Pronunciation of English as a second or foreign language learners [sic] : the reexamination of teaching pronunciation

Yukari Ryu

The University of Montana
The University of Montana

Permission is granted by the author to reproduce this material in its entirety, provided that this material is used for scholarly purposes and is properly cited in published works and reports.

**Please check "Yes" or "No" and provide signature**

Yes, I grant permission  

No, I do not grant permission  

Author's Signature:  

Date: 5-29-02  

Any copying for commercial purposes or financial gain may be undertaken only with the author's explicit consent.
Pronunciation of English as a Second or Foreign Language Learners:
The reexamination of Teaching Pronunciation

By

Yukari Ryu

B.A. The Prefectural University of Kumamoto, Japan, 1999

Presented in partial fulfillment of the requirements

For the degree of

Master of Arts

The University of Montana

2002

Approved by:

Dr. Irene Appelbaum, Chair

Dr. David Strobel, Dean, Graduate School

5-29-02

Date

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Adults learning a second or foreign language often have difficulty mastering native-like pronunciation in the target language (Archibald, 1998; Major, 2001). This adult learner's difficulty is often perceived as a foreign accent. In fact, the acquisition of second language pronunciation is often cited as the last level of proficiency to become categorized as native or native-like. This thesis examines the problem of adult language learners' pronunciation. Specifically, I argue that although foreign accent in adult learners' pronunciation is partially caused by biological constraint, pronunciation can nevertheless be improved by motivation and instruction. I also criticize the traditional neglect of pronunciation instruction in second language teaching, arguing instead that the acquisition of oral communication skill is recognized as important today. When sub-components of second language phonology are closely investigated at the segmental level, quantitative and qualitative input seems to be necessary for improving learners' pronunciation (Mochizuki, 1980; Zimmerman et al., 1984; Flege et al., 1995; Riney et al., 2000). Furthermore, a number of studies suggest that suprasegmental properties, including stress, rhythm, and intonation, have a great influence on learner's pronunciation. The evidence includes findings that deviance in suprasegmentals tends to be perceived as foreign accent more than deviance in segments (Anderson-Hsieh, et al., 1992; Munro, 1995) and that suprasegmentals can signal implicit messages in various ways which syntax or lexical choice cannot (Crystal, 1986; Chun, 1988). Regardless of its significant role in pronunciation, the acquisition of suprasegmentals, especially rhythm, is believed to emerge at the most advanced stage in second language learning. Based on these findings, I propose several ways to teach suprasegmentals in the second language classroom.
Acknowledgements

Many people have helped me complete this thesis. I would like to take this opportunity to express my warmest thanks to Dr. Irene Appelbaum, chair of my committee, for her encouragement, support, and critical evaluation of my work. I would equally like to thank Dr. Tully Thibeau and Dr. Olivier W. Rolfe, for patient guidance, helpful comments, and participation. Thanks go to my fiancé, Casey Zednick, who shared laughs and tears while I worked on this thesis, and to my family in Japan for encouraging my interest in English and supporting me in pursuing my degree. Thanks also go to my friends in the Linguistics Program for their academic and personal help.

Finally, I would like to thank Dr. David Tomlinson, who first got me interested in Linguistics and who encouraged me to pursue this M.A. degree. He passed away before finding out I had been accepted. I dedicate this thesis to him.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title page</td>
<td>i</td>
</tr>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>v</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
<tr>
<td>Chapter 1</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>9</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>19</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>23</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>38</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>47</td>
</tr>
<tr>
<td>Chapter 7</td>
<td>66</td>
</tr>
<tr>
<td>Appendix I</td>
<td>69</td>
</tr>
<tr>
<td>Bibliography</td>
<td>72</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 4.1: Summary of Japanese vowels .......................................................... 25
Table 4.2: Summary of English vowels .......................................................... 25
Table 4.3: Summary of Japanese consonants ............................................... 26
Table 4.4: Summary of English consonants .................................................. 26
Table 6.1: The chat show ................................................................................... 60
Table 6.2: The casual conversation ................................................................. 60
LIST OF FIGURES

Figure 6.1: Rhythm typology ................................................................. 48

Figure 6.2: Silent correction chart ...................................................... 64
Chapter 1

Introduction

In second language acquisition studies, it has been claimed that language learners beginning as adults have difficulty mastering second language phonology. As a result adult language learners' speech production is easily detectable as non-native. In terms of second language phonology, Archibald describes an adult's accent in second language (L2 henceforth) in the following way: "When we look closely at the construct of accent...we realize that it is highly complex. A second language learner must learn to present and implement information related to such things as the segmental inventory, phonotactics, syllable structure, stress, rhythm, and intonation of the language in question" (1998, p.37). These L2 phonological features interact with first language (L1 henceforth) phonology and are assumed to cause adult learners difficulty in acquiring native or nativelike L2 phonology. Because of this, the acquisition of L2 pronunciation is often cited as the last thing that adult learners reach a level of proficiency in that can be categorized as native or nativelike. This fact motivates my work on this thesis. Why is L2 pronunciation so difficult for adult learners? Is there any way to improve their foreign accent?

In this thesis, I claim that, although adult L2 learners' foreign accent in pronunciation is partially caused by the critical period, the foreign accent can be reduced by learners' motivation and instruction. Based on the findings from various studies on L2 pronunciation, I will also emphasize the importance of teaching pronunciation, especially suprasegmentals, in L2 pronunciation. The paper includes many studies on Japanese learners of English, since I am a native speaker of Japanese who speaks English...
as a second and foreign language, and am interested in approaches for improving Japanese learners' mastery of English in oral communication skills.

1.0 Why Pronunciation?

In this section, I will briefly explain two important characteristics which distinguish pronunciation from other components in L2 learning.

Language acquisition is one of the most impressive and remarkable aspects in human development. In particular, being able to speak is a vital component of human communication, which differentiates human beings from all other creatures. However, most of us hardly think about the role of speech in communication, i.e. how the message being conveyed on the stream of sounds produced by one person is comprehended by the other. This is because a speaker and a listener with homogenous L1 backgrounds communicate with each other orally, they usually do not experience difficulty in communication. Even if they do, for example due to regional dialect differences which present vocabulary variations, they are still capable of continuing oral communication by guessing or asking for clarification. This overall knowledge of how a language is used, which native speakers of a language possess, is called *communicative competence* (Hymes, 1972; cited by Richards and Rodgers, 1986). Communicative competence is not only the knowledge of grammar but also how to use language appropriately in social occasions (Cook, 1991). A detail analysis of communicative competence, done by Canale and Swain (1980), identified four dimensions of communicative competence: grammatical competence, sociolinguistic competence, discourse competence, and strategic competence:

*Grammatical competence* refers to what Chomsky calls linguistic competence and what Hymes intends by what is "formally possible." It is the domain of
grammatical and lexical capacity. Sociolinguistic competence refers to an understanding of the social context in which communication takes place, including role relationships, the shared information of the participants, and the communicative purpose for their interaction. Discourse competence refers to the interpretation of individual message elements in terms of their interconnectedness and of how meaning is represented in relationship to the entire discourse or text. Strategic competence refers to the coping strategies that communicators employ to initiate, terminate, maintain, repair, and redirect communication. (Richards et al., 1986, p.71)

The recent widespread adoption of a communicative approach in L2 teaching which emphasizes the development of communicative competence has brought urgency to the reexamination of pronunciation (Celce-Murcia and Goodwin, 1991). The role of speech is significant when it is considered in terms of these four dimensions, especially in terms of sociolinguistic competence and discourse competence. When interlocutors who have heterogeneous L1 backgrounds engage in oral communication in L2, a crucial factor in determining the success of oral communication in L2 would be whether the interlocutors have sociolinguistic competence or discourse competence in L2. This is because sound produced by a speaker not only carries a message which is interpreted word for word by a listener but also can include an unspoken message. Pronunciation, which physically shapes the speech and carries the spoken and unspoken message, is, thus, an important component to be mastered in L2 learning.

Another quite unique characteristic which cannot be seen in any components of L2 learning other than pronunciation is that the acquisition of pronunciation involves neuro-cognitive-motor skill. People cannot learn it by looking at textbooks. Learning pronunciation requires physical practice, just like a child cannot learn how to ride on a bicycle by reading a book on how a bicycle works. Due to its role as a basis of oral communication and uniqueness involving neuro-cognitive-motor skills, I find
pronunciation a very attractive component to be studied in L2 teaching. In the following section, the definition of “good” pronunciation will be described.

1.1 What Is “Good” Pronunciation?

Before discussing pronunciation in L2 learning, it is important to consider how good adult L2 learners’ pronunciation should be. At what level is adult L2 learners’ pronunciation regarded as good? Should it be regarded as good, if they can make themselves understood in a supermarket in an L2 speaking community, or if they are complimented on their pronunciation by native speakers of L2? In fact, as Leather and James state, there has been no consensus on the definition of “good” pronunciation (1991, p.307). It is quite arbitrary: When L2 learners’ pronunciation is auditorily evaluated by the native speakers, considerable differences in judgment are often seen. For example, consider when a native speaker of English (speaker A), usually not having any linguistic contact with nonnative speakers of English, has an opportunity to talk to a nonnative speaker (speaker B). Speaker A may judge subjectively that B’s English pronunciation is good just because A could understand what B meant to say. On the contrary, when another native speaker of English (speaker C), professionally teaching English as a Second Language who has seen many nonnative speakers, talks to the same speaker B, speaker C might find B’s pronunciation not that good, compared to other learners he or she has seen. Further, the listener’s patience level toward the foreign accented speech by the nonnative speaker would also reflect that pronunciation evaluation. Those who are more tolerant toward foreign accented pronunciation might judge one’s pronunciation good, on the other hand, those who are impatient might judge the pronunciation of the same speaker as bad. Thus, the definition of “good”
pronunciation is very variable according to individual listeners' experience with nonnative speakers and attitude to the accent of nonnative speakers' pronunciation.

Nevertheless, in order for adult L2 learners to be able to communicate, there should at least be a threshold level of pronunciation in a language such that if L2 learners' pronunciation falls below this level, he or she will be inhibited for communicating orally no matter how good his or her control of grammar and vocabulary might be in that language (Hinofotis and Bailey, 1980; cited by Celce-Murcia et al., 1991). Although the definition of such threshold level requires further investigation, here I will define the general L2 learners' goal as the quality of pronunciation that will not detract from their ability to communicate (Celce-Murcia et al., 1991). To put it another way, this is the level at which the speakers engaging in communication do not have to ask regularly for repetitions or clarification due to foreign accent. However, it should be noted that, for L2 learners who set higher goals due to occupational requirements (such as a teacher in L2, a telephone operator, or an air-traffic controller) or simply high motivation, they may be satisfied at the threshold level, but should be encouraged to set higher goals.

1.2 Phonological Systems of a Language

Now the question is what constitutes the speech of a language. Such components are the features that need to be mastered when people learn a language. Phonological systems can be viewed as comprising features of several types of phenomena: segments, syllabic structures, and suprasegmentals (Major, 2001).

1.2.0 Segments
Segments influence the presence and phonetic feature values associated with each consonant and vowel (Crystal, 1986). Learning the segments of a language requires one not only to be able to articulate and perceive acoustically, but also to work out which properties of the sound segments that constitute words are predictable in that language. One example is that native speakers of English, not knowing why explicitly, recognize that forms such as *slish* and *screnk* are possible sound sequence in English, though forms such as *srish* and *screpk* are not (Dobrovolsky, 1997). Goodluck elaborates this possible-word recognition process as follows: “Non-predictable properties will be entered in the lexical entry for the word; predictable properties will be spelled out by phonological rules that work on and change the basic lexical entry for the word” (1991, p.23). Language learners must learn what type of combination of segments can occur in their L1 based on the lexicon they have developed.

1.2.1 Syllabic Structures

Before looking at suprasegmentals, it is necessary to look at the syllabic structure of a language. This is because, structurally, the syllables function to combine the segments of a language to form words, but the permitted syllabic structures in a language have a strong effect on the suprasegmentals of the language. Here the permitted structure means that different languages adopt different phonotactic constraints (i.e. the rules that characterize permissible syllable structure in a language), and most of this difference seems to be captured by the number of consonants permitted in a syllable (Finegan, 1999). It should be noted that this level, which structurally connects segments and suprasegmentals, is sometimes interpreted differently depending on the researchers. Some researchers include syllabic structures in suprasegmentals, which can also be called
prosody (Crystal, 1985; James, 1988; Goodluck, 1991), whereas others separate it from suprasegmentals (Riney and Anderson-Hsieh, 1993). As seen above, I distinguish among segments, syllabic structures, and suprasegmentals, since I believe that segments and syllabic structures which affect lexical level should be dealt with separately from suprasegmentals which influence on sentential level.

1.2.2 Suprasegmentals

Suprasegmentals include pitch, stress, length, tone and intonation, and rhythm and timing of a language.

