THIRD LANGUAGE PROSODY: EVIDENCE FROM CANTONESE-ENGLISH-GERMAN TRILINGUALS

Yanjiao Zhu [1], Aoju Chen [2], Stefan Sudhoff[2], and Peggy Mok[1]
[1]The Chinese University of Hong Kong, [2]Utrecht University

ABSTRACT

This study investigates the acquisition of prosody in a third language (L3) by speakers of first language (L1) Cantonese and second language (L2) English, with the goal of understanding the factors characterising L3 prosody. Recordings of 13 trilinguals' read speech in their two non-native languages were compared with those of German and English native speakers in 15 suprasegmental measures. Results show that the L3 German was more syllable-timed, slower, less fluent and contained more IP-final rises than native German was. L3 speech rhythm exhibited transfers effects from both the L1 and the L2 interlanguage, as well as developmental characteristics. In contrast, features such as speaking rate, pause and IP-final rises reflected more of speakers' general proficiency. Therefore, L3 prosodic acquisition should be viewed as a complex dynamicity of various interplaying factors.

Keywords: Third Language, Prosody, Rhythm, German, English, Cantonese

1. INTRODUCTION

Prosodic deviation in non-native speech is notoriously persistent [1] and can substantially contribute to non-native speakers' foreign accent [2]. Nevertheless, most studies on non-native prosody are based on bilinguals, while hardly any study has examined trilingual speakers. Different from L2, imperfection in L3 prosody can be subject to various sources of transfer such as one's L2 [3], a joint force of L1 and L2 [4] [5], or the language that is typologically closer to L3 [6]. An investigation of L3 prosody should therefore consider characteristics of speakers' L1 and L2. The present study on L3 prosody examines the non-native production of passages by Cantonese-English-German trilinguals, whose L3 German is similar to L2 English but different from L1 Cantonese in prosody. In terms of speech rhythm, English and German are "stress-timed" languages that have a substantial contrast between stressed and reduced vowels and allow a complex syllable structure [7], [8], while Cantonese is a typical "syllable-timed" language that does not have stress-related vowel lengthening or reduction, and only has a simple syllable structures [9]. As for intonation, English and German are non-tonal languages that use pitch to express intonational meaning, while Cantonese is a tone language in which pitch is additionally used to distinguish lexical meaning. Given these mismatches, it would be interesting to ask how the L3 prosody will be when Cantonese-English bilinguals learn an additional L3 German. Will they carry the L1 Cantonese prosody to L3, or will the experience of L2 English help them acquire the L3 German, because of the higher similarity between German and English than between German and Cantonese in terms of prosody?

2. METHOD

The study collected read speech samples from trilinguals and native speakers, and then calculated multiple measurements for the quantification of speech prosody. Not only trilinguals' L3 German speech, but also their L2 English speech were analysed because the trilinguals were non-native L2 English speakers and any L2 influence on L3 should be interpreted from trilinguals' L2 interlanguage rather than canonical standard English.

2.1. Participants

Participants of this study were 13 Cantonese-English-German trilingual speakers (F = 7, M = 6, M age = 22.5 years, *SD* = 0.74), 13 English native speakers (F = 8, M = 5, M age = 26.1 years, *SD* = 3.13), and 13 German native speakers (F = 11, Male = 2, M age = 26.4 years, *SD* = 5.44).

The trilinguals were more proficient in L2 English than in L3 German. They started learning L2 English at age 3.0 (*SD* = 0.67), and L3 German at age 18.4 (*SD* = 0.93). Their L2 English proficiency corresponded to IELTS 6.8 (*SD* = 0.91), and their L3 German proficiency was confirmed as pre-intermediate (A2-B1 in Common European Framework). Since the target trilinguals were Hong Kong youngsters who were taught British English in schools but were increasingly affected by American pop culture, most of them tried to approach a British accent while some preferred an American accent. To set the native norm for this group of English learners, the study recruited ten of the native English speakers from the United Kingdom and three from the United States. The German native speakers were all from Germany and spoke Northern Standard German.

