Concentric: Studies in Linguistics
36.2 (July 2010):183-212

Second Language Timing Patterns and Their Effects on Native Listeners' Perceptions*

Hsueh-Chu Chen
The Hong Kong Institute of Education

This study, an extended study of Chen and Chung (2008), investigates the difficulties encountered by Taiwanese learners in English speech timing patterns and identifies critical variables that affect native listeners' perceptions of foreign accents. A total of 30 Taiwanese learners and 10 native speakers of American English were requested to read an article. Six variables—syllable duration, vowel reduction, pause duration, linking, consonant cluster simplification, and speech rate—were acoustically analyzed in five sentences. Ten English listeners were recruited to rate these speech samples. The results of this study showed that the Taiwanese learners displayed very different speech patterns according to the six acoustic variables from those of native English speakers. The perceptual ratings of the six individual variables exhibited a very strong positive correlation with the overall ratings, suggesting that timing patterns were a more holistic impression rather than a discrete component. Speech rate was the primary predictor determining native listeners' perception of foreign accent. If this overall fluency variable was excluded, then vowel reduction and linking duration became the two most heavily weighted variables. A temporal perception model was tentatively proposed to account for the effects of timing variables on native English listeners' judgment of foreign accents.

Keywords: timing phenomena, rhythm patterns, acoustic phonetics, interlanguage phonology, production and perception

1. Introduction

This study, an extended study of Chen and Chung (2008), attempts to investigate the relationship between the production and perception of a foreign accent, and to ascertain the extent to which language transfer factors affect the timing performance of L2 learners and English native listeners' perceptions of foreign accents. The present study extends Chen and Chung's (2008) research in two ways: (1) in the acoustic part, aside from the adoption of four phonetic variables—syllable duration, vowel reduction, linking duration, and consonant cluster duration—another two variables, pause and speech rate, are added to do further comparison in the same study; (2) the present study broadens the use of the variability index (VI) proposed in Chen and Chung (2008).

The major goal of this research is to explore six basic components of the foreign accent of English spoken by English as a foreign language (EFL) learners in Taiwan. The term "foreign accent" might be readily characterized as the subjective impression of a native listener or an advanced student of a foreign language. The precise nature of a foreign accent still remains mostly unexplored even though it has intrigued a number of second language acquisition (SLA) researchers (Anderson-Hsieh et al. 1992, Derwing

* I am indebted to Prof. Raung-fu Chung for his helpful comments. I would also like to thank two anonymous reviewers of *Concentric: Studies in Linguistics* for their constructive suggestions.

et al. 2004, Dickerson 1989, Flege et al. 1995, Munro and Derwing 1995, 1998, 2001, Riney et al. 2005, Sole 1997, Tajima et al. 1997, to name only a few). The findings in the literature, however different they might be, have shed considerable light on this issue. Based on these findings, we can identify six basic components of foreign accent: (a) syllable duration, (b) vowel reduction, (c) pause duration, (d) linking duration, (e) consonant cluster duration, and (f) speech rate. These six components are common in nature, for each of them is closely related to timing.

1.1 Second language (L2) timing features

There have been several studies on the production of L2 rhythm, and, in particular, L2 timing (Anderson-Hsieh and Venkatagiri 1994 and Ueyama 2000). In terms of timing patterns, languages have been classified as either syllable-timed, such as Chinese, or stress-timed, such as English (Abercrombie 1967 and Pike 1945). In a syllable-timed language, every syllable is perceived as taking up roughly the same amount of time, though the absolute length of time depends on the prosody. This tends to give syllables approximately equal stress, and does not generally produce reduced vowels. In a stress-timed language, syllables may last different amounts of time, but there is perceived to be a fairly constant amount of time between consecutive stressed syllables. Although this distinction may seem rather basic, it has been widely adopted in the SLA literature (Celce-Murcia et al. 2001, Cruttenden 2001, Marks 1999, Prator and Robinett 1985, and Roach 2009). Chinese 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 (1968) system, is a bit longer than other tones. In Howie's (1974) study, the falling-rising tone (Tone 3) is longest in vowel duration, while the falling tone (Tone 4) is shortest. However, it is usually simplified to a falling tone 21 in phonetics unless it is located in the final position of an utterance. Fon et al. (2004) found potential similarities between T2 and T3 of Taiwan Mandarin. The contour features that distinguish T2 from T3 have become smeared in certain aspects. Thus the duration of each syllable in Chinese is likely to be similar (Chao 1968, Cheng 1973, and Chung 1990).

Chen and Chung (2004) found that Taiwanese learners of English frequently drop the final consonant or insert an additional schwa between consonants or following the final consonant. This behavior can be attributed to the difference in syllable structure between Chinese and English. The maximum syllable template for a Chinese syllable is CGVN or CGVG, where C denotes nasal consonant, V denotes vowel, G denotes glide, and N a nasal. The coda position in Chinese only allows nasal coda. English can accommodate up to three consonants in the onset and four in the coda position. Both insertion and deletion strategies affect the speech rhythm flow. Kelly (2000) observed that Taiwanese learners tend to have problems in English with vowel reduction and connected speech modification. Taiwanese learners find it difficult to reduce unstressed syllables as well as link and blend words in the same thought group, referring to a discrete stretch of speech that forms a semantically and grammatically coherent segment of discourse. The thought groups or chunks are groups of words which go together to express an idea or thought. In English, pauses and low pitch are usually used to mark the end of thought groups. As observed in Kelly (2000), when Taiwanese learners speak, they are not able to divide speech up into small chunks to help the listener understand messages in an appropriate way. Consequently, the English speech timing patterns spoken by Taiwanese learners have been found to be more syllable-timed.

The way in which Taiwanese learners acquire the timing patterns of a new language, like English, still remains unclear. Most studies have examined one or two rhythmic phenomena (Deterding 2001 and Low et al. 2000). The general mechanism of timing transfer could be better understood by considering multiple timing phenomena within the same study. Chen and Chung (2008) attempted to explore the issue and investigated four timing properties of English spoken by Taiwanese learners at the same time. A new variability index (VI) was proposed and calculated. The results showed a significant disparity in syllable duration and syllable-to-syllable variation between American English and English spoken by Taiwanese learners. Taiwanese learners' consonant-vowel linking duration was much longer than that of the native speakers. Some word boundaries were maintained through the insertion of short pauses and glottal stops. The dominant strategy for consonant cluster simplification across the four groups was consonant deletion, but the EFL low achievers evinced a specific preference for consonant epenthesis. In Chen and Chung's study, however, two important timing factors, pause duration and speech rate, were not explored, which in fact play some roles in native listeners' perception of intelligibility and foreign accent (Anderson-Hsieh and Venkatagiri 1994, Munro and Derwing 1995, 1998, 2001). To explore the effects of timing variables on native English listeners' judgment of foreign accents in a more comprehensive way, six variables altogether are investigated.

1.2 Perceptions of foreign-accented timing

Many claims have been made in the ESL literature that inappropriate timing patterns account for a major part of the intelligibility deficit in L2 learners' utterances (Adams 1979 and Kenworthy 1987). A number of studies have also shown that native speakers of English have greater intolerance of prosodic deviance than of segmental errors when they are asked to assess intelligibility, an accent's acceptability, or overall pronunciation in foreigners' speech (Anderson-Hsieh et al. 1992 and Johansson 1978). Tajima et al. (1997) further demonstrated that nonnative timing patterns in the speech of Chinese learners of English had a significant, negative impact on their intelligibility. When differences in timing were digitally corrected, the learner's speech became significantly more intelligible even though errors in segmental quality persisted.