1.2.2.0 Pitch: Tone and Intonation

*Pitch* is the perceptual property of a sound that permits us to measure a sound in a range from low to high. A pitch may fall, rise, stay level, or do some combinations of these things within a given phonological unit (e.g., fall-rise on a syllable) (Crystal, 1986,). *Tone* refers to a language feature in which differences in pitch signal difference in meaning. For example, in the tonal language Mandarin, speakers differentiate the four meanings of /ma/ by changing the pitch of /a/. *Intonation* refers to the pitch movement in spoken utterances that do not affect the word level but the sentential level. For instance, in English falling pitch at the end of the sentence implies that the speaker's intention that the utterance is complete, while the rising pitch indicates incompleteness.

1.2.2.1 Length

*Length* is shown in vowels and consonants whose articulation takes longer relative to that of other vowels and consonants (Dobrovolsky, 1997). For example, many languages (e.g., Italian, Hungarian, German, Cree, and Finish) have long and short
vowels. Also long and short consonants are found in many languages (i.e., Finnish, Turkish, and Hungarian).

1.2.2.2 Stress

In general, stress refers to syllabic segments that are perceived as more prominent than other segments (Dobrovolsky, 1997). Pitch, loudness, and length may all contribute to stress.

1.2.2.3 Rhythm

Rhythm and timing are the repetitive patterns of stress and length (Major, 2001). Languages are traditionally classified into three basic rhythmic types: syllable-timing, stress-timing, and mora-timing.

In summary, these sub-components of phonology are to be learned when a language learner becomes capable of producing speech. In the next chapter, I will look at a possible explanation for why native or native-like attainment of L2 pronunciation is believed to be extremely difficult for many adult L2 learners.
Chapter 2

Constraints on the Acquisition of L2 Pronunciation

2.0 Fossilization

In this chapter, I will describe the issue of what causes difficulty in L2 pronunciation acquisition. Consequently I will argue that, although L2 learners' foreign accent in pronunciation is partially caused by the critical period, foreign accent can be reduced by learners’ motivation and instruction.

As adult L2 learners advance, their proficiency in L2 gets increasingly L2 nativelike. It has been observed that, although adult L2 learners often attain native-like proficiency in syntactic, morphological, and lexical systems of L2, they rarely or never master the complete L2 sound system (Scovel, 1969; cited by Boatman, 1990). Even with the substantial amount of exposure to L2, sometimes there seems no advance in the adult learner’s Interlanguage phonology. When adult L2 learner’s Interlanguage stops developing toward the L2 forms and persists in non-L2-like forms, the phenomenon is called fossilization (Selinker 1972, cited by Archibald, 1998; and Selinker, 1992). The typical example of L2 phonological fossilization can be illustrated by the “Joseph Conrad phenomenon,” the famous author whose syntax and writing skill in English were undoubtedly native-level, or more than native-level, but who retained a heavy Polish accent all his life (Tarone, 1987, p.80). Thus, it is necessary to pursue the issue of why attainment in L2 phonology is exceptionally difficult for adult L2 learners while other areas in L2 acquisition such as syntax, morphology, and vocabulary seem less problematic.
With respect to fossilization in L2 learning, many researchers have investigated the reality of fossilization and its validity in L2 acquisition (Gass and Selinker, 1992; Hill, 1970; Lenneberg, 1967; Neufeld, 1977; Walsh and Diller, 1981). Broadly speaking, some researchers argue that biological constraints influence the L2 learners’ fossilization, while other researchers support psychological and social constraints as the reason for fossilization.

2.1 Biological Constraints: Critical Period Hypothesis

2.1.0 Brain Lateralization

It was Lenneberg (1967) who first proposed that "... there is a biologically determined 'critical period' for language learning, extending approximately from the age 2 years to puberty" (cited by Leather and James, 1991, p.306). This hypothesis is based on the well-known belief that children have an advantage over adults when it comes to language learning. Scovel (1969) later attempted to apply the data in the critical period and L1 acquisition to L2 acquisition by claiming that a cortical lateralization of the brain which occurs around puberty limits the ability for L2 learners to master the sound system of the language.

Although the Critical Period Hypothesis has been cited in many studies as a possible reason for fossilization, the available evidence is far from convincing (Flege, 1986; Archibald, 1998). First, the arguments differ over when the process of lateralization is thought to be complete (Major, 2001). Krashen (1973) claims that lateralization completes by age five, not by puberty. According to Major (2001), dichotomous listening tests also indicate that brain lateralization may be complete before puberty. In that case it is questionable how to account for that children between age five...
to puberty can master native or nativelike L2 speech. Scovel has responded to this by stating that it is important to pay attention to the difference between the emergence of lateralization and completion of lateralization: if lateralization is not completed until puberty, he claims it is possible to use it as evidence of the Critical Period Hypothesis (1981; cited by Archibald, 1998). However, conflicting evidence also reports that lateralization does not increase with age (Segalowitz, 1983; cited by Archibald, 1998).

Brain damage studies suggest that "... in right-brain damaged patients, there are more language disturbances in children (even up to age 10) than adults, thus indicating that the children's right-brains had more language functions than adults" (Major, 2001, p.7).

In contrast, a limited but increasing number of studies indicate that the Critical Period does not exist (Major, 2001). Flege (1986) argues that Critical Period Hypothesis was originally intended to describe animal behavior but not human behavior and several studies conducted showed evidence that adults could perceive and produce second language sounds as well as or better than children.

Furthermore, Major argues that even though research indicates that there is a Critical Period in phonology, the point when the Critical Period ends is not agreed upon: "Long (1990) claimed it is 6 or 7, Patkowski (1994, critiquing Long) said it is slightly later, and Scovel (1988) claimed it lasts through puberty" (2001, p.9). These arguments against the Critical Period Hypothesis suggest that the relation between fossilization and age in language learning cannot be explained in a straightforward way. Even proponents of the Critical Period Hypothesis who claim that the lateralization causes fossilization disagree over when and how lateralization is completed. Considering all these findings,
brain lateralization does not seem very persuasive as an argument to explain the relation between fossilization and age in language learning.

2.1.1 The Loss of Neural Circuits Plasticity

The other position which supports the Critical Period Hypothesis focuses on the difficulty of phonological acquisition in L2 learning. Walsh et al. (1981, cited by Archibald, 1998; and Flege, 1986) specifically argue adult L2 learners’ problem of pronunciation by suggesting “... a possible neurological basis for ‘difficulty’ in eliminating foreign accent after childhood” (Walsh et al., cited by Flege, 1986, p.163). They claim that, although adult L2 learners may achieve an advanced stage in areas other than pronunciation, it is impossible for them to master L2 pronunciation completely since:

Lower-order processes such as pronunciation are dependent on early maturing and less adaptive macroneural circuits, which makes foreign accents difficult to overcome after childhood. High-order language functions, such as semantic relations, are more dependent on late-maturing neural circuits, which may explain why college students can learn many times the amount of grammar and vocabulary that elementary school students can learn in a given period of time. (Walsh et al., 1981; cited by Archibald, 1998, p.22)

I find Walsh et al.’s proposal more persuasive than the hypothesis based on the brain lateralization. This hypothesis, stating that the acquisition of sub-components of a language interact with neural circuits at a different pace, might be also attributed to the findings in L1 learning. In L1 learning until approximately eight months old, infants are able to perceive universal phones but after that they lose the ability because their sound perception is adapted to their immediate linguistic environment (De Boysson-Bardies, 1999). Also when we look at the linguistic developmental stages in L1 learning, children learn basic L1 sounds quicker than any other sub-components of the target L1 grammar
(e.g., syntax, morphology, etc), though it takes the next several years for children to establish the complete sound system of the language (De Boysson-Bardies, 1999; Ingram, 1976). This evidence in L1 learning indicates that the acquisition of phonology might be quite different from the acquisition of other sub-components of a language. Thus, the early maturing and adaptive neural circuits which are already operative in the L1 sound system acquisition could hinder one's learning of L2 phonology as adults. In other words, since the neural circuits responsible for phonological acquisition become adapted to L1 at a very early time in one's life, the neural circuits are not effective for mastering new sounds any more by the time people learn L2 as adults. In summary, the difficulty of ultimate attainment in L2 pronunciation adult L2 learners face can be attributed to the fact that L2 phonological acquisition is related to neural circuits that are less adaptive than other neural circuits that are responsible for acquisitions of syntactic, morphological, and lexical systems in L2.

Considering the common observation that children have an advantage over adults in language learning, I believe that there must be a certain degree of biological constraint on second language acquisition, especially in phonology. The evidence found in L1 acquisition such as Genie (Rymer, 1993), who was found at age 13 having been deprived of nearly all sensory stimulus, especially linguistic and who never subsequently achieved nativelike competence in English, indicates that a person needs to be exposed to a language by a certain period of time in order to be able to master a language natively. Even though the controversy over whether the Critical Period exists in L2 acquisition remains undecided, the evidence is too overwhelming to ignore that age matters in language learning. As Major (2001) states, even if there is some small number of older
learners that successfully acquire nativelike L2 phonology, the vast majority of studies show that the younger the learner the more nativelike the pronunciation. Although in a natural learning setting children have an advantage over adults in L2 acquisition, this does not necessarily mean that adults cannot attain nativelike L2 pronunciation. I claim that in adult L2 learning, psychological and sociological constraints are more powerful than biological constraints in determining how successful the learner can be in L2 pronunciation. When most adult L2 learners have a biological disadvantage in learning L2 pronunciation, what determines the proficiency of the learner is how he or she attempts to overcome that disadvantage. In other words, adult L2 learners’ psychological and sociological factors can contribute to compensate for their biological disadvantages. For children, motivation does not seem to be relevant to whether they learn an L2; rather they seem to acquire it unconsciously in a natural setting. However, for most adults L2 learning does not progress as smoothly as L1. As mentioned above, I argue that the difference in L2 proficiency among adult learners is due to psychological and sociological factors. With respect to phonology, adult L2 learners seem to be able to acquire minimally adequate L2 pronunciation without being exposed to formal pronunciation instruction. After going through this average-learner-pronunciation stage, the next question is whether the learner is motivated to master the more intelligible or nativelike pronunciation.

2.2 Psychological and Sociological Constraints

The other position which attempts to explain what causes phonological fossilization points out that the learner’s individual psychological and sociological factors, such as motivation, cultural empathy, and desire to sound like a native speaker.
may influence the learner's progress and ultimate achievement in the acquisition of L2 phonology (Flege, 1986; Leather and James, 1991). This approach claims that child-adult difference in performance may derive from the quantity or quality of L2 speech input they receive.

2.2.0 Psychological Habit Formation Hypothesis

Neufeld (1977) argues that the problem in the acquisition of L2 phonology derives from "... the inappropriate learning situations where they form inaccurate acoustic images of the target language sound patterns" (cited by Tarone, 1987, p.81). The idea behind his proposal is that "[o]nce formed, those acoustic images are set, and so are the learner's pronunciation patterns" (cited by Tarone, 1987, p.81). In his experiment, no explicit instruction about pronunciation was given to Chinese, Japanese, and Eskimo young adults who studied English as an L2 and they watched videotape lessons consisting of "100 stock phrases" to have the accurate input and to form a correct acoustic image in their mind (cited by Tarone, 1987, p.82). The subjects were not allowed to speak at first, but later they were allowed to trace intonational contours of the utterances they heard, to whisper them, and finally to repeat the utterances in a normal voice. The result indicated that more than half of the subjects were judged as native or near-native pronunciation by English native speakers. Even though Tarone (1986) points out that Neufeld's argument did not mention why adults are influenced negatively by inaccurate acoustic images and children are not, his experiment is very interesting in terms of pronunciation teaching and cast a reconsideration of the importance of the pedagogical effect in L2 learners' pronunciation. This study leads to a further
investigation of whether these adult second language learners can maintain such nativelike pronunciation for real communication.

2.2.1 Social and Cultural Factors

Hill (1970) claimed that fossilization is not inevitable but a product of social and cultural factors (cited by Archibald, 1998; Flege, 1986; and Tarone, 1986). He presented examples of native people such as the Vaupes, Indians of the Amazon, and the Siane of New Guinea, whose society requires them to marry people who speak a different language, actually acquire several languages, and achieve the native fluency as adults.

Schumann (1976) also argued that the affective factors such as political, cultural, technical, or economic distance between the learner’s L1 and L2 groups play an important role in the learner’s mastery of the L2. He claimed the greater the social distance between the learner’s L1 and L2 groups, the more difficult it is for the learner to acquire the L2, and emphasized the importance of the learner’s empathy toward the L2 group as a successful factor in L2 acquisition. Similar to Schumann, Gardner and Lambert (1972) described motivation as a crucial factor in determining how successfully an L2 will be acquired. They classified motivation into two types: integrative and instrumental. According to them, the learner who wants to learn the L2 to meet its speakers and to learn their culture is called integratively motivated, while the learner who wants to learn the L2 to achieve social or professional advancement is called instrumentally motivated. It can be deduced from Schumann (1976) and Gardner et al. (1972) that the differences in empathy or motivation may lead to the differences in amount of L2 input: Those who feel more empathy toward the L1 group or are integratively motivated would have more communicative input through interaction with
native speakers than those who do not feel empathy and are instrumentally motivated may prefer the minimum contact with L2 to achieve the social or professional advancement.