2.2. Material and procedure

The materials were two equivalent self-introduction passages of 106 words in English and 104 words in German. The sentence structure and vocabulary were designed to be as simple as possible to elicit natural connected speech from nonproficient speakers.

Participants read the materials in a clear and natural manner at a comfortable speech rate in a sound-attenuated room. The trilingual group read English and German passages in a counterbalanced order, and the two native control groups read in their L1s. Recordings were made with a portable recorder at a 44.1kHz/16bit sampling rate. After the main task, participants completed a language background survey and were paid for the participation.

3. STATISTICAL ANALYSIS AND RESULTS

Recordings were first divided into intonational phrases (IPs) by the first author. For reliability, 30% of the materials were processed by a native English speaker and a native German speaker (the third author) with the interrater agreement rates being 97.6% for English and 99.3% for German. In Praat [10] TextGrids, the IPs were segmented into phonemes and silent pauses using the automatic aligner WebMAUS [11] and hand-corrected by listening to the audio signal and by visual inspection of the waveforms and spectrogram. From the phonemic annotations, vocalic and consonantal intervals were derived following the principles in Grabe and Low [7].

Table 1 lists the measurements taken by this study. Among them, %V, VarcoV, VarcoC, nPVI-V, rPVI-C, speech rate, articulation rate, final word proportion, IP duration and pitch range were measured by each IP, whereas the number of pauses, mean pause duration, the number of IPs, the number of IP-final rises, and the degree of IP-final rise in semitone were measured by each participant.

For group and language comparisons, separate linear mixed-effects models were built on each of the prosodic measure using the "lme4" package in R Studio [14]. All of the models first included Language (English vs. German), Group (Native group vs. Trilingual group) and their interaction as fixed effects. For variables measured per IP, the models included by-subject and by-IP random intercepts and random slopes for language. For variables measured per speaker, the models included by-subject random intercepts and by-subject random slopes for language. The best fit models were selected with backward elimination that removed insignificant predictors based on Maximum Likelihood Chi-squared tests. Post-hoc comparisons with Tukey adjustment were made between native English and L2 English, native German and L3 German, as well as between trilinguals' L2 English and L3 German through the R package "lsmean" [15]. The results were shown in Figure 1 and described below.

Table 1: Summary of prosodic measurements

Measures	Description
1. Speech rhythm	
%V	The sum of vocalic interval duration divided by the total duration of vocalic and consonantal intervals and multiplied by 100. [12]
VarcoV	The standard deviation of vocalic interval duration divided by the mean vocalic interval duration and multiplied by 100. [12], [13]
VarcoC	The standard deviation of consonantal interval duration divided by the mean consonantal interval duration and multiplied by 100. [12], [13]
rPVI-C	The raw pairwise variability index for consonants.[7]
nPVI-V	The normalized pairwise variability for vowels. [7]
2. Speaking rate	
Speech rate	The number of syllables per second.
Articulation rate	The number of syllables per second excluding silent pause time.
IP duration	The duration of an intonational phrase.
3. Intonation	
Pitch range	The distance between the highest point and the lowest point on the pitch contour.
Number of IP-final rises	The number of IPs that ended with a rising pitch. IP-final rises were identified by two annotators and the interrater agreement was 98.5% for English and 97.4% for German.
Degree of IP-final rise	The average distance between the lowest and the highest points on the rising pitch contour for IPs that ended with a rising pitch.
4. Boundary division	
Final word proportion	The ratio of the duration of the last word to the total IP duration multiplied by 100.
Number of pauses	The total number of silent pauses above 100 ms.
Mean pause duration	The total duration of pauses above 100 ms. divided by the number of pauses.
Number of IPs	The total number of intonational phrases