Anderson-Hsieh and Koehler (1988) found that an increase in L2 speech rates is generally problematic for native listeners. Since L2 speech tends to require more processing time than L1 speech (Munro and Derwing 1995), the use of rapid speech rates may place an additional processing burden on the listener. In contrast, Munro and Derwing (1998) found that the English speech of native speakers of Mandarin was more accented if slowed and that some speakers' accents appeared weaker if they quickened their speech rate. Although the effect of speech rates on foreign accent remains controversial, it seems rather clear that the speech rates do significantly influence listeners' perceptions of a foreign accent.

Based on the aforementioned reviews, several issues relating to second language timing pattern and foreign accent have not been fully discussed. First, since listeners generally perceive foreign accents holistically, it is important to analyze the contribution of various phonetic factors to foreign accent perception in order to improve our understanding of the L2 acquisition process and, possibly, L2 teaching. Prior studies have considered only part of the important dimensions of timing patterns, such as vowel duration, syllable duration, vowel reduction, or pauses (e.g. Anderson-Hsieh and Venkatagiri 1994, Chen 2005, Chen and Chung 2008, and Deterding 2001). Furthermore, most studies of foreign accents have discussed global rating impressions in general terms or as a single factor (such as speech rate) within a comparative analysis (e.g. Anderson-Hsieh and Koehler 1988, Jilka 2000, and Munro and Derwing 1995, 1998, 2001). No studies have comprehensively investigated all six basic components—syllable duration, vowel reduction, pause duration, linking, consonant cluster simplification, and speech rate.

In addition, few studies have investigated the timing patterns of a large number of non-native speakers (the sample sizes were frequently under 20). No studies have even considered the developmental stages of speech timing by comparing learners of varying proficiency levels and learning environments, e.g., learning English as a second language (ESL) versus learning English as a foreign language (EFL). It would be of great value to determine whether and in what way ESL learners with extensive experience in English speaking countries can significantly outperform EFL learners with minimal experience abroad, in terms of the six basic components of foreign accent.

Moreover, little empirical evidence is available in the form of acoustic measurements, which physically characterize the difficulties non-native speakers may experience with English speech timing patterns. Few researchers have considered understanding the nature of foreign accents from the perspective of native listeners and assessed the perceptual significance of several phonetic and phonological variables of foreign accented speech. Finally, not many studies have explored the relationships and interactions among articulatory, acoustic, and perceptual aspects of nonnative English speech in terms of inter-language phonology.

However, two recent studies took these factors into account. Derwing et al. (2004) attempted to determine whether untrained raters' assessments of fluency in low-proficiency Mandarin speakers' speech are related to temporal measures and whether they varied across tasks. Their results demonstrated that the rating data parallel the speech measurements; however, listeners' perception of "the goodness of prosody" does not vary across tasks. Riney et al. (2005) identified some phonetic parameters (e.g. intonation, fluency, speech rate, /l/, and /r/) that correlate with Japanese speakers' perception of English accent and investigated Japanese listeners' perception of English from the standpoint of World Englishes. The results suggested that native English speakers and Japanese listeners perceive the degree of accent in English in fundamentally different ways, each based on different phonetic parameters. The results of these two studies are based on the performance of both Mandarin (unspecified whether from China or Taiwan) ESL learners in Canada and of Japanese learners in Japan. The research in inter-language phonology mainly centers on the developmental nature of the learner's inter-language and on investigating the universality of phonological acquisition patterns across age and language groups. The timing patterns of different languages would behave in different ways. Second language learners with different language backgrounds may have different phonological difficulties. The difficulties in producing native-like timing patterns may vary with acquisition stage. The findings of these studies therefore may not readily apply to Taiwanese learners of English in Taiwan. These two recent studies investigated only a single level of proficiency and ignored the developmental trends, thus the patterns of acquisition stages still remain largely unknown.

There might be several methodological problems in the studies by Derwing et al. (2004) and Riney et al. (2005). Neither study provided detailed descriptions of the acoustic measures. It is doubtful that in Derwing et al.'s (2004) findings the rating data parallel speech measurement. Moreover, it has been claimed that the relationship between human ratings and speech rate is non-linear (Munro and Derwing 2001). However, it is not clear whether these two studies applied non-linear transformation before computing the Pearson correlation and linear regression, which require a linear relationship. Furthermore, like other studies in the literature, these two studies utilized traditional statistics to describe the acoustic measures. This paper adopts the statistical method proposed by Chen and Chung (2008) to measure acoustic variation, which, as shall be illustrated in the method section, outperforms prior methods in many ways. Finally, Riney et al. (2005) recruited only five native speakers, and Derwing et al. (2004) did not include any native English speakers (in the normative sample). Therefore, the human ratings on nonnative speakers in both studies and a distinct lack of reference points might yield inaccurate results. Consequently, the significance of foreign accents' role and the interaction among these variables require further investigation.

This study, an extended study of Chen and Chung (2008), in attempting to provide some resolutions of the above problems, employs the perspective of interlanguage phonology to investigate the relationship between the production and perception of a foreign accent, and ascertains the extent to which language transfer factors affect the timing performance of L2 learners. Four major research questions were raised:

1. In terms of production, what distinctions between Taiwanese EFL learners and native speakers can be made based on acoustic findings?
2. In terms of perception, what distinctions between Taiwanese EFL learners and native speakers can be made based on accent ratings?
3. What is the relationship between the production and perception of Taiwanese foreign accented speech?
4. In terms of the acquisition of second language phonology, to what extent do Taiwanese ESL learners, Taiwanese EFL low achievers, and Taiwanese EFL high achievers differ in their speech timing patterns?

2. Method

This study consists of two interrelated parts. The acoustic part was conducted to explore six timing patterns. The acoustic data obtained from the Taiwanese learners was then compared with the norm of English native speakers to identify difficulties in and specific to English speech timing patterns used by Taiwanese learners. Part of the findings in this section has been published in Chen and Chung (2008). Therefore, some detailed procedures of data collection were not repeated here. The perceptual rating portion as conducted to ascertain the relationship between the acoustic measures and the perceptual ratings so as to determine the phonetic and phonological components of foreign accented speech.

2.1 Participants

For the acoustic part, the speech samples of 40 participants were acoustically analyzed. To ensure diversity of speech patterns, four groups of participants were recruited to speak a long English passage. These included ten Taiwanese learners with a low English proficiency level (denoted as EFL-L), ten Taiwanese learners with a high English proficiency level (denoted as EFL-H), 10 ESL Taiwanese learners (denoted as ESL), and ten native speakers of English (denoted as NS). The EFL-L group consisted of students from two-year junior colleges 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 GEPT (equivalent to 213 computer-based TOEFL scores), although they had never studied in English-speaking countries. The ESL group consisted of undergraduate students from the University of California at Berkeley who came from Taiwan. They had a TOEFL score above 600 points (equivalent to 250 computer-based-test points) (M = 273), or a SAT verbal score 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 undergraduate students from UCB who were native speakers. They were monolingual English speakers and all were Californians from the US West Coast. Their speech samples were treated as the norm to which those of the other three groups were compared.

In the perceptual rating part, ten native speakers of American English rated the speech samples of the 40 speakers. They were experienced ESL teachers in California (6 men and 4 women, aged between 28 and 42). They had received their master degrees in applied linguistics or related fields, and had a basic understanding of phonetics. They had been teaching English as a second language for five years on average, ranging from three to eight years.