2.3 Argument

To summarize, I believe if the learner feels a need to improve his or her pronunciation, whether it is the personal desire to assimilate oneself in an L2-speaking community or the occupational requirement for exceptional intelligibility in L2 pronunciation, the learner would pay more attention to his or her own pronunciation. On the other hand, if the learner is satisfied with the ability to communicate, as the majority of the L2 learners are, with a goal of “comfortably intelligible” (Abercrombie, 1963; cited by Leather et al., 1991, p.308), he or she would be less aware of or less concerned about how he or she sounds to a listener. Moyer (1999) who investigated L2 German phonological performance by American graduate student instructors in German found that, although the instructional and motivational factors seem unable to override the impact of age, the qualitative instruction (in this case, the instructor’s feedback in segmentals and suprasegmentals) and the motivation are certainly relevant to the outcome of the L2 pronunciation. This finding is consistent with my central claim: Motivation and the qualitative instruction might be able to compensate for the biological disadvantages of adult L2 learners. When these two factors, the motivation and the qualitative instruction, are the key potential for the adult L2 learners’ success in L2 pronunciation, language teachers might be vulnerable with respect to motivation because motivation is particularly learner-dependent, as well-known proverb says “A man may lead a horse to the water, but he cannot make it drink.” However, language teachers can and should give
a qualitative instruction to the learners to make the learner’s pronunciation acquisition smoother. In the next chapter, I will discuss the need for L2 pronunciation instruction in L2 teaching.
Chapter 3

Revisiting Pronunciation Instruction in the L2 Classroom

3.0 Neglect of L2 Pronunciation Instruction

Unfortunately, regardless of L2 learners’ desire or need to improve L2 pronunciation, both teachers and researchers in L2 acquisition seemed to have neglected L2 pronunciation. Morley summarizes four myths of “misguided conventional wisdom” as reasons that the need of pronunciation instruction has been denied in L2 teaching and criticizes them (1996, p.146): (1) *Pronunciation isn’t important:* Morley states that “[t]his belief is patently false from any perspective” (1996, p.146). It is the speaker’s verbal message riding the wave stream of speech that enables communication between the speaker and listener. She emphasizes the importance of intelligible pronunciation as an essential component of communicative competence. (2) *Students will pick it up on their own:* Morley criticizes this point saying that many L2 learners will not pick up accurate pronunciation by themselves. Also, withholding pronunciation instruction takes away L2 learners’ opportunity to improve their oral communication skills systematically and to develop personal oral language learning strategies. (3) *Pronunciation is too hard to teach:* Morley uses Marks’ (1986) remark to cast a question:

> Few teachers, probably, would claim that they do not teach grammar or vocabulary, on the grounds that they are either too difficult or else not sufficiently important. Yet these are the kinds of comment which many teachers make with regard to the teaching phonology [ pronunciation]. (cited by Morley, 1996, p.146)

This attitude seems to derive from the fact that the characteristics of phonology are quite different from those of other sub-components in L2 learning. As stated above, when we compare phonology with other sub-components such as syntax and
morphology, the acquisition of phonology involves a very complex neuro-motor-cognitive human behavior, whereas the acquisition of syntax or morphology which does not require motor skills. In this sense, average teachers who do not teach pronunciation might show their reluctant attitude that they even do not want to try it because it is time consuming and arduous to improve the learners' pronunciation. Nevertheless, as stated before, human speech is one of the most basic components for oral communication. The myth of "Too hard to teach" sounds too unprofessional. (4) *I don't have the training to teach it, so I just won't bother (and, I'll just say pronunciation isn't important).* Morley states that this view can be eliminated from for two reasons: First, the importance of pronunciation as a primary feature of communicative competence is well recognized among language teaching professionals; and second, both the redesigning of L2 language teacher preparation programs to include "appropriate modern-day theory and practice" in pronunciation, and the appearance of many excellent teacher reference books and student texts have been seen over the last ten years (1996, p.147).

I agree that these four myths are typical excuses among L2 language teachers who avoid pronunciation instruction. I would add that the institutional policy which underestimates the importance of pronunciation might deprive language teachers of pronunciation instruction. I was taught English as a foreign language as one of the required subjects in junior high school and high school for six years in the 1990's in Japan. The English education I received reflected the fact that there exist competitive entrance examinations for advanced schools in Japan. The content of the instruction was focused on what would be asked on the entrance examinations on which students are expected to get high scores in order to be admitted to their-first-choice high schools or
colleges. As a result, the instruction I received at that time was similar to traditional Grammar-Translation method (focusing on reading and writing) which requires only a paper and a pencil. Through six years of taking English classes in junior high and high school, it was only during the first semester in junior high school that I received pronunciation instruction. Fortunately I had a teacher who was very enthusiastic about teaching pronunciation and spent approximately 50-hours training on English segments. That was the first and last time that I had pronunciation instruction in Japan. Recent uniform guidelines in teaching (which are proposed by the Japanese government every ten years, stating what needs to be taught from elementary school through high school) appear to have reconsidered the fact that most Japanese are unable to communicate in English even with six years of learning and are beginning to emphasize the importance of students' development in oral communicative skill and communicative competence.

With the rise of recognition of the importance of oral communication skills in L2 teaching, I suggest that teaching pronunciation to adult L2 learners should be given more emphasis for the following reasons. First, speech is the most basic component of oral communication. Without speech, interpersonal oral communication cannot happen. Second, since English is considered an international language today, oral communication skills become more important than ever. In order for many English speakers from heterogeneous linguistic backgrounds to communicate orally, intelligible pronunciation becomes more essential. As Morley (1996) warns, unintelligible speech may place L2 learners at serious risks educationally, occupationally, professionally, and socially. It should be remembered that the disadvantages an L2 learner may encounter because of unintelligible speech are determined not only by the L2 learner but also by a listener's
goodwill to be patient and supportive with the L2 learner's speech. When this L2 learners' challenge can be expected to happen outside of the classroom, the fact that teaching pronunciation has been neglected in L2 teaching should be reconsidered. Teachers should give qualitative pronunciation instruction for L2 learners, explain the importance of pronunciation to the L2 learners for their sake, and integrate more pronunciation instruction on various occasions in the classroom. My intention has been to suggest that L2 language teachers need to understand that intelligible pronunciation is crucial for L2 learners in order to communicate orally with people on social occasions. Thus it is the teachers' job to incorporate more pronunciation instruction in a classroom. Next, let us look at the content of pronunciation instruction.
Chapter 4

Segments and Syllabic Structures in L2 Pronunciation

So far I have emphasized the importance of teaching pronunciation in the L2 classroom. In this chapter, I discuss what the qualitative instruction should be. The examples I present in the following sections are based mostly on the pronunciation of Japanese learners of English. I claim that, although pronunciation instruction has been focusing on segments if any, suprasegmentals are as important as segments, or perhaps even more important. In order to support my claim, I propose four points regarding effective pronunciation instruction for the development of oral communication skills: (i) The L2 learners’ mastery of segmental articulation appears not to be so difficult in a natural setting; (ii) Deviance in suprasegmentals tends to be perceived as foreign accent more than deviance in segments; (iii) The mastery of suprasegmentals seems to emerge at very advanced stages in L2 pronunciation; and (iv) Unlike segmentals, suprasegmentals function in various ways in a discourse.

4.0 Contrastive Analysis Hypothesis

According to Major (2001), the mastery of the phonology of a language involves (a) individual segments, (b) combinations of segments, which produce syllables, (c) prosody (i.e., suprasegmentals), and (d) global accent, or the overall accent of the speaker. He states that a nonnative combination of (a), (b), and (c) result in (d) (a learner’s global foreign accent); thus, for a learner who wants to master nativelike pronunciation, it is necessary to master not only one or two but all three levels. However, the traditional approach in teaching pronunciation has largely focused on teaching the accurate production of segments. When segments are the target in pronunciation...
instruction, what L2 language teachers and researchers are likely to adopt is Contrastive Analysis (CA henceforth) proposed by Lado in 1950's. The extreme version of CA is supposed to predict errors of the L2 learners by comparing and contrasting the learner's L1 and L2: Consider 1,000 Japanese who have never learned English. It is possible to predict that "a great number of them . . . will experience difficulties with English liquids and syllable structures" (Major, 2001, p.34). In contrast, the moderate version considers degrees of similarity between the L1 and L2 (Oller and Ziahosseiny, 1970; cited by Major, 2001). Although the extreme version of CA has faced challenges in predicting difficulty for an L2 learner, a moderate version of CA has been adopted widely in L2 phonology research (Major, 2001). The reason the moderate version has been used in L2 phonology is that in phonology it is easy to distinguish the notions of similar and dissimilar between the learner's L1 and L2, compared to other fields such as semantics and discourse (Major, 2001). For instance, when we apply the moderate version of CA to compare English segments /b/ and /v/ and the Japanese segment /b/ and to examine the degree of similarity, it is possible to say that Japanese /b/ is similar to English /b/ and dissimilar to English /v/ with respect to the manner and the place of articulation. The studies of pronunciation by Japanese learners of English I reviewed also heavily rely on CA; therefore, I present the basic structural descriptions between Japanese and English below. By presenting these studies, I propose to show that mastery of segmental articulation appears to be not so difficult for Japanese learners of English in a natural ESL setting. In other words, when the quantity and quality of input are sufficient, L2 learners can master segments without much trouble.
First, let us see what L2 language teachers or researchers may do to teach segments to L2 learners: they take a look at the phonetic differences between the learner’s L1 (here Japanese) and L2 (English).

<table>
<thead>
<tr>
<th>front</th>
<th>central</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>ika “squid”</td>
<td>iti “one”</td>
<td>usiro “behind”</td>
</tr>
<tr>
<td>mid</td>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>“painting”</td>
<td>eki “station”</td>
<td>usagi “rabbit”</td>
</tr>
<tr>
<td>low</td>
<td>a</td>
<td>oka “hill”</td>
</tr>
<tr>
<td>asa “morning”</td>
<td>asi “leg”</td>
<td>oto “sound”</td>
</tr>
</tbody>
</table>

Table 4.1: Summary of Japanese vowels

(adopted from Tsujimura, 1996, p. 18)

<table>
<thead>
<tr>
<th>front</th>
<th>central</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>Pete, beat</td>
<td>pit, bit</td>
<td>pool, boot</td>
</tr>
<tr>
<td>mid</td>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>late, bait</td>
<td>about, sofa</td>
<td>poke, boat</td>
</tr>
<tr>
<td>low</td>
<td>æ</td>
<td>o</td>
</tr>
<tr>
<td>pat, bat</td>
<td>a</td>
<td>port, father</td>
</tr>
<tr>
<td>pat, bat</td>
<td>a</td>
<td>park (Boston dialect)</td>
</tr>
</tbody>
</table>

Table 4.2: Summary of English vowels

(adopted from Tsujimura, 1996, p. 18)

<table>
<thead>
<tr>
<th>Stops:</th>
<th>[+]V</th>
<th>[−]V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops:</td>
<td>[+]V</td>
<td>[−]V</td>
</tr>
<tr>
<td>Fricatives:</td>
<td>[+]V</td>
<td>[−]V</td>
</tr>
<tr>
<td>Fricatives:</td>
<td>[+]V</td>
<td>[−]V</td>
</tr>
<tr>
<td>Affricates:</td>
<td>[+]V</td>
<td>[−]V</td>
</tr>
<tr>
<td>Approximants:</td>
<td>[+]V</td>
<td>[−]V</td>
</tr>
<tr>
<td>Approximants:</td>
<td>[+]V</td>
<td>[−]V</td>
</tr>
<tr>
<td>Approximants:</td>
<td>[+]V</td>
<td>[−]V</td>
</tr>
<tr>
<td>Nasals:</td>
<td>[+]V</td>
<td>[−]V</td>
</tr>
<tr>
<td>Nasals:</td>
<td>[+]V</td>
<td>[−]V</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frontal</th>
<th>Alveolar</th>
<th>Alveo-palatal</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stopped</td>
<td>[+]V</td>
<td>[−]V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricated</td>
<td>[+]V</td>
<td>[−]V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affricated</td>
<td>[+]V</td>
<td>[−]V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximated</td>
<td>[+]V</td>
<td>[−]V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximated</td>
<td>[+]V</td>
<td>[−]V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasaled</td>
<td>[+]V</td>
<td>[−]V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasaled</td>
<td>[+]V</td>
<td>[−]V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[-]V = voiceless
[-]V = voiced

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 4.3: Summary of Japanese consonants

(adopted from Tsujimura, 1996, p.16)

<table>
<thead>
<tr>
<th></th>
<th>bilabial</th>
<th>labio-dental</th>
<th>interdental</th>
<th>alveolar</th>
<th>alveo-palatal</th>
<th>palatal</th>
<th>velar</th>
<th>labio-velar</th>
<th>glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops:</td>
<td>[+V]</td>
<td>b</td>
<td>d</td>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-V]</td>
<td>p</td>
<td>r</td>
<td>k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricatives:</td>
<td>[+V]</td>
<td>v</td>
<td>θ</td>
<td>z</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td>h</td>
</tr>
<tr>
<td></td>
<td>[-V]</td>
<td>f</td>
<td>θ</td>
<td>s</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td>h</td>
</tr>
<tr>
<td>Affricates:</td>
<td>[+V]</td>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-V]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximants:</td>
<td>liquid</td>
<td>[+V]</td>
<td>r</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>glide</td>
<td>[-V]</td>
<td></td>
<td>y</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasals:</td>
<td>[+V]</td>
<td>m</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td>η</td>
</tr>
</tbody>
</table>

[+V] = voiced
[-V] = voiceless

Table 4.4: Summary of English consonants

(adopted from Tsujimura, 1996, p.12)

With regard to segmentals, English has more vowels and consonants than Japanese. Thus, the segments of English that Japanese does not have, /æ, α, ɪ, ɛ, ʌ, θ, s, f, v, θ, δ, r, l/, are the ones for which Japanese learners of English are required to learn to pronounce.