3.1. Speech rhythm

The %V model included Language ($\chi^2(1) = 32.74$, $p < 0.001$) and Language × Group interaction ($\chi^2(2) = 30.33$, $p < 0.001$) as fixed effects. Post-hoc test shows that trilinguals' L3 German production had higher %V than English natives ($p < .001$) but their L2 English and native English did not differ significantly ($p = .81$). Trilinguals' L3 had higher %V than L2 ($p < .001$); the VarcoV model with Group ($\chi^2(1) = 6.5$, $p < .05$) and Language × Group interaction ($\chi^2(2) = 14.86$, $p < 0.001$) shows that L3 German did not differ significantly from native German ($p = .99$), while L2 English had higher VarcoV than native English ($p < .001$); the model on VarcoC with Group ($\chi^2(1) = 6.52$, $p < .05$) and Language × Group interaction ($\chi^2(2) = 14.86$, $p < .001$) indicates that L3 German had higher VarcoC than native German ($p < .001$) and L2 English ($p < .001$); the model on nPVI-V with effect of Group ($\chi^2(1) = 15.84$, $p < .001$) shows that in both English and German, trilinguals produced significantly lower nPVI-V values than natives ($p < .001$); the model on rPVI-C with Group ($\chi^2(1) = 27.53$, $p < .001$) and Language × Group interaction ($\chi^2(1) = 13.96$, $p < .001$) as fixed effects suggests that trilinguals had higher rPVI-C in L3 German than native speakers ($p < .001$), while trilinguals' L2 English did not differ significantly from natives in rPVI-C ($p = .99$).

3.2. Speaking rate

The model on speech rate with Language ($\chi^2(1) = 37.78$, $p < .001$), Group ($\chi^2(1) = 9.72$, $p < .001$) and their interaction ($\chi^2(1) = 12.75$, $p < .001$) as fixed effects shows that trilinguals had lower speech rate than natives in L3 ($p < .001$), but not in L2 ($p = .73$). Speech rate of trilinguals' L3 was lower than that of L2 ($p < .001$); the model on articulation rate with main effects of Language ($\chi^2(1) = 13.88$, $p < .001$) and Group ($\chi^2(1) = 25.14$, $p < .001$) indicates that L2 English and L3 German were articulated more slowly than native English ($p < .01$) and native German ($p < .001$), respectively; the model on IP duration with Language ($\chi^2(1) = 75.49$, $p < .001$) and Language × Group interaction ($\chi^2(1) = 53.51$, $p < .001$) as fixed effects shows that IPs in L3 German took longer time than those in native German ($p < .001$) and in L2 English ($p < .001$). In contrast, the duration of IPs in L2 English was shorter than that in native English ($p < .01$).