2.2 Materials

The materials employed in the acoustic part included a short questionnaire about personal background and a diagnostic passage for acoustic measurement. A widely used diagnostic passage with 14 sentences was selected from *Teaching Pronunciation* (Celce-Murcia et al. 2001) as the reading material. This passage contained five types of sentences (i.e. Wh- questions, Declarative sentences, Yes-No questions, Tag questions, and Closed-choice alternative questions) which can help eliminate or counterbalance the effects of different sentence types on the timing patterns produced by Taiwanese learners. The five sentences are illustrated as follows:

1. Wh questions
   Why do people usually have an accent when they speak a second language?
2. Declarative sentences
   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. Data collection procedure

The participants were informed of the research goal and procedure of recording before the session began. Each participant then read the passage individually in a quiet room and was recorded on a notebook computer with the Praat software (Version 4.0.48), developed by Boersma and Weenink (2003). Each participant was asked to read the diagnostic passage and allowed to request help and practice words he/she was not familiar with before recording began. It took about fifteen minutes for each participant. The author then analyzed the speech data acoustically and auditorily.

2.4 Measurement of the six timing variables

The speech data were segmented into syllables and subsequently analyzed for duration using the Praat software. The total number of syllables analyzed was 3,320 (4 groups × 10 participants × 83 syllables in 5 sentences). The whole syllable duration rather than the vowel duration was measured because of a tendency for schwa deletion in fast speech and the case of syllabic consonants.

The segmentation of the syllables generally followed two basic principles. First, the duration of the syllables was measured based on the placement of 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. Second, when it was difficult to ascertain syllable boundaries, the phonotactic constraints and maximum onsets principle (MOP) were followed. According to Roach (2009), this principle states that where two syllables are to be divided, any consonants between them should be attached to the right-hand syllable as far as possible.

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 words 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 syllables and the latter being among the shortest (Bolinger 1986 and Katamba 1989). It was hypothesized that native speakers of English would produce a greater duration in tonic syllables than unstressed syllables, whereas the Taiwanese speakers would not differentiate one from the other. Ratios, instead of absolute values, were used to compare syllable durations. Ratios were computed by dividing unstressed syllables and tonic syllables by the sentence duration. A mean ratio of unstressed syllables and tonic syllables was subsequently computed for each of the four groups of speakers.

“Pause” is defined as an interval of the oscillographic trace, where the amplitude is indistinguishable from that of the background noise (Duez 1982:13). Only those pauses greater than 100 ms, the cut-off for pause (Griffiths 1991), were considered as a “pause.” Pauses were identified by ear perception and the acoustic cue of intensity. Detailed acoustic descriptions have been discussed in Chen and Chung (2008).

As Hieke (1987) observed, linking can occur in English between two consonants, between a consonant and a vowel, or between two vowels. Due to the high frequency of CV linking present in daily speech, this study is limited to CV linking only. For each participant, a computation was made based on the measurement of linking 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 participant, a computation was made, based on the measured value of the final consonant cluster duration.

Speech rate is defined as the total duration of a sentence (including pauses). The most common measurements of speech rate are syllable per second (sps) and words per minute (wpm) (Buck 2001). The total duration and SPS were calculated. Since measurement of syllable duration is the major focus in this study and the selected speech samples from the four groups were short and identical, the measurement of WPM was not employed here.

2.5 Calculations of the six timing variables

A total of six timing variables—syllable duration, vowel reduction, pause duration, linking duration, consonant cluster duration, and speech rate—were measured. We did not employ the commonly used indices of VI_1 and VI_2 (see Appendix for details) to describe the six variables, because these two indices are actually self-referenced and thus cannot accurately represent the difference in the timing variables between native speakers and non-native speakers. Instead, an index was adopted from Chen and Chung (2008) for this study which makes a norm-referenced comparison possible and is referred to as variability index-3:

(1) $$VI_3 = \sqrt{\sum_{k=1}^{K} (X_k - E_k)^2 / K}$$

where X_k is the k-th component, and 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. VI_3 is a norm-referenced variability index because it involves the norm (E_k). When a speaker gives exactly the same timing pattern as the norm, VI_3 will equal zero. The larger the VI_3, the greater the speaker's timing pattern will deviate from the norm. It was expected that non-native speakers would have a larger VI_3 than native speakers.

2.6 The perceptual rating part

In this part of the study, a brief questionnaire about personal background, an instruction sheet for rating speech samples, and a rating homepage were used. The instruction sheet for human rating was provided to the ten native listeners in the orientation session. They then assessed the degree of foreign accent in the speech samples made by the 40 speakers along a seven-point rating scale (1 = very heavy foreign accent and 7 = no foreign accent at all) for each of the six timing variables as well as a variable of overall rhythm. In order to eliminate possible order effects, the orderings of the speech samples given to the native listeners were randomized and counter-balanced between the four groups of participants, between the 10 participants (within groups), and between the five sentences (within participants). Each speech sample was rated by three listeners. In order to provide a flexible and comfortable rating environment, all the work was conducted via the Internet. This also had the advantage of facilitating the presentation of sound files (in the WAV format) and the compilation of results. The perceptual rating task was self-paced. The native listeners were allowed to leave and re-enter the rating sheet after each completed task.

3. Results

3.1 Acoustic measures

This section measures and compares the six timing variables from the four groups. In these calculations the mean timing pattern of the native speakers serves as the norm, and those of the Taiwanese learners are then compared to the norm. The syllable durations of the four groups are first measured and then compared.

Table 1 illustrates the mean, standard deviation, minimum value, and maximum value of VI_3 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_3 than NS (M = 0.059). EFL-L generated more than twice the deviance in syllable duration from the norm than 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 participants deviated from the norm far more than NS participants. Moreover, 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 standard deviation for these groups. The small within-group variation in NS indicated that these 10 native speakers were highly clustered around their mean and therefore justified that the mean could serve as the "norm."

Table 1. Means, standard deviations (SD), maximum values (Max), and minimum values (Min) of VI_3 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

The score patterns of the other five timing variables are very similar to those of syllable duration. Due to their similarity and space constraints, a detailed summary of the statistics is not reported here, but is available in Chen (2006). The mean scores of VI_3 statistics and their ratios to NS for the six timing variables are in turn summarized in Table 2. According to the ratios on the six variables, EFL-L performed worst in speech rate (the largest ratio), followed by pause, linking, vowel reduction, consonant cluster simplification, and finally syllable duration (the smallest ratio). Since there was a high correlation between the six variables, and EFL-L could not perform well with one variable, it is reasonable to assume that they would not perform well with the other variables, either.

Table 2. Mean VI_3 statistics and ratios to the native speakers on the six timing variables

	Syllable Duration		Vowel Reduction		Pause Duration		Linking Duration		Consonant Cluster		Speech Rate	
Group	Mean	Ratio	Mean	Ratio	Mean	Ratio	Mean	Ratio	Mean	Ratio	Mean	Ratio
EFL-L	0.142	2.41	0.387	4.03	0.241	7.77	0.320	5.93	0.175	2.92	3.547	11.41
EFL-H	0.084	1.42	0.207	2.16	0.093	3.00	0.146	2.70	0.069	1.15	1.373	4.41
ESL	0.088	1.49	0.184	1.92	0.086	2.77	0.096	1.78	0.073	1.22	0.839	2.70
NS	0.059	1*	0.096	1*	0.031	1*	0.054	1*	0.060	1	0.311	1*

Note: *treated as the reference

EFL-H and ESL performed quite differently on speech rate and linking, but similarly on syllable duration, vowel reduction, pause, and consonant cluster duration. This is also shown in Figure 1, in which the slopes decreased rapidly from EFL-H to ESL only on speech rate and linking. This finding might reveal that in English acquisition, syllable duration, vowel reduction, pause, and consonant cluster duration are much easier to pick up than speech rate and linking. There are several possible explanations. First, EFL-H’s greater explicit phonological knowledge, stemming from the result of formal instruction in English linguistics along with additional incentives, may have contributed to their performance in the identification of the English timing patterns. Consequently, they performed equally well—compared to ESL—in syllable duration, vowel reduction, pause, and consonant cluster duration. Nevertheless, non-native EFL teachers probably provided deficit language input in appropriate speech rate and linking, or possibly gave fewer chances for EFL learners to speak out. These retard the progress of EFL learners seeking to improve speech rate and linking. Moreover, ESL could speak more fluently, as reflected in the linking of segments and acceleration of speech rate, because of greater exposure to native English speakers in everyday life.

