HOW NATURAL IS CHINESE L2 ENGLISH PROSODY?

Jue Yu
Tongji University,
Shanghai, China
erinyu@126.com

Dafydd Gibbon
Universität Bielefeld,
Bielefeld, Germany
gibbon@uni-bielefeld.de

ABSTRACT

Standard varieties of Chinese and English have major typological prosodic differences, which present considerable difficulties for Chinese L2 learners of English at all levels: first, differences in the phonotactic foundations of prosody (syllable and syllable sequence patterns); second, the difference between lexical tone language and lexical stress-accent language; third, timing differences in the prosodic hierarchy, including the timing of grammatical units. We compare Chinese L2 and English native speakers in respect of temporal distribution patterns at the phonetics-phonology interface. The SPPAS and TGA phonetic analysis tools are used. Results indicate clear relations between timing patterns at different L2 proficiency levels and native patterns.

Keywords: Timing, grammar, L2, Chinese, English

1. OBJECTIVES AND BACKGROUND

Standard native varieties of Chinese and English have several major typological prosodic differences, which present considerable difficulties for Chinese L2 learners of English. Native varieties of Chinese differ from native varieties of English: (1) differences in the phonotactic base for prosody (syllable and syllable sequence patterns); (2) the difference between tone language and stress-accent language; (3) timing differences at all levels in the prosodic hierarchy, including differences in prosody-grammar mapping. We interpret L1-L2 performance differences of these types as differences in 'naturalness'; this use of the term is similar to its use in speech synthesis evaluation.

The first two of these prosodic issues are well known and do not figure further in the current study. The third issue, timing in the L2 context has been widely researched (cf. [1], [7]) in disciplines from phonetics through psycholinguistics and discourse analysis to the speech technologies, using a variety of methods. We investigate this third area and introduce new distributional methods for timing analysis which contrast with (1) global duration dispersion measures in traditional phonetics e.g. standard deviation, pairwise variability [10], [9], which used to be regarded as 'rhythm metrics' but only address the rhythm property of near-isochrony, and not the complementary rhythm property of alternation [6]; (2) cognitive and oscillator models of production, perception or storage [2], [8].

The distributional method introduced here in the L2 timing context treats sequences of the pairwise duration differences used in some previous metrics as 'temporal *n*-gram' patterns, by analogy with *n*-gram phonotactic or morphotactic sequences. The basic units (unigrams) are pairwise *shorter*, *longer* and *equal* duration relations between adjacent syllables; digrams are, for example, *shorter-longer*, *longer-shorter*, *shorter-shorter*, *longer-longer*, etc., sequences. These temporal *n*-grams are investigated for two properties: (1) temporal pattern distribution as evidence for language differences; (2) the relation between temporal patterns and grammatical units.

The primary objective is to obtain new findings on strategies underlying temporal patterning in Chinese L2 English. The practical objective is to provide criteria for creating general guidelines on timing for L2 teaching, diagnosis, self-monitoring and testing.

2. METHODS AND DATA

2.1. Methods

Speech recordings of Chinese L2 speakers and English native speakers were automatically annotated using the SPPAS tool [3], manually post-edited and stored in standard Praat long format [4]. The data time-stamps in the annotation files were further investigated for temporal properties and temporal structures using an online tool which provides heuristics for automatically investigating

temporal properties and distributions in annotated data, including global, local and structural (sequential and hierarchical) timing properties, (TGA online tool: cf. [5]). English proficiency levels of the L2 speakers were tested, and temporal properties of the speech of all speakers were compared with their proficiency levels.

2.2. Data

The reading aloud genre is used because it is a standard feature of EFL teaching material. Data are from the AESOP-CASS Chinese EFL learner corpus [14], which contains both speech and speaker proficiency evaluations, with readings of the well-known and widely-used IPA standard text, Aesop's fable *The North Wind and the Sun*.

Recordings of 10 male and 10 female adult Chinese EFL learners (mostly college students) and of a baseline set of 6 English native speakers were used. The length of readings depends on proficiency and phonostylistic characteristics of the reader, and averages just over 1 minute.

3. RESULTS AND DISCUSSION

3.1. Proficiency evaluation

The pre-evaluations were done by 4 Chinese English teachers and 4 native speaker teachers:
(1) The 20 L2 speakers are graded on a scale of 5 by general impression (*excellent*, *good*, *average*, *poor*, *unintelligible*).
(2) Speakers are further evaluated on a 5 point scale by 6 detailed criteria, performance with segments, intonation, stress, rhythm etc. There is no necessary correlation between quantitive evaluation and general impression, as there may be other general factors than these detailed criteria.
(3) In the final score, quantitative evaluation counts 60% and general impression 40%.
(4) Based on the final score, the 20 L2 speakers are classified into 3 subgroups (*advanced*, *medium* *poor*).
(5) To test validity, the final evaluation results are compared between and within Chinese English and native English teacher groups Both are highly correlated.

