# A Rhythmic Analysis on Chinese EFL Speech

*Aijun LI, Zhigang YIN*
Institute of Linguistics,
Chinese Academy of Social Sciences
liaj@cass.org.cn

*Yiqing ZU*
MOTOROLA Research Center China
Yiqing.Zu@motorola.com

## Abstract

This paper, based on a phonetic experiment, depicts a contrastive study on the rhythmic pattern of Chinese learners of English as a foreign language (CL2) as compared with that of the native speakers of both standard British and American English (EL1) in their respective pitch accent distribution patterns, prosodic structures and duration patterns.

## 1. Introduction

Phonetic analyses of Chinese EFL (English as foreign language) learners' spoken English are mostly qualitative. However little has been studied to unveil their distinct acoustic features on segmental and supra-segmental aspects, which are a result under the influence from their mother tongue.

Chinese is a tone language and also called a syllabic language. Much research has been carried out on its prosody, rhythm and stress, e.g. Duanmu San offers a phonological analysis of rhythm and stress of Chinese [1, 2]. There are even more acoustic analyses on Chinese prosody, rhythmic structure and stress. The interplay of the F0 and duration to the stressed syllables is also investigated [ 4,9] .

English is a stressed language featuring word stress and pitch accent. Here pitch accent is a post-lexical stress, which is realized by its relative prominence in pitch, duration and intensity. There have been many research endeavors on English intonation, stress and rhythm [3,5, 6, 7, 8].

There is much research on CL2 learners' segmental acquisition, such as Joanna Radwanska's investigation of the insertion or deletion of the final plosives and his interpretation using the OT theory [10]. However, because there are so many dialects in China, people from different dialectal areas will speak English with different dialectal features, which complicates the segmental acquisition. If a certain phoneme in the target language is absent in the speaker's mother tongue, she/he is likely to substitute it with a similar one in his/her mother tongue. For example, many native Cantonese speakers cannot tell the difference between [n] and [l], usually pronouncing 'night' as 'light'. Another example is that there are only voiceless instead of voiced vs. voiceless stops in Beijing dialect, which is why they usually pronounce 'bus' as [pʌs]. This mistake is rare in the Wu dialect, which has the voiced vs. voiceless opposition in stops.

Researchers of second language (L2) acquisition have undertaken many studies on prosodic structure and intonation acquisition [5，11，12，13，14], but less research is based on the perspective of acoustic phonetics. We know that CL2 learners will bring patterns of Chinese stress, rhythm and intonation to their spoken English, such as a flatter F0 range, pitch accent misplacement and quite different realization of pitch accent. In [15] the author carries out some phonetic analysis on Chinese ELF learners' intonation from the perspective of language teaching. The experiments in this study focus on contrasting the rhythmic patterns between CL2 learners and EL1 speakers, including patterns of stress distribution, prosodic structure and duration structure.

## 2. Corpus and annotations

The standard English materials for this study is from the Motorola China Research Center's speech synthesis corpus, consisting of 200 utterances selected for each of the 4 EL1 speakers who are TV or radio announcers, with two from the US (1male and 1 female) and two from the UK (1male and 1 female). The utterances cover as many sentence patterns as possible, including declaratives, exclamatives and interrogatives. Declarative vs. interrogative pairs by some of the speakers are also recorded for comparison. Average sentence length is 9.6 words.

Table 1: *The proficiency level of CL2 as evaluated by a Chinese and an English experts*

| Info. | Chinese expert's rating | | English Expert's rating | |
|---|---|---|---|---|
| TZD | A | fairly good pronunciation and intonation | A | a fairly neutral accent, a competent job of overall intonation patterns. |
| LB | B | fairly good intonation | A | |
| WX | B | fairly clear performance with fairly good rhythm and intonation | A | |
| JY | C | pronunciation not clear in some part | B | problems of pronunciation, stress, and tone patterns such as to impede comprehension at times. |
| WL | D | vague pronunciation and broken intonation | B | |
| ZL | D | inaccurate pronunciation and problematic intonation | B | |