Finally, the learning plateau may provide another possible explanation for this phenomenon. It is commonly observed that there are different stages in the language learning curve. The initial stage of the curve rises slowly as learners become familiar with basic components of the language. The steep ascending phase occurs when learners have enough experience with rudimentary components to start putting them all together. Rapid progress follows until the skill “hits a ceiling” or stabilizes at a higher proficiency level. Accordingly, the improvement of phonetic timing patterns may not seem hard in the beginner stage, but once a certain degree of proficiency is attained, progress becomes more difficult. The ESL learners in the current study had hit an intermediate-high level of language proficiency, and their progress of timing patterns thus slowed down. For this reason the ESL group did not greatly surpass the EFL-H group.

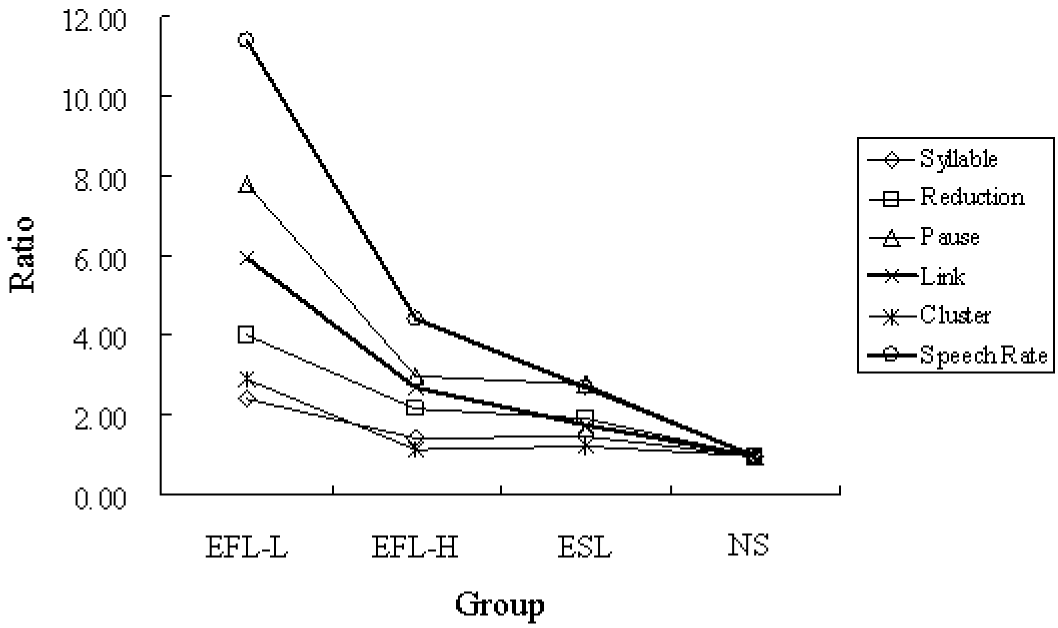

Figure 1. Ratios of VI_3 on the six major variables of the four groups

3.2 Perceptual ratings

3.2.1 Inter-rater reliability

Ten native speakers of English rated the speech samples of the 40 speakers on six specific variables and overall performance along a 7-point rating scale. The six variables were: (a) syllable duration (H-SD), (b) vowel reduction (H-VR), (c) pause (H-PA), (d) linking (H-LI), (e) consonant cluster (H-CC), and (f) speech rate (H-SR). The raters were also asked to give an overall rating of the speech, which is referred to as the human overall rating (H-OR). In this study, every subject was asked to read the same passage. Five sentences were selected from it and each sentence was assessed by three independent raters. Reasonably high inter-rater reliability was thus critical to the success of this study.

According to Erwin (1991), reliability in performance rating scales can be defined as the degree to which observers agree when rating the same performance independently. An inter-rater reliability of .70 is usually considered a minimum requirement and .80 as satisfactory. According to Table 3, the inter-rater reliabilities in this study, which were the mean Pearson correlation coefficients between all possible pairs of raters , varied between .63 and .84 (*M* = .78). It would seem that these raters worked fairly harmoniously, except for rater 10, who had an inter-reliability of .63. Fortunately, because every speech sample was rated by three independent raters, the impact of the low inter-rater reliability by rater 10 could be minimized.

Table 3. Correlations between raters and inter-rater reliabilities for the 10 raters

Rater	1	2	3	4	5	6	7	8	9	10
1		.86	.82						.86	.64
2	.86		.69	.80						.51
3	.82	.69		.87	.90					
4		.80	.87		.82	.85				
5			.90	.82		.76	.72			
6				.85	.76		.73	.71		
7					.72	.73		.81	.82	
8						.71	.81		.78	.71
9	.86						.82	.78		.66
10	.64	.51						.71	.66	
Inter-Rater Reliability	.80	.72	.82	.84	.80	.76	.77	.75	.78	.63

3.2.2 Perceptual ratings

Table 4 summarizes the means and standard deviations of human ratings on the six timing variables and the variable of overall rating. Figure 2 shows box-plots of the overall rating. EFL-L (*M* = 2.80), EFL-H (*M* = 4.50) and ESL (*M* = 5.44) had smaller overall mean ratings than NS (*M* = 6.94). A one-way ANOVA showed that these four groups had a statistically significant difference in the overall rating ($F_{3, 36} = 49.19$, $p < .001$). A post-hoc Tukey's honestly significant difference test also revealed that the four groups had a statistically significant difference from each other. The ratings given to the NS group served as a reference to check whether the listeners used the rating scales appropriately. The ratings of these items were of interest only in the sense that they gave us a means of identifying listeners who may not have followed the instructions and whose data might therefore be unsuitable for analysis. A mean rating of 6.94 on a 7-point scale for NS thus implied that the ratings were very reliable.

Table 4. Means and standard deviations of human ratings on the six timing variables and the overall rating

		Human Ratings						Overall Human Rating
Group		H-SD	H-VR	H-PA	H-LI	H-CC	H-SR	H-OR
EFL-L	Mean	3.12	2.98	3.42	3.10	2.82	2.90	2.80
	SD	0.90	0.86	0.89	0.89	0.88	0.76	0.83
EFL-H	Mean	4.52	4.34	4.72	4.64	4.60	4.72	4.50
	SD	0.85	0.84	1.01	0.81	1.01	0.98	0.74
ESL	Mean	5.66	5.48	5.36	5.46	5.52	5.56	5.44
	SD	0.83	0.97	0.88	0.95	0.98	1.03	1.09
NS	Mean	6.92	6.94	6.86	6.92	6.92	6.88	6.94
	SD	0.14	0.13	0.21	0.14	0.10	0.19	0.13

Note: H = human rating; SD = syllable duration; VR = vowel reduction; PA = pause; LI = linking; CC = consonant cluster; SR = speech rate; OR = overall.

