Concentric: Studies in Linguistics
34.1 (January 2008): 79-108

Interlanguage Analysis of Phonetic Timing Patterns by Taiwanese Learners*

Hsueh-chu Chen
WuFeng Institute of Technology

Raung-fu Chung
National University of Singapore

This study investigates the timing properties of English spoken by Taiwanese learners, which has been predicted as syllable-timed, and American English, which is usually assumed to be stress-timed. Syllable duration, vowel reduction, linking duration, and consonant cluster duration were acoustically measured and compared. A new variability index (VI) was proposed and calculated. The results show a significant disparity in syllable duration and syllable-to-syllable variation between American English and English spoken by Taiwanese learners. The higher frequency of reduced syllables with a schwa in American English contributes to the difference between American English and that spoken by Taiwanese learners. Taiwanese learners' consonant-vowel linking duration is much longer than that of the native speakers. Some word boundaries are maintained through the insertion of short pauses and glottal stops. The dominant strategy for consonant cluster simplification across the four groups is consonant deletion, but the Taiwanese learners with low English proficiency evinced a specific preference for consonant epenthesis.

Key words: syllable-timing, stress-timing, speech prosody, rhythm patterns, acoustic phonetics, interlanguage phonology

1. Introduction

Contrastive analysis hypothesis has played a very important role in the development of second language acquisition (SLA) theories. Briefly speaking, in acquiring an L2, one can hardly avoid the interference of the L1. In this study, the L1, unless otherwise noted, refers to Mandarin spoken in Taiwan, and the L2, to English. Apart from segment substitutions like [s] for [θ] in *thing* [sɪŋ], or [ŋ] for [n] in *sing* [sɪn] (Chung 2006), a very interesting issue would be the transfer of the prosodic patterns. In terms of prosodic patterns, there has been a simple classification of languages as either syllable-timed, such as Chinese, or stress-timed, such as English (Pike 1945 and Abercrombie 1967). Oversimplified this classification may be, it has been widely adopted in the SLA literature (Prator and Robinett 1985, Roach 1991, Marks 1999, Celce-Murcia, Brinton and Goodwin 2001, and Cruttenden 2001). Yet, in either Pike (1945) or Abercrombie (1967), there is no in-depth discussion of this

* This paper is part of the first author's dissertation, which was awarded the Outstanding Dissertation of the Year by The Language Training and Testing Center (LTTC) in 2006.

classification, nor have phonetic properties been revealed from the classification.

A stress-timed language can be distinguished by the speakers' psychological reality of where the stress is and what syllable should be stressed in an utterance. The number of stresses does not have very much to do with the number of syllables. In addition, the vowel of an unstressed syllable in an utterance is reduced, giving rise to a schwa [ə]. In some English dialects, the unstressed vowel is realized as [ɪ] (Ladefoged 2005). For example, in each of the following English sentences there are only three stresses regardless of the number of syllables:

(1) a. Récords shów prógress.
b. His récords shów his prógress.
c. His récords can shów him his prógress.
d. His récords have shówn him his prógress.
e. His récords have been shówing him his prógress.

In (1a), three syllables receive three stresses. When the number of syllables increases from three to eight (1e), the number of stressed syllables remains constant. This is the specific property of a stress-timed language. Simply put, the number of stresses is the critical component in the determination of rhythm.

Some phonetic properties are inherent to a stressed-time language. First of all, the vowel of a stressed syllable is usually longer than that of an unstressed syllable. It further suggests that the timing of each syllable in an utterance is quite uneven; some syllables are longer, while some are shorter. Second, an unstressed syllable might be omitted or weakened to such an extent that it is no longer explicitly perceived in phonetics. For example, the two unstressed words *have been* in (1e) are hardly audible in the normal speech of a native speaker of English.

On many occasions, linking occurs when a stressed syllable precedes an unstressed one, as exemplified by the following sentences:

(2) a. They find *it* interesting.
b. I don't believe *him* anyway.

By linking,[1] we mean that the coda of the preceding syllable has been syllabified as the onset of a vowel-initial syllable (2a) or that the onset of a syllable has been omitted due to rapid speech (2b). The duration of consonant clusters in an utterance

[1] The linking in English is somewhat different from the liaison in French. The latter is segmental assimilation irrespective of the stressed or unstressed syllables. This is due to the fact that French is a syllable-timed language, according to Abercrombie (1967). For more discussion on the differences between linking and liaison, see Clements and Keyser (1983) and Lamontagne (1993).

varies and is subject to occurrence of linking.

On the other hand, a syllable-timed language is so named because of the role of the syllable in the utterance. Chinese, for instance, is a tone language in which the syllable receives a tone and each tone is more or less equally long. The third tone of which the pitch value is 214 in Chao's system is a bit longer than the others. However, it is usually simplified to a falling tone 21 in phonetics unless it is at the final position of an utterance. Thus the statement that the duration of each syllable in Chinese is equal is true (Chao 1968, Cheng 1973, and Chung 1990). Consequently, the timing of an utterance is made up of syllable duration:

(3) a. cheng^{2} ji^{1} xian^{3} shi^{4} jin^{4} bu^{4} le^{0}
'The grade has improved.'
b. ta^{1} cheng^{2} ji^{1} xian^{3} shi^{4} jin^{4} bu^{4} le^{0}
'His grade has improved.'
c. ta^{1} cheng^{2} ji^{1} ke^{3} yi^{3} xian^{3} shi^{4} jin^{4} bu^{4} le^{0}
'His grade could be improved.'

There are seven syllables in (3a), eight in (3b), and ten in (3c), which means that the duration of (3c) will be longer than that of (3b), which in turn is longer than that of (3a). It is clear that what matters here is the number of syllables. In contrast with English, Chinese has several noteworthy phonetic properties in an utterance. First, the duration of each vowel or syllable is usually identical (Tseng 1990). Second, there is no distinctive vowel reduction in Chinese. There is no distinctive linking in Chinese with the exception of few specific phrases like /tʂə yi/ → [tʂe] ('this one'). For vowel reduction in Chinese, only sparse syllables in a few specific phrases are reduced, which incurs syllable contraction, much like linking in English. The most noticeable case is r-retroflexation, like /pɑi/ 'card' + /ər/ 'suffix' → [pɑr] (Cheng 1973).

From the foregoing discussion, it would seem that Chinese and English are in complementary distribution in terms of prosodic patterns. The comparison of Chinese and English is of great value in language acquisition because both languages are typologically very different in their prosodic properties: Chinese is a tone language with a syllable-timed oriented rhythm, while English is a stress-timed language. Due to the differences between these two timing patterns, further exploring how these differences influence intelligibility of learners' second language becomes a critical issue. It would be of great value to reach a better understanding of how timing patterns come to be transferred from Chinese to the English spoken by Taiwanese learners.

Previous research in timing patterns generally manifests a number of limitations.

(a) Timing patterns have seldom been analyzed at the sentence level or over longer stretches of speech. Most studies examined the timing of isolated words or highly selected syllables. (b) Few studies have considered possible developmental stages of speech timing patterns by comparing learners of different proficiency levels. (c) Little is known about how non-native speakers realize stress and rhythm in the new language when their native language signals stress in a different way from their second language or when the phonetic cues for stress in the second language are used for different major linguistic functions in their native languages. (d) The number of subjects has usually been very small. Although acoustic analyses take a tremendous amount of time, it would improve the reliability and generalizability of the results to include a larger number of subjects.

This paper attempts to investigate multiple aspects of English timing patterns for Taiwanese learners. Speech samples of Taiwanese and English speakers will be compared with respect to difficulties in producing English timing. Four acoustic correlates of timing in English will be analyzed: syllable duration, vowel reduction, linking duration and consonant cluster duration.

Two Research questions are raised as follows:

1. To what extent do Taiwanese learners display specific patterns on the four acoustic timing variables (syllable duration, vowel reduction, linking duration, and consonant cluster duration) that deviate from those of native speakers of American English (the norm)?
2. To what extent do ESL speakers, EFL low achievers, and EFL high achievers differ in terms of acoustic timing patterns?

2. Method

2.1 Participants

The speech of 40 subjects was analyzed. To ensure diversity of speech patterns, four groups of subjects were recruited to produce long English speeches. Both English as a Foreign Language (EFL) students and English as a Second Language (ESL) students were selected for this study. These included 10 EFL Taiwanese learners with low English proficiency (EFL-L), 10 EFL Taiwanese learners with high English proficiency (EFL-H), 10 ESL Taiwanese learners (ESL), and 10 native speakers of English (NS). The three groups of Taiwanese learners were distinguished from one another based on two criteria—their exposure to an English-speaking environment (namely, ESL and EFL learning environments) and their general level of proficiency in English. Each group consisted of an equal number of male and female speakers, aged between 19 and 25 years old.