4.1 The Acquisition of Segments

Among these new English segments for Japanese learners of English, the liquids /r/ and /l/ have received more attention than any other segments (Goto, 1971; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujiwara, 1975; Cochrane, 1980; Mochizuki, 1980; MacKain, Best, and Strange, 1981; Sheldon and Strange, 1982; Zimmerman. Price,
and Ayusawa, 1984; Flege, Takagi, and Mann, 1995; Riney and Flege, 1998; Riney, Takada, and Ota, 2000). One of the possible reasons why the liquids have drawn so much attention is because Japanese lacks /r/ and /l/ in its inventory and the phonetically closest segment, Japanese /ɾ/, seems to “... occupy a position in phonological space that is somewhere between English /l/, /ɾ/, and /d/ (and possibly /w/)” (Flege et al., 1995, p.26). Therefore, learning how to articulate the English segments /ɾ/ and /l/ requires Japanese learners of English to reconfigure phonological space. Flege et al. purport that researchers have attempted to explain the phonetic realization of Japanese /ɾ/ in various ways:

Although the /ɾ/ phoneme of Japanese is often referred to as a “liquid” and is usually represented with the phonetic symbol “ɾ”, its articulatory characteristics led Koutsoudas and Koutsoudas (1983) to represent it with the phonetic symbol “l”. Jones (1967) claimed that phonetically distinct variants of Japanese /ɾ/ are produced “indiscriminately” as a sound resembling English /ɾ/, as a lingual flap (/ɾ/), as a “kind of” retroflex /d/, a “kind of” /l/, or something “intermediate”. In a study of Sekiyama and Tohkura (1993), native English listeners identified the initial consonant in Japanese /ra/ syllable as /l/ (51% of judgements), /gr/ (17%), /dl/ (10%), “Spanish /ɾ/” (10%), or /wl/ (2%). (Flege, et al., 1995, p.26)

Of all the explanations, I believe the explanation by Vance (1987, cited by Flege et al., 1995, p.26) is the most appropriate: “Japanese /ɾ/ is usually realized as an apico-alveolar tap /ɾ/, but may be palatalized when spoken in the context of /i/ and /y/... when the tip of the tongue (which is held tightly against the alveolar ridge for Japanese /ɾ/) is released rapidly, native English-speaking listeners may hear Japanese /ɾ/ as /d/.”

I have a strategy for dealing with the articulatory realization of Japanese /ɾ/, which is compatible with Vance’s description of Japanese /ɾ/. Since both my first and
last name include this Japanese /r/, I have had a hard time explaining to English speakers how this /r/ is articulatorily realized in my native language. When “Ryu” [ryu:] and “Yukari” [yukari] are produced (i.e., the /y/ and /i/ follow the /r/, as Vance presents) in Japanese, the tip of the tongue taps the alveolar ridge and produce a sound almost like /d/. Although some researchers presented above say the Japanese /r/ is like English /l/, the area of the tongue which touches the alveolar ridge to produce /r/ is more similar to /d/ (stop) than to /l/ (liquid). Thus, after struggling with it, I came to realize the best way to have native speakers of English accurately pronounce my family name “Ryu” is to tell them to pronounce like “dew” as in Mountain Dew, and my first name “Yukari” as “yukadee”.

The reason this confusion of the /r/ sound among different languages occurs might be the arbitrary correspondence between the phonetic realization of /r/ and its orthographies in different languages. Although many languages use /r/ not in the sense of International Phonetic Alphabets but in their writing system to represent “their own” /r/ sound, the way English, Japanese, Spanish, and French speakers pronounce the /r/ shows variations. Thus, when these speakers who have different backgrounds of /r/ see a character r in a foreign language, they may interpret the r in the same way as their own language’s pronunciation. The example of my name containing the r shows that Japanese, lacking an English retroflex /r/ and using r in its orthography to represent “an apico-alveolar tap /r/” (Vance, 1987; cited by Flege et al. 1995, p.26) is misunderstood by English speakers as an English /r/ existing in Japanese.
The studies on the pronunciation of English segments by Japanese learners of English have addressed whether or not they can accurately articulate new segments (Cochrane, 1980; Mochizuki, 1980; MacKain et al., 1981; Sheldon et al., 1982; Zimmerman et al., 1984; Flege et al., 1995; Riney et al., 1998; Riney et al., 2000). Mochizuki’s study conducted in a university in the United States was aimed at “... examin[ing] the identification of /r/ and /l/ by Japanese and Americans” and “... compar[ing] the linguistic behaviors of the two groups in response to natural speech and synthesized speech” (1980, p. 283). Her Japanese subjects were either graduate students or the wives of graduate students at a university in the United States, who had had formal instruction in English in Japan for at least eight years. The length of their stay in the US varied from six months to four years. The results show that, though the position of /r/ and /l/ in a word largely influence the Japanese’ perceptive and productive abilities of these two sounds, they were capable of identifying them auditorily and articulatorily at a rate much better than chance in most positions. It is also reported that in synthesized speech, Japanese subjects can identify /ra-la/ syllables as well as American subjects, although some Japanese showed poorer performance than those syllables in natural speech. Mochizuki’s study shows that the quantitative and qualitative English input that Japanese subjects received in a university in an ESL setting may help improve their perceptive and productive abilities in English segments.

Zimmerman et al. (1984) also looked at the articulation of /r/ and /l/ by one American English speaker and two Japanese learners of English. Of the Japanese subjects, one was a student at a university in the United States who had less experience in English with eight years of formal instruction in Japan, and had resided in the US for
three years. Another one who had more experience in English was majoring in Speech Pathology and Audiology at the same university and had the same instruction experience as the former Japanese subject in Japan. However, this subject had spent two years in Malaysia where English is used as a medium of communication by local people, had taken a year long course in intensive English including speaking and listening, and had resided in the US for about seven years. They investigated the articulation of the sounds with high speed cinefluography (i.e. a device to “track radiopaque markers secured with a dental adhesive . . . to the lips, jaw, tongue dorsum, and tongue tip”) and, it is reported that “… the productions were also on magnetic tape along with an octal code to allow alignment of kinematic and acoustic events” (Zimmerman, et al., 1984, p.188). The findings showed that there were great tongue positioning differences among the subjects. The less experienced Japanese subject seemed to substitute Japanese flap /ɾ/ for English /r/ and /l/.

Although this tendency of substituting was seen in the more experienced Japanese subject too, the experienced subject produced articulations that approximated those of the American’s. Zimmerman et al.’s study shows that, of two subjects who have the same experience in English formal instruction in an EFL setting, the one having the greater amount of exposure to English and more phonological training in English in an ESL setting had better articulations. This phenomenon may mean that the quantity and the quality of input in an ESL setting may contribute to a learner’s articulation skills.

Flege et al. (1995) also examined the production of /ɾ/ and /l/ by adult native speakers of Japanese who had studied English at school in Japan, but who were not exposed to English until they arrived as adults in the United States. The Japanese were
divided into two groups according to their length of residence in the United States (2 vs. 21 years) and were tested to produce English minimal-pair words which start with /r/ and /l/. The results indicate that the two sounds produced by Japanese who stayed in the United States for a longer period were identified correctly by native English listeners and received "only slightly lower ratings" than the ones produced by native speakers of English (Flege et al., 1995, p.51). In contrast, the sounds produced by the Japanese who stayed in the United States for a relatively short period were often misjudged by native English listeners and were described as strongly foreign accented. This study reveals that in an ESL setting the quantity of input the L2 learners receive has an influence on L2 learners' proficiency in L2 pronunciation.

Riney et al. (2000) also investigated the substitution of the Japanese /r/ for English /r/ and /l/ by 11 Japanese university students in a longitudinal (42 months) study. All of the subjects received 18 hours of pronunciation instruction after Time 1, and 2-3 hours were used for /r/ and /l/. The findings show that there is a strong correlation between the learners' substitution of Japanese /r/ and global foreign accent. The authors stated that the more a Japanese subject substitutes English /r/ and /l/ with Japanese /r/, the less the subject's pronunciation sounds nativelike. They also reported that between Time 1 and Time 2, there was no significant decrease in global foreign accent as a group, except for two out of three students who spent a year in the United States, who showed a great improvement in foreign accent. The authors finally concluded that Japanese /r/ substitution does not improve or is not gotten rid of in a short period of time, and
Japanese EFL speakers have a higher tendency to substitute Japanese /r/ for English /l/ than for English /r/ (Riney et al., 2000).

In fact, Riney et al. (1999) investigated the correlation between global foreign accent and voice onset time produced by 11 Japanese speakers of English and 5 native speakers of English in a university in Japan longitudinally, which is the same longitudinal experiment conducted by Riney et al. (2000). They stated while English word initial voiceless stops (/p/, /t/, and /k/) involve relatively long aspiration with a range of 58 to 80 ms (an average of 69.3ms), the Japanese corresponding set of voiceless stops is found to be intermediate value between aspiration and unaspiration with a range of 30 to 66ms (an average of 45.7ms). Their study looked at whether L1 Japanese intermediate lag of VOT change to L2 English long lag overtime and hypothesized if it does not, it may be an indication that the VOT difference between Japanese and English stops is too small to be perceived and adjusted to in the course of L2 learning. The findings showed that VOT of Japanese learners of English did not change over time as a whole: One Japanese subject had VOT values increase in length, one had them decrease, and the rest remained the same. Riney et al. attributed this result to “the constraints imposed by phonological similarity and pedagogical neglect” (1999, p. 298) and stated that VOT correlated with global foreign accent. It can be inferred from these two studies that not the quality but the quantity of input does not seem to be enough: The substantial amount of input and having to use L2 outside of the classroom appear to be the important factors for developing L2 pronunciation. Also the reexamination of the quantity or the quality of pronunciation instruction given to subjects might be necessary by the teachers.
Finally, Flege et al.'s study (2001) also supports my hypothesis that the quantity and the quality of L2 input have a great influence on L2 learners, although they did not specifically investigate the L2 learners' articulation. Flege et al. compared groups of Chinese adults living in the United States who differed in length of residence and occupation (students vs. nonstudents). The subjects were categorized as a group of (a) short length of residence & nonstudents, (b) short length of residence & students, (c) long length of residence & nonstudents, and (d) long length of residence & students, and took (1) an auditory identification of word-final English consonant test, (2) a 144-item grammaticality judgement test, and (3) a 45-item listening comprehension test. The results showed that the students with relatively long residence outperformed the students with relatively short residence, but there were no distinctive differences between the groups of nonstudents who differed in length of residence. Flege et al. analyzed these data that students with long residence appear to have not only quantitative input but also qualitative input by college education. These findings are valuable in that, the quantity and the quality of input seem to imbue L2 learners with better performance in aspects other than pronunciation in L2 language learning.

4.2 The Input and the Acquisition of L2 Segments

A body of studies on the segmental acquisition of L2 English by L2 learners, mostly Japanese learners, was presented. The common findings from these studies are: (i) For Japanese subjects, all had at least 6 years (in junior high schools and high schools) of formal instruction which focused on reading and writing in Japan (Mochizuki, 1980; Zimmerman et al., 1984; Flege et al., 1995; Riney et al., 2000); (ii) in an EFL setting in Japan, improvement in pronunciation is difficult sometimes even when there is
pronunciation instruction (Riney et al., 1999; and Riney et al. 2000); (iii) many subjects, who have resided in an English-speaking community with educational experiences, demonstrated improvement in pronunciation (Mochizuki, 1980; Zimmerman et al., 1984; Flege et al. 1995; Riney et al. 2000) or in other areas in L2 proficiency (Flege et al., 2001). From these findings, it can be suggested that not only the quantity but also the quality of L2 input (i.e., the naturalistic interaction with L2 users) seem to be relevant to improving performance at the segmental level. In other words, at the segmental level a greater amount of L2 learners' interaction with native speakers of the language in a naturalistic setting (i.e., an ESL setting) provides better input and facilitates better pronunciation level for L2 learners than input in a formal setting (i.e., an EFL setting). As noted above, for Japanese learners of English, it is unusual for them to receive pronunciation instruction in middle schools in Japan. Then, without much formal training in pronunciation, residing and going to school in an English-speaking community seem able to lead L2 learners to better pronunciation at the segmental level. Since L2 learners can improve segmental pronunciation through increased input (i.e., quantitative and qualitative) by native speakers at the segmental level, it can be assumed that the acquisition of segments is not so difficult. In other words, the acquisition of segmentals is not a barrier in second language acquisition.