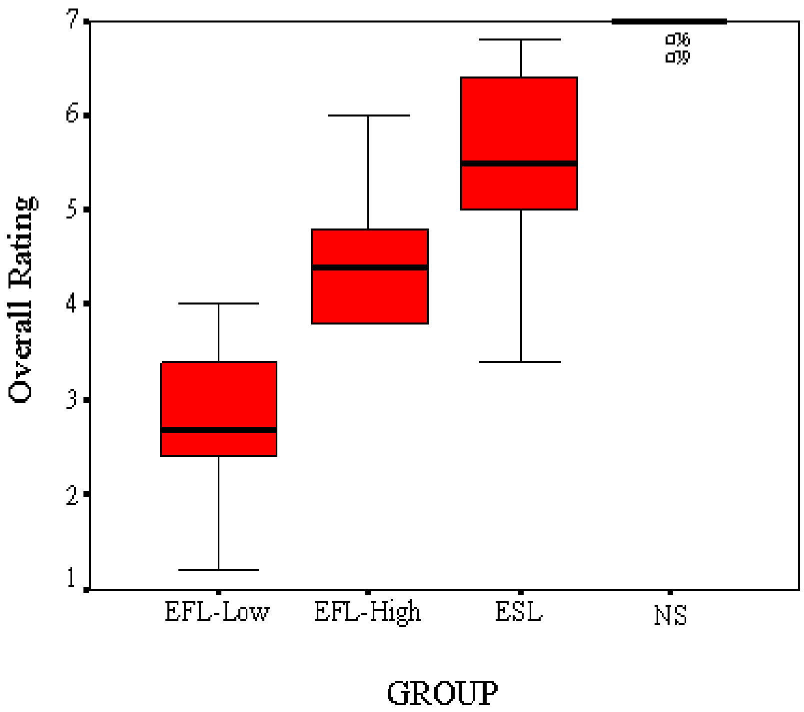

Figure 2. Box-plot for the overall ratings of the four groups

Table 5 reveals that the ratings of six variables had strong positive correlations (from .80 to .95, with a median of .86) with the overall rating. Among the six variables, speech rate (H-SR) was the most highly correlated with the overall rating (r = .95). Although the correlation pattern is reasonable, it also demonstrates the limitation of subjective impressions. Indeed, these raters were not always able to differentiate the specific ratings from the overall ratings. The subjective impression of timing patterns was consequently more holistic than discrete. As each of the five sentences was rated by three raters, each variable had 15 data points. When averaging these 15 data points to yield a mean value, it was found that the correlations among the six variables and the overall ratings were between .96 and .99 (with a median of .98). Hence, the overall rating was considered as the best index of human perception.

Table 5. Correlations among human ratings on the six timing variables and the overall judgment

	H-SD	H-VR	H-PA	H-LI	H-CC	H-SR
H-VR	.95					
H-PA	.85	.87				
H-LI	.86	.86	.85			
H-CC	.84	.84	.80	.85		
H-SR	.90	.90	.88	.86	.85	
H-OR	.91	.92	.87	.88	.86	.95

Note: H = human rating; SD = syllable duration; VR = vowel reduction; PA = pause; LI = linking; CC = consonant cluster; SR = speech rate; OR = overall.

3.3 Correlation between acoustic variables and overall human rating

Prior to the computation of the Pearson correlation between the six acoustic variables and the overall human rating (denoted as H-OR), we examined their relationship with a scatter plot. It was found that the six acoustic variables were nonlinearly rather than linearly correlated with H-OR. Hence, the logarithm transformation was made onto H-OR (hereafter referred to as LH-OR). After the transformation, the relationship between the six acoustic variables and LH-OR became linear such that the subsequent Pearson correlation and multiple regression could be applied (these two methods assume a linear relationship) to them.

Table 6 presents the correlations among the six acoustic variables and LH-OR. The correlations between the overall human rating and the six acoustic variables were moderate to high ($-.90 \leq r \leq -.67$; $p < .01$). It is also important to note that the correlations among all the six acoustic variables were moderately to highly correlated ($.56 \leq r \leq .95$; $p < .01$). The rankings (in absolute value) of the correlations between the six acoustic variables with LH-OR are: speech rate > syllable duration > linking > pause > vowel reduction > consonant cluster. The strongest correlation among the six acoustic variables occurred between speech rate and linking ($r = .95$). The second strongest correlation occurred between speech rate and pause ($r = .92$). The strong negative correlation between speech rate and LH-OR suggested that the longer the duration of the speech produced, the lower the rating given to it by the native listeners.

Table 6. Correlations among the six acoustic timing variables and the overall human judgment

	A-SD	A-VR	A-PA	A-LI	A-CC	A-SR
A-VR	.70					
A-PA	.86	.65				
A-LI	.86	.56	.89			
A-CC	.76	.60	.71	.74		
A-SR	.92	.62	.92	.95	.77	
LH-OR	-.87	-.74	-.85	-.86	-.67	-.90

Note: A = Acoustic; SD = syllable duration; VR = vowel reduction; PA = pause; LI = linking; CC = consonant cluster; SR = speech rate; LH-OR = logarithm transformed overall human rating.

These findings are consistent with those of Anderson-Hsieh et al. (1992), who observed that prosody, a larger phonological component, rather than vowel, consonant and syllable structure, is the most decisive factor in the apprehension of a foreign accent. This would also be consistent with Munro and Derwing’s (1998, 2001) conclusion that slowing down non-native speech is unlikely to have a positive impact on listener’s perception, and may actually result in a worse rating. A listener may be inclined to notice phonological errors in very slow speech and assign poorer ratings to the speech. On the other hand, a slight increase in speech rate of non-native speakers may improve their comprehensibility. That is, as long as the speech rate remains manageable from a processing standpoint, listeners should benefit from the acceleration of speech rate.

Although speech rate may appear to be a global phenomenon across speech flow on the surface, it is in fact very much an integrated part of prosody organization constrained by each prosodic level. The reorganization of syllable structure, such as consonant deletion, vowel insertion or re-syllabification, may cause a change in utterance timing, and thus slow down the speech rate. That is why speech rate has the strongest correlation with perceived foreign accents. It should be noted that followed by speech rate are syllable duration, linking duration, and pause duration in that order. All these three variables have almost the same correlation values with overall human ratings as well as being highly correlated with one another. Relatively speaking, smaller components, such as vowel reduction and consonant cluster, have weaker correlations with overall human ratings.

The ratings given for each of the six variables demonstrated a very strong positive correlation (from .96 to .99) with the overall human ratings. One interpretation for this correlation might be that ESL teacher raters cannot effectively differentiate the specific ratings from the overall ratings. However, the results suggest that timing patterns, or prosody in general, are a more holistic impression than a discrete component of speech. It would appear then that human beings encounter greater difficulty in differentiating one variable from another.