The EFL-L group was composed of two-year college students from the WuFeng Institute of Technology in Taiwan. They had passed the first phase (listening and reading) of the elementary level of the General English Proficiency Test (GEPT), which is regularly administered in Taiwan by the Language Training and Testing Center, but failed the second phase (speaking and writing). Therefore, none of them were truly beginners.

The EFL-H group consisted of students from four-year colleges in Taiwan. They had all passed the high-intermediate level of the GEPT (equivalent to a score of 213 on the computer-based TOEFL), although they have never studied in an English-speaking country. In this research, they are referred to as EFL students with a high level of English proficiency.

The ESL group included university students from the University of California, at Berkeley (UCB) in America. The members of this group were ESL learners from Taiwan, who were enrolled in a variety of undergraduate programs at UCB. They had either TOEFL scores above 600 points (equivalent to a score of 250 on the computer-based-test) (M=273) or SAT verbal scores above 550 (M=608). The length of their residency in the United States ranged from five to thirteen years (M=7.2 years).

The NS group consisted of university students from UCB who were native speakers. Their speech samples were treated as the norm to which those of the two EFL groups and the ESL group were compared. They were monolingual English speakers and all were Californians from the US West Coast.

2.2 Materials

The materials employed in the acoustic study included a short questionnaire about personal background and a diagnostic passage for acoustic measurement. A well-known diagnostic passage with 14 sentences was selected from *Teaching Pronunciation* (Celce-Murcia et al. 2001) as the reading material. This passage had several obvious advantages. First, it limited vocabulary, grammar, sound segments, and consonant clusters, thereby enabling listeners to make more reliable comparison of speakers. The sentences contained a high concentration of word-boundary consonant clusters that provided many opportunities for linking and consonant cluster simplification. Second, the sentences in the passage were carefully designed to focus on a particular theme, and avoided sequences that were hard to syllabify or segment. The passage dwelled on the problems encountered by second language learners. It was expected that learners' familiarity with the topic would reduce their anxiety while reading. Third, the passage contained five types of sentences (*Wh*-questions, Declarative sentences, Yes-No questions, Tag questions and Closed-choice alternative

questions) in order to eliminate or counterbalance the effects of different sentence types on timing patterns produced by the Taiwanese learners. The following five sentences corresponding to these sentence types were selected for further acoustic analysis:

1. *Wh*-questions
   Why do people usually have an accent when they speak a second language?
2. Declarative sentence
   Most native speakers of English can, for example, recognize people from France by their French accents.
3. Yes-No questions
   Does this mean that accents can't be changed?
4. Tag questions
   Old habits won't change without a lot of hard work, will they?
5. Closed-choice alternative questions
   Will you manage to make progress, or will you just give up?

2.3 Measurement of syllable duration

The speech data was segmented into syllables and subsequently analyzed for duration using Praat software. The total number of syllables analyzed was 3,320 (4 groups × 10 subjects × 83 syllables in 5 sentences). In contrast with Low, Grabe and Nolan (2000), this paper measured speech rate and whole syllable duration rather than merely vowel duration, excluding consideration of the final syllable in the analysis, and calculated normalization based on the whole utterance (excluding the final syllable).

The whole syllable duration rather than the vowel duration was measured because of a tendency for the schwa deletion in fast speech and the case of syllabic consonants. For example, in Low et al.'s (2000) study, the schwa is sometimes deleted in the word *just*. The /m/ in *Morocco* and /l/ in *people* are syllabic consonants. However, judgment about whether a vowel has been deleted or not remains an issue, and this highlights the need for additional guidelines to circumvent the subjectivity inherent in the measurement process.

The segmentation of the syllables generally followed two basic principles. First, the duration of the syllables was measured based on the placement of the syllable boundaries. Given that the test sentences were largely composed of monosyllabic words (42 out of 61 words), syllable boundaries usually coincided with word boundaries. Wherever possible, the syllable boundary was placed at a distinct acoustic event that could reliably and consistently be identified. Second, when it was difficult to determine syllable boundaries, the maximum onsets principle (MOP) and

phonotactic constraints were followed. In MOP, the consonants are preferred in the onset and thus allow no coda consonants except for the word final position (Goldsmith 1990:128).

2.4 Measurement of vowel reduction

In order to measure vowel reduction, two syllable types at sentence level were investigated: (a) lexical stress plus tonic syllables (sentential stresses with full vowels occurring before a clause or sentence boundary), and (b) no lexical stress on a reduced vowel in polysyllabic word and unstressed monosyllabic function words. These two syllable types represent the extremes of duration in English rhythm, the former being among the longest type of English syllable and the latter being among the shortest (Bolinger 1986 and Katamba 1989). It was predicted that native speakers of English would demonstrate greater duration in tonic syllables than unstressed syllables while the Taiwanese speakers would not differentiate one from the other.

The influence of schwa or absence of vowels on syllable duration was considered first. All syllables expected to be pronounced with a schwa in American English were identified. Roach's guideline (Roach 1991:103-109) was consulted to predict the most likely weak form of an assumed function word (i.e., auxiliary verb, preposition, conjunction, and article). Therefore, both *the* and *to* are assumed to have a schwa before a consonant, but not before a vowel, where they will more often have /i/ and /u/.

Tonic syllables can also be determined by the sentential prominence of four types of acoustic evidence: loudness, length, pitch, and quality (Roach 1991:86). These four factors generally work in combination, but syllables can become prominent through only pitch or length.

Ratios instead of absolute values were used to compare syllable durations. These ratios were determined by dividing the unstressed and tonic syllables. Mean ratios (over subjects) of unstressed syllables and tonic syllables were subsequently computed for the four groups. The durations for each were expected to be approximately equal for unstressed and tonic syllables (U/T ratio=1.0) in the productions of Taiwanese learners, as would be expected if the syllable-timed patterns from Chinese phonology were used in English timing patterns.

2.5 Measurement of linking duration

As Hieke (1987) observed, linking can occur in English between two consonants, between a consonant and a vowel, or between two vowels. In consonant-to-consonant (CC) linking, (e.g. *that time*) in which the consonants are identical, only one

consonant is manifest and may be slightly prolonged. Consonant-to-vowel (CV) linking (e.g. *kind of*) involves the assignment of the final consonant of a word to the following, vowel-initial syllable. Vowel-to-vowel (VV) linking (e.g. *say it*) occurs when a word-final tense vowel is followed by a word-initial vowel. This kind of linking often involves glide attraction. Due to the high frequency of CV linking occurring in daily speech, this paper is limited to CV linking only. For each subject, a computation was made based on the measurement of linking duration.

2.6 Measurement of consonant cluster duration

A consonant cluster is a group or sequence of consonants that appear together in a syllable without a vowel between them. In the first tested sentence, for example, *nt* in *accent* and *nd* in *second* are the CC final consonant clusters. For each subject, a computation was made, based on the measured value of the final consonant cluster duration. The simplification strategies between vowel epenthesis and final consonant deletion were also investigated. This paper, however, does not examine the word initial consonant clusters, the length, or segmental composition of the clusters.

2.7 Calculations of acoustic data

In this paper, a total of four timing variables—syllable duration, vowel reduction, linking duration and consonant cluster duration—were measured. The timing patterns of the Taiwanese speakers were then compared to those of the native speakers. The timing patterns of the native speakers, specifically, functioned as the norm. The author devised a new index that meets this demand, which is referred to as the variability index (VI):

(4)

$$\mathrm{VI} = \sqrt{\sum_{k=1}^{K} (X_k - E_k)^2 / K}$$

where X_k is the k-th component, E_k is the mean of the k-th component over the 10 native speakers of English (treated as the norm), and K is the number of components in the sentence. Because VI involves the norm (E_k), it is a *norm-referenced* variability index. If a subject gives exactly the same timing pattern as the norm, VI will equal zero. The larger VI is, the greater the subject's timing pattern will deviate from the norm. It was expected that the native speakers would have smaller VI than the non-native speakers.

To demonstrate the computation of VI, we may consider, for example, syllable

duration. Table 1 shows the acoustic measures of a native speaker and an EFL-L subject for the sample sentence: "*Does this mean that accents can't be changed?*" The first row of Table 1 lists the mean duration for the 10 native speakers, which was E_k in Equation (4). The second and the third rows list the data for the native speaker and the EFL-L subject, respectively. From Equations (5) and (6), we find that the VI value of the EFL-L subject (0.128) was about twice as large as that of the native speaker (0.061).