4.3 The acquisition of Syllabic Structure

For the acquisition of syllabic structure, let us look at the syllabic patterns of Japanese and English. The basic syllabic structures Japanese adopts are V and CV syllables, although there is one exception that a CVC pattern could happen when the final consonant is a geminate or nasal /n/ which has allophonic variations. In contrast, English
permits not only the following syllable types, VC, V, CCV, and CVC, but also CVCC (past [pæst]) and CCVC (queen [kwin]) (Finegan, 1999). As seen, the most distinctive difference between Japanese and English syllabic structure is that English permits consonant clusters, while Japanese does not. This difference in syllabic structure is believed to neatly affect the pronunciation of L2 learners. Consider Japanese learners whose L1 permits only CV, V, and CVC patterns are in the process of learning English syllabic structure which permits more consonants on the edges of syllables. It is obvious that those Japanese learners of English need to learn this English syllabic structure somehow, but beginners tend to employ “epenthesis” (i.e., the insertion of a vowel to break up a consonant cluster). This is proved by the Japanese words which were originally borrowed from English:

I. Double consonant clusters

‘study’ [s(u)tadi]
‘slide’ [s(u)raydo]
‘place’ [pures(u)]

To account for the general rule of the epenthetic vowels by Japanese learners of English, it is necessary to categorize the consonant clusters of these borrowed words from English in more detail.

II. Double onset consonant clusters (#CC-)

i. stops + liquids -
‘plastic’ [pučas(u)ciku]
‘provider’ [pučobayda:] 
‘black’ [bura2ku]
‘brown’ [baru:un]
‘drive’ [doraybu]
‘train’ [toren]
‘class’ [kuras(u)]
‘group’ [guru:pu]
ii. fricatives + stops -
   'study' [s(u)tadi]
   'special' [s(u)peʃaru]
   'ski' [s(u)ki]
   'flower' [ʃurawə:]  
   'thrill' [s(u)riɾu]

III. Triple onset consonant clusters (#CCC-)
   fricatives + stops + liquids -
   'stripe' [s(u)toraypu]
   'spray' [s(u)purə:]  

IV. Word internal consonant clusters (-CCC-)
   'English' [inguris(u)]
   'homepage' [ho:mupeji]
   'option' [opuʃon]

V. Single coda consonant (-C#)
   'towel' [taoruu]
   'room' [ru:mu]
   'dog' [doʔgu]
   'cat' [kyaʔto]
   'bed' [be ʔdo]
   'cake' [keki]
   'steak' [s(u)te:ki]
   'salad' [sarada]

VI. Double coda consonant clusters (-CC#)
   'first' [ʃa:s(u)to]
   'third' [sa:do]
   'handle' [handoru]

(Note: Japanese tend to drop a vowel after /s/, /ʃ/, however Japanese orthography always adopts a letter which phonetically stands for [su].)

As presented above, the rule of epenthesis in consonant clusters in Japanese is very general: the insertion of /u/ happens to break up the clusters. After alveolar stops (/t/ or /d/), a vowel /o/ is inserted, as in 'dry' [doɾay] and 'cat' [kyaʔto], though the exception
such as ‘salad’ [sarada] exists. Further, after alveo-palatal affricates (/ɕ/ or /ʃ/) and a voiceless velar stop (/k/) in coda position, /i/ is inserted. In summary, the generalizations of epenthesis for consonant clusters are straightforward.

- General Japanese epenthesis rule
  After alveolar stops: /t/ or /d/ → /t/ or /d/ + /o/
  After alveo-palatal affricates: /ɕ/ or /ʃ/ → /ɕ/ or /ʃ/ + /i/
  After a voiceless velar stop in coda: /k/ → /k/ + /i/ /__# /
  In consonant clusters and elsewhere: C(CC) → C + /u/ (C + /u/C)

This epenthetic rule Japanese exhibits has a great influence on suprasegmentals, which will be discussed in the next chapter.
Chapter 5

Suprasegmentals in L2 Pronunciation

5.0 Shift from Segments to Suprasegmentals

In the previous section, it was argued that the acquisition of segments does not seem to be very difficult in L2 acquisition when L2 learners receive quantitative and qualitative input in a natural setting. When pronunciation is taught in a formal instruction setting, average L2 teachers without adequate knowledge of the nature and multi-function of suprasegmentals were (or are) likely to concentrate on teaching only the articulation of segments in isolation, probably hoping to reduce the learners' foreign accent and to develop native-like pronunciation at post-lexical or sentential level. However, it has come to be commonly recognized that it is not word-for-word accurate articulation but suprasegmentals that make the learner's speech more nativelike. Due to this fact, as Pennington and Richards (1986) state, L2 language teachers and researchers' emphasis has begun to shift from a narrow focus on accuracy at the segmental level to a broader focus on non-segmental information called suprasegmentals. In the following section, I discuss why suprasegmentals have been increasingly recognized as an important part of L2 language teaching, emphasizing the necessity of suprasegmental instruction in L2 classroom.

5.1 L2 Learner's Deviance in Suprasegmentals

One of the reasons suprasegmentals have drawn attention from L2 language teachers and researchers seems to derive from the fact that “... it is now known that accurate production of segmental features does not in itself characterize native-like pronunciation, nor is it the primary basis for intelligible speech” (Pennington et al., 1986,
In fact, it is reported that the deviance in suprasegmentals tends to be perceived as foreign accent more than deviance in segments does. Anderson-Hsieh, Johnson, and Koehler (1992) investigated the relationship between the nonnative pronunciation and the native listener judgments of deviance in segments, prosody (i.e. suprasegmentals), and syllabic structure. The nonnative English subjects were from 11 language groups, and the judges were ESL teachers and were also experienced raters of the SPEAK Test (i.e. a test used widely for evaluating the speaking proficiency of International Teaching Assistants at universities throughout the United States) (Anderson-Hsieh et al., 1992). The subjects were asked to read a passage from the SPEAK Test, and judges focused on intelligibility and acceptability as criteria, using a scale with seven points. The results revealed that prosody seems to have a greater effect on pronunciation rating than either segments or syllables, regardless of the subjects’ linguistic backgrounds. It can be deduced from this study that the correct usage of stress, rhythm, intonation, and phrasing with pauses determine the degree of intelligibility and acceptability.

Munro (1995) also examined the influence of nonnative English speakers’ suprasegmentals on the listeners’ judgment in foreign accent. His subjects included 10 Mandarin learners of English who resided in Canada as university students (the mean length of residence was 3.1 years) and 10 native speakers of Canadian English. The Mandarin learners were asked to read aloud several short sentences in English. After the recorded utterances were low-pass filtered (i.e., segmental information was removed), native speakers of English who were linguistically untrained judged using a 4-point scale whether the filtered utterances sounded like native or not. Munro found that the information remaining in the filtered speech such as pitch, intonation contours, locations
of stressed syllables, rate of speech, and durations of pause was prominent enough for the listeners to distinguish nonnative speakers from native speakers. He also reported that the Mandarin speakers' non-reduction of [d] in “and” [ænd] and non-flapped [t] in “sitting” [sɪnɪŋ] may have resulted in two noticeable effects as nonnative. Anderson-Hsieh, et al. and Munro’s studies suggest that the acquisition of suprasegmentals contributes to native-like pronunciation more than the accurate production of isolated forms of segments and words do. The mastery of segments in isolation is undoubtedly prerequisite for the acquisition of suprasegmentals, but the focus on segments in isolation is not sufficient for the overview of natural speech in L2.

5.2 The Mastery of Suprasegmentals

Although suprasegmentals are clearly important in native-like pronunciation, the acquisition of suprasegmentals has been considered extremely difficult. James who examined the development of L2 sound structure argues that “there is an ordering relationship between the different sub-representations of phonological structure as manifested in acquisition” (1988, p.158):

... the three main sub-representations of phonological structure – the lexical, the prosodic and the rhythmic... are ordered in the emergence in second language phonological development such that the lexical precedes the prosodic which precedes the rhythmic. (James, 1988, p.105)

James’ main claim is that the emergence of three main sub-representations of phonological structure toward the norm of L2 is commonly observed in the order of lexical (including phonemic) representation of L2 sound structure first, then prosodic representation, and finally rhythmic representation. It is notable that James in fact makes his own distinction between rhythmic representation from prosodic representation. In his
approach, since prosodic representation organizes linear structure of lexical representation (i.e., segments) by giving it the patterns of strong/weak nodes, prosodic representation is considered hierarchically higher than lexical representation. Furthermore, since rhythmic representation subsequently distributes the value of strong/weak binary nodes with respect to the rest of the sentence, rhythmic representation is hierarchically higher than prosodic representation.

In an account of the developmental process from lexical properties to prosodic properties, and from prosodic properties to rhythmic properties, James (1988) claims that the acquisition of higher properties presupposes the existence of lower properties, or at least presupposes that these properties are being established. More specifically, the existence of phonological properties of phonemes and lexical word units are the prerequisite for the emergence of prosodic representation to provide strong/weak nodes, and rhythmic representation packages these nodes according to the structural units combining prosodic strong/weak nodes (James, 1988). I agree with him. Before mastering the structurally higher properties, I believe L2 learners need to master lower properties.

For example, an L2 English learner may find it easy to say “one” /wʌn/ in English, because the word includes only three phonetic segments; thus the production may sound just like a native speaker of English. When the learner reads aloud a paragraph from literature written in English for five minutes, the learner’s production might not be able to avoid being perceived as nonnative. Consider as an example in which an L2 English learner reads aloud an English passage word for word. That speech would probably sound nonnative. This difficulty in producing native-like production is
that although the L2 learner succeeds in mastering accurate segments of L2 at the lexical level, the learner may face difficulty in producing nativelike running speech containing assimilations of neighboring sounds and suprasegmentals on which various morphological and syntactic operations are performed. In other words, L2 learners need to learn to produce not an abrupt combination of words but a smooth stream of sounds with proper stress, clitics, rhythm, and intonation of L2 to sound like a native speaker.

5.3 The Functions of Suprasegmentals

Another fact that explains the complexity and the difficulty of suprasegmental acquisition in L2 pronunciation is that, unlike segments, suprasegmentals are multifunctional in running speech and discourse. There are several ways in which prosodic features (i.e., suprasegmentals) function in discourse (Crystal, 1986; Chun, 1988):

5.3.1 Grammatical Function

Crystal states “... the prosodic feature(s) signal a contrast, the terms of which would be conventionally recognized as morphological or syntactic in a grammar” (1986, p.177). For example, the distinction between positive/negative, singular/plural, and statement/question, can be identified by the prosodic variations in a speech especially in tone languages. In the case of English, intonation can identify grammatical distinctions already overt in word order or morphology as the followings (Crystal, 1986, p.177).

\[(1) a. \text{You are coming / ARE’NT you. } \uparrow \text{ (asking)}
\quad \quad \text{b. You are coming / ARE’NT you. } \downarrow \text{ (telling)}
\quad \quad \text{c. I’ll ask the FIRST question / and you ask the SE\textsc{COND} one.}\]

Moreover, English intonation determines whether the utterance in declarative form is meant as a statement or a polite request, and whether the utterance invites or inhibits a response (Chun, 1988). L2 learners need to learn how syntactic operations signal
differences in meaning, but it is also important for them to realize that variation in intonation can alter the meaning of the sentence.

5.3.2 Semantic Function

Suprasegmentals also comprise a speaker’s “... organization of meaning in a discourse” and signal which parts of the utterance are most important and which are parenthetical (Crystal, 1986, p.177). Intonation variations signal that the information being used in a discourse is known or unknown between the speaker and the listener in a discourse. Also stress and intonation can emphasize specific elements which require the listener’s attention (Brown and Yule, 1983a; cited by Pennington et al., 1986).

5.3.3 Attitudinal Function

A speaker’s emotions such as anger, puzzlement, and surprise with regard to the subject matter or context of an utterance can be also signaled by suprasegmental variations. In particular, rhythmic variations which involve the speaker’s speech volume and rate may reflect the speaker’s attitude or emotional state: Fast speech with a loud voice is commonly associated with the view that the speaker is upset, whereas moderate volume and slow speech may indicate that the speaker is in an emotionally calm state.

5.3.4 Psychological Function

According to Crystal (1986), several experiments have indicated that a speaker’s speech with different prosodic variations has an influence on the degree of listener’s performance in short-term memory recall, perception, and other variables.

