How to Teach English Phonetic Components to Speakers of Japanese: A Review of Previous Studies

Shoji Mitarai[1] & Norie Moriyoshi[2]

日本人学習者に対する第二言語としての英語音声学教授法について

御手洗　昭　治[1]　　森　吉　紀　江[2]

Abstract

The purpose of this paper is to compare some salient suprasegmental and segmental differences between English and Japanese, to identify difficulties that Japanese learning English may encounter, and to revisit effective pronunciation teaching based on previous research findings. Current tendency is to focus on integration of suprasegmental and segmental features rather than choosing either of them. Regarding setting priority in teaching pronunciation, a consensus between experienced teachers on relatively important pronunciation features for Japanese learning English was recently provided by Saito (2014). As for teachers' roles and responsibilities, there are many factors (e.g., learners' goal, proficiency level, and development stage) which pronunciation teachers need to take into account. With this in mind, teachers are responsible for selecting appropriate approaches, instructional materials, and learning activities from a wide range. Above all, phonemic distinction exercise, articulatory descriptions or diagrams seem beneficial especially for beginners. Form Focused Instruction (FFI) and social interaction with others are regarded to be effective approaches to improve phonetic abilities in spontaneous communication.

1. Introduction

In today's global society, command of English is required in many fields and mastering English would expand future career opportunities. MEXT (i.e., the Ministry of Education, Culture, Sports, Science and Technology) has generally revised the courses of study for elementary and secondary schools once every ten years. With regards to criticism that Japanese do not have sufficient communicative skills in English, the following contents include in the revision of 2008: junior high school students are not only expected to "speak accurately to the listener(s) about one's thoughts and feelings or facts," but also "to become familiar with the basic characteristics of English sounds

所属：
[1] 札幌大学地域共創学群教授・藤女子大学非常勤講師
[2] 札幌大学地域共創学群非常勤講師
[1] School of Regional Collaborative Studies at Sapporo University & Part-time Instructor at Fuji Women's University
[2] School of Regional Collaborative Studies at Sapporo University

such as stress, intonation and pauses and pronounce English sounds correctly" (MEXT, 2008). However, English skills of junior high and high school students in Japan have failed to meet the government's targets and a ministry official commented that the number of schools conducting English speech and discussion classes is still low (The Japan Times, 2016).

Approximately sixty years ago, Edwin O. Reischauer (see Appendix A), professor emeritus at Harvard University and the former American ambassador to Japan under the Kennedy administration aptly pointed out the major problem area in English language teaching in Japan. In view of a usable facility of communication, he put it this way:

> The chief problem area, however, has been in the area of communication. Of the many, many thousands, now really millions of students who have gone through the English language learning process in this country, only very few have emerged what a usable facility of communication in the English language, an ability to express their thoughts and make themselves clearly understood in English. I think that would all agree that this aspect of English language teaching has been the least satisfactory, although in the present world, this has perhaps become the most important aspect of English language learning.

Concerning issues associated with phonetics for the Japanese learners of English, Reischauer asserted:

> Actually phonetics must be learned early. What one learns in the first stages often determines what one will always do after that. It is extremely difficult to learn the sounds of a language incorrectly at first and then learn them better later on. One should learn correct phonetics from the beginning, and therefore, I believe Japan has to put more emphasis on correct pronunciation from the very beginning if students are going to achieve a mastery of spoken English. They very different syntax of English, also, I think, poses a similar problem. There has to be a strong effort from the very beginning to try to get students to express themselves in English so that they acquire a mastery of this very different type of organization of their thoughts. What often happens in Japanese education, I am afraid, is just the opposite. (Reischauer, 1961)

To date, it has been believed that acquiring English sounds properly is regarded as challenging for Japanese learning English due to the different phonetic systems between English and Japanese (Saito, 2007). "Pronunciation, which encompasses dimensions associated with linguistic attributes of spoken language (e.g., prosody, segmental accuracy), is arguably one of the hardest skills to acquire" (Trofimovich, Isaacs, Kennedy, Saito, Crowther, 2016; p.4-5). This paper, thus, reviews some prominent suprasegmental and segmental differences that many Japanese learners typically encounter in the process of their English learning based on contrastive analysis. Following previous studies, the present paper also presents resulting pedagogical implications for pronunciation teaching.

Contrastive analysis

Current theory is that L2 speech production is influenced by universal and cross-linguistic factors (Lowie, 2013), indicating that Contrastive Analysis which is the structural comparison of two languages may not always become a predictor to detect errors made by learners (Cai & Lee, 2015) and it may not always be able to clarify that observed errors by learners are attributed to interference by L1. Even though learners' first language (L1), age, and educational background are

identical, there can be discrepancy of phonetic abilities among learners (Yoshida, 2006). However, L1 plays a role in phonetic learning and the phonology of the learner's L1 should not be overlooked particularly in EFL contexts where a group of learners is often homogeneous (e.g., Carey, Sweeting, & Mannel, 2015; Noguchi, 2014; Ohata, 2004). "As student issues often correspond to typical errors made by learners from particular language backgrounds, teachers can rely on contrastive analyses of the L1 and target language to ascertain where such problems might lie" (Celce-Murcia, Brinton, & Goodwin, 2010, p.43). Nation and Newton (2009) also state that speakers whose first language is the same tend to make the same kinds of substitutions and show similar patterns of pronunciation. There seems to be a reasonable degree of predictability in the types of relationships between L1 and L2.