Table 1. Computation of VI for a native speaker and an EFL-L subject on the sample sentence, "*Does this mean that accents can't be changed?*"

	does	this	mean	that	ac	cent s	can't	be	changed	VI
$E_{k\ (Norm)}$	0.153	0.201	0.211	0.151	0.181	0.370	0.256	0.109	0.591	
NS	0.144	0.156	0.179	0.197	0.160	0.295	0.179	0.138	0.467	0.061
EFL-L	0.294	0.287	0.435	0.31	0.17	0.46	0.415	0.102	0.696	0.128

(5)

$$\begin{aligned} \mathrm{VI}_{\mathrm{NS}} &= \sqrt{\sum_{k=1}^{K}(X_k - E_k)^2 / K} \\ &= \sqrt{[(0.144-0.153)^2 + (0.156-0.201)^2 + \cdots (0.467-0.591)^2]/9} \\ &= 0.061 \end{aligned}$$

(6)

$$\begin{aligned} \mathrm{VI}_{\mathrm{EFL\text{-}L}} &= \sqrt{\sum_{k=1}^{K}(X_k - E_k)^2 / K} \\ &= \sqrt{[(0.294-0.153)^2 + (0.287-0.201)^2 + \cdots (0.696-0.591)^2]/9} \\ &= 0.128 \end{aligned}$$

3. Results and discussion

The four timing variables—syllable duration, vowel reduction, linking duration, and consonant cluster duration—performed by the Taiwanese speakers are compared with those of the native speakers of English, using VI. Because VI involves the norm (E_k), it is a norm-referenced variability index. If a subject gives exactly the same timing pattern as the norm, his or her VI will equal zero. As the value of the VI increases, the subject's timing pattern will deviate more from the norm. It was expected that the native speakers would have smaller VI than the non-native speakers. The difficulties in English speech timing patterns encountered by the Taiwanese learners could then be identified.

3.1 Syllable duration

This section measures and compares the syllable durations from the four groups. In these calculations the mean timing pattern of the native speakers served as the norm and those of the Taiwanese learners were then compared to the norm.

The distributions of VI in syllable durations for the four groups are plotted in Figure 1. For EFL-H, ESL, and NS, all the data fell within the inner fences, indicating there were no outliers. Subjects 5 and 10 in EFL-L were located beyond the inner fences (circle sign). These two subjects produced extremely longer syllable durations. Subject 5 had a VI of 0.221, which was nearly twice the mean value of EFL-L (0.142), and four times the mean of NS (0.059). For the outliers, we had a follow-up interview and we double checked their data. There was no evidence suggesting that these two extreme values were due to coding errors or the subjects' low motivation. These values simply reflected their English speaking characteristics. Hence, their data were not removed to maintain the sample representative and its generalization validity. Actually, even if these two subjects were removed and the data were recomputed, the major finding would not change significantly: EFL-L still had the largest VI among the four groups.

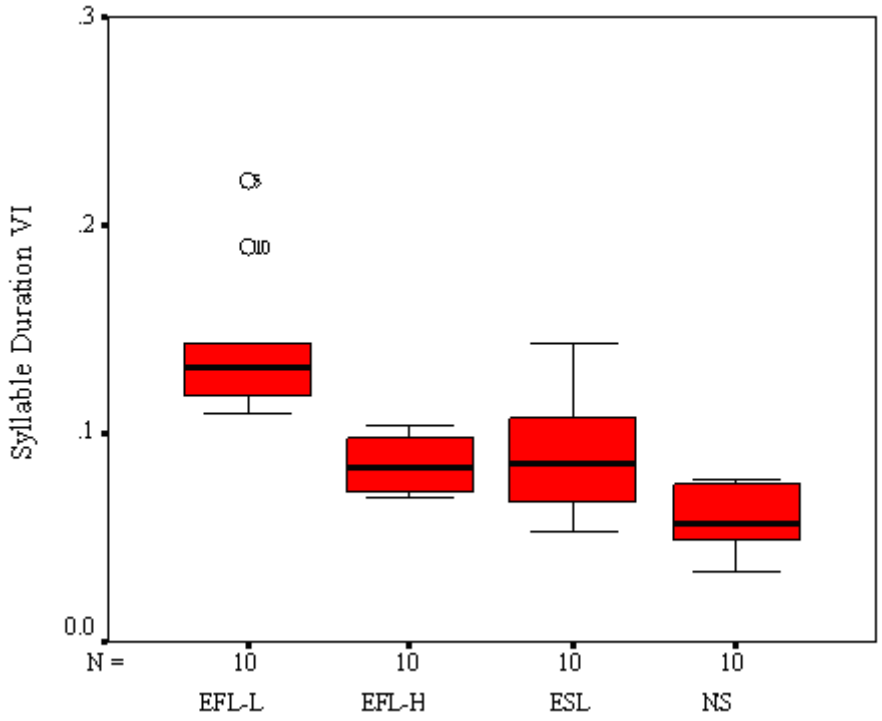

Figure 1. Box-plot of VI in syllable duration for the four groups

Table 2 illustrates the means, standard deviations, minimum values, and maximum values of VI in syllable duration for the four groups. On average, EFL-L (*M*=0.142),

EFL-H (*M*=0.084) and ESL (*M*=0.088) had larger VI than NS (*M*=0.059). EFL-L generated more than twice the deviance in syllable duration from the norm as NS did. The range for NS was 0.034~0.078, whereas that for EFL-L was 0.109~0.221. The minimum value for EFL-L was even larger than the maximum value for NS, suggesting that all the EFL-L subjects deviated from the norm far more than any NS subject. In addition, the range of EFL-L was 2.5 times the range of NS, which means that EFL-L had much larger within-group variation than NS. This can be seen in the standard deviations for these groups. The small within-group variation in NS indicted that these 10 native speakers were highly clustered around their mean, which justified that the mean could serve as the "norm".

Table 2. Means, standard deviations (SD), maximum values (Max), and minimum values (Min) of VI in syllable duration for the four groups

Group	n	Mean	SD	Max	Min
EFL-L	10	0.142	0.036	0.221	0.109
EFL-H	10	0.084	0.013	0.103	0.069
ESL	10	0.088	0.029	0.144	0.052
NS	10	0.059	0.015	0.078	0.034

English is generally classified as a stress-timed language, where the fundamental isochronous unit or timing is the stress foot (Pike 1945 and Abercrombie 1967). In contrast, Chinese is usually classified as a syllable-timed oriented language, in which the syllable is the basic timing unit. In this study, all the EFL-L subjects had limited English experience and proficiency, and stayed at the initial stage of phonological acquisition; that is, they performed the syllable-timing nature of rhythm patterns by different degrees. This disparity can be attributed to the native language transfer (i.e., the syllable-timed nature in Chinese), which strongly influenced the syllable duration produced by EFL-L.

The one-way ANOVA showed that VI syllable duration for these four groups was statistically significant ($F_{3, 36}$=19.37, p<.001). Based on the post-hoc Tukey's honestly significant difference test (Figure 2), there was a significant difference between EFL-L and the other three groups (if the groups are connected by a line, then they do not differ statistically significantly at the .05 nominal level). No statistically significant differences were found among EFL-H, ESL and NS.

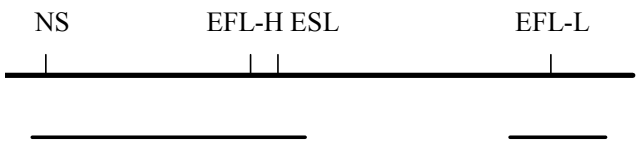

Figure 2. Post-hoc group comparison in VI in the syllable duration according to Tukey's honestly significant difference test

Figure 2 shows that NS had the smallest mean, EFL-L had the largest mean, and ELF-H and ESL fell in between these two ends and were not substantially different from each other. These results suggest that it is appropriate to place Chinese and English along a scale,[2] on which Chinese is located at one end (maximally syllable-timed) and English at the other end (maximally stress-timed). English learning for Chinese learners is a progressive process, starting from syllable-timed Chinese toward stress-timed English. In the development of this progressive process for learning stress-timed English, EFL-L subjects were at the initial stage and suffered from the negative transfer impact from their native language of Chinese, while those in the EFL-H and ESL groups were farther advanced at the middle stage and approaching the stress-timing patterns of English.

3.2 Vowel reduction

This section investigates the durations of two types of syllables at the sentence level. The first type is a syllable with lexical stress plus tonic stress (sentential stress). The second type is unstressed syllables in polysyllabic words as well as in unstressed monosyllabic function words. Figure 3 presents the mean ratios of the five types of sentences, mentioned in Section 2.2, and their composite mean ratios over the five sentences (called total sentence) for the four groups. The figure clearly demonstrates that for the NS the unstressed syllables are much shorter than the tonic syllables (the ratio is smaller than 1) while EFL-L did not differentiate one from the other (the ratio is equal to or larger than 1). That is, EFL-L performed worse and could clearly be differentiated from NS. EFL-H and ESL performed similarly (though ESL performed slightly better). The small ratio between unstressed syllable duration and tonic syllable duration in American English accounts for weaker vowel reduction in American English than in English spoken by Taiwanese learners. This lack of vowel reduction contributes to the impression of syllable timing in English produced by Taiwanese learners.

