categorical speech rhythm typology, most studies have viewed speech rhythm as a continuum between stress-timing and syllable-timing. As a result, in most current research, the measurement of speech rhythm is based on three related prosodic correlates: stress, vowel reduction, and syllable complexity. In this conception, a greater amount of these three factors characterises languages traditionally considered stress-timed, such as English; and a smaller amount characterises those traditionally considered syllable-timed, such as Cantonese. To measure these correlates, several durational metrics have been developed over the last 20 years. The first two, Pairwise Variability Indices (PVI) and Varco metrics, are measures of pairwise and global durational variability, respectively, for three different intervals: consonantal, vocalic, and syllabic (PVI-C, PVI-V, PVI-S, VarcoC, VarcoV, VarcoS [4], [5], [9], [12], [17], [21]). For each of these metrics, a higher score indicates more durational variability, i.e. characteristic of stress timing. An additional metric used in the present study is PercentV [17], which measures the percentage of vocalic content in an utterance. A higher score in PercentV reflects characteristics of syllable timing. Although there are rhythmic difference between native varieties of English, all varieties are generally more stress-timed than English spoken by Cantonese-L1 speakers.

2.2. Language Experience (LE)

LOR is quite uninformative when considered in isolation. It is a simplistic measurement of time spent in a new country, but does not include any details about *how* that time was spent. In order to obtain a

fuller understanding of the participants in the present study, a LE survey was adapted from Freed et. al. [8], which was a detailed inquiry of their L1 and L2 interactions during the observation period. To the best of our knowledge, this is the first study of its kind that has examined both speech rhythm and LE.

3. METHOD

All participants attended the same Hong Kong secondary school prior to emigration. Table 1 details their AOA and destinations.

Table 1: Participant information

Participant	AOA (years;months)	Destination
CanGirl 1	17;9	Markham, ON
CanGirl 2	17;10	Toronto, ON
CanBoy	17;11	Markham, ON
CanUSABoy	18;5	Comox, BC
USAGirl	16;7	Wausau, WI
AusBoy	20;5	Sydney, NSW
UKBoy	16;11	Cambridge

3.1. Recordings

Six participants were recorded at approximately six-month intervals beginning with their pre-emigration (T1) recording. (Because of the differences between the school schedules of Hong Kong and Australia, AusBoy was recorded at 3 months, 9 months, 15 months, and two years after moving to Australia.) The interviews were sometimes conducted in person and sometimes remotely over Skype. In every case, the participants were recorded on a Zoom H2 recorder, with digital sampling at 44.1 Hz, placed approximately 20 centimetres from their mouths. They were recorded in free conversation with the first author (the various topics were related to things they had experienced during their time abroad), and reading three passages: “The North Wind and the Sun”, “The Rainbow”, and fourteen sentences composed by the authors.

3.2. Selection, Segmentation, and Analysis

Utterances were selected from the reading passages, based on the following criteria: first, utterances had to be at least five syllables in length and within the same breath group; second, utterances with pauses were rejected; finally, the utterance had to meet these criteria across all five recordings. In other words, the utterances for each participant were identical from T1 to T5. (The justification for identical, read utterances in speech rhythm investigations is stated convincingly in [22].)

After selection, utterances were segmented in Praat [1] on two tiers. The first demarcated vocalic (vowel) and consonantal boundaries; syllable boundaries were established on the second tier. Syllables were segmented according to the Maximum Onset Principle [10]; however, this did not preclude a number of judgment calls based on careful listening and observation of spectrograms. Most of these were cases of the final coda consonant resyllabified across a word boundary. In these instances, the final consonant was considered part of the first syllable in the second word.

After segmentation, the speech rhythm (according to the metrics in Section 2.1) and speech rate in syllables per second (s/s) were measured for each utterance. These results were then analyzed with five paired comparisons of means between the following time points: T1-T2; T2-T3; T1-T3; T3-T5; T1-T5. More comparisons were examined between T1 and T3 because greater changes were expected during that time (see Section 2). As such a large number of comparisons increases the chance of a Type 1 error, we applied a Bonferroni correction, dividing our threshold p value of .05 by the number of comparisons. Our requirement for significance, therefore, is .01.