Hence, the notion that L1 somewhat interferes with L2 speech sound (Flege, Frieda, & Nozawa, 1997; Fledge, Bohn, & Jang, 1997) still plays a role and cross-linguistic analyses have been conducted in order to detect features that may affect learners' comprehensibility[1] and/or intelligibility[2] (e.g., Saito, 2013; 2014; 2015).

Age

Whether the pronunciation attainment is the result of age-related factors or not is still the subject of much debate, because the differences in the target languages, subjects, and research methodologies reach consensus difficult. Some theoretical accounts lay emphasis on critical period hypothesis—a notion that only language learners who have not passed the "critical language development age" are possible to acquire second language sound system, or cognitive aging hypothesis—a claim that various abilities including speech sound processing gradually decline associated with cognitive aging and affect second language acquisition (Hakuta, Bialystok, Wiley, 2003; Piske, MacKay, & Flege, 2001) and advise to start young (Lowie, 2013).

Yet according to psychological perspectives, the notion that age and attainment of pronunciation may somewhat correlate does not consistently apply to everyone. Pronunciation is a part of learner's personality and the more they become older, the more they become protective of their personality and unwilling to change it (Guiora, Beit-Hallami, Brannon, Dull, & Scovel, 1972). In addition, other studies have demonstrated that even adult L2 learners can alter their perception performance (Iverson, Hazan, & Bannister, 2005; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994) and can improve speech sound production (Bongaerts, Van Summeren, Planken, & Schils, 1997).

Goal of pronunciation learning

Each learner may have different phonetic learning goals and they should certainly be respected; however, many researchers have stressed that the goals of pronunciation learning should aim at improving comprehensibility rather than aiming at native-like accent (e.g., Jenkins, 2000; Levis, 2005; Munro & Derwing, 1999; Saito, 2012; Ur, 1991). "The ultimate goal of L2 instruction is to lead students to attain comprehensible speech and enhance their interlocutors' successful comprehension" (Saito, 2012, p.851) rather than eliminating accents. The reasoning behind is that a variety of

---

[1] The definition of comprehensibility is a listener's perception of how difficult it is to understand the utterance or message (Derwing and Munro, 2005, p.385).

[2] The definition of intelligibility is the extent to which a listener actually understands an utterance or message (Derwing and Munro, 2005, p.385).

English dialects have acknowledged as lingua franca or World Englishes (e.g., Celce-Murcia et al., 2010; Kirkpatrick, 2007; Matsuda, 2002) and it is recognized that most adult learners who learn a foreign language tend to have some sort of accent (Leaver, Ehrman, & Shekhtman, 2005). In fact, "inaccurate pronunciation has more potential to seriously interfere with understanding" (Mackey, Gass, & McDonough, 2000, p.493); however, comprehensibility does not necessarily correlate with the degree of accentedness[3] of speech and having a foreign accent does not definitely lead to communication breakdowns (Munro & Derwing, 1999).

2. English and Japanese phonetic features

The phonetic system comprises two main components: suprasegmentals[4] and segmentals (phonetic features). Suprasegmentals which "extend over more than one sound segment in an utterance, over longer stretches of speech" (Chun 2002, p.3) generally refer to elements of prosody including pitch, intonation, loudness, stress, duration and rhythm (Cutler, Oahan, & Donselaar, 1997; Seikel, King, & Drumright, 2016), whereas segmentals refer to individual sounds such as consonants and vowels.

2.1. Suprasegmentals

One of the most interesting characteristics of suprasegmental differences between Japanese and English is the amount of time to say a sentence. The time differs in Japanese depending on the number of syllables instead of the number of stressed vowels the sentence contains. Consequently, Japanese people may unconsciously know that the same amount of time must be spent for each Japanese syllable regardless of whether the syllable is stressed or unstressed (i.e., mora-timed language or syllable-timed language). Japanese is also a pitch accent language in which words differ in pitch patterns (Kondo, 2009; Tsujimura, 2014).

English is, in contrast, categorized as a stress-timed language (Ohata, 2004) and is also a stress-accent language (Graham, 2013) where the accent is expressed by a combination of pitch, duration, intensity and vowel quality (Kondo, 2009, p.105).