[2] Languages, as pointed out by Miller (1984), could be better described along a continuum of syllable- and stress-timing rather than a rigid classification of types.

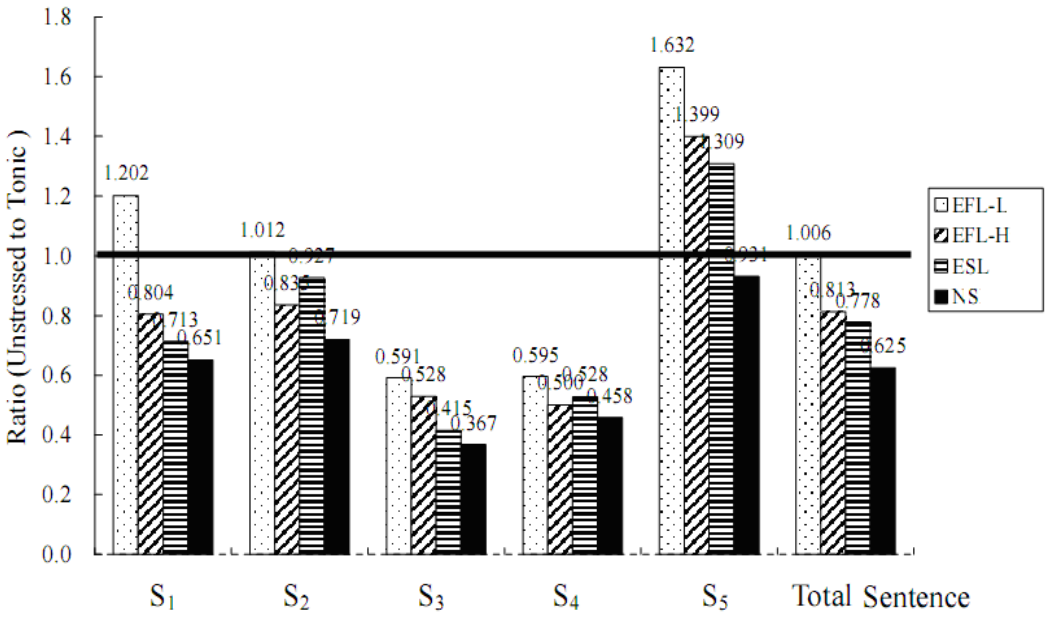

Figure 3. Ratios of the five types of sentences and the total sentence for the four groups

Table 3 shows the means and standard deviations of the ratios of the total sentence for the four groups. The one-way ANOVA showed that the statistical difference between the four groups in the ratios of all the five types of sentences and the total sentence was significant (all the *p*-values were smaller than .05).

Table 3. Means and standard deviations of the ratios on the total sentence for the four groups

Group	Mean	SD
EFL-L	1.006	0.185
EFLH	0.813	0.068
ESL	0.778	0.090
NS	0.625	0.054

In order to make a cross-section comparison among the four timing variables (i.e. syllable duration, vowel reduction, linking duration, and consonant cluster duration), the ratio of unstressed to tonic syllables VI was also calculated. Table 4 and Figure 4 reveal that EFL-L, EFL-H and ESL had larger VI in ratios of unstressed to tonic syllables (0.387 for EFL-L, 0.207 for EFL-H, 0.184 for ESL) while NS had the smallest VI (0.096). The one-way ANOVA showed that the difference in VI of these four groups was statistically significant ($F_{3, 36}$=12.62, p< .005).

Table 4. Means, standard deviations, maximum values and minimum values of VI in ratio of unstressed to tonic syllables for the four groups

Group	Mean	SD	Max	Min
EFL-L	0.387	0.189	0.785	0.213
EFL-H	0.207	0.072	0.303	0.059
ESL	0.184	0.069	0.317	0.094
NS	0.096	0.039	0.161	0.042

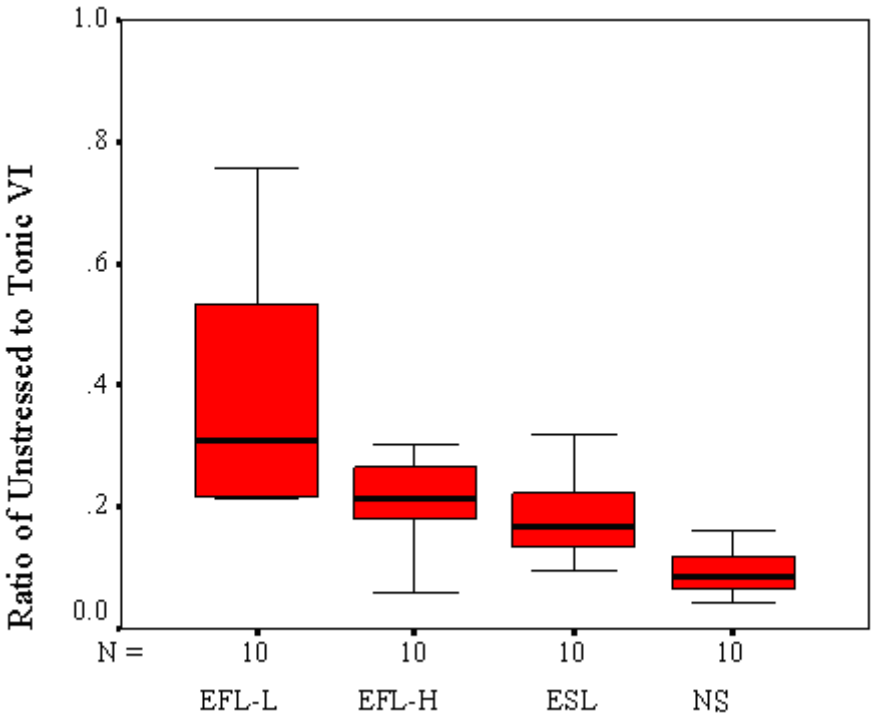

Figure 4. Box-plot of VI in ratio of unstressed to tonic syllables for the four groups

It appears that unstressed syllables pose a great challenge to Taiwanese learners. Taiwanese learners usually stress every syllable with great emphasis. For example, in the question "*Does this mean that accents can't be changed?*", a Taiwanese EFL-L learner produced syllables of longer duration and placed greater intensity on *does* and *this*, whereas an American speaker placed less emphasis on these two function words. Another example shows that the word *accents* was produced by a Taiwanese EFL-L learner with longer duration and greater intensity on both *ac-* and *-cents*. Vowel insertion at the word final position was also demonstrated. On the contrary, a sample of an American speaker showed relatively shorter duration and less energy on these two syllables. A glottal stop occurs between *ac-* and *-cents*. It was observed that Taiwanese EFL-L learners' whole strategy of stress assignment was to lengthen the monosyllabic word or the first syllable of every polysyllabic word, using tonic stress

on virtually every lexical item, whether it is semantically important or just a function word.

The Taiwanese students reduced their vowels mainly in the definite and indefinite articles. They were not able to reduce /o/ to [ə] in the unstressed syllables, such as *cog* in *recognize*. They also rarely reduced vowels in words such as *you*. This failure to reduce vowels may arise at least in part from native language transfer, since Chinese vowels retain their pure quality and are hardly ever centralized as they are in English.

Archibald (1997) described a case in which the subjects' L1s are tone languages (e.g. Chinese), and argued that these learners do not compute stress placement in English, but rather store it lexically. In our study, it was observed that Taiwanese learners did assign tones to syllables, lengthened the duration, and therefore had difficulty in reducing unstressed syllables. They produced less differentiation between stressed and unstressed syllables than did English speakers. They had problems realizing the alternating stressed and unstressed syllables in English, particularly when the weak syllables consisted of unstressed function words.

In light of the results in this paper, the authors proposed that, on one hand, there might be languages with abundant monosyllabic words that strongly favor the one-syllable-one-stress structures as is the case for Chinese. On the other hand, there might be languages that tolerate a wider range of permissible unstressed syllables, including long stretches of unstressed syllables, as is the case for English. Note that the latter is also inclusive of the former. From this perspective, the syllable timing patterns of Taiwanese speakers could be construed as one type of stress pattern, which is indeed simpler and more unmarked. When Taiwanese learners learn a language like English which has more complicated and marked syllable variations, their primary task is to learn the various numbers of unstressed syllables within a foot and a various number of prosodically weak syllables at a higher metrical level. Without a doubt, it is very challenging for Taiwanese learners to learn English speech rhythm.