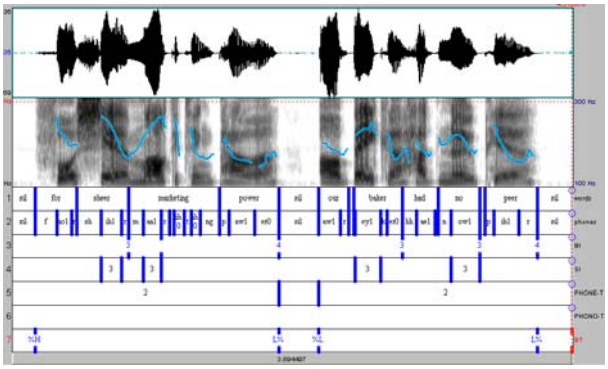

Figure 1: *An annotated utterance "For sheer marketing power, our baker had no peer", in which BI stands for the break index tier, 3 for intermediate phrase boundary (pitch accent ID as in EVIA[16]), 4 for intonational phrase boundary; SI for pitch Accent stress tier annotated on the prominent vowels, and BT for the boundary tone tier.*

6 Chinese graduate students (3 male and 3 female) speaking Standard Chinese are recruited and paid for the recording of 250 English utterances from each of them (with 50 more

interrogative vs. declarative pairs). The recordings include two channel signals: speech and laryngeal (Lx) signals at a sampling rate of 16KHz. Altogether 250*6+200*4=2300 utterances are obtained.

Then the proficiency level of the CL2 speakers above are respectively evaluated by a Chinese and an English experts. The results scaling from A to D are shown in Table 1.

Utterances of EL1 speakers in the synthesis corpus are automatically segmented into words and phones, as shown in Figure 1 while the CL2 learners' utterances are only segmented into words. Prosodic annotations include tags of pitch accent, boundary tones, and intermediate phrase and intonational phrase boundaries.

## 3. Prosodic boundary

### 3.1. Intermediate phrase and intonational phrase

Table 2 gives the statistic results on prosodic boundaries: the average numbers of intermediate phrase (ID) and intonational phrase (IP) boundaries for each utterance, and the average number of ID in each IP. We classify them into two classes respectively (marked by bold italic font vs. norm font; the following parts will follow this marker).

Table 2: *The average numbers of intermediate phrase (ID) and intonational phrase (IP) for each utterance, and the number of ID in each IP*

| Speakers | ID | | IP | | |
|---|---|---|---|---|---|
| | Ave per utter. | Sd. | Ave per utter. | Sd. | ID No. in IP |
| UKM | ***6.303*** | 4.044 | 2.505 | 1.573 | ***2.516*** |
| UKF | 5.391 | 3.149 | ***2.030*** | 1.022 | 2.655 |
| USM | 5.577 | 4.003 | 2.447 | 1.667 | ***2.279*** |
| USF | ***6.613*** | 4.046 | 2.508 | 1.452 | 2.636 |
| LB | 5.601 | 3.691 | ***2.020*** | 1.138 | 2.772 |
| WL | ***7.15*** | 5.172 | 2.689 | 1.864 | 2.658 |
| TZD | 5.545 | 3.863 | 2.319 | 1.552 | ***2.391*** |
| WX | 6.195 | 4.391 | 2.426 | 1.568 | 2.553 |
| JY | 5.848 | 4.445 | 2.159 | 1.472 | 2.708 |
| ZL | 5.896 | 4.035 | ***2.406*** | 1.602 | ***2.450*** |

The average numbers of prosodic phrase in an utterance for CL2 and EL1 are ranging from 5.3 to7.2 for ID, and from 2 to 2.7 for IP. We find that the average numbers of two level prosodic boundaries can not be classified into two distinctive classes, the average numbers of intermediate phrases in an intonation phrase are from 2.27 to 2.77, which can't be distinguished into two groups either.

### 3.2. Boundary agreement

Agreement rates of two level boundaries are shown in table 3, data of first 4 columns are the agreement rates among EL1, data of the last 6 columns show the agreement rates between CL2 and EL1. Most of the boundary agreement rate among EL1 are higher than CL2 learners', but it is not related to the English proficiency level directly, for instance, TZD has lower agreement rate than JY, but his English proficiency level is higher than JY.