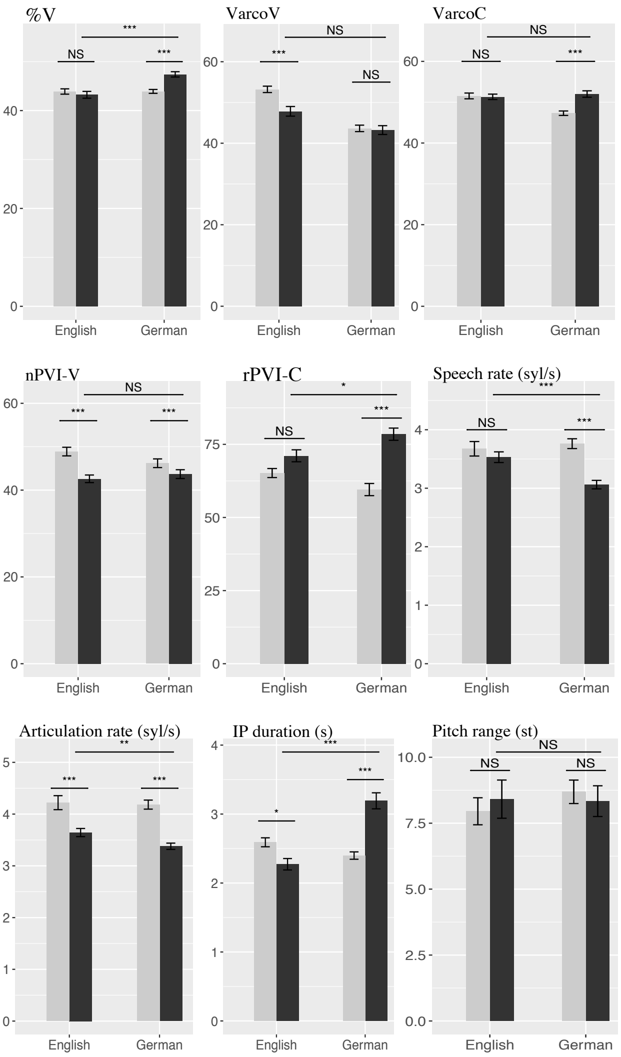
3.3. Intonation

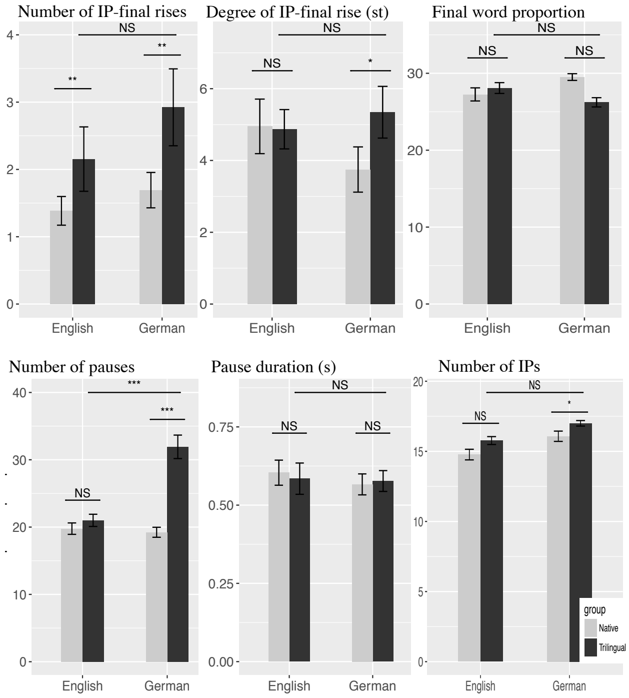
No group or language difference was found in pitch range; the model on the number of IP-final rises with Group ($\chi^2(1) = 8.07$, $p < .01$) as the fixed effect shows that trilinguals implemented more rises at IP boundaries than native speakers in both English and German ($p < .01$); the model on the degree of IP-final rise with Language × Group interaction ($\chi^2(3) = 8.40$, $p < .03$) shows that L3 German had a larger scale of rise at IP boundaries than native German ($p < .05$).

3.4. Boundary division

As an estimation of final lengthening, final word proportion did not differ between native and non-native speech or between German and English speech; the model on number of pauses with Language ($\chi^2(1) = 8.69$, $p < .01$), Group ($\chi^2(1) = 21.63$, $p < .001$), and their interaction ($\chi^2(1) = 25.61$, $p < .001$) as fixed effects shows that more pauses were produced in trilinguals' L3 German than in native German ($p < .01$); no significant difference was found between English and German, or between natives and non-natives in the duration of pauses; the model on the number of IPs produced with fixed effects of Language ($\chi^2(1) = 11.67$, $p < .001$) and Group ($\chi^2(1) = 9.27$, $p < .01$) suggests that L3 German utterances were divided into more IPs than native utterances were ($p < .001$).

Figure 1: Group comparisons of the prosodic measures





NS = not significant.
* $p < .05$; ** $p < .01$; *** $p < .001$.