Moreover, from the acoustic data, the performance of ESL learners was closer to that of NS. For example, the durations of the word *to* between the groups were similar (ESL: 0.15ms vs. NS: 0.135ms). Both of the groups reduced the vowel quality (ESL: /tu/ to [tə] vs. NS: /tu/ to [t^h]). In term of Universal Grammar (UG) principles, this ESL group of learners was attempting to reset their parameters to better reflect English syllable length and stress patterns, and escape from L1 interference. As for EFL-H, their interlanguage performance was between that of the ESL and EFL-L groups. EFL-H produced similar duration to that of EFL-L but reduced the vowel /u/ to schwa, which is more like the performance of ESL. As for EFL-L, they were still struggling with L1 interference. This group of learners clearly produced the citation form of "to" /tu/ rather than reduced it to schwa, or turned it to aspirated /t/. The

duration EFL-L produced is the longest (0.389 ms). Thus, we may argue that the L2 learners' interlanguages are a combination of Universal Grammar principles, correct L2 parameter settings from resetting, and incorrect L1 parameter setting from transfer.

3.3 Linking duration

This section examines C-V linking and tapping duration between words in connected speech, comparing native English speakers with Taiwanese speakers of different levels. It is noted that tapping is a phonological process found in many dialects of English, especially in American English, by which prevocalic /t/ and /d/ surface as the alveolar tap [ɾ]. The reasons for looking at this particular allophone are that it can differ so markedly from the parent phoneme, and that it occurs so frequently that it is very important to be recognized. The reason for the tap is gestural economy: it is much easier, and takes less time, to tap the tongue briefly against the roof of the mouth in contrast with the normal closure, release, and optional aspiration cycle of the plosive. Since the tap is against the alveolar ridge, the alveolar quality of the sound is preserved.

Alveolar tapping is a common articulatory "weakening" process in American English that can cause the neutralization of underlying lexical contrasts. Tapping also occurs when it spans a word boundary, as in "*got it*". There is often a V-shaped falloff in total energy, with the alveolar flags and a mini-plosion just before the resumption of the full vowel after the tap. The mini-plosion occurs when the tongue leaves the alveolar ridge. Therefore, the acoustic cues of alveolar tapping are different from those of regular alveolar stops.

Table 5 and Figure 5 reveal that EFL-L, EFL-H and ESL had larger overall VI in linking duration (0.320 for EFL-L, 0.146 for EFL-H, 0.096 for ESL) while the NS had the smallest VI (0.054). The one-way ANOVA showed that the differences between these four groups were statistically significant ($F_{3, 36}$=19.63, p<.001). The post-hoc Tukey's honestly significant difference test showed a significant difference between EFL-L and the other three groups. This indicates that EFL-H and ESL used tapping and linking like NS, while EFL-L had not yet mastered tapping and CV linking.

Table 5. Means, standard deviations, maximum values and minimum values of VI in linking duration for the four groups

Group	Mean	SD	Max	Min
EFL-L	0.320	0.156	0.650	0.159
EFL-H	0.146	0.040	0.209	0.066
ESL	0.096	0.042	0.166	0.045
NS	0.054	0.011	0.074	0.040

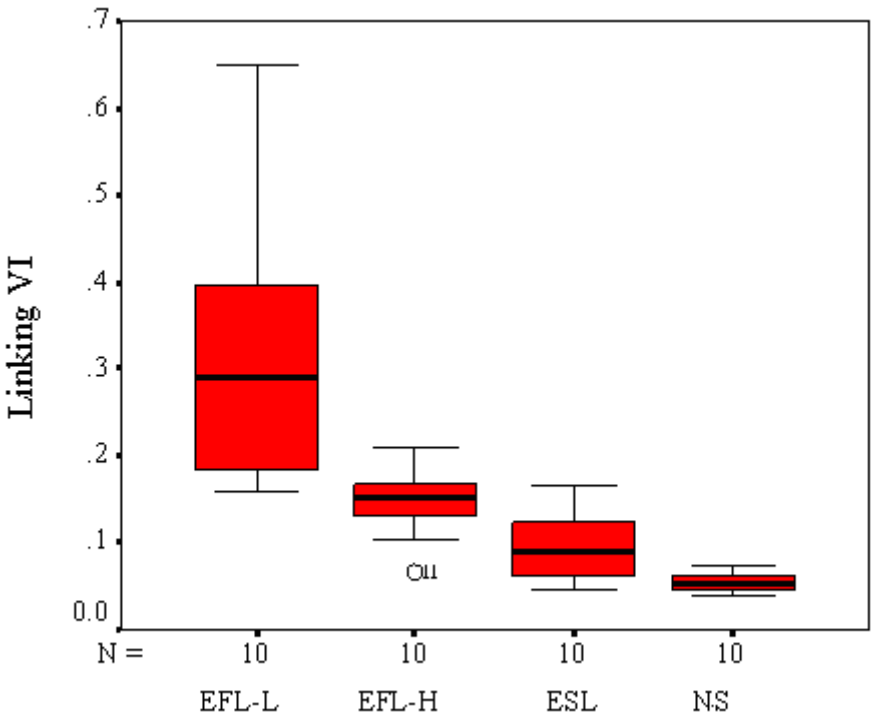

Figure 5. Box-plot of VI in linking duration for the four groups

Table 6 shows the means, standard deviations, and maximum values and minimum values of each group's tapping duration. For Taiwanese learners, another factor contributing to inadequate tapping was the frequency with which word final stop consonants had an audible burst or insertion (which would lengthen the duration). More specifically, an examination of the forms occurring in EFL-L data revealed a tendency to use an aspirated alveolar stop instead of the tap, and to insert a glottal stop or a schwa before the vowel beginning the following word: "…*that accent*…", "…*but old*…", "…*out a*…", "…*lot of*…". The extra aspirations or the insertions produced by the Taiwanese learners caused their overall linking duration to be longer than that of native speakers of English.

Table 6. Means and standard deviations of each tapping duration for the four groups

Tapping Tokens	thaT-ACcent		buT-OLD		ouT-A		loT-OF	
Groups	Mean	SD	Mean	SD	Mean	SD	Mean	SD
EFL-L	0.457	0.233	0.342	0.187	0.320	0.092	0.306	0.081
EFL-H	0.346	0.138	0.387	0.202	0.270	0.074	0.224	0.100
ESL	0.307	0.078	0.239	0.069	0.240	0.039	0.180	0.039
NS	0.239	0.052	0.238	0.043	0.170	0.017	0.160	0.020

Table 7 shows the means and standard deviations of each group's linking duration. EFL-L evinced a strong tendency to keep word boundaries intact by inserting a short pause or a glottal stop before the word-initial vowel in the second word as in "*have [ʔ]an accent*" for "*have an accent*". EFL-H and ESL also showed the same tendency to insert glottal stops as EFL-L, although they did so less often. All of the insertions produced by the Taiwanese learners caused their overall linking duration to be longer than that of native speakers of English.

Table 7. Mean and standard deviations of each linking duration for the four groups

Linking Tokens	haVE-AN-AC		speaK-A		speakerS-OF		foR-EXample		frenCH-ACcents		giVE-UP	
Groups	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
EFL-L	0.708	0.356	0.659	0.234	0.523	0.262	0.397	0.098	0.212	0.061	0.649	0.257
EFL-H	0.420	0.089	0.452	0.088	0.271	0.075	0.291	0.058	0.227	0.041	0.401	0.143
ESL	0.374	0.066	0.433	0.181	0.280	0.076	0.290	0.088	0.206	0.035	0.276	0.042
NS	0.288	0.046	0.328	0.088	0.212	0.078	0.218	0.050	0.180	0.039	0.269	0.055

In summary, EFL-L used tapping and linking significantly less often than NS, and the quantitative differences were dramatic; EFL-H and ESL demonstrated a large, but not significant difference in means, as is evident in C-V linking and tapping duration. Thus while EFL-H and ESL were almost native-like in their rate of tapping and linking, EFL-L lagged behind in most cases. The analysis of forms for tapping and linking showed that EFL-L demonstrated a tendency to preserve word boundaries. The absence of linking between segments accentuates the impression of syllable timing in Taiwanese-accented English. In C-V linking, the word boundaries were maintained by the insertion of pause and glottal stops. It is possible that the tendency to keep vowels intact may be related to a concern for intelligibility.

In agreement with the observation that extensive segment overlapping or linking is cross-linguistically a more marked pattern (Eckman 1977), EFL-L did not extensively use the English linking pattern. EFL-L preferred to carry over their Chinese timing patterns when they switched from L1 to L2, because the articulatory English pattern is more marked. According to the Markedness Differential Hypothesis (MDH) (Eckman 1977), the difficulty that a language learner encounters in producing some systematic pattern in the target language corresponds with the degree to which that pattern is cross-linguistically undesirable. Both groups of non-native speakers of English in Zsiga's (2003) study preferred on average an articulatory pattern with little overlap

between consonant closures, and thus more clearly articulated final consonants, which support the word integrity constraint of Cebrian (2000) and the recoverability principle of Weinberger (1994). That is, the unmarked pattern favors separation between words (Cebrian 2000), and clear phonetic cues to contrasts (Weinberger 1994).