Table 3: *Boundary agreement rate*

| | UKM | UKF | USM | USF | LB | WL | TZD | WX | JY | ZL |
|---|---|---|---|---|---|---|---|---|---|---|
| ID | 0.210 | 0.201 | 0.203 | 0.208 | 0.218 | ***0.170*** | 0.192 | ***0.151*** | 0.211 | ***0.159*** |
| IP | 0.301 | 0.313 | 0.270 | 0.279 | 0.285 | ***0.217*** | ***0.257*** | ***0.252*** | 0.272 | ***0.253*** |

From boundary analysis, we can't tell any detailed discrepancy on prosodic structure between CL2 and EL1 except EL2 have more agreeable prosodic boundaries.

## 4. Duration and word number of prosodic units

### 4.1. Duration of words

The word lengths of the 10 speakers are between 300ms to 390 ms (as shown in table 4). The relative speech rates(RS) of the 10 speakers are calculated and shown in the final column by setting the average length 340ms as the based parameter.

Table 4: *Average duration of words of the 4 EL1 learners and 6 CL2 speakers (S) and relative speech rate(RS)*

| Speakers | Ave.duration of words | Dev. | RS |
|---|---|---|---|
| UKM | 0.38 | 0.29 | 1.16 |
| UKF | 0.30 | 0.19 | 0.9 |
| USM | 0.32 | 0.20 | 0.99 |
| USF | 0.35 | 0.21 | 0.98 |
| LB | 0.36 | 0.22 | 1.06 |
| WL | 0.32 | 0.17 | 0.98 |
| TZD | 0.34 | 0.22 | 0.99 |
| WX | 0.39 | 0.23 | 1.08 |
| JY | 0.31 | 0.18 | 0.91 |
| ZL | 0.33 | 0.19 | 0.96 |

### 4.2. The average word number and duration of intermediate phrase

Table 5 shows the average word number in an ID, duration of intermediate phrase and the duration by removing speech rate (/RS?). The average word number of IPs is between 1.3 to 2.05, with EL1 learners (average 1.857) having more lexical words than CL2 speakers (average 1.613). But word number doesn't seem to be related with the English proficiency level, e.g. IPs of CL2 speaker WX who has a higher English level contain fewer lexical words.

As for the ID duration, both CL2 and EL1 have no distinctive classifications ranging from 0.6s to 0.86s by removing the speech rate (/RS). The ID length doesn't seem to be related with the English proficiency level as well.

Table 5: *Statistic data on intermediate phrases*

| Speakers | Ave. Word No. | Ave. dur. (S) | Std.(s) | Dur. (/RS) | Std. (/RS) |
|---|---|---|---|---|---|
| UKM | 1.700 | 0.785 | 0.371 | ***0.677*** | 0.320 |
| UKF | ***2.055*** | 0.770 | 0.322 | 0.856 | 0.358 |
| USM | ***1.993*** | 0.761 | 0.806 | 0.769 | 0.814 |
| USF | 1.680 | 0.730 | 0.310 | 0.745 | 0.317 |
| LB | 1.716 | 0.738 | 0.307 | ***0.696*** | 0.290 |
| WL | 1.351 | ***0.587*** | 0.291 | ***0.599*** | 0.296 |
| TZD | 1.773 | 0.776 | 0.321 | 0.784 | 0.324 |
| WX | 1.565 | 0.797 | 0.295 | 0.738 | 0.274 |
| JY | 1.641 | ***0.667*** | 0.326 | 0.733 | 0.358 |
| ZL | 1.634 | ***0.678*** | 0.272 | ***0.706*** | 0.284 |

The number of lexical words is not related to speech rate ($r^2$=0.003 for EL1, $r^2$=0.448 for CL2). In other words, the determination of IP boundary of both CL2 learners and EL1

speakers are not affected by speech rate，which demonstrates that the intermediate phrases contain rather steady syntactic components.

The duration of ID is no related to the speech rate for both CL2 and EL1 ($r^2$=0.309; 0.217). This tells us that their rhythmic patterns are quite stable at intermediate phrase level that the duration of pitch accent ID isn't affected by speech rate .

### 4.3. The average word number and duration of intonational phrase

Table 6 shows the average word number, duration of intonational phrase and the duration by removing speech rate. The average lexical number in an IP is between 3.5 to 5.5, with EL1's IPs having more lexical words (average 4.638) than those of CL2 (average 4.126). But it doesn't seem to be related to the English proficiency level, e.g. IPs of CL2 speaker WX who has a higher English level but contain fewer lexical words.