3.4 Multiple regression analysis

The next issue that concerned us was the question: If speech rate, a relatively holistic variable, was excluded, what would be the most important variables that affect human overall rating? Multiple regression analyses were thus conducted to assess the predictive power of the acoustic variables on LH-OR, in which LH-OR is the criterion variable and the five acoustic variables are the predictive variables. The speech rate variable was not included as a predictor because it was a combination of the other five acoustic variables. Thus, its effect on LH-OR could be assessed through the effects of the other five acoustic variables. When all the five acoustic variables were entered as the predictors, the coefficient of determination (R^2) was .86 and the adjusted R^2 was .84. As these five predictors were highly correlated, some predictors became statistically insignificant, and were thus removed. The final model contained only two statistically significant predictors: linking and vowel reduction, with a R^2 of .84 and an adjusted R^2 of .83, which was statistically significantly different from zero ($F_{2,37} = 96.73$, $p < .001$). Obviously, the final model with only two acoustic variables had a predictive power practically identical to the model with all the five acoustic variables. Hence, the other three variables did not provide additional predictive information about LH-OR, given that linking and vowel reduction were in the model. The non-standardized and standardized equations of this final model are

$$(2) \quad \text{LH-OR} = 2.065 - (2.070 \times \text{linking}) - (1.029 \times \text{vowel reduction})$$

$$(3) \quad \text{LH-OR} = (-.654 \times \text{linking}) - (.374 \times \text{vowel reduction})$$

respectively. The standardized regression coefficients for linking and vowel reduction were -.654 and -.374, indicating that linking was more predictive than vowel reduction.

When a participant has a level of linking and vowel reduction exactly identical to that of an average participant in EFL-H (0.146 and 0.207, respectively), according to Equation 2, we expect that the participant will receive an overall rating of 4.71 points on the 7-point scale:

(4) $\text{LH-OR} = 2.065 - (2.070 \times 0.146) - (1.029 \times 0.207) = 1.55$

(5) $\text{H-OR} = \exp(1.55) = 4.71$

where exp is the exponential function and is roughly equal to 2.718. This example demonstrates the strong predictive power of this model relative to the overall human rating, because the predicted value was 4.71, which was very close to the observed mean value of the overall human rating for the EFL-H group (M = 4.50).

In short, the most important factor in predicting the perception of a foreign accent is speech rate. If the overall fluency variable is excluded, then it would appear that native speakers give the most weight to a combination of vowel reduction and linking in the detection of a foreign accent.

4. Discussion and conclusion

The speech of Taiwanese learners has long been described as "bullet-like" when they speak English. However, little empirical evidence exists to explain what exactly makes their speech timing patterns different from those of native English speakers. In Chen and Chung's (2008) study, they acoustically measured and discussed four timing variables produced by Taiwanese learners. Pauses and speech rate, two timing factors playing a role in the perception of intelligibility and foreign accent (Anderson-Hsieh and Venkatagiri 1994 and Munro and Derwing 1995, 1998, 2001), however, were not examined. To fully ascertain the effects of timing variables on native English listeners' judgment of foreign accents, this study compares the production of English timing patterns of Taiwanese learners and English speakers by analyzing six acoustic variables, and explores the ways in which these patterns contribute to the perception of foreign accented speech. Based on its findings, two issues of prosodic transfer and general fluency are worth further discussion.

4.1 Prosodic transfer

The most likely explanation for Taiwanese learners' difficulty in the production of English speech timing patterns would be language transfer. This study finds that the acquisition of 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 suggests that as phonological transfer occurs, the postlexical processes of the L1, rather than the lexical alternations, cause the most interference (Cebrian 2000 and Weinberger 1994).

This study explores two aspects of language negative transfer: (a) the absence of vowel reduction, and (b) the absence of linking. First, one must distinguish between full and potentially reduced vowels in American English and the near absence of vowel reduction observed amongst Taiwanese English learners. In fact, Taiwanese learners differ markedly from American speakers (see Figure 2). Since most syllables are produced with full tones in Taiwan Mandarin, Taiwanese learners tend to assign tones to syllables. Taiwanese learners thus may encounter difficulty in the reduction of unaccented syllables, including those that bear some stress and those that are completely unstressed. Sentences containing long stretches of weak syllables may therefore pose greater rhythmic difficulty than those containing strong syllables.

Second, the timing patterns of connected speech should be more difficult than words in isolation. The results of this study corroborate the notion that Taiwan Mandarin patterns of word-to-word articulation are different from linking and consonant cluster durations in native English. Taiwanese learners employ strategies to maintain the acoustic cues that maximize lexical contrasts among the word-final consonant clusters (Chen and Chung 2008). In American English, however, word-to-word timing is not organized so as to promote the preservation of a lexical contrast. Native English speakers generally place contrasts between word-final clusters in the lexicon and in segmental articulation, but often eliminate them through overlapping, flapping or linking the segments.

4.2 Speech rate, pause and general fluency

Among the six acoustic variables, the perception of accent generated more statistically significant results for speech rate and pause than the other four variables. Temporal aspects, such as speech rate and pause, can be correlated with general fluency to some extent (Riazantseva 2001). Fluency refers either to the quantity or amount of language produced within a given time or to the rapidity and smoothness of speech output (Lennon 1990). When non-native speakers employ a pausing pattern outside the norm of their target language, their speech might be perceived as less fluent than it could be because it lacks native-like rapidity and smoothness. Taiwanese learners' performance within the context of the four variables (syllable duration, vowel reduction, linking, and consonant cluster simplification) thus accounts for the results of L1 negative transfer and universal markedness. However, the scope of their exposure to and practice of English best explains the measurements of their speech rate and pause duration. If one accepts this explanation as reasonable, the different performance of the

EFL-H group and the ESL group can then be readily explained. Consequently, EFL teachers should strive to expose their students to a considerably wider variety of situations that demonstrate native patterns of communication. These students can then better mimic native speech patterns.

4.3 The temporal perception model

A variety of studies have shown that simple counts of segmental errors and prosodic assessments correlate well with listeners' ratings of L2 speech for such dimensions as foreign accent and comprehensibility (Anderson-Hsieh et al. 1992, Brennan and Brennan 1981, and Munro and Derwing 1995). Although it is clear that speaking rate, pause, linking, stressed syllable duration, vowel reduction, and consonant cluster simplification can influence the perception of L2 speech, the available evidence does not permit us to ascertain the extent to which these timing variables influence the perception of accent.

The purpose of multiple regression analysis is to obtain an equation that is simple enough while at the same time providing enough power of prediction. The simplest equation we find is the use of speech rate as the only predictor for foreign accent, because speech rate has the highest correlation with foreign accent. The other variables are so highly correlated with speech rate that they can only make little "additional" contribution to the prediction of foreign accent, given that speech rate is in the equation.

In order to clarify the contribution of the other variables, we have to remove speech rate from the equation and use the other variables as predictors. In doing so, we find that vowel reduction and linking duration are the most powerful predictors, and the other three variables (pause, stressed syllable duration, and consonant cluster duration) cannot make substantially "additional" contribution, given that vowel reduction and linking duration are already in the equation.

Although pause, stressed syllable duration, and consonant cluster duration are not included in both equations, it does not mean that they cannot predict foreign accent. As shown in Table 4, stressed syllable duration and consonant cluster duration are also highly correlated with foreign accent. They are not included in equations because they cannot make substantially additional contribution to the prediction, given that the other variables are already in the equations.

Given that vowel reduction and linking duration are the predictors, pause, stressed syllable duration, and consonant cluster duration become less important to the prediction of foreign accent. Therefore, they are not included in the equation.

It should be noted that pause is more important to the prediction than vowel reduction and linking duration, when speech rate is already a predictor. This is mainly because pause, stressed syllable duration, and consonant cluster duration are less highly correlated with speech rate than vowel reduction and linking duration.