In terms of linking patterns of English, it reflects "ease of articulation" (e.g. Lass 1984). Rather than being easy, it may also be that the English pattern of linking sacrifices segmental contrast in order to convey prosodic information. Thus, the extensive linking between vowel and consonant at word boundaries seen here in native English may well serve to signal a close prosodic relationship between the words. For example, the linking in such a case as *a lot of* demonstrates the bounded relationship among these three words. Without the appropriate linking, native listeners may not be able to understand the meaning that the non-native speakers attempt to convey.

3.4 Consonant cluster duration

Preference for syllable types may be regarded as a domain separate from that of speech timing, but it may still influence the temporal organization of speech. The presence or absence of a segment has direct impact on articulatory timing to the extent that the elements comprising a syllable have durations associated with them.

The difference between Chinese and English syllable structure raises two questions about consonant cluster simplification. First, was the total consonant cluster duration longer for the three Taiwanese groups than for NS? Second, was the simplification type of the three Taiwanese groups the combination of consonant deletion and vowel epenthesis categories, while that of NS a reflection of only consonant deletion?

3.4.1 Consonant cluster duration

This section examines consonant cluster duration, comparing native English speakers with Taiwanese learners at different levels of English competence. In order to compare with the norm of native speakers and make a cross-section comparison among four timing variables, consonant cluster VI was calculated. Table 8 and Figure 6 reveal that EFL-L, EFL-H and ESL had larger overall VI in consonant cluster duration (0.175 for EFL-L, 0.069 for EFL-H, 0.073 for ESL) while NS had the smallest VI (0.060). The one-way ANOVA showed that the differences between these four groups were statistically significant ($F_{3, 36}$=10.18, p<.005). Based on the post-hoc Tukey's honestly significant difference test, there was a significant difference between

EFL-L and the other three groups. No significant differences were found among the EFL-H, ESL and NS.

Table 8. Means, standard deviations, maximum values and minimum values of VI in consonant cluster duration for the four groups

Group	Mean	SD	Max	Min
EFL-L	0.175	0.096	0.374	0.072
EFL-H	0.069	0.020	0.092	0.039
ESL	0.073	0.022	0.112	0.051
NS	0.060	0.038	0.166	0.041

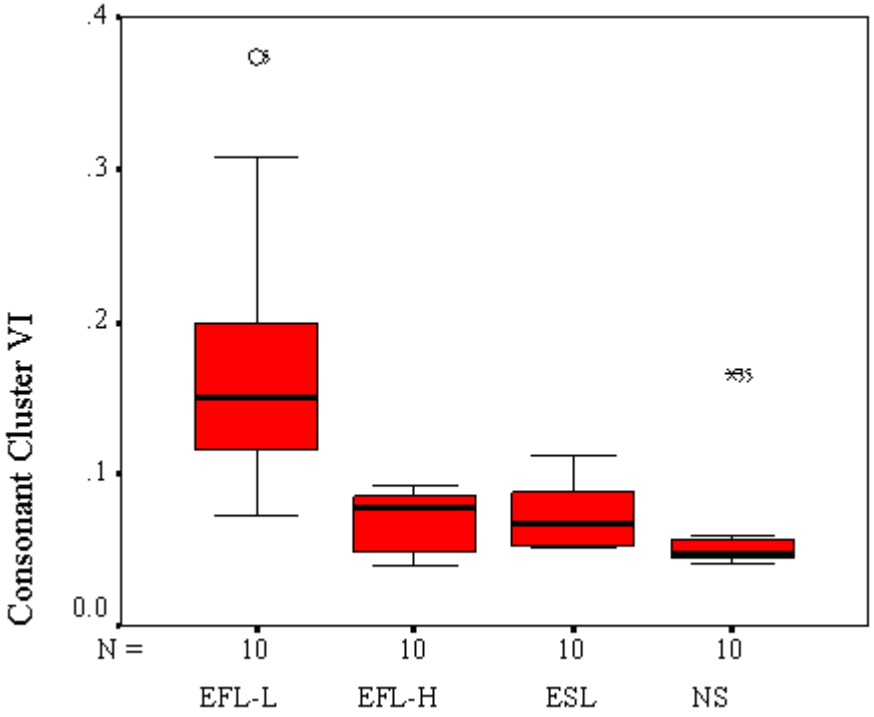

Figure 6. Box-plot of VI in consonant cluster duration for the four groups

These results indicate that the lower the English proficiency of the learners, the longer the duration of the consonant cluster that they might produce. Examples from three Taiwanese learners and a native speaker showed that for the word *most* the duration was 0.478 seconds (EFL-L), 0.425 seconds (EFL-H), 0.3432 seconds (ESL), and 0.229 seconds (NS).

Among the 14 consonant clusters, (i.e. *accent*, *second*, *most*, *France*, *accents*, *habits*, *won't*, *change*, *work*, *accents*, *can't*, *changed*, *just*, *French*), one-way ANOVA showed that these four groups differed in three coda durations "*most*", "*French*", and "*won't*" and were statistically significant at the 0.001 level, as shown in Table 9. It should be noted that the higher rate of vowel insertion produced by EFL-L resulted in

this disparity.

Table 9. Means and standard deviations of three coda durations, "moST", "FreNCH", and "woN'T" for the four groups

	moST		FreNCH		woN'T	
Group	Mean	SD	Mean	SD	Mean	SD
EFL-L	0.329	0.047	0.294	0.106	0.248	0.122
EFL-H	0.265	0.051	0.240	0.069	0.127	0.033
ESL	0.190	0.061	0.162	0.061	0.122	0.031
NS	0.118	0.021	0.160	0.025	0.110	0.022

3.4.2 Simplification strategies: deletion or insertion

The second question concerned the strategy for simplification—would Taiwanese speakers use epenthesis or deletion more often? Table 10 lists the frequency of deletion, epenthesis and no modification among the four groups.

Table 10. Frequency of deletion, epenthesis and no modification among the four groups

Group	Deletion	Epenthesis	No Modification	Total Tokens
EFL-L	51 (36%)	35 (25%)	54 (39%)	140
EFL-H	65 (46%)	4 (3%)	71 (51%)	140
ESL	56 (40%)	2 (1%)	82 (59%)	140
NS	57 (41%)	0 (0%)	83 (59%)	140

The combined rate of deletion and insertion revealed a higher rate of simplification for the EFL-L group (36% + 25% = 61%) than for the NS group (41%); the dominant strategy for simplification for all groups was consonant deletion (EFL-L: 36%, EFL-H: 46%, ESL: 40%, and NS: 41%) but not epenthesis; the EFL-L group used epenthesis (25%) more often than the EFL-H (3%) and ESL groups (1%). These results confirm the assumption that the CV syllable is the least marked, or the optimal syllable type (Carlisle 1994, Chomsky and Halle 1968, and Hooper 1976). In general, Taiwanese learners employ both deletion and epenthesis as simplification strategies to break up the more complex, or more marked, CVCC (i.e. *most*) or CCVCC syllables (i.e. *French*).

For consonant deletion, the NS group displayed more deletion than the ESL and

EFL-L groups. The consonants deleted most often by all the groups were *accent* and *second* although this tendency was observed more often in the NS group. Instead of final /t/ air burst, /t/ turned to be unreleased and merged with /n/. Those consonants form a natural class owing to the same phonological feature [+coronal]. However, the Taiwanese groups also deleted consonants not deleted by the NS group, for example, final consonant clusters in *work*.

Their preference of deletion to epenthesis may arise from the fact that deletion is a more natural or universal process than epenthesis in the sense that children use it more often than epenthesis to simplify consonant clusters when acquiring their first language (Oller 1973). Broselow, Chen and Wang (1998) also observed that the simplification strategies used to modify English codas by native speakers of Mandarin could be unmarked, and this was also later claimed by McCarthy (2002). In McCarthy's framework of Optimality Theory (OT), when a low-ranked constraint, usually a markedness constraint that is generally not decisive in a language, comes to be the determining constraint in an evaluation, the consonant cluster deletion that surfaces as optimal is referred to as the emergence of the unmarked. OT provides a ready explanation for the source of second language patterns that are attributable to neither the L1 nor the L2.

Moreover, this result challenges the findings of Weinberger (1987), Major (1994), Lin (2001), and Chen and Chung (2004), in which the deletion or insertion strategies for consonant cluster simplification were construed as learners' errors. In these studies, the speech samples from native speakers were not used as the norm. In fact, some deletion cases cannot be counted as errors. These instances of deletion (e.g. *nt* and *nd*) have long been observed in English phonology. In the present study, the native speakers demonstrated similar rates of deletion to the nonnative learners even though some instances differed in nature from those of the nonnative speakers.