As for the IP duration, CL2 and EL1 have no distinctive classifications ranging from 1.6 to 2.3s by removing the speech rate(/RS). The IP length doesn't seem to be related with the English proficiency level as well.

Table 6: *Statistic data on intonational phrases*

| Speakers | Ave. Word No. | Ave. dur.(S) | Std.(S) | Dur. (/RS) | Std. (/RS) |
|---|---|---|---|---|---|
| UKM | 4.207 | 1.977 | 0.884 | 1.704 | 0.762 |
| UKF | ***5.433*** | 2.046 | 0.819 | ***2.274*** | 0.910 |
| USM | ***4.508*** | ***1.736*** | 1.181 | 1.753 | 1.193 |
| USF | 4.404 | 1.927 | 0.791 | ***1.966*** | 0.807 |
| LB | ***4.700*** | 2.046 | 0.983 | ***1.930*** | 0.928 |
| WL | 3.553 | 1.564 | 0.642 | 1.596 | 0.6560 |
| TZD | 4.108 | ***1.857*** | 0.820 | 1.875 | 0.828 |
| WX | 3.979 | 2.036 | 0.885 | 1.885 | 0.819 |
| JY | 4.433 | 1.807 | 0.715 | ***1.986*** | 0.785 |
| ZL | 3.982 | ***1.662*** | 0.736 | 1.731 | 0.767 |

The number of lexical words in an IP is not related to speech rate of CL2s ($r^2$=0.01), but it is related to the speech rate of EL1s ($r^2$=0.63). This tells us that the determination of CL2s' IP boundary is not affected by speech rate whereas EL1s' speech rate is in positive proportion of the length of lexical words in his IPs. It contains rather stable syntactic component for CL2, but it contains more syntactic components for EL1 at faster speech rate than at lower speech rate.

The average duration of IP is not affected by speech rate ($r^2$=0.43 for CL2; $r^2$=.001 for EL1). This tells us that their rhythmic patterns are both stable on the IP level.

Sections 4.2 and 4.3 indicate that prosodic structures of CL2s and EL1s mainly differ at IP level. As for rhythmic pattern, both CL2s and EL1s show stable ID and IP duration patterns.

## 5. Duration of silent pause

In the case of CL2 learners (as shown in table 7), the pause after intermediate phrase shows no significant difference from those of EL1 speakers while the pause after the intonational phrase is mostly longer than those of EL1s, indicating that most CL2s except WX use the same pause strategies after ID as EL1s, but most CL2 (except UKM and ZL) use different pause strategies after IP from EL1 speakers. In other words, they show different acoustic cues to signal intonational boundary.

The statistic data show that CL2s always use longer pause to signal IP boundary, which is exactly how native Chinese speakers display rhythms in their mother language.

Table 7: *Silent pause (SP) after prosodic boundaries*

| Speakers | SP after ID (s) | SP after IP (s) | SP after ID (/RS) | SP after IP (/RS) |
|---|---|---|---|---|
| UKM | 0.058 | ***0.650*** | 0.050 | ***0.561*** |
| UKF | 0.038 | 0.401 | 0.042 | 0.446 |
| USM | 0.041 | 0.271 | 0.042 | 0.273 |
| USF | 0.041 | 0.321 | 0.042 | 0.327 |
| LB | 0.044 | ***0.623*** | 0.042 | ***0.587*** |
| WL | 0.048 | ***0.563*** | 0.049 | ***0.574*** |
| TZD | 0.053 | ***0.701*** | 0.054 | ***0.708*** |
| WX | ***0.085*** | ***0.617*** | ***0.078*** | ***0.571*** |
| JY | 0.039 | 0.481 | 0.043 | ***0.528*** |
| ZL | 0.052 | 0.460 | 0.054 | 0.479 |

## 6. Analysis on pitch accent

### 6.1. Duration of pitch accent

Table 8: *Duration of the vowels bearing pitch accent*

| Speakers | Ave. dur | Std. | Ave. dur(/RS) | Std.(/RS) |
|---|---|---|---|---|
| UKM | 0.139 | 0.064 | 0.120 | 0.055 |
| UKF | 0.124 | 0.054 | 0.138 | 0.061 |
| USM | 0.131 | 0.056 | 0.132 | 0.057 |
| USF | 0.141 | 0.057 | 0.144 | 0.059 |
| LB | ***0.102*** | 0.042 | ***0.096*** | 0.040 |
| WL | ***0.057*** | 0.023 | ***0.058*** | 0.024 |
| TZD | 0.139 | 0.059 | 0.141 | 0.060 |
| WX | 0.167 | 0.066 | 0.155 | 0.062 |
| JY | ***0.075*** | 0.034 | ***0.082*** | 0.038 |
| ZL | ***0.090*** | 0.036 | ***0.093*** | 0.038 |