In an online **LE survey**, the participants were asked to estimate the number of days per week, and the number of hours per day that they communicated in their L2 and L1 during the observation period. In addition, they gave details about their interlocutors, as well as their living situations, etc.

For the **Native Speaker Judgments**, a collection of utterances was rated by 26 American university students for foreign accent, comprehensibility, and intelligibility. Four utterances, two read and two extemporaneous, were chosen for each participant at each time point (Mean Length of Utterance - 10.28 words; Standard Deviation (SD) - 2.58). The design of this rating system was based on Munro and Derwing [14]. Foreign accent was judged on a Likert Scale from 1 to 9, in which 1 denoted no foreign accent and 9 denoted a very strong foreign accent. (Ratings of accent are, of course, highly subjective. One concern about the students who rated the participants was their familiarity with native English accents from outside of North America. To address this concern, the instructions emphasized that they were not to rate the accents based solely on their own accent as a baseline. Instead they were asked to adopt a view of native English accents that was as objective as possible.) An identical scale was used for the comprehensibility rating, where utterances rated 1 were extremely easy to understand, and those rated 9 were impossible to understand. Finally, the intelligibility rating asked respondents to type the utterances that they had heard. This rating was quantified by a ratio in which the aggregate number

of correct words per utterance was divided by the aggregate number of words (words per utterance × number of respondents).

4. RESULTS

Because of limited space, an overview of related highlights from the Speech Rhythm, Language Survey, and Native Speaker Judgments results will be provided below. Readers who are interested in a full report of the results are welcome to contact the first author.

4.1. Speech Rhythm Results

Three metrics will be presented in order from the highest to lowest number of significant increases among the seven participants: speech rate, followed by VarcoC, and, finally, PVI-V. (When we refer to speech rhythm changing in the “expected” direction, it means that the L2 English has become more stress-timed, since Cantonese-L1 speakers tend to have greater syllable-timing in English when compared to English-L1 speakers.)

First, the speech rate (s/s) of six participants increased significantly during their first year after immigration. Table 2 shows the mean (M) speech rates and SD for T1 and T2. The exception was UKBoy, who had no significant change between these time points, and whose T1 speech rate was the highest among all participants.

Table 2: T1-T2 Speech Rate Comparison

Participant	T1 (s/s)	T2 (s/s)	p
	M (SD)	M (SD)	
CanGirl 1	4.22 (.72)	4.83 (.78)	< .001
CanGirl 2	4.56 (.65)	4.93 (.66)	< .001
CanBoy	4.06 (.61)	5.32 (.84)	.006
CanUSABoy	3.71 (.8)	4.96 (1.16)	.006
USAGirl	3.86 (.52)	4.61 (.62)	< .001
AusBoy	4.67 (.67)	5.43 (.87)	< .001
UKBoy	4.77 (.77)	4.48 (.46)	.19

Among the VarcoC rhythm scores, there were just two participants with significant increases: CanGirl 1’s VarcoC increased significantly between T1 (M = 42.4, SD = 11.24) and T3 (M = 50.31, SD = 14.49) ($t(17) = -2.917, p = .01$); and USAGirl’s VarcoC increased from T1 (M = 49.58, SD = 12.92) to T2 (M = 59.75, SD = 16.12), $t(17) = -3.194, p = .005$.

Finally, only one participant had a significant increase in PVI-V. USAGirl’s PVI-V score increased from T1 (M = 39.86, SD = 12.11) to T3 (M = 53.41, SD = 15.32), $t(17) = -4.194, p = .001$.

The selected results above illustrate a common theme in the speech rhythm scores: across all metrics, USAGirl had the greatest number of significant changes in the expected direction.

4.2. LE Survey Results

To gain a better understanding of the patterns in the rhythm scores, the LE survey was examined for any apparent differences between USAGirl the other participants. A selection of her answers is shown in Tables 3 and 4. According to her own estimates, USAGirl had the highest amount of L2 interaction and lowest amount of L1 interaction during the first year after emigration. Section 5 gives further details of the results from the LE survey.