As expected, epenthesis did not occur for the NS group. Within each of the three Taiwanese groups, the deletion rate was higher than the epenthesis rate. The EFL-L group, however, used more epenthesis than the EFL-H and ESL groups. In Heyer's (1986) study on Chinese speakers' production of English final obstruents, the learners' higher percentage of epenthesis to monosyllabic words than to polysyllabic words was construed as their preference for "bi-syllabicity." Wang (1995) proposed that a preference for bisyllabicity is a factor in Chinese interlanguage. Broselow et al. (1998) further proposed the "word binarity" constraint to capture the Mandarin preference for disyllabicity. These results agree with the earlier studies of second language phonology which have shown that native language transfer is more apparent in the earlier stages of second language development (Dickerson 1974 and Major 1987). Taiwanese learners modify syllable-final codas in favor of shorter and less

marked codas. L1 transfer has an effect on syllable structure production, especially in the early stages of acquisition (i.e. EFL-L learners in this paper). L1 transfer can also positively affect acquisition, and may determine which sounds the learner substitutes for the English sound. L1 developmental effects appear to interact with L1 transfer in a sequential relationship, increasing in the later stages of acquisition after interference effects decrease.

The results of this phonetic study give credence to Weinberger's (1994) recoverability principle: nonnative speakers prefer configurations in which contrastive information is maintained. As we previously noted, Weinberger's finding is that nonnative speakers prefer epenthesis to deletion in difficult clusters. In this paper, instead of the deletion strategy, the unexpected insertion may be interpreted by native listeners as insertion of an extra vowel (e.g., schwa in *accent[ə]* or [i] in *French[i]*) or lengthening the sibilant consonants (e.g. *France*). Tajima, Port and Dalby (1997) also noted that the presence of unexpected release bursts on final consonants could be construed by native English listeners as additional unstressed syllables, rendering individual words less intelligible and creating a garden path effect in the interpretation of running speech.

4. Conclusion

An accurate timing pattern has been considered one of the most important phonetic aspects for the auditory comprehension and intelligible oral production of English (Wong 1987, Dickerson 1989, Anderson-Hsieh, Johnson, and Koehler 1992, Anderson 1993, Culter 1994, Flege, Munro, and Macky 1995, Sole 1997, and Tajima et al. 1997). However, Taiwanese learners of English are frequently reported to experience difficulties with English timing patterns (Anderson-Hsieh and Venkatagiri 1994, Hua 2003, and Chen 2005). Little empirical evidence exists to explain what exactly makes their speech timing patterns different from those of native English speakers. Few studies have explored in depth the ways in which Taiwanese learners with different proficiency levels produced the timing patterns. This study acoustically investigates the patterns of syllable duration, vowel reduction, linking duration, and consonant cluster duration, produced by Taiwanese learners. All of these four components are common in nature, for each of them is closely related to timing.

In this study two research questions were raised. The first research question relates to whether the four acoustic timing variables (syllable duration, vowel reduction, linking duration, and consonant cluster duration) in Taiwanese learners' output are similar to or deviate from those of native speakers of English (the norm). The findings are summarized as follows.

First, an analysis of the four acoustic timing variables discussed in this study

suggests that Taiwanese learners display specific speech patterns that deviate from those of native speakers of English. The difference between American English and English spoken by Taiwanese learners may in part stem from the fact of the greater frequency of reduced syllables with a schwa in American English. Vowel reduction occurs rarely amongst Taiwanese learners. In fact, the findings suggest that Taiwanese learners may fail to differentiate the duration between stressed and unstressed syllables. Moreover, Taiwanese learners performed much longer linking duration than is commonly encountered among native English speakers. In C-V linking, the word boundaries are maintained through the insertion of short pauses and glottal stops. The tendency to keep vowels intact may possibly be related to a speaker's concern with intelligibility. The dominant strategy for speech simplification across the four groups (i.e. EFL-L, EFL-H, ESL, and NS) is consonant deletion, but EFL-L learners commonly resort to vowel epenthesis instead of consonant deletion.

Second, the inter-syllable timing patterns of Taiwanese learners differ from those of American English speakers. This study has demonstrated that a significant disparity exists in syllable duration and syllable-to-syllable variation between American English and English spoken by Taiwanese learners. The acoustic timing patterns in American English, as expected, can be considered stress-timed, while those in English spoken by Taiwanese learners (especially by Taiwanese EFL-L learners) are more syllable-timed.

The second research question addresses the issue of the timing patterns for Taiwanese ESL and EFL learners engaged in second language acquisition. The ESL group, the EFL-L group, and the EFL-H group exhibit different acoustic timing patterns in the phonological acquisition of a second language. Both the ESL group and EFL-H group acoustically outperform the EFL-L group. All three groups can be contrasted within the context of the four timing variables. One surprising result is that the ESL group, with its extensive experience of living in English speaking countries, does not significantly outperform the EFL-H group with only minimal experience abroad.

This study finds that phonetic timing may transfer from L1 to L2 in the same way as phonological patterns are hypothesized. The acquisition of the English timing patterns by Taiwanese learners is an example of L2 acquisition of a marked linguistic feature. In acquiring this marked L2 feature, such as vowel reduction, linking and consonant clusters, Taiwanese learners must overcome their L1 mechanisms for coping with that marked linguistic feature. In accordance with other research, this study observes that, as phonological transfer occurs, the postlexical processes of the L1, rather than the lexical alternations, cause the most interference (James 1976, Hammarberg 1990, and Weinberger 1994). Cebrian (2000:7) suggested that, "the

difficulty in changing the fossilized articulatory timing habits" in learners' speech makes the transfer of postlexical processes commonplace.

This study makes several contributions. First, it takes into consideration timing patterns over whole sentences. Prior studies have usually compared stressed and unstressed syllables using isolated words, or highly selected syllables from utterances. This study examines complete sentences and thus enables us to gain a more realistic view of the timing patterns.

This study also uses a larger sample size (40 subjects) than most other studies. Due to the time constraint required for measurement and analysis, it is typical for acoustic studies to have limited samples (e.g., usually under 10 subjects). Since larger individual variations are commonly observed in the study of inter-language, the reliability of results based on a larger sample, as in this work, should be higher.

This study also develops an innovative and theoretically sound method for calculating timing patterns, VI, to make cross-sectional comparisons. VI takes into account the characteristics of syllable duration, vowel reduction, linking duration, and consonant cluster duration, and thus enables one to capture the differences between the average performances of native English speakers and those of the other three groups of Taiwanese learners.

Furthermore, this study takes English proficiency into consideration as a potential independent variable in the acquisition of speech timing patterns. It contributes to our understanding of the acquisition of second/foreign language timing patterns and especially to the way in which second/foreign language learners develop new timing patterns. In comparing English learners at different proficiency levels, one is able to investigate the difficulties in timing patterns at different stages of language acquisition.

Although this study makes several contributions to the field, several issues remain to be thoroughly investigated. Further research on the second language acquisition of speech timing may proceed in several directions. First, it is of great interest to investigate a comparison of the bidirectional acquisition of second language phonology. Further research can explore the way in which English speakers learning Chinese acquire the Chinese rhythm patterns. The results in this study suggest that English leaning for Taiwanese learners is a progressive process that begins with the syllable-timed Chinese and proceeds toward stress-timed English. In the development of the progressive process for learning stress-timed English, EFL-L was at the initial stage and suffered from negative transfer impact from their native language of Chinese, whereas EFL-H and ESL were at the middle stage and approaching the stress-timing patterns of English. It is worth further investigating whether English speakers learning Chinese proceed along the opposite path, from stress-timed English

toward syllable-timed Chinese.

Second, in this study only Californian native speakers of English from the American West Coast comprise the norm-referenced group. Although selecting such a homogeneous group can reduce variation in the norm, one should be cautious in the applicability of its conclusions within very narrow parameters. From a sociolinguistic perspective, further studies may select native speech samples with different accents from different regions, such as the American East Coast, the United Kingdom, New Zealand or Australia. Furthermore, speech samples of non-native speakers in this study are limited to Taiwanese learners with syllable-timed oriented speech patterns. Several studies (Hua 2003 and Jun 2005) have claimed that Beijing Mandarin is more stress-timed (or foot-timed) oriented than Taiwan Mandarin, so it would be of great value to compare the performances of learners from Taiwan with those from Beijing, Mainland China.

Finally, despite the fact that the use of synthetic speech to simulate accented speech is artificial and different from naturalistic speech, such manipulation of each timing variable might help us understand the different angles of timing patterns. It is also of interest to ascertain the way in which the insertion or deletion of each variable in one utterance may influence the timing pattern of that utterance.

References

Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Anderson, P. 1993. Defining intelligibility parameters: The secret of sounding native. Paper presented at the XXVII TESOL Annual Convention, Atlanta.