Based on the findings, this study takes the first step in developing a tentative model that demonstrates the effect of six timing variables on native English listeners' judgments of foreign accents. This model is referred to as the temporal perception model of foreign accents, shown in Figure 3. According to this model, when non-native speakers produce accented speech, native listeners would perceive the utterance through the prism of a foreign accent. In processing the phonological aspects of language, speech rate will first be detected, followed by the combination of linking and vowel reduction, and finally the rest of the timing variables. It should be noted again that pause is placed in the third candidate group because when speech rate is selected to be the most critical factor in the perception of a foreign accent, the impact of pause decreases. In contrast, the relative impact of the combination of linking and vowel reduction becomes greater than that of other variables. The ranking of consonants and vowels also merits some consideration even though it was not formally evaluated in this study. Prior studies have shown that deviance in segments rather than prosody was found to be least significantly related to the global ratings (Anderson-Hsieh et al. 1992, Magen 1998, and Major 1986). This explains the reason why the segments, both vowel and consonant, are placed in the last candidate group. The temporal perception model of foreign accent is a daring proposal and of course needs further validation from more samples of cross-dialect and cross-language speech that can be employed in future studies.

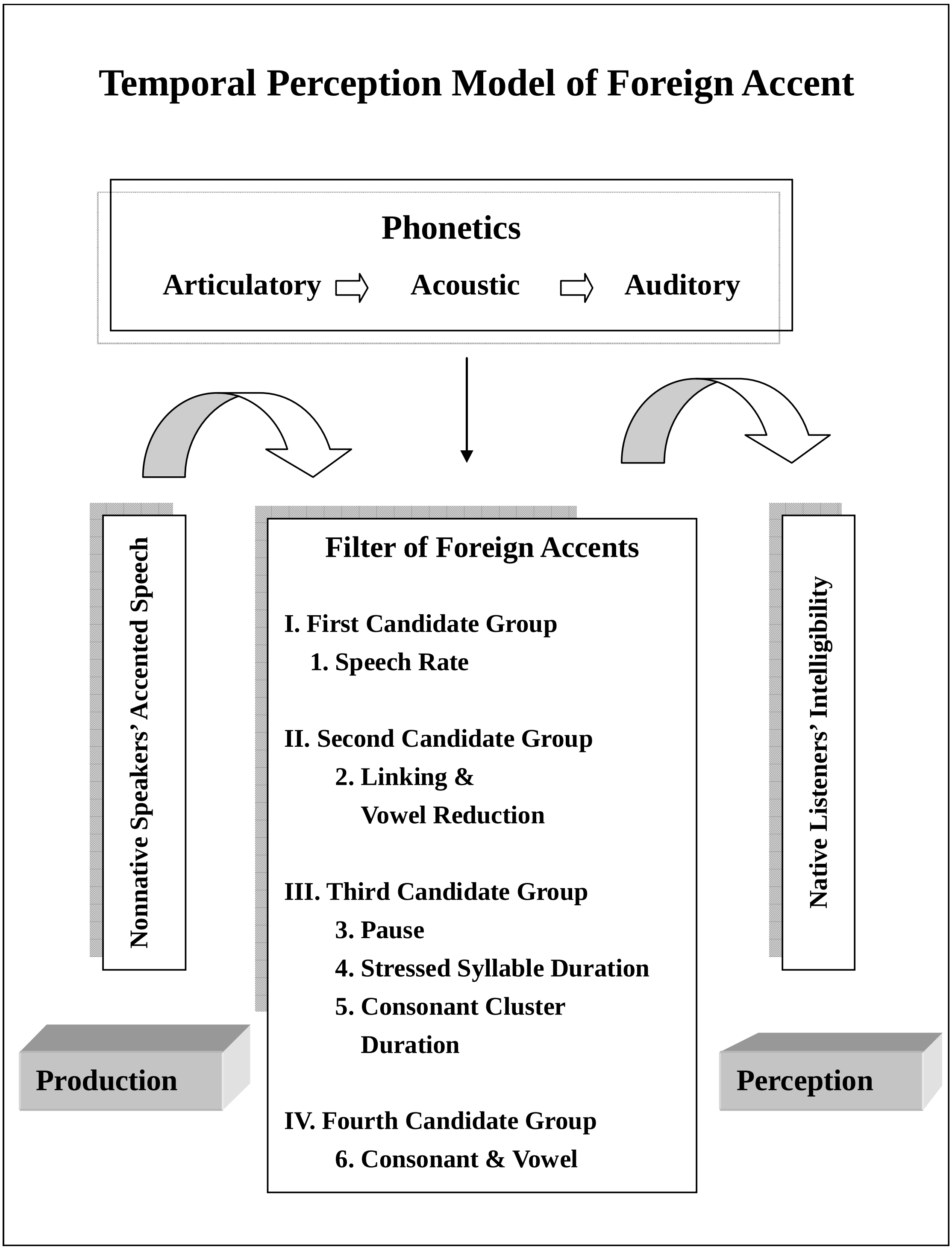

Figure 3. The temporal perception model of foreign accents

In terms of pedagogical implication, a better understanding of the phonetic and phonological factors that weigh most heavily in native speaker reactions to non-native speech can be used to establish a hierarchy of priorities for teaching pronunciation to second or foreign language learners. The results are expected to help sensitize ESL/EFL language professionals to the specific difficulties that Taiwanese learners encounter in producing new speech rhythm. In effect, this study hopes to encourage language teachers and researchers to reexamine the ways in which timing patterns are introduced in contemporary classroom practices and to further develop appropriate teaching materials for Taiwanese learners.

Appendix: Comparison of the three variability indices

In the literature, two indices have been employed to ascertain the variability of syllable duration (Deterding, 2001). The first variability index, denoted as VI_1, is defined as:

$$\text{(A.1)}\quad \mathrm{VI}_1 = \sqrt{\sum_{k=1}^{K-1} (X_k - \overline{X})^2 / (K-2)}$$

where X_k is the k-th syllable duration; $\overline{X}$ is the mean syllable duration for that sentence; and K is the number of syllables. In statistical terms, VI_1 is equivalent to "standard deviation." Hence, a large VI_1 indicates a large variation among the targeted variables, which is syllable durations in a specific sentence for a specific participant in this study. Deterding (2001) argued that VI_1 has certain limitations. If a speaker produces several syllables quickly, followed by several syllables slowly, VI_1 would be high, even though the speaker has been using syllable-timed rhythm with a sudden change of pace in the middle.

To resolve this problem, Deterding proposed another index to take into account the change in duration between adjacent syllables, which is referred to as VI_2:

$$\text{(A.2)}\quad \mathrm{VI}_2 = \frac{1}{K-2} \sum_{k=1}^{K-2} \left| X_{k+1} - X_k \right|$$

where X_k and X_{k+1} are the k-th and k+1-th syllable durations, respectively; and K is the number of syllables. In both VI_1 and VI_2, the last syllable has to be excluded in order to eliminate the final lengthening effect. Moreover, in order to avoid confounding each speaker's speech rate, all utterances have to be divided by the speech rate of that participant. A smaller value of VI_1 or VI_2 indicates a more consistent rhythmic pattern within a sentence for that particular participant. Unlike VI_1, which focuses on the overall variation among measures, VI_2 focuses on the variation between two adjacent measures. In theory, VI_1 and VI_2 can yield very different values, yet they often produce very similar values in practice (Chen 2005).

Note that VI_1 and VI_2 are both "self-referenced" measures, in that by definition they involve only the specific participant himself/herself, independent of the others. They do not describe the degree to which a non-native speaker deviates from the norm of native speakers. Moreover, VI_1 and VI_2 are designed specifically for measuring syllable duration. They are not applicable to the other five acoustic variables that are examined in this study. For these five variables, another statistical index is required. To resolve these problems, we propose a new variability index called VI_3:

$$\text{(A.3)} \quad \mathrm{VI}_3 = \sqrt{\sum_{k=1}^{K} (X_k - E_k)^2 / K}$$

where X_k is the k-th component, and E_k is the mean of the k-th component over the native speakers of English (treated as the norm); and K is the number of components in the sentence. As VI_3 includes the norm (E_k), it is a "norm-referenced" variability index. If a speaker gives exactly the same timing pattern as the norm, VI_3 will equal zero. The larger the value of VI_3, the greater the speaker's timing pattern will deviate from the norm.