Table 3 Estimated time speaking English to L1 English speakers

Participant	Year 1 (hours/week)	Year 2 (hours/week)
CanGirl 1	14	21
CanGirl 2	4	4
CanBoy	9	2
CanUSABoy	28	10
USAGirl	35	35
AusBoy	24	20
UKBoy	14	7

Table 4: Estimated time speaking Cantonese

Participant	Year 1 (hours/week)	Year 2 (hours/week)
CanGirl 1	8	21
CanGirl 2	35	35
CanBoy	9	28
CanUSABoy	16	35
USAGirl	5	35
AusBoy	8	8
UKBoy	10	21

4.3. Native Speaker Judgements

The results from the 26 respondents indicated a clear trend across all participants, and support a finding of [14]: the intelligibility ratings of T1 utterances was highly accurate (> 90%), and contrasted with lower ratings for accent (M = 6.31, SD = 1.78) and comprehensibility (M = 3.8, SD = 2.13) from the same time point. In most cases, these intelligibility ratings were similar across all time points.

In the T2 responses, there were two notable significant reductions in the ratings for Foreign Accent when compared to T1: USAGirl and CanGirl 2. USA Girl’s One-Way Anova result for all time points was $F(3,410) = 37.994, p < .001$; CanGirl 2’s was $F(3,409) = 10.789, p < .001$. Their Tukey post hoc test results between T1 and T2 are shown in Table 5.

Table 5: Tukey post hoc test results for accent

Participant	T1	T2	p
USAGirl	5.88 ± 1.7	4.63 ± 2.2	< .001
CanGirl 2	4.62 ± 1.9	3.70 ± 1.6	.001

On the other hand, while some of the ratings for Comprehensibility were significantly better between T1 and T2, they did not seem to align with the ratings for Foreign Accent. On the contrary, after USAGirl's One-way Anova suggested significant changes ($F(3,406) = 19.453, p < .001$), a Tukey post hoc test found that she was rated significantly less comprehensible at T2 ($3.97 \pm 2.6, p < .001$) in comparison to T1 (2.83 ± 1.4). Generally, the differences between accent and comprehensibility ratings is another finding that mirrors [14].

In summary, the selected results do not tell the whole story, but do suggest some significant rhythmic changes in the expected direction. These changes were especially prevalent for USAGirl, whose LE survey results suggested limited exposure to Cantonese and extensive exposure to English during her first year after emigration. USAGirl's rhythmic scores also seem to parallel her rating for foreign accent, which was rated as significantly reduced during the first six months after immigration; however, during the same time period, her comprehensibility was rated significantly worse.

5. DISCUSSION

The initial results suggest that L2 English speech rhythm may change in the expected direction after immigration to an English-speaking country, but only under certain conditions. In the present study, an important factor seems to be the population and demography of the city to which one immigrates (see Table 1). USAGirl, the participant with the greatest amount of rhythmic changes in the expected direction, immigrated to the small city of Wausau, Wisconsin, population 40,000. In her interviews, she told us that she encountered not a single Cantonese speaker while living in this city. Additionally, she was living with an English-speaking host family, with whom she communicated very frequently. Her interviews also suggest that she is generally an outgoing person who enjoys conversation a great deal, and that she adapted very quickly to the secondary school she attended. As a result, her ratio of L2 to L1 use was very high, and her communication in Cantonese was limited to online conversation with her friends and family back in Hong Kong.

In contrast, several of the participants had much lower ratios of L2 to L1 communication. In some cases, this was probably because they moved to a city in which Cantonese communication was prevalent. CanGirl 1 and CanBoy both moved to Markham,

Ontario, Canada, where a sizable minority of the population speaks Cantonese. In other cases, it was more likely due to Cantonese-speaking roommates. CanGirl 2, for example, lived in a condominium with her two Cantonese-speaking siblings. What seems clear is that there is a possible relationship between L2 to L1 ratio and the changes in speech rhythm during their first year in an English-speaking country.