Anderson-Hsieh, J., R. Johnson and K. Koehler. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42: 529-555.

Anderson-Hsieh, J., and H. Venkatagiri. 1994. Syllable duration and pausing in the speech of Chinese ESL speakers. *TESOL Quarterly* 28: 807-812.

Archibald, J. 1997. The acquisition of second language phrasal stress. *Focus on Phonological Acquisition*, ed. by S. J. Hannahs, and M. Young-Scholten, 263-289. Amsterdam: John Benjamins.

Bolinger, D. 1986. *Intonation and Its Parts*. Palo Alto, CA: Stanford University Press.

Broselow, E., S. I. Chen, and C. Wang. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20: 261-280.

Carlisle, R. S. 1994. Markedness and environment as internal constraints on the

variability of interlanguage phonology. *First and Second Language Phonology*, ed. by M. Yavas, 223-249. San Diego: Singular.

Cebrian, J. 2000. Transferability and productivity of L1 rules in Caltalan-English interlanguage. *Studies in Second Language Acquisition* 22: 1-26.

Celce-Murcia, M., D. M. Brinton and J. M. Goodwin. 2001. *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. New York: Cambridge University Press.

Chao, Y.-R. 1968. *A Grammar of Spoken Chinese*. Berkeley: UC Press.

Chen, H.-C. 2005. The acquisition of American English rhythm by Taiwanese EFL learners. *English Teaching and Learnin*g 29: 1-23.

Chen, H.-C., and R.-F. Chung. 2004. Effects of task formality on consonant cluster simplification for Taiwanese learners with limited English proficiency. *English Teaching and Learning* 28: 67-93.

Cheng, C.-C. 1973. *A Synchronic Phonology of Mandarin Chinese*. Mouton: The Hague.

Chomsky, N., and M. Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Chung, R.-F. 2006. *Introduction to Modern Linguistics*. Taipei: Wu-nan

Chung, R.-F. 1990. Contemporary phonological theories and Chinese phonology. *The World of Chinese Studies* 5: 58-60.

Clements, G. N., and J. Keyser. 1983. *CV Phonology: A Generative Theory of the Syllable*. Cambridge, Mass: MIT Press.

Cruttenden, A. 2001. *Gimson's Pronunciation of English*. 6th edition. London: Edward Arnold.

Culter, A. 1994. The perception of rhythm in language. *Cognition* 50: 79-81.

Dickerson, L. J. 1974. *Internal and External Patterning of Phonological Variability in the Speech of Japanese Learners of English: Toward a Theory of Second Language Acquisition*. Doctoral dissertation, University of Illinois.

Dickerson, W. 1989. *Stress in the Stream of Speech: The Rhythm of Spoken English*. Urbana, IL: University of Illinois Press.

Eckman, F. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315-330.

Flege, J. E., M. J. Munro and I. R. Macky. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of Acoustical Society of America* 97: 3125-3134.

Goldsmith, J. 1990. *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell.

Hammarberg, B. 1990. Conditions on transfer in phonology. *New Sounds 90: Proceedings of the 1990 Symposium on the Acquisition of Second Language*

Speech, ed. by A. James, and J. Leather, 198-215. Dordrecht: Foris.

Heyer, S. C. 1986. English final consonants and the Chinese learner. MA thesis, Southern Illinois University.

Hieke, A. 1987. Absorption and fluency in native and non-native casual speech in English. *Sound Patterns in Second Language Acquisition*, ed. by A. James and J. Leather, 41-58. Dordrecht: Foris.

Hooper, J. 1976. *An Introduction to Natural Generative Phonology*. New York: Academic Press.

Hua, T. F. 2003. *The Acquisition of English Speech Rhythm by Adult Chinese ESL and EFL Learners*. Doctoral dissertation, University of Hawaii.

James, E. 1976. The acquisition of prosodic features using a speech visualizer. *International Review of Applied Linguistics in Language Teaching* 14: 227-243.

Jun, S. A. 2005. *Prosodic Typology: An Approach through Intonational Phonology and Transcription*. Oxford: Oxford University Press.

Katamba, F. 1989. *An Introduction to Phonology*. London: Longman.

Ladefoged, P. 2005. *Vowels and Consonants*. 2nd edition. Oxford: Blackwell.

Lamontagne, G. A. 1993. *Syllabification and Consonant Co-Occurrence Conditions*. Doctoral dissertation, University of Massachusetts, Amherst.

Lass, R. 1984. *Phonology*. New York: Cambridge University Press.

Lin, Y.-H. 2001. Syllable simplification strategies: A stylistic perspective. *Language Learning* 51: 681-718.

Low, E. L., E. Grabe and F. Nolan. 2000. Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech* 43: 377-401.

McCarthy, J. 2002. *A Thematic Guide to Optimality Theory*. New York: Cambridge University Press.

Major, R. C. 1987. Phonological similarity, markedness, and rate of L2 acquisition. *Studies in Second Language Acquisition* 9: 63-82.

Major, R. 1994. Chronological and stylistic aspects of second language acquisition of consonant clusters. *Language Learning* 44: 655-680.

Marks, J. 1999. Is stress-timing real? *ELT Journal* 53: 191-199.

Miller, M. 1984. On the perception of rhythm. *Journal of Phonetics* 12: 75-83.

Oller, D. K. 1973. The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* 54: 1235-1247.

Pike, K. 1945. *The Intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Prator, C. H., and B. W. Robinett. 1985. *Manual of American English Pronunciation*. New York: Holt, Rinehart, and Winston.

Roach, P. 1991. *English Phonetics and Phonology: A Practical Coursebook*.

Cambridge: Cambridge University Press.

Sole, M. J. 1997. Timing patterns in the production of a foreign language. *Views on the Acquisition and Use of a Second Language: Proceedings of the European Second Language Association 7th Annual Conference*, ed. by L. Diaz, and C. Perez, 539-551. Barcecelona: Universitat Pompeu Fabra.

Tajima, K., R. Port and J. Dalby. 1997. Effects of temporal correction on intelligibility of foreign accented English. *Journal of Phonetics* 25: 1-24.

Tseng, C.-Y. 1990. *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Taipei, Taiwan: Institute of History & Philology Academia Sinica.

Wang, C. 1995. *The Acquisition of English Word-Final Obstruents by Chinese Speakers*. Doctoral dissertation, State University of New York at Stony Brook.

Weinberger, S. 1987. The influence of linguistic context on syllable simplification. *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, ed. by G. Ioup, and S. Weinberger, 279-291. Rowley, MA: Newbury House.

Weinberger, S. 1994. Functional and phonetic constraints on second language phonology. *First and Second Language Phonology*, ed. by M. Yavas, 283-302. San Diego, CA: Singular Publishing Group.

Wong, R. 1987. *Teaching Pronunciation: Focus on English Rhythm and Intonation*. Englewood Cliffs, NJ: Prentice Hall.

Zsiga, E. C. 2003. Articulatory timing in a second language. *Studies in Second Language Acquisition* 25: 399-432.

[Received 19 September 2007; revised 2 January 2008; accepted 11 January 2008]

Department of Applied English
WuFeng Institute of Technology
Chia-yi, TAIWAN
Hsueh-chu Chen: hcchen@mail.wfc.edu.tw

Department of Chinese Studies
National University of Singapore
SINGAPORE
Raung-fu Chung: chscr@nus.edu.sg

台灣學習者之中介語語音時長類型分析

陳雪珠　　鍾榮富
吳鳳技術學院　　新加坡國立大學

本論文旨在分析台灣英語學習者對習得美式英語語音時長類型，從聲學測量討論四個時長變項，分別為音節長度、母音弱化、連音時長以及子音串時長，並提出新的變異指標以比較變項間的差異，以此診斷出學習者習得英語語音時長之困難。在語言類型分類上，國語是聲調語言，以音節時長為導向；而英語是重音語言，以重音時長為依據。由於國語和英語韻律節奏上的差異，分析二者間之中介語對語言習得有重要價值。主要研究結果如下：英語母語人士之音節時長的變異性大於台灣英語學習者，此結論支持了先前之假設：英語較偏向重音時長之韻律，台灣英語學習者受母語影響，較偏向音節時長之韻律。台灣英語學習者之音節時長類型和英語母語人士極為不同，美式英語較常出現弱化母音，而台灣英語學習者較偏好使用全母音，因而造成韻律之差異。此外，台灣英語學習者的說話速度較慢，子音母音間的連音時長較英語母語人士長，單字與單字之間常插入短暫停頓或聲門塞音。從四組受試者中發現其子音串簡化的主要策略為子音刪除，而在台灣學習之低成就英語學習者則特別偏好子音插入。

關鍵詞：音節時長、重音時長、語流韻律、節奏型態、聲學語音學、中介語音韻學