As for the duration of pitch accent, EL1 vowels bearing pitch accent have longer duration than most CL2 ones as shown in table 8. CL2s' proficiency level seems to be related to the closeness of their pitch accent realization as compared to that of EL1s. For example, CL2 learners TZD and WX display similar duration of vowels bearing pitch accent to that of EL1 speakers.

### 6.2. Pitch accent number vs. the utterance length

As shown in Figure 2, the relationship between the number of pitch accents in an utterance and the length of this utterance are also different in the cases of CL2s and EL1s. If an utterance contains less than 12 words long, both CL2 and EL1 cases show a similar correlation between the two parameters.

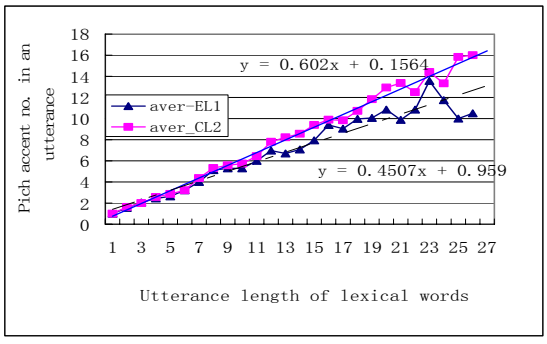


Figure 2: *The correlation of pitch accent number with the utterance's length*

But with the increase of the utterance length, the increased

number of pitch accents is smaller for EL1 than for CL2 . This indicates that CL2 try to segment the utterances into more IPs than EL1 and that their respective prosodic structures differ more significantly in longer utterances. The data of CL2 reveals that those with a higher level of English proficiency display a more similar pattern as that of EL1, as in the cases of TZD, LB and WX.

### 6.3. Distribution of pitch accent

With a comparison of the placements of pitch accents in the same 200 utterances produced by each speaker, we consider the two pitch accents placed on the same word as the same placements. Table 9 describes the placement agreement rates among EL1s and between EL1s and CL2s, with the last column as the average agreement rate.

Table 9 also shows that the placements of pitch accents are highly agreeable among EL1 speakers (>0.6), with a much higher agreement rate than that between CL2 learners and EL1 speakers. The highest rate of the latter is 0.514 for speaker TZD, followed by speakers WX and JY. This indicates that the more similar to EL1 speakers in the placement of pitch accent for CL2, the higher their English proficiency level.

Table 9: *The placement agreement rates of pitch accents*

| SPK | UKM | UKF | USM | USF | AVE. |
|---|---|---|---|---|---|
| UKM | | 0.7 | 0. 641 | 0. 715 | **0.685** |
| UKF | 0.7 | | 0. 577 | 0. 638 | **0.638** |
| USM | 0.641 | 0.577 | | 0. 585 | **0.601** |
| USF | 0.715 | 0.638 | 0.585 | | **0.646** |
| LB | 0.453 | 0.434 | 0.4 | 0.47 | *0.439* |
| WL | 0.479 | 0.411 | 0.396 | 0.484 | *0.442* |
| TZD | 0.552 | 0.51 | 0.455 | 0.541 | ***0.514*** |
| WX | 0.523 | 0.479 | 0.434 | 0.52 | ***0.489*** |
| JY | 0.497 | 0.462 | 0.423 | 0.508 | *0.472* |
| ZL | 0.429 | 0.388 | 0.369 | 0.437 | *0.406* |

## 7. Concluding remarks

The preliminary conclusions are:

(1) The occurrence times of two level prosodic boundaries can not be classified into two classes, the average number of intermediate phrases in one intonation phrase can't be distinguished into two parts either. Boundary agreement rate among EL1 is higher than CL2 learners', but it is not related to the English proficiency level directly. This implies that the general prosodic structure displays no distinctive discrepancy.