The Chinese English teachers tend to give higher scores for prosodic criteria than English teachers, indicating either shared L1 and L2 features, or less focus on prosody in teaching.

Table 1 shows the proficiency of the female learners to be higher than that of the males, $F(1,18) = 4.73$, $p<0.05$. Language proficiency does not correlate with years of learning, $r^2 = 0.214$, $p>0.05$.

Table 1: Chinese learners' proficiency in English.

		English proficiency		
		advanced	**medium**	**poor**
Gender	**male**	0	5	5
	female	3	4	3
Learning experience (years)	**<10**	1	0	2
	10	1	4	1
	>10	1	5	5

3.2. Global measures: rate and variability

Speech rates of the Chinese L2 learners and the native speakers differ. Most of the learners' speech rate is 2-4 syllables per second, much lower than the natives, above 5 syllables per second. This overall difference (males and females) is significant, $F(1, 28) = 29.693$, $p<0.01$. Moreover, there is a significant correlation between speech rate and language proficiency, $r^2 = 0.575$, $p<0.01$.

Table 2 shows mean pairwise syllable duration variability and mean syllable per second rate per group. Because of some extreme outliers in learner syllable durations, speech rates are calculated as inverses of median syllable durations. There were no male advanced learners.

Table 2: Summary of mean variability and mean syllable rate for female (F) and male (M) reader groups.

	Ch L2 poor	**Ch L2 medium**	**Ch L2 advanced**	**Eng native**
F: *nPVI*	56	62	73	73
F: syll rate	4.2	4.7	6.3	5.3
M: *nPVI*	59	65	-	73
M: syll rate	4.3	4.9	-	4.8

The variability of both male and female Chinese learner groups is clearly a function of proficiency level, possibly a declining effect of substrate Chinese L1. Males and females are comparable. The male native speakers have a rather low syllable rate; the female native speakers had a lower syllable rate than the advanced L2 speakers. The syllable rate values are not clearly related to proficiency.

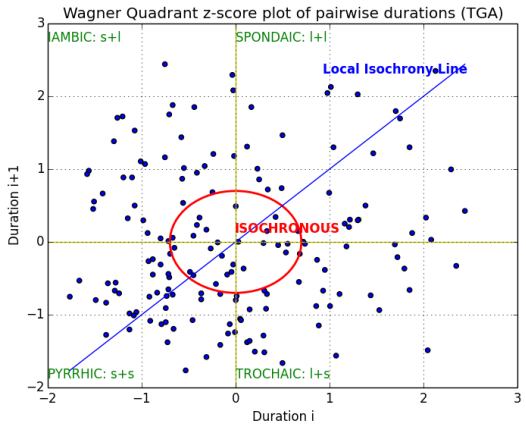
3.3. Temporal dispersion: Wagner Quadrants

A more informative technique than the older global metrics is the Wagner Quadrants method [11], shown for 3 speakers in Figure 1, Figure 2 and Figure 3. The scatter plots are of *z*-scores of duration pairs. The relations are labelled in the quadrants of the plots around zero as *s+s*: *shorter-shorter*; *s+l*: *shorter-longer*; *l+s*: *longer-shorter*; *l+l*: *longer-longer*. We hypothesise that the poor Chinese reader (Figure 1) will show a Wagner Quadrants distribution which is less similar than that of the advanced Chinese reader (Figure 2) to the English native speaker distribution (Figure 3).

Visual inspection of the sample figures show that this tendency is indeed present: the low proficiency speaker shows a random distribution of values through the four quadrants.

The English native speaker, on the other hand, tends to cluster values in the *shorter-longer* and *longer-shorter* quadrants, reflecting the tendency of English to anisochronous syllable patterning, in which stressed or strong syllables alternating with unstressed or weak syllables tend to be longer and shorter, respectively. There are many *shorter-shorter* quadrant cases which reflect non-binary patterns of a longer syllable followed by more than one shorter syllables. The distribution is approximately "L-shaped". This tendency is also evident in the case of the advanced Chinese speaker. The shapes of these patterns are as predicted by our initial hypothesis.

Figure 1: Chinese L2 English, poor, female.