VI_3 has several theoretical and practical advantages over VI_1 and VI_2. First, VI_3 is a norm-referenced measure that is in accordance with the nature of individual perception of foreign accents, whereas VI_1 and VI_2 are self-referenced measures independent of the norm. Regardless of the division by the speech rate or not, VI_1 and VI_2 are self-referenced and their interpretations rely on self comparison and cannot be directly compared to the norm. Second, VI_3 can be applied to all the six acoustic variables, whereas VI_1 and VI_2 can be applied to syllable duration only. Third, without VI_3, one must compile a variety of statistics for the other five acoustic variables, which may not be rooted in the same base. The comparison of these six acoustic variables thus becomes more complicated.

References

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

Adams, Corinne. 1979. *English Speech Rhythm and the Foreign Learner.* The Hague: Mouton.

Anderson-Hsieh, Janet, and Kenneth Koehler. 1988. The effects of foreign accent and speaking rate on native speaker comprehension. *Language Learning* 38:561-613.

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

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

Boersma, Paul, and David Weenink. 2003. Praat: doing phonetics by computer. Software version 4.0.48. Retrieved October 28, 2003, from http://www.praat.org.

Bolinger, Dwignt. 1986. *Intonation and Its Parts: Melody in Spoken English.* Palo Alto: Stanford University Press.

Brennan, Eileen M., and John Stephen Brennan. 1981. Accent scaling and language attitudes: Reactions to Mexican American English speech. *Language and Speech* 24:207-221.

Buck, Gary. 2001. *Assessing Listening.* Cambridge: Cambridge University Press.

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

Celce-Murcia, Marianne, Donna Brinton, and Janet Goodwin. 2001. *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages.* Cambridge: Cambridge University Press.

Chao, Yuen-ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chen, Hsueh-chu. 2005. The acquisition of American English rhythm by Taiwanese EFL learners. *English Teaching and Learning* 29:1-23.

Chen, Hsueh-chu. 2006. Interlanguage Phonetic Timing Patterns and Their Effects on Native Listeners' Perceptions. Doctoral dissertation, National Kaohsiung Normal University, Kaohsiung.

Chen, Hsueh-chu, and Raung-fu Chung. 2004. Effects of task formality on consonant cluster simplification for Taiwanese learners with limited English proficiency. *English Teaching and Learning* 28:67-93.

Chen, Hsueh-chu, and Raung-fu Chung. 2008. Interlanguage analysis of phonetic timing patterns by Taiwanese learners. *Concentric: Studies in Linguistics* 34:79-108 .

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

Chung, Raung-fu. 1990. Dangdai yinyun lilun he hanyu yinyun xue [Contemporary phonological theories and Chinese phonology]. *Guowen Tiandi* [*The World of Chinese Studies*] 5:58-60.

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

Derwing, Tracy M., Marian J. Rossiter, Murray J. Munro, and Ron I. Thomson. 2004. Second language fluency: Judgments on different tasks. *Language Learning* 54:655-679.

Deterding, David. 2001. The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics* 29:217-230.

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

Duez, Danielle. 1982. Silent and non-silent pauses in three speech styles. *Language and Speech* 25:11-28.

Erwin, T. Dary. 1991. *Assessing Student Learning and Development: A Guide to the Principles, Goals, and Methods of Determining College Outcomes.* San Francisco: Jossey-Bass.

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

Fon, Janice, Wen-yu Chiang, and Hintat Cheung. 2004. Production and perception of the two dipping tones (Tone 2 and Tone 3) in Taiwan Mandarin. *Journal of Chinese Linguistics* 32:249-281.

Griffiths, Roger. 1991. Pausological research in an L2 context: A rationale and review of selected studies. *Applied Linguistic*s 12:345-364.

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

Howie, John M. 1974. On the domain of tone in Mandarin. *Phonetica* 30:129-148.

Jilka, Matthias. 2000. Testing the contribution of prosody to the perception of foreign accent. *New Sounds* 4:199-207.

Johansson, Stig. 1978. *Studies of Error Gravity: Native Reactions to Errors Produced by Swedish Learners of English.* Goteborg: Acta Universitas Gothoburgensis.

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

Kelly, Gerald. 2000. *How to Teach Pronunciation.* Harlow: Longman.

Kenworthy, Joanne. 1987. *Teaching English Pronunciation.* London: Longman.

Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 40:387-417.

Low, Ee-ling, Esther Grabe, and Francis Nolan. 2000. Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English, *Language and Speech* 43:377-401.

Magen, Harriet S. 1998. The perception of foreign-accented speech. *Journal of Phonetics* 26:381-400.

Major, Roy C. 1986. Paragoge and degree of foreign accent in Brazilian English. *Second Language Research* 2:53-71.

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

Munro, Murray J., and Tracy M. Derwing. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45:73-97.

Munro, Murray J., and Tracy M. Derwing. 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48:159-182.

Munro, Murray J., and Tracy M. Derwing. 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition* 23:451-468.

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

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

Riazantseva, Anastasia. 2001. Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition* 23:497-526.

Riney, Timothy J., Naoyuki Takagi, and Kumiko Inutsuka. 2005. Phonetic parameters and perceptual judgments of accent in English by American and Japanese listeners. *TESOL Quarterly* 39:441-466.

Roach, Peter. 2009. *English Phonetics and Phonology*. Cambridge: Cambridge University Press.

Solé, Maria Josep. 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 Lourdes Diaz and Carmen Perez, 539-551. Barcelona: Universitat Pompeu Fabra.

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

Ueyama, Motoko. 2000. Prosodic Transfer: An Acoustic Study of L2 English vs. L2 Japanese. Doctoral dissertation, University of California, Los Angeles.

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

[Received 7 October 2009; revised 14 May 2010; accepted 30 June 2010]

Department of English
The Hong Kong Institute of Education
Hong Kong, P.R.C.
Hsueh-Chu Chen: hsuehchu@ied.edu.hk

第二語言時長類型以及其對母語人士感知之影響

陳雪珠
香港教育學院

本論文爲陳和鍾（2008）之延伸研究，旨在探討台灣英語學習者對習得美式英語語音時長類型之困難，並瞭解影響英語母語人士評定外國口音之關鍵變項。計三十位台灣英語學習者以及十位英語母語人士，各自唸出一篇診斷式文章，從而擷取其中五句不同句型之句子加以分析六個時長變項，其中包括音節長度、母音弱化、停頓時長、連音時長、子音串時長、以及說話速度。另十位母語人士評分者根據收集的四十人的語音資料，分別對六個時長變項及一個整體的口音印象加以評分。主要結果顯示從聲學測量得知，台灣英語學習者之音節時長類型和英語母語人士極爲不同。根據人工感知評分，六個個別時長變項和整體的口音印象有非常高的正相關，顯示語音時長類型或韻律是整體的印象，人類的聽覺難以抽離爲單獨成分。藉由聲學測量的說話速度是最主要外國口音的預測變項，當不包含這個有關整體流利度的變項時，母音弱化和連音時長二者爲母語人士偵測外國口音最重要的依據。根據以上發現，本研究嘗試提出外國口音時長感知模式，藉以釐清影響母語人士對非母語人士口音反應最重要的語音或音韻因素。

關鍵詞：時長現象，節奏型態，聲學語音學，中介語音韻學，產出與感知