(2) CL2's English proficiency level is correlated to his/her performance of pitch accent placements. Those CL2s with more similar pitch accent positions to those of EL1s are perceived as having a higher level of English proficiency.

(3) CL2s' English proficiency is correlated to his/her way to acoustically manifest pitch accent. CL2's vowels bearing pitch accent are significantly shorter than those of EL1's. The vowels bearing the pitch accent produced by those CL2s with a higher English level are significantly longer than those produced by CL2s with a lower level, close to those by EF1s.

(4) The analysis of durations of prosodic units and the number of lexical words in the prosodic units demonstrates that the respective EL1 and CL2 prosodic structures differ at the level of intonational phrase (IP). As for the rhythmic pattern, both CL2 and EL1 have stable intermediate and intonational phrases.

(5) An analysis of the pause after prosodic units reflects that CL2s have longer pause after IP group than EL1s, which is probably caused by Chinese rhythmic structure. The English IP is realized by pitch accent and boundary tone, with its rhythm mainly presented by pitch accent distribution, while the Chinese rhythm is presented mainly by inter-phrase pauses.

(6) In English, the number of pitch accent in an utterance is correlated to its length. The longer the utterance, the more the pitch accents. But the number of pitch accent in CL2 utterances increased significantly more than that of EL1. This further shows that difference between rhythmic structures is prominent in longer utterances (> 12 words).

Although there exist differences between the US and the UK speakers, more consistency is displayed than between them and CL2s. The silent pause after IP is far shorter for the US speakers than the UK ones, which may indicate their rhythmic pattern discrepancy. But more data are needed to arrive at this result.

This study is only pilot investigation on rhythmic patterns, which will be followed by further and more detailed studies, considering the following issues: rhythm pattern on syllabic or phone level, F0 of pitch accent in different sentence types or syntactic constructions; studies of prosodic patterns on the discourse level. Instead of isolated utterances, discourse and spontaneous dialogues will be used as reading material in order to provide more reliable data for L2 acquisition, teaching and automatic accent evaluation systems.

***Acknowledgment***: this project is supported by Motorola research center and CASS key project "Phonetics and spontaneous discourse processing".

## 8. References

[1] San Duanmu, Chinese rhythm, in Contemporary linguistics, Vol. 4 2000.

[2] San DuanMu, *ThePhonology of Standard Chinese*, Oxford University,2000.

[3] D. Robert Ladd, *Intonational phonology*, Cambridge University Press.

[4] Xiaobo Zhong, *Perception on SC stress and its acoustic cues*, CAS doctor thesis,2000, 6.

[5] Dorothy M. Chun, 2002. *Discourse intonation in L2- from theory to practice*, John Benjamin Publishing Company.

[6] Yallop, *An Introduction to Phonetics and Phonology*, Foreign Language Teaching and Research Press & Blackwell Publishers Ltd.

[7] William J. Hardcastle and John Laver, Prosody of speech– Melody and Rhythm, from *The handbook of phonetic sciences*,, Blackwell publishers.

[8] John Clark and Colin Yallop, *An introduction to phonetics and phonology*, Blackwell publishers.

[9.] Aijun Li, Prosodic analysis on conversations in standard Chinese., *ZhongGuo YuWen* 2002.

[10] Joanna Radwanska-Williams & Josephine P.S. Yam, The Acquisition of English Plosives by Chinese Learners, *PTLC2000*.

[11] Hiroyuki Obari, Ryousuke Tomiyama.etc., Differentiation of English utterances of Japanese & native speakers by several prosodic parameters, *O-COCOSDA* 2005.

[12] Martha Young-Scholten, *The acquisition of prosodic structure in a second language*, Max Niemeyer Verlag Tübingen 1993.

[13] Lisa Lim & Tan Ying Ying, How we are stressed? Phonetic correlates of stress placement in Singaporean English, *PTLC2000*.

[14]Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. 1999. Training American listeners to perceive Mandarin tones, *J. Acoust. Soc. Am*. 106, 3649–3658.

[15] Hua Chen, The Study on Chinese EFL Learners' English Intonation Patterns in Read Speech, *doctor thesis of Nanji*ng Univ. 2005.

[16] Yuan Jia and Aijun Li, IViE: An introduction to the intonation annotation system, *Acoustic technology*, Vol 24, 3.