4. DISCUSSION

In terms of rhythmic classification, both trilinguals' L3 and L2 productions showed higher degrees of syllable-timing than natives. L2 English had lower vocalic variabilities measured in VarcoV and nPVI-V. Similarly, trilinguals' L3 German produced lower vocalic variabilities in nPVI-V and %V than natives. These findings are consistent with the general observation that non-native speech is less stress-timed than native speech in English [16]–[19].

Counterintuitively, increased VarcoC and rPVI-C in L3 German suggest a larger consonantal variability in L3 speech than in native speech. Careful re-examination of the original recording revealed that trilinguals uttered every consonant in consonant clusters one-by-one, while German natives more often coarticulated consonant clusters so as to reduce the durational differences between singleton consonants and consonant clusters. In fact, consonant-based metrics are suggested to be less reliable than vowel equivalents due to connected speech processes which affect consonants more than vowels [20]. Thus, VarcoC and rPVI-C values are viewed as artefacts that do not always truly reflect the reality of speech rhythm.

Three factors are identified to result in L3 speech timing: universal developmental effects, L1 transfer, and L2 interlanguage transfer. Larger degrees of syllable-timing in L3 might be due to the speakers' syllable-timed L1 Cantonese, or to a universal trajectory of developing from syllable-timed to stress-timed rhythm [21]. Transfer from speakers' L2 interlanguage is seen in the high consistency between speakers' L2 and L3 productions in most of the rhythmic measurements except for %V and rPVI-C. Meanwhile, if L3 speech is only affected by L1 transfer and the unmarked syllable-timed rhythm, with pre-intermediate proficiency and limited language experience, the L3 German should have been extremely syllable-timed, which is inconsistent with the actual patterns in this study, hence there should be an additional facilitative effect of L2 on L3 acquisition.

Speaking rate and pause correspond closely with speakers' proficiency levels. Trilinguals' less proficient L3 German was slower and contained more pauses, more IPs than native German, while their more proficient English did not demonstrate such divergences from English natives. The results are consistent with previous studies suggesting that low proficiency speakers tended to speak slowly [22] and pause frequently [23] due to processing difficulties.

Trilinguals' L3 and L2 had more instances of IP-final rises than natives. This could possibly be resulted from the increased number of IPs produced by trilinguals' than natives, as more IPs imply more IP-final continuation rises. Another possibility is related to the expression of paralinguistic intonational meaning. According to Ohala [24], [25], high or rising pitch has a social meaning of lack of confidence, so trilinguals have possibly used a rising pitch to express their uncertainty when speaking in their non-native languages.

5. CONCLUSION

In this study, Cantonese-English-German trilinguals' prosody in German and English was compared to the prosody of selected models of native German and English regarding speech rhythm, speaking rate, pitch pattern, and intonational phrasing. Among these aspects, rhythmic properties see effects of L1 transfer, L2 interlanguage transfer, and developmental characteristics, while fluency aspects such as speaking rate, pause, and continuation rises demonstrate more of developmental traits. Thus, prosodic acquisition by L3 speaker is a result of the interaction between various factors. Such complexity and dynamicity make trilinguals a unique group that deserves more future research.

6. REFERENCES

[1] W. Grosser, “Aspects of intonational L2 acquisition,” in *Current issues in European second language acquisition research*, B. Kettemann and W. Wieden, Eds. Tübingen: Gunter Narr Verlag, 1993, pp. 81 – 94.

[2] J. Sereno, L. Lammers, and A. Jongman, “The relative contribution of segments and intonation to the perception of foreign-accented speech,” *Appl. Psycholinguist.*, vol. 37, no. 2, pp. 303–322, 2014.

[3] C. Bardel and Y. Falk, “The role of the second language in third language acquisition: the case of Germanic syntax,” *Second Lang. Res.*, vol. 23, no. 4, pp. 459–484, 2007.