The axis lengths of the Wagner Quadrants are normalised to actual z-scores. The range of the poor speaker is higher than that of the advanced speaker, perhaps showing that randomness of the inexperienced reader of English overrides the syllable near-isochrony of the native language.

Figure 2: Chinese L2 English, advanced, female.

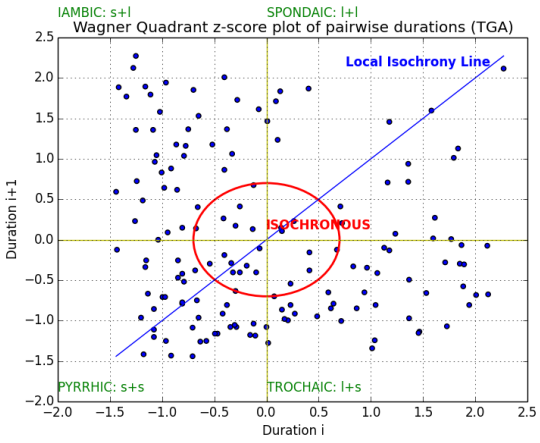
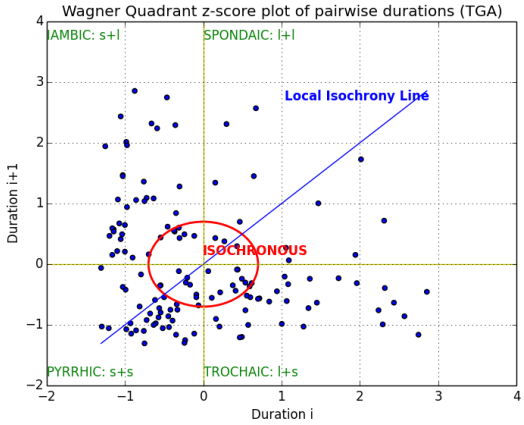


Figure 3: Wagner Quadrant for female native speaker (USA).



Quantitative analysis of each of the four quadrants for all speakers is planned for future studies in addition to visual inspection.

3.4. Temporal *n*-grams

Traditional methods for quantifying the timing of speech provide single global indices of near-isochrony based on the dispersion of duration differences, thereby factoring out the structure of timing patterns. The temporal *n*-gram method was developed solely to examine the sequential structure of binary alternation patterns in syllable sequences. The other main defining factor of rhythm, relative isochrony, covered by the global metrics discussed above, is factored out by treating the duration difference *n*-grams as categorical; at this stage, ternary and other rhythm patterns are not dealt with.

Syllable duration alternation patterns are predicted to differ between Chinese and English. Single alternations between adjacent syllables are too simple: at least three beats, with four alternations, are needed to constitute a recognisable syllabic rhythm, for example: *dum-di-dum-di-dum* (three beats, four alternations, five syllables). For this reason, sequences of four and five alternations (temporal quadgrams and quingrams) were automatically extracted with the TGA software. Percentages for purely alternating quadgrams and quingrams at the top two *n*-gram frequency ranks were calculated for each speaker (Table 3). Purely alternating *n*-grams are represented /\/\, \/\/, /\/\/, \/\/\, where / and \ stand for *longer-shorter* and *shorter-longer* duration relations (top left and bottom right quadrants).

Table 3: Temporal quadgram and quingram alternation.

	Chinese poor	**Chinese medium**	**Chinese advanced**	**English native**
F: 4-gram	4.5	8.5	9.5	13.1
F: 5-gram	1.7	4.3	2.3	8.5
M: 4-gram	5.1	5.8	-	12.2
M: 5-gram	2.6	2.4	-	9.8

The number of strict quadgram alternations appears as a function of proficiency. Quingrams show no obvious tendency. The non-natives have far fewer strictly alternating sequences than the English native speakers. In contrast, temporal digrams and trigrams did not show any differences between learners and natives. Male and female speakers show similar tendencies. Future work will need to take a wide range of duration difference limens for *n*-gram formation into account, as well as ternary and other alternation patterns.