5.3.5 Social Function

Crystal (1986) states that a speaker’s prosody reflects his or her sociolinguistic characteristics, such as gender, class, and professional status.
It is important to note that these functions of suprasegmentals are mutually dependent on each other in a discourse. For example, the suggestion "You might want to close the door" could be uttered by a sister to her brother who just came into the room and let the cold air in (Chun, 1988). In this case, the sister’s intention is "I am telling you to close the door," and her monotone intonation might reflect her irritation and the close relationship between her and her brother. As seen, suprasegmentals used by interlocutors in discourse reflect different kinds of information which cannot be interpreted by examining the syntactic forms of the utterances in the discourse.

Among the suprasegmentals, the function of intonation in a discourse is reported to be very influential in English. Chun (1988) emphasizes the influence and the importance of intonational functions in the context of discourse. She introduces not only the various functions of intonation already presented above, but also Brazil’s (1975) findings that intonation signals the "intension of the speaker to complete a turn" and "desire or expectation of the speaker for the hearer to reply or assume the floor (or, on the other hand, an attempt by the speaker to discourage or inhibit a response by the hearer)". It can be inferred from these observations of intonation that the speaker’s choice of intonation in utterance-final position plays an important role in determining the listener’s next behavior in the discourse.

Native speakers of a language engaging in a discourse may be unconscious of the fact that their utterance in fact reflects these complicated functions of suprasegmentals of the language. Considering the complexity of the suprasegmental functions in a language, we can easily assume that a discourse between someone who speaks the language as L1 and another who speaks it as L2 can cause confusion for them in interpreting what the
other meant to say because the L1 speaker’s unconscious cues in speech may not be properly interpreted by the L2 listener. Then, multi-functional characteristics of suprasegmentals show us a couple of implications in L2 teaching. First, from the point of view of language as a system for communication, it is necessary for L2 language teachers to teach L2 learners that not only syntax or lexical choice but also suprasegmentals in a discourse heavily affect the interlocutor’s interpretation of the information being conveyed. Second, L2 language teachers need to be aware that the same suprasegmental feature could provide different interpretations in different languages. For example, in one language, an utterance with low intonation might signal politeness, while in another language it might signal impoliteness. When native speakers of these two languages have a conversation, one’s intention might be misunderstood by the other due to the different function of the same suprasegmental feature in these two languages. To avoid this confusion between interlocutors from different linguistic backgrounds in oral communication, L2 language teachers should teach L2 learners the basic intonation contours and their functions in L2 and that deviations from native speakers’ norm could imply an unspoken message.

On the other hand, L2 language teachers might be able to benefit from an examination of the cross-linguistically common function of suprasegmentals. Bolinger (1978) found that intonation is cross-linguistically used to “... form closures (descending lines, clause-final falls and non-falls) and to form accents (obtrusions for prominence, mainly upward). Terminals are almost universally low or falling for finality and assertion, and high or rising for lack of assertiveness and nonfinality, including yes-no questions” (cited by Chun, 1988, p.297). This intonational function which is commonly
shared by the L2 learner's L1 and L2 might be something that can be easily and quickly acquired by the learner. In summary, for communicative competence development, the suprasegmentals are a vital component of strategy that the L2 learner needs to learn for "negotiating meaning, managing interaction, and achieving discourse coherence" in the context of the natural discourse (Chun, 1988, p.295).

To conclude this section, I would like to suggest L2 language teachers incorporate suprasegmental teaching in the classroom. This overview on suprasegmentals has shown that an L2 language learner's deviance in suprasegmentals is more likely to be perceived as a foreign accent than the deviance in segments by a native listener; that the L2 learner's acquisition of suprasegmentals is one of the most difficult features to be learned in L2 language learning, so it happens at the very advanced stages in L2 learning; and that for real discourse purposes, the L2 learner needs to be aware of the multi-functional role of suprasegmentals and master them for a natural interaction in L2. As Chun (1988) states, unfortunately students have not been taught how to use suprasegmentals for discourse purposes in real communication, and typical beginning L2 learners are probably unaware of the multi-functional role of suprasegmentals. With the rise of L2 as a communication tool, the role of suprasegmentals in L2 pronunciation is tremendous. Considering this fact, as Cook (1996) claims, pronunciation instruction should be seen not as a matter of learning accurate segments but of learning the way people interpret each other in oral communication.
Chapter 6

Pedagogical Implications

6.0 What Feature Should the Priority in Teaching Suprasegmentals Be?

In order to compensate for adult L2 learners’ biological disadvantage in L2 pronunciation, L2 language teachers need to incorporate the teaching of pronunciation in the L2 classroom, so that L2 learners will be understood in real communication. To put it another way, L2 language teachers need to instruct L2 learners with the goal of reducing learners’ foreign accent enough so as not to distract the listener from the speaker’s message in communication. What we need to consider next is how to incorporate suprasegmental training in the classroom. With respect to teaching English as a second or foreign language, I believe rhythm and intonation should be the first features to be taught in suprasegmentals. This claim is based on the research findings that English rhythm, which exhibits stress-timing, seems to be one of the most difficult features to be learned by L2 English learners regardless of linguistic background, and that English intonation has a great influence on the interpretations of the utterance between interlocutors. These two aspects of the suprasegmentals of English are also reported as "...two major organizing structures that native speakers rely on to process speech," and they also draw the listener’s attention to the central information being conveyed in discourse (Wong, 1987, p.21). As Wong states, due to their major roles in communication, teaching rhythm and intonation can benefit L2 learners in discourse in social situations. Moreover, since there are always time constraints on both teachers and learners for teaching and learning, it is reasonable to work on the aspects that have the
greatest influence on communication. Teachers should thus give priority to teaching on rhythm and intonation.

6.1 Rhythm

Findings from a number of studies on L2 rhythmic acquisition have suggested that “. . . an accurate rhythm has been considered to be one of the most important phonetic aspects for the auditory comprehension and intelligible oral production of English” (Prator, 1971; Brown, 1977; Morley, 1979; Savingnon, 1983; Faber, 1986; Anderson, 1993; cited by Flores, 1993, p.151). One may have the experience of hearing a language spoken from a distance, and even though one cannot hear the sounds, one can distinguish whether the speaker is a native or nonnative speaker of one’s language. I have experienced it: When I was in a computer lab on campus while many students were talking, I could easily detect someone who was a nonnative speaker of English, even though I could not understand what the speaker was saying. The stream of sounds produced by the nonnative speaker of English seemed less smooth than native speakers.

6.1.1 Rhythmic Typology

With respect to the rhythm of a language, it is characterized by “. . . the timing pattern of successive syllables” (Wong, 1987, p.22). Major reports, “[l]anguages are traditionally classified into three basic types: syllable-timing, stress-timing, and mora-timing” (2001, p.18) This three-rhythm typology can be illustrated as follows:

```
Rhythm
   /     \
  /       \
Stress-timing Unstress-timing
   |       |
  |         |
Syllable-timing Mora-timing
```

Figure 6.1: Rhythm typology
First, languages can be classified as to whether they are stress-timed or unstress-timed. Unstressed-timed languages are further classified as either syllable- or mora-timed languages.

Stress-timed languages, exemplified by German, English, and Brazilian Portuguese, focus on the difference between stressed and unstressed syllables. The stressed syllables are much longer than the unstressed syllables and the durations between major stressed syllables are approximately equal in length, though there exists a variation in the number of intervening syllables (Major, 2001). Major states this length equality between stressed syllables is captured by lengthening stressed syllables and shortening or reducing unstressed syllables “... often to schwa [ə] as in give it to [tə] me” (2001, p.18) This centralization of a vowel in an unstressed syllable (i.e. schwa) is a key feature which characterizes stress-timing language rhythm.

On the other hand, syllable-timed languages, such as French and Spanish, carry syllables that are nearly of equal length in spite of stress. Therefore, the durations between stressed and unstressed syllables are almost equal in Spanish. Major (2001) reports that this type language gives the impression to a listener that “the language has equal beats and in rapid speech it sounds like rapid machinegun fire.”

Japanese is a prime example of a mora-timed language. This type of language has the unit of timing called the mora. Mora is defined as:

... a unit of phonological weight that captures differential behaviour of certain syllable types cross-linguistically. With respect to phonological phenomena like stress, it is often the case that not all syllable are treated equally. For example, it may be the case that a CVC syllable attracts a stress in a way that a CV syllable does not. For this reason, a distinction is made between light and heavy syllables. A light syllable is associated with one unit of quantity while a heavy syllable is associated with two units of quantity ... Thus, a light syllable is monomoraic while a heavy syllable is bimoraic ...
Vowels must project mora . . . Languages vary as to whether the coda consonants are able to project a mora. (Archibald, 1998, p.39)

To account for the characteristics of mora in Japanese, Tsujimura (1996) introduces three points: First, the basic syllables Japanese has, CV and V, are each counted as one mora, because vowels must project mora. Thus the Japanese word *okashi* “snack, sweets” has three morae, *o-ka-shi* ([shi] is monomoraic), while the word *okashii* “funny, weird” has four morae, *o-ka-shi-i* ([shii] is bimoraic) (for other examples, see Major 2001). Second, when Japanese permits CVC syllabic structure when (1) a geminate occurs and (2) a nasal occurs in coda. *Gakki* /gaki/ “musical instrument” including the geminate, first /k/, is considered as one mora. Therefore, *gakki* has three morae ga-k-ki. As in *ginkoo* “(money) bank,” the nasal in coda, /n/ is also counted as one mora, thus *ginkoo* has four morae gi-n-ko-o. These examples show that consonants in coda are able to project mora in Japanese. In summary, all permitted syllables and segments in Japanese, CV, V and (CV)C (i.e. a geminate or a nasal in coda) serve as morae in equal duration in Japanese.

6.2 The Acquisition of L2 Rhythm

A central issue for us is how rhythmic differences among languages affect adult L2 learning. Traditionally the difficulty of English rhythmic acquisition in L2 was attributed to the transfer from L1 to L2: If the learner’s L1 is syllable-timed and L2 is stress-timed English, he or she would use syllable-timing in English. For example, Wenk (1986) investigated the English rhythmic acquisition by French learners of English at beginning, intermediate, and advanced level. In order to examine the acquisition of vowel reduction (i.e. schwa) in English, the beginners were asked to imitate and read aloud an excerpt from an introductory English textbook, whereas the intermediate and
advanced learners were examined in reading, repetition, and description tasks (Wenk, 1986). The native judges rated the subjects’ speech based on whether the vowels had been produced in the native manner, reduced or non-reduced. In other words, the vowel reduction in unstressed syllables was taken as “... a meaningful sign of the ongoing acquisition of English speech rhythm” (Wenk, 1986, p.125). Wenk found that beginners’ speech sounded consistently heavy due to the non-reduced vowels, while advanced learners’ exhibited vowel reduction in unaccented syllables. Wenk attributes this French learners’ difficulty in mastering L2 English stress-timing to the transfer of L1 French syllable-timing. It also should be noted that this study affirms James’ (1988) finding that the acquisition of rhythm emerges at the latest stages in L2 acquisition.

With respect to the stressed and unstressed syllables in English nonnative speakers, Fokes and Bond (1989) studied the durations for the vowel production and argued that nonnative speakers rely on the timing patterns of their native language to speak English. The subjects included five foreign students who are native speakers of Farsi, Japanese, Spanish, Hausa, and Chinese in a university in the United States. It should be noted that, although Fokes et al. explain that Japanese and Spanish each presents mora-timing and syllable-timing, they did not provide rhythmic explanations on other languages, Farsi, Hausa (tonal language which contrasts long and short vowels), and Chinese (tonal language). Their production of two-syllable words (e.g., confess), three-syllable words (e.g., conclusion), and four-syllable words (e.g., confirmation) which all had the same prefix were examined. They reported that “many of the nonnative speakers had difficulty with the four syllable words, producing a variety of vowels in the first {con} syllable and failing to reduce the vowel of the second syllable” (Fokes et al.,
They also added that nonnative speakers had a tendency to produce stressed vowels too short and unstressed vowels too long; therefore they did not adopt reduced vowels for unstressed syllables. This study suggests that the longer the word is, the harder it is to learn the proper stresses, and that vowel lengthening and shortening are difficulties that L2 English learners have, regardless of the learners' L1 backgrounds.