In some cases, the willingness of the other participants to communicate in their L2 also seemed to be reflected in ratings by the native English speakers, even when there were no corresponding changes to speech rhythm. During his first year abroad, UKBoy lived with a host family in Cambridge, and enjoyed speaking with his cohabitants regularly. In his second year, however, he moved into a student dormitory at the University of Exeter. In this situation, he was severely alarmed by the dipsomaniacal tendencies of many of his English-speaking cohabitants, and so tended to interact more frequently with Cantonese speakers. (During the second year, Cantonese communication increased among several participants. The wide variety of reasons for these increases are too detailed to include here.) This change in communication patterns is evident in Tables 3 and 4, but his ratings for comprehensibility also seem to mirror the decline in the ratio of L2 to L1. UKBoy's One-way Anova for ratings of comprehensibility suggested significant changes during the observation period ($F(3,407) = 19.563, p < .001$). Surprisingly, the Tukey post hoc tests indicated both significant improvement and deterioration in comprehensibility. After one year abroad, his rating was significantly more comprehensible at T3 ($3.16 \pm 1.8, p < .001$) when compared to T1 (5.22 ± 2.2). After two years, however, he received a significantly worse rating for comprehensibility at T5 ($5.22 \pm 2.6, p < .001$) when compared to T3.

While the small number of participants in the present study may preclude definitive conclusions about L2 speech rhythm development, there are several possible correlations with LE that have been suggested by the results. Presently, we are preparing for mixed model regression analysis that will take all results into account and find the factors that have the largest effects on the L2 Speech Rhythm of participants during the observation period.

6. REFERENCES

- [1] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5, 341-345.
- [2] Boula de Mareüil, P., Vieru-Dimulescu, B. 2006. The contribution of prosody to the perception of foreign accent. *Phonetica* 63, 247-267.
- [3] Dauer, R. M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.

- [4] Dellwo, V. 2006. Rhythm and speech rate: A variation coefficient for ΔC . In: Karnowski, P., Szigeti, I. (Eds.), *Language and language processing*. Peter Lang: Frankfurt am Main, 231-241.
- [5] Deterding, D. 2001. The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, 29, 217-230.
- [6] Flege, J. E. 1988. Factors affecting degree of perceived FA in English sentences. *Journal of the Acoustical Society of America* 84, 70-79.
- [7] Flege, J. E., Munro, M. J., MacKay, I. R. A. 1995. Factors affecting strength of perceived FA in a second language. *Journal of the Acoustical Society of America* 97, 3125-3134.
- [8] Freed, B., Dewey, D., Segalowitz, N., Halter, R. 2004. Language contact profile. *Studies in Second Language Acquisition* 26, 349-356.
- [9] Grabe, E. and Low, E. 2002. Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology 7*, 515-546.
- [10] Kahn, D. 1976. Syllable-based generalizations in English phonology. *Doctoral dissertation*, MIT.
- [11] Kawase, S., Kim, J., Davis, C. 2016. The influence of second language experience on Japanese-accented English rhythm. *Proceedings of Speech Prosody 2016*, 746-750.
- [12] Mok, P., Dellwo, V. 2008. 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.
- [13] Munro, M. 1995. Nonsegmental factors in foreign accent: ratings of filtered speech. *Studies in Second Language Acquisition* 17, 17-34.
- [14] Munro, M., Derwing, T. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45, 73-97.
- [15] Munro, M., Derwing, T. 2008. Segmental acquisition in adult ESL learners: a longitudinal study of vowel production. *Language Learning* 58, 479-502.
- [16] Quene, H., Orr, R. 2014. Long-term convergence of speech rhythm in L1 and L2 English. *Proceedings of Speech Prosody 7*, Dublin, Ireland, 342-345.
- [17] Ramus, F., Nespors, M., Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265-292.
- [18] Riney, T. J., Flege, J. E. 1998. Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition* 20, 213-244.
- [19] Saito, K. 2015. Experience effects on the development of late second language learners' oral proficiency. *Language Learning* 65, 563-595.
- [20] Tajima, K., Port, R., Dalby, J. 1997. Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics* 25, 1-24.
- [21] White, L., Mattys, S. L. 2007. Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35(4), 501-522
- [22] Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., Mattys, S. 2010. How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustical Society of America* 127, 1559-1569.