3.5. Time-tree and grammar correspondences

The digram-based Time Tree method [6] was used to create and recursively combine syllables into hierarchies based on *shorter-longer* ('iambic') duration difference digrams (iambic trees show closer timing-grammar relations than *longer-shorter* 'trochaic' trees in Mandarin [12], [13]), using varying difference limens 0...100 ms. Within this range, the generated Time Tree constituents have a better agreement with linguistic units, and the number of basic Time Tree constituents remains stable. Time-tree/grammar-tree match percentages between Chinese L2 learners and native speakers were compared in respect of tree-match and proficiency. The following example from one speaker shows an iambic Time Tree (in bracket notation) of the English utterance "*then the north wind blew as hard as he could"*, and a grammatical bracketing of the utterance. For duration difference limen 10ms, TGA generated a local Time Tree: (((*the* (*north*)) (*wind* (*blew* ((*as hard*) (*as* (*he could*)))))) PAUSE). The grammatical bracketing is (*then* ((*the* (*north wind*)) (*blew* ((*as hard*) (*as* (*he could*)))))). The lowest level pairs (*as hard*), (*he could*) match grammatical constituents, but not (*north wind*). The percentage of agreement illustrated by this example is thus 2 out of 3, i.e. 67%. Results for all speakers are shown in Table 4; matchings and proficiency correlate, $r^2 = 0.955$, $p<0.01$.

Table 4: Average time-tree/grammar correspondences.

	Chinese poor	**Chinese medium**	**Chinese-advanced**	**English native**
female	65.8	72.4	75.4	77
male	67.08	69.2	-	76.95

4. CONCLUSION

Each method used in this study yields independent results placing L2 learners on 'naturalness' scales in relation to L1 speakers. Most of the results showed performance as a function of proficiency.

The persistence of Chinese substrate timing patterns [12] even in those readers who were pre-evaluated as more highly proficient indicates that the problem of temporal patterning in Chinese L2 pronunciation is not only difficult, but perhaps not focussed enough in L2 learning processes. This may be confirmed by the higher prosody ratings given by Chinese teachers than by native speakers.

We anticipate applications of the methods we have demonstrated in identifying prosodic problems in the pronunciation of Chinese leareners. The encouraging results justify further work on larger datasets and wider parameter ranges (e.g. difference limens). Further integration of SPPAS-type automatic segmentation and labelling techniques with timing pattern measures is planned, with the ultimate goal of providing an automatic indicator of timing proficiency for diagnostic, self-monitoring and testing purposes in L2 teaching. The efficacy of such a device is likely to be higher where L1 and L2 are typologically very different in terms of their timing patterns.

5. REFERENCES

[1] Arvaniti, A. 2009. The usefulness of metrics in the quantification of speech rhythm. *Phonetica* 66, 46-63.

[2] Barbosa, P. 2009. Measuring speech rhythm variation in an oscillator-based framework. *Proc. Interspeech*, Brighton, UK, 1527-1530.

[3] Bigi, B., Hirst, D. 2012. SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody. *Proc. Speech Prosody*, Shanghai, 1-4.

[4] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5, 9/10, 341-345.

[5] Gibbon, D. 2013. TGA: a web tool for Time Group Analysis. In: Hirst, D. Bigi, B., eds. *Proc. Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop*, Aix en Provence, 66-69.

[6] Gibbon, D. 2006. Time Types and Time Trees: Prosodic Mining and Alignment of Temporally Annotated Data. In: Sudhoff, S., Lenertová, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N. and Schließer, J., eds. *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter, 281–209.

[7] Gut, U. 2012. Rhythm in L2 speech. In: Gibbon, D., Hirst, D., Campbell, N., eds. *Rhythm, Melody and Harmony in Speech*. *Speech and Language Technology: Studies in Honour of Wiktor Jassem* 14/15, 83–94.

[8] Inden, B., Malisz, Z., Wagner, P., Wachsmuth, I. 2012. Rapid entrainment to spontaneous speech: A comparison of oscillator models. *Proc. 34th Annual Conference of the Cognitive Science Society*, Sapporo, Japan. Austin, TX: Cognitive Science Society, 1721-1726.

[9] Low, E. L., Grabe, E., Nolan, F. 2001. Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech* 43 (4), 377-401.

[10] Roach, P. 1982. On the distinction between 'stress-timed' and 'syllable-timed' languages. In: Crystal, D., ed. *Linguistic Controversies: Essays in Linguistic Theory and Practice*. London: Edward Arnold, 73-79.

[11] Wagner, P. 2006. Visualizing levels of rhythmic organisation. *Proc. ICPhS* 2007, 1113-1116.

[12] Yu, J., Gibbon, D. 2012. Criteria for database and tool design for speech timing analysis with special reference to Mandarin. *Proc. Oriental COCOSDA*, Macau, 41-46.

[13] Yu, J., Gibbon, D. Klessa, K. Computational annotation-mining of syllable durations in speech varieties, *Proc. Speech prosody 7*, Dublin, 443-447.

[14] Yuan J., Wang, M., Zhai, H., Li, A. 2011. Construction of Speech Corpus of AESOP-SD. *Proc. Oriental COCOSDA*, Hsinchu, Taiwan.