[4] G. De Angelis, “Multilingual speech production,” in *Third or Additional Language Acquisition*, D. A. Gessia, Ed. Clevedon: Multilingual Matters, 2007, pp. 72–94.

[5] S. Flynn, C. Foley, and I. Vinnitskaya, “The Cumulative-Enhancement Model for language acquisition: Comparing adults’ and children’s patterns of development in first, second and third language acquisition of relative clauses,” *Int. J. Multiling.*, vol. 1, no. 1, pp. 3–16, 2004.

[6] J. Rothman, “Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered,” *Bilingualism*, vol. 18, no. 2, pp. 179–190, 2015.

[7] E. Grabe and L. Low, “Durational variability in speech and the rhythm class hypothesis,” in *Laboratory Phonology 7*, C. Gussenhoven and N. Warner, Eds. New York: Mouton de Gruyter, 2002, pp. 515–546.

[8] R. M. Dauer, “Stress-timing and syllable-timing reanalyzed,” *J. Phon.*, vol. 11, no. 1, pp. 51–62, 1983.

[9] P. K. P. Mok, “On the syllable-timing of Cantonese and Beijing Mandarin,” *Chinese J. Phonetics*, vol. 2, pp. 148–155, 2009.

[10] D. Boersma, P., & Weenink, “Praat: doing phonetics by computer.” 2015.

[11] T. Kisler, R. U. D., and F. Schiel, “Multilingual processing of speech via web services,” *Comput. Speech Lang.*, vol. 45, pp. 326–347, 2017.

[12] F. Ramus, M. Nespor, and J. Mehler, “Correlates of linguistic rhythm in the speech signal,” *Cognition*, vol. 73, no. 3, pp. 265–292, 1999.

[13] V. Dellwo, “Rhythm and Speech Rate: A Variation Coefficient for deltaC Rhythm and Speech Rate: A variation coefficient for C,” *Lang. Lang. Process.*, no. May, pp. 231–241, 2006.

[14] D. Bates, M. Maechler, B. Bolker, and S. Walker, “Fitting linear mixed-effects models using lme4,” *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, 2015.

[15] Lenth Russell V., “Least-Squares Means: The R Package lsmeans,” *J. Stat. Softw.*, vol. 69, no. 1, pp. 1–33, 2016.

[16] P. K. P. Mok and V. Dellwo, “Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English.,” *Speech prosody 2008*, no. January 2016, pp. 423–426, 2008.

[17] H. Jian, “On the Syllable Timing in Taiwan English,” *Speech Prosody 2004*, pp. 2–4, 2004.

[18] A. Tortel, D. Hirst, and A. Université, “Rhythm metrics and the production of English L1 / L2,” in *Proceedings of Speech Prosody 5*, 2010.

[19] U. Gut, “Non-native speech rhythm in German,” in *15th ICPhS Barcelona*, 2003, pp. 2437–2440.

[20] R. A. Knight, “Assessing the temporal reliability of rhythm metrics,” *J. Int. Phon. Assoc.*, vol. 41, no. 3, pp. 271–281, 2011.

[21] A. Li and B. Post, “L2 acquisition of prosodic properties of speech rhythm,” *Stud. Second Lang. Acquis.*, vol. 36, pp. 223–255, 2014.

[22] M. J. Munro and T. M. Derwing, “The Effects of Speaking Rate on Listener Evaluations of Native and Foreign-Accented Speech,” *Lang. Learn.*, vol. 42, no. 2, pp. 159–182, 1998.

[23] A. Riazantseva, “Second language proficiency and pausing: A study of Russian speakers of English,” *Stud. Second Lang. Acquis.*, vol. 23, no. 4, pp. 497–526, 2001.

[24] J. J. Ohala, “Cross-Language Use of Pitch: An Ethological View,” *Phonetica*, vol. 40, pp. 1–18, 1983.

[25] J. J. Ohala, “An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice,” *Phonetica*, vol. 41, pp. 1–16, 1984.