The tendency to produce English with mora-timing is also reported as a typical characteristic of beginning Japanese learners of English. As already mentioned, Japanese speakers tend to follow the Japanese epenthesis rule to break up consonant clusters in a word, which results in the word having a CV pattern with equal length in duration. Tsujimura presents the example of how speakers of English and Japanese divide words into smaller units: "If we ask a native speaker of English how many parts there are in the word London, (s)he would be most likely to answer 'two'. If we ask a native speaker of Japanese the same question, however (s)he would probably say 'four'" (1996, p.64). Both speakers appear to follow their own language's phonotactic constraints and rhythm: The English speaker follows the CVC syllable and stress-timing of English, thus he or she breaks the word as lon-don. In contrast, the Japanese speaker is influenced by the CV syllable and mora-timing, he or she answers lo-n-do-n, which consists of four morae with equal duration (Tsujimura, 1996). This difference gives us a hint as to how Japanese learners of English might store English words mentally, which might eventually cause the L1 Japanese mora-timing to transfer to the production of English stress-timing. Furthermore, since morae in Japanese are of equal duration, the English syllables to which the epenthesis rule has been applied by a Japanese learner will be of equal duration.
At a glance, Wenk (1986), Fokes et al. (1989), and Tsujimura's (1996) work all seem to support the traditional idea that the rhythmic structure difference between L1 and L2 (i.e., learners' L1s are non-stress-timed languages and L2 is stress-timed English) is the cause of difficulty in the acquisition of English rhythm. However, an overwhelming number of studies on English rhythm suggests that the acquisitional difficulty of English rhythm by foreign learners is more than just a transfer of L1 rhythmic structure. Rather it seems to be a universal tendency for stresses to recur regularly which results in a failure to make sufficient difference in length between the vowels in stressed and unstressed syllables (Adams and Munro, 1978; Anderson, 1993; Taylor, 1981; cited by Flores, 1993; Fokes et al., 1989). Therefore, in fact whatever the L2 English learner's L1 rhythmic typology is (i.e., a stress-timed, syllable-timed, or mora-timed), those who learn English as L2 appear to have a problem producing appropriate vowel durations in both stressed and unstressed syllables in English. In this connection, consider the argument by James (1988) stated above: Rhythmic representation distributes the value of prosodic strong/weak nodes; thus the acquisition of rhythmic representation emerges after the acquisition of prosodic representation. Here the binary nodes, strong/weak, are equivalent to the notion of stressed and unstressed syllables. The frequent observation that L2 learners' tendency to produce vowels in unstressed syllables that are too long and vowels in stressed syllable that are too short indeed indicates that those L2 learners have already mastered the prosodic representation with strong/weak nodes, but have not attained the proper distribution of strong/weak value with respect to the rest of speech.

This idea also seems very interesting when we look at the canonical CV syllable produced by infants cross-linguistically. Young-Scholten (1993) argues that, in terms of
language acquisition, learners prefer CV patterns in both L1 and L2 learning: In L1 acquisition, children tend to use this syllable at the beginning of the phonological developmental stages, also in L2 acquisition, L2 learners tend to prefer this CV structure at the beginning stage (Young-Scholten, 1993). She further states “[w]hen we say a learner prefers a certain syllable type, what we mean is that the learner attempts to bring syllables in the target language into conformity with either the canonical syllable structure of the L1 or with a universal unmarked CV syllable” (1993, p.13).

Then, how do we know where the CV syllable produced by an L2 English learner comes from? Taylor’s study (1981, cited by Flores, 1993) which examined speech and reading by experienced nonnative teachers of English with various linguistic backgrounds showed that a dichotomy theory of transfer between the stress-timed language and unstress-timed language cannot explain a case in which the subjects with “non-syllable-timed native languages” (i.e., stress-timed languages) (Flores, 1993, p.155) had problems in English stress-timing rhythm. When those whose L1 is a stress-timed still tend to produce syllable-timing reflecting its CV syllabic structure, one may be able to say that in order to preserve this universal CV syllabic structure, L2 English learners unconsciously produce equal vowel duration in stressed and unstressed syllables in English. Then, it might be possible to say that the L2 English learners’ difficulty in vowel duration is not caused by transfer from L1, but a product of language learning that everybody goes through regardless of L1.

6.3 Teaching English rhythm

How can teachers make use of these findings regarding the difficulty of English rhythmic acquisition? First of all, one can say that teachers should recognize that the L2
English learners' difficulty in producing appropriate vowel duration in stressed and unstressed syllables is typical; thus this aspect should be emphasized in teaching. As explained above, in order to master proper English stress-timing, what L2 learners need to learn is that time for the unstressed syllables is squeezed so that it can give more time to the preceding stressed syllables, which provides equal time duration among stressed syllables. This phenomenon is realized by the vowel reduction in unstressed syllables to schwa. Bolinger (1981) who analyzed vowel types shows that three reduced vowels can occur: the mid-central vowel (i.e., schwa) /ə/; the front vowel /ɪ/ (as in the final sound found in some speakers’ pronunciation of the word city); and the back vowel /ʊ/ (as in the final sound found in some speakers’ pronunciation of the word potato) (cited by Wong, 1987). Although there exist three kinds of reduced vowels, it is unrealistic to teach the learners to distinguish them. This is because the differences in these reduced vowels are difficult to be perceived auditorily in the first place, since they are vowels in unstressed syllables. Rather, for learners, the acquisition of schwa, which is the most frequently addressed feature in various rhythmic textbooks, is an easier and more realistic goal.

Now how should schwa be taught to L2 English learners? This is the question that teachers face in teaching pronunciation. As mentioned above, although some teachers might be aware that some aspects of pronunciation need to be taught, it is another issue how to integrate them in the classroom. In that case, teachers can refer to teaching materials for English rhythm. I will briefly summarize the approach found in those teaching materials (See Appendix I. Bowen, 1975; Adams, 1979; Kenworthy, 1987; Morley, 1987; Wong, 1987; Flores, 1993; Dale & Poms, 1994).
First, English sample sentences are presented to familiarize learners with English stress-timed rhythm patterns. This is usually realized by the introduction of rhythm in general, such as lyrics from songs, the introduction of English poetry, a paragraph from English literature. Teachers may produce orally or play the recordings of these English rhythm examples, and have the learners use their body (e.g., tap the toe, clap hands, or use rubber bands) to physically realize the metrical foot they hear.

After this stage, various approaches can be found in different teaching materials. Basically, some teaching materials suggest teaching learners stress at the word level (words containing two- or three-syllables) first and then move on to the clause or sentential level (Morley, 1979; Kenworthy, 1987; Morley, 1987; Wong, 1987; Flores, 1993; Dale et al., 1994), while others approach the sentential level first and introduce stress at the word level (Bowen, 1979). When teachers begin with the word level, they may pronounce words including more than two syllables by themselves, or play recordings to have learners get used to the stress pattern of English and practice them. In contrast, when they begin with the sentential level, they need to prepare dialogues or paragraphs which include thematic words being stressed. At this point, learners are required to actively participate in classroom activities. In addition to this, outside classroom assignments should be included because there is too little time to be spent on every student’s pronunciation practice in the classroom. Learners’ individual practice should involve repetition of words or playing dialogues which require them to carefully attend to which syllable must be given stress primary. After this stage, usually pauses, linking, and assimilation phenomena are introduced as the last modification to master English rhythm patterns (Adams, 1979; Morley, 1987; Wong, 1987; Dale et al., 1994).
As Brinton and Goodwin (2001) argue, I believe it is more productive to teach learners to focus on stressed syllables than to de-emphasize unstressed ones. Although it is necessary to tell learners that there is a reduced vowel called schwa, since it is not taught to learners usually, and because many learners even do not know about it, there are several reasons for focusing on stressed syllables. First, it is stressed syllables that form the framework for rhythm and that communicate meaning. Second, asking learners to “think about something that should be made less noticeable” rather seems to be a contradiction (Brinton et al., 2001).

One way to have the learners focus on stressed syllables is to tell learners that they do not have to pronounce everything. The vowel in the unstressed syllable in a word of more than two syllables is rarely heard anyway (as in the second and third syllables of reasonable [riːənəbl]): Thus, instead of teaching learners this vague “vowel in an unstressed syllable,” it might be easier just to tell them to ignore the vowel in unstressed syllable, unless it occurs word-initially as in “about” /əbəwt/.

As to which level (word or sentence) should be introduced first to teach rhythm, based on the argument stated above, I think that the word level needs to be worked on first. As James (1988) claims for rhythmic acquisition, segments feed into prosodic representation for the realization of strong/weak nodes, and these binary nodes are the prerequisite for the actual distribution of rhythm in a sentence. To be consistent with this learning process, I believe that it is easier for learners to start with the word level by paying attention to each word’s primary stressed syllable, and move on to the sentential level to realize that according to the sentential structure, the location of stress has to be adjusted. Suppose that the stress on the sentential level was introduced first. Learners
would find it very difficult to track down which syllable in a word is primarily stressed in isolation especially when they need to modify the stress pattern in a sentence. Compare the following:

(2) th'irteen | Mississippí | législator

(3) Th'irteen Mississippí législator visited the White House.

(Salting, 2002)

When the same words are pronounced in isolation, the stress is assigned to its original location. However, when they are adopted in a sentence, the location of the stress originally seen in isolated words is distributed differently. Since rhythmic representation is always adjusted depending on the sentence structure, it is necessary to work on stress at the word level to have learners master the basic stress patterns first, and move on to the sentential level to learn how to adjust the patterns with respect to the rest of the sentence.

6.4 Teaching English intonation

As mentioned in the section on suprasegmental functions, people use intonation to signal how they perceive a particular situation and to send messages about this to listeners. Although there seem to be intonational universals among languages which could be very beneficial for L2 learners in mastering L2 intonation, it is dangerous for language teachers to set up definite intonation rules for learners because of the complexities of the intonational system which reflects individual and situational variations. For example, one may state that lexical words such as nouns, verbs, adjectives, and adverbs are usually produced prominently in a sentence, whereas functional words such as auxiliary verbs and prepositions do not. However, counterexamples can be often found to confuse the issue (Adams, 1979; Wong, 1987).
Let us see the following examples (Adams, 1979) (Capital letters show the stressed words):

(4) Its' a VERY good BOOK. \( \rightarrow \) an adverb *very* is stressed.
   It's not a very KIND thing to DO to them. \( \rightarrow \) an adverb *very* is unstressed.
(5) What is it USED FOR? \( \rightarrow \) a preposition *for* is stressed.
   It's for CLEANING the CAR. \( \rightarrow \) a preposition *for* is unstressed.

These examples indicate that the same adverb or preposition gets prominence on one occasion and not the other. As these examples demonstrate, "[t]here is not a set of exact complex alternatives which must be exactly imitated to achieve a desired effect" (Kenworthy, 1987, p.44). Thus, when language teachers integrate intonation teaching in a classroom, intonational meanings or functions should be introduced after the clear suggestion that these meanings or functions are not absolute but just tendencies. Bearing this in mind, teachers should encourage learners to explore and experiment with these tendencies outside the classroom (Kenworthy, 1987).

In order to avoid too complex and abstract patterns of intonation, I will focus on the teaching of rising and falling tone patterns on English questions, which is one of the most frequently taught characteristics of intonation (Bowen, 1975; Kenworthy, 1987; Wong, 1987; Dale et al., 1994; Thompson, 1995; Cauldwell and Hewings, 1996).

According to Thompson, a typical formulation of English rising and falling tone patterns seen in language textbooks is this (1995, p.236):

Questions with the answer *yes* or *no* go *up* at the end.
Questions starting with a *wh*-word (e.g. *what, where, which, who, how*, etc.) go *down* at the end.

With respect to this formulation, Thompson (1995) states that the formulation in fact does not work in a number of cases based on her observation of chat show or casual
conversation. She claims that, although the counterexample of wh-question pattern with a rising tone is very rare, that of yes-no question pattern with a falling tone is often seen:

<table>
<thead>
<tr>
<th>Tone choice</th>
<th>No. of occurrences</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>wh- + fall</td>
<td>6</td>
<td>what do you do then as a couple now</td>
</tr>
<tr>
<td>wh- + rise</td>
<td>1</td>
<td>how much more therapy have you got to go through now</td>
</tr>
<tr>
<td>yes/no + fall</td>
<td>4</td>
<td>were you prepared for it to come back now</td>
</tr>
<tr>
<td>yes/no + rise</td>
<td>5</td>
<td>has it brought you closer to God</td>
</tr>
</tbody>
</table>

Table 6.1: The chat show

<table>
<thead>
<tr>
<th>Tone choice</th>
<th>No. of occurrences</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>wh- + fall</td>
<td>12</td>
<td>what in fact did you do</td>
</tr>
<tr>
<td>wh- + rise</td>
<td>0</td>
<td>are you going to America</td>
</tr>
<tr>
<td>yes/no + fall</td>
<td>8</td>
<td>are you going to America</td>
</tr>
<tr>
<td>yes/no + rise</td>
<td>6</td>
<td>d'you teach</td>
</tr>
</tbody>
</table>

Table 6.2: The casual conversation

(adopted from Thompson, 1995, p237)

These tables show that the fact that a falling tone on yes-no questions is frequently used in communication is not compatible with what is introduced in English pronunciation textbooks. Thompson reports that “... [There is] only one instance ... of a falling intonation pattern on yes-no questions exemplified even as an exception to the general 'rule'” (1995, p.238), and analyzes this neglect:

It is possible that this neglect is due to the lack of an appropriate pedagogic explanation of the function of falling tone on yes/no questions, or it may be because designers of EFL teaching materials are unaware of the frequency of this pattern. Another contributing factor may be the understandable ‘intertextuality’ of many EFL materials, such that materials writers are strongly influenced by the models used by previous generations of writers. (Thompson, 1995, p.238)
Regardless of these possible reasons for the neglect of a falling tone on yes-no questions, I believe a falling tone on yes-no questions should not be ignored in intonation teaching. Although we still need to be cautious to establish a straightforward rule which explains intonation form and its meanings or functions, an alternative approach in teaching a rising and a falling tone is actually seen (Hudson, 1975; and Tench, 1988; cited by Thompson, 1987; Brinton et al., 2001). This approach sees the function of tonal choice on yes-no questions with respect to the notion of “conductive” and “non-conductive”:

Conductive questions are those to which the speaker thinks he or she already knows the answer, and expects the addressee to confirm expectations, while non-conductive questions are those to which the speaker does not think he or she already knows the answer, but thinks that the addressee does. (Thompson, 1995, p.239)

Thompson states that research findings by Brown, Currie, and Kenworthy (1980) and Tench (1988) are consistent with her analysis on a rising and a falling tonal usages: “... a rising tone on a yes/no question indicates a non-conductive question (a ‘real’ or ‘open’ question) whilst a falling tone suggests a conductive (a ‘leading’ or ‘checking’) question” (1995, p.239). I find this approach highly valuable and effective in teaching since it can be applied not only on yes-no questions but also on declarative utterances and tag questions. For example, declarative sentence

(6) You are coming to my house.

can signal different meanings with contrastive tones: a rising tone means that the speaker does not know if the listener is coming (non-conductive question), while a falling tone means that the speaker already kind of knows that the listener would come, but making sure just in case. For tag questions, the same analysis can be made:
(7) She is nice, isn’t she?

With a rising tone, again it signals the speaker is asking the listener because he or she is curious about the woman’s personality, while with a falling tone, it implies the speaker expects the listener to confirm what he or she said. From these observations, it is possible for teachers to formulate the following rule: When you ask a yes-no question, if you don’t know the answer, use a rising tone, but if you think you may know the answer, use a falling tone.

In summary, the overall view of teaching intonation suggests that generalizations of intonation are quite difficult. No matter how explicitly the examples of sentences or dialogues presented in textbooks are, there are no definite “if-then” rules that learners can rely on all the time. Although there is a relatively simple tendency such as a rising and a falling tone on yes-no questions and its meanings, in real communication learners need to learn how to acknowledge information given by their interlocutor in their responses (Wong, 1987). For this purpose, teachers should tell learners to pay attention to which elements are the focus in the conversation and what the interlocutor expects from their speech. In order to do so, teachers can begin by introducing the function of intonation, by providing speech samples that illustrate the dynamics of intonation, and teaching learners how to perceive them, as Wong (1987) states. This is probably the most reasonable and realistic approach to begin with in teaching intonation.

6.5 When should teachers give feedback?

We now come to one of the most important issues in pronunciation instruction, which is when and how to give effective feedback to learners. First, it is necessary to consider who could provide the feedback. Celce-Murcia et al. (1991) suggest several
possible sources for feedback that learners need: students themselves, peers, tutor, and teacher. Self-correction by learners themselves or feedback by peers can be very effective. Self-correction encourages the learner to be autonomous. Peer feedback, especially by those who share the same L1 backgrounds helps learners to develop their own monitor through oral activities such as speeches, interviews, or role-plays. However, one needs to remember that these kinds of feedback can happen only when learners have already developed correct representation of L2 pronunciation. In other words, learners need to be aware of their own errors. Teachers (or tutors), then, need to step in to make learners aware of errors that they are yet able to distinguish. Although criteria on when and how to give the feedback could vary depending on the teaching setting, I will talk about the case of pronunciation instruction in an EFL or ESL classroom here.

First, let us consider for what kinds of errors teachers need to give feedback on. Celce-Mulcia et al. point out four criteria to correct learners’ errors:

1. Does the error cause a breakdown in communication?
2. Is this a recurring pattern or an isolated mistake?
3. Does this error stigmatize the student?
4. Can it be corrected easily?  

(1991, p.147-8)

In fact, these all can be applied to the feedback on other language skills, but in L2 pronunciation, the first and the third criteria should be the priority in determining whether to correct errors, because these errors cause a detrimental outcome in communication on the spot. Thus, one may say that errors that affect the interlocutor’s response need the teacher’s feedback on the spot. For the second and the fourth criteria, they require teachers to keep error samples for a certain amount of period to become to be able to
analyze the characteristics of those errors. When teachers gather a body of mistakes which seem to be recurring, those mistakes should be corrected. This approach might be effective especially when learners' L1 is homogenous.

Next, it is necessary to consider how to give the feedback. As stated above, for those errors which inhibit the communication on the spot, teachers need to interrupt and point out what the problem in the learner’s pronunciation is and subsequently provide a specific solution to the problem. With respect to feedback on the spot, the way to correct the error might be done orally or, as Celce-Murcia et al. suggests, “silent correction” which uses signs or symbols posted in the classroom to point errors:

![Silent correction chart](image)

Figure 6.2: Silent correction chart

(adopted from Celce-Murcia et al., 1991, p.148)

I recommend using silent correction first so that teachers do not interrupt the learners’ speech interaction, but if the learner cannot recognize what the problem is by the signs, teachers need to give correction orally.

On the other hand, for errors which do not hinder communication but which occur frequently, teachers may take note and give the correction later providing a periodical handout on pronunciation. Regardless of whether the error inhibits communication on the spot, teachers must avoid overcorrection of pronunciation at any rate, since, overcorrection tends to discourage learners from improving their pronunciation, and it interrupts the learners’ oral activities.
To conclude this section, I claim that pronunciation feedback should not be done in such a way as to change the learners' pronunciation because it is a deviated norm. Such an objective is unrealistic for English which is considered an international language today. Rather pronunciation feedback should focus on teaching how to speak more clearly and effectively so that it smoothes the oral communication.
Chapter 7

Conclusion

In this thesis, I have claimed that the importance of pronunciation in L2 teaching needs to be recognized as the communicative approach is widely adopted in language teaching. Traditionally, teaching pronunciation was a neglected area in L2 teaching. If it existed at all, it focused on an articulatory segmental approach based on CA. Although there seems to be a biological disadvantage in learning L2 for adult learners, I have argued that individual motivation and instruction largely determine how well the L2 learner's pronunciation will be. Especially when we look at the fact that teaching pronunciation was rarely done, it is highly expected that there is room for improvement in L2 pronunciation that L2 teachers can contribute to L2 learners.

Many studies on segmental acquisition in L2 pronunciation report that, quantitative and qualitative input that adult L2 learners receive in a natural setting seem to imbue the learners with better articulation (Mochizuki, 1980; Zimmerman et al., 1984; Flege, et al., 1995; Riney et al., 1998; Riney et al., 2000). This finding indicates teaching implications in a formal setting that L2 pronunciation instruction should not be seen as easy task for both L2 teachers and L2 learners: L2 teachers need to provide pronunciation instruction which effectively leads learners to a better pronunciation in a relatively long duration, and L2 learners need to pay attention to how segments, syllables, and suprasegmentals should be produced in as many occasions as possible.

Recent studies on L2 pronunciation indicate that suprasegmentals, rather than segments, should be focused on. The supporting evidence is: (1) Deviance in suprasegmentals is more likely to lessen the intelligibility and comprehensibility of the
speech than deviance in segments (Anderson-Hsieh, et al. 1992; Munro, 1995); (2) The acquisition of suprasegmentals, especially rhythm, appears to be mastered at the most advanced stages in L2 pronunciation suggesting that suprasegmentals need to be taught due to their difficulty (James, 1988); and (3) Suprasegmentals can signal implicit messages in various ways which syntax or lexical choice cannot (Crystal, 1986; Chun, 1988). For teaching implications in L2 pronunciation, I prioritized rhythmic acquisition due to its difficulty, and intonation due to its complexity. Although further research on pronunciation teaching is necessary, I would like to recommend that L2 teachers explore many helpful guidance materials in teaching L2 pronunciation which are available.

As I progressed on the thesis, one thing that I became aware of is the complexity of suprasegmentals. The notion of stress, rhythm, length, and intonation is indeed inseparable from one another whenever one of them is accounted for, which made it really difficult for me to describe teaching implications on rhythm and intonation in Chapter 6. This fact implicitly suggests to us that this interrelated nature of suprasegmentals is the very reason that mastery of suprasegmentals is arduous and intricate. The context-dependent sensitivity of suprasegmentals also makes the mastery difficult for L2 learners. In terms of this pragmatic nature of suprasegmentals, oral activities focusing on suprasegmental forms and their function in various situations are important in teaching suprasegmentals in the L2 classroom. If permitted, learners should be encouraged to explore and experiment with what they learned in the classroom in social situations (Kenworthy, 1987). In that situation, learners' socio-psychological constraints may influence the quality of their interaction with L2 users: L2 learners who feel more empathy toward the L2 group or are integratively motivated are more likely to
apply what they learned in the L2 classroom to social situations than those who do not feel empathy and are instrumentally (Schumann, 1976; Gardner et al., 1972). The learners’ real-situation experiences would give them confidence in oral communication, help them recognize the way suprasegmentals are used on social occasions, and eventually automatize the process of applying suprasegmentals form and their function.

Finally I hope this thesis helps L2 teachers and learners who are interested in pronunciation improvement understand the overview of the acquisition of pronunciation. As a native speaker of Japanese, I cannot help hoping that this thesis provides tips for ESL or EFL teachers who have Japanese learners in teaching them pronunciation.
Appendix I

Outline of Teaching English Rhythm

(Bowen, 1975; Adams, 1979; Kenworthy, 1987; Morley, 1987; Wong, 1987; Flores, 1993; Dale et al., 1994)

I. Bowen, J.D. (1975)

1. Recognition of English rhythm in sentences
2. Repetition and observation of rhythm pattern of English sentences
3. Comparison of English rhythm patterns - sentences
4. Comparison of English rhythm patterns - words
5. Stress shift by verb alternation to noun by suffix /-ən/
6. Rhythmic shift of strong stress in two-syllable and three-syllable modifiers

II. Adams, C. (1979)

1. Recognition of stress-timed rhythm
2. Explanation of principles of English speech rhythm and comparison with syllable-timed rhythm
3. Audio-visual presentation of basic English rhythm patterns
4. Extension of patterns to non-metrical material
5. Analysis of word forms normally stressed in connected utterance
6. Organization of the sense group
7. Disjuncture and the pause in phrasing
8. Demonstration of the principle of gradation
9. Timing of the rhythm unit
10. Explanation of the speech stress profile - degrees of stress
11. Explanation of the rules governing liaison, assimilation, and juncture
12. Relationship between stress, rhythm, and intonation

III. Kenworthy, J. (1987)

1. Recognition of stress at word level in place and people names
2. Identification of stress and alternation in strong and weak beats in rhymes, verses, limericks, children’s games, and lyrics of songs
3. Stress placement at clause or sentential level by focusing on important words
4. Shifting stress in dialogues
5. Recognition of weak forms by comparison with meaning words and grammar/functional words:
that – vowel reduction
and – omit of /d/ as in fish ’n chips
of – omit of /f/ as in cuppa (cup of)
(6) Recognition of schwa in unstressed syllables and weak forms

IV. Morley, J. (1987)

(1) Introduction to stress, rhythm, and intonation.
(Key sentences for stress, rhythm, and intonation)
(2) Accented/unaccented syllables
(two- and three-syllable words, and reduced syllables and schwa)
(3) Syllables and suffixes
(past tense and plural)
(4) Sentence sense: rhythm and stress
(rhythm and sentence stresses, rhythm and reduced words, and rhythm and linking)
(5) Elisions and assimilations
(one and two-word contractions and linking, contractions and sound changes)
(6) Intonation
(final rising/falling intonation, nonfinal intonation)

V. Wong, R. (1987)

(1) Introduction of English rhythm: rhythm in general to rhythm in English
(2) Syllable length in three-syllable words by showing length as a set of lines or circles
and rubber bands
(3) Stressed syllables and syllable length by using dictionary
(recognition of primary stress and secondary stress)
(4) Full and reduced vowels and syllable length
(introduction of schwa in reduced syllables)
(5) Relative syllable length (cap vs. cab), minimal dialogue pairs
(6) Pauses and thought groups
(7) Rhythm and linking sounds

VI. Flores, B.C. (1993)
*Lenguas Modernas* 20: 151-164.

(1) Introduction to English rhythm
(two rhythmic patterns graphically presented and echoed with the nonsense syllable ti for
unstressed syllables and TA for stressed ones)
(2) The auditory perception of the rhythmic patterns
(a number of words, phrases, and sentences with rhythmic patterns)
(3) The oral production of the rhythmic patterns

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
(repetition of a number of words, phrases, and sentences with rhythmic patterns)
(4) The oral production of dialogues with rhythmic patterns

VII. Dale et al. (1994)

(1) Stress at the word level (two-syllables words)
(2) Stress in noun/verb pairs
(3) Stress in dialogues
(4) Stress in English poem
(5) Stress within the sentence (content words vs. function words)
(6) Rhythm (contractions, linking and word reductions, and sound changes)
Bibliography


Arbor, Michigan: University of Michigan Press.