A case in point is that Japanese pronounce "McDonalds as "マクドナルド" ("Makudonarudo") with more or less equal length and stress. This indicates that Japanese accent is manifested solely by pitch. When some syllables are given more prominence, Japanese is tied more in with pitch than stress. On the other hand, in English, the accented syllable is pronounced with a strong stress, such as "Michael"—not "マイケル" ("Maikeru"). The stressed syllables are marked by making vowels louder than other relatively not important sounds. In addition, English vowels in a phrase or a sentence are not pronounced with the same duration and stressed syllables are pronounced longer.

To be more specific, three main differences between the stress-timed language and syllable-timed language are syllable structure, vowel reduction and lexical stress (Nation & Newton, 2009). Firstly, stress-timed languages allow several patterns of consonants (C) and vowels (V): VC, V, CCV, CV and CVC, both open and closed syllable types including complex clusters.[5] A word can end with

---

[3] The definition of accentedness is listener's perception of how different a speaker's accent is from that of the L1 community (Derwing and Munro, 2005, p.385).

[4] A synonym for prosody

[5] Some examples of English open and closed syllables are: sea (CV), sit (CVC), spin (CCVC), spill (CCVCC), spring (CCCVCC)

a consonant as well as initial and final consonant clusters such as: twin (CCVC), stress (CCCVCC), past (CVCC).   Japanese, syllable-timed language, in contrast, permits open syllables only,[6] indicating that it basically permits syllables of the forms CV and V, as well as CVC (Finegan, 2004).   The consequences of such cross-linguistic differences are likely to cause a problem that Japanese tend to insert a vowel between consonants, and end up saying "torai," "dorinku," and "torein" instead of "try," "drink," and "train."   Secondly, unlike syllable-timed languages, "stress-timed languages are more likely to use centralized vowels in unstressed syllables and vowels may be shortened or omitted" (i.e. vowel reduction) (Nation & Newton, 2009, p.90), such as unstressed /ə/, in "pizza" (i. e., pitsə) or "Brazil" (i.e., brəzɪl).   Thirdly, stress-timed languages usually have word level stress (i. e. lexical stress)" (p.90), which is a phonemic realization of the stress.   Some examples are: "PREsent" used as a noun (i.e., gift) VS. "preSENT" used as a verb (i.e., to show or give something formally), "ADdress" used as a noun (i.e., the location of a building) VS. "adDRESS" used as a verb (e.g., to speak to a group of people).   These sets of words have the same spelling but different syllable stress.

Given the differences between stress-timed languages and syllable-timed languages, Kondo (2009) examined the differences in vowel reduction between eight Japanese and four American English speakers.   The result showed that native English speakers reduced vowel duration significantly more than the Japanese speakers did in unstressed vowels and native English speakers' unstressed vowel were centralized.   Likewise, Meng et al. (2009) demonstrated that Japanese also had problems in learning English lexical stress.   A reason for those results possibly lies in lack of lexical stress as well as vowel reduction in Japanese.

Furthermore, Smith (2012) analyzed a sample dialogue between two Japanese and it was compared with the one produced with General American (GA) accent.   The study showed that Japanese speakers tended to employ word stress differently in words consisting of two or more syllables in length, such as "Manchester," "happening," and "kettle."   Regarding sentence stress, they produced "Oh, let me put the kettle on" with a very flat-sounding.   In particular, Japanese does not have a secondary stress unlike English.   Besides, Japanese exhibits a large number of words without accent in contrast to English which all content words have at least one stress to indicate the prominence of the words (Tsujimura, 2014).

2.2. Segmentals
2.2.1. Vowels[7]
Apparently, there are more vowels present in English than in Japanese.   Japanese contains only five monophthongal vowels (/i/, /ɛ/, /a/, /ɯ/, /ɔ/).[8]   In contrast, American English (AE) contains twelve vowels[9] (i.e., /i/, /ɪ/, /e/, /ɛ/, /æ/, /ə/, /ʌ/, /u/, /ʊ/, /o/, /ɔ/, /ɑ/) or thirteen vowels (i.e., /i/, /ɪ/, /e/, /ɛ/, /æ/, /ə/, /ʌ/, /u/, /ʊ/, /o/, /ɔ/, /ɑ/, /a/) and three diphthongs (i.e., /aw/, /

---

[6] Some examples of Japanese open syllables are: 蚊ka (CV), 彼kare (CVCV), 枯れるkareru (CVCVCV)

[7] Depending on analysis of linguists or phoneticians, the number of vowels identified in English and Japanese can differ (Ohata, 2004).

[8] Japanese /ɔ/ is not the same as English sound /ɔ/. Each of Japanese vowels has a greater range of variation than English vowels.

[9] Phoneme inventory varies depending on classification and some contains other sounds (e.g., /ɜ /). In England and in certain parts of the United States, including New York City, sixteen distinct vowels and diphthongs exist. However, in other areas of the United States, fewer distinct vowel sounds exist because no distinction is made between the vowels of bought VS. pot or caught VS. cot or hawk VS. hock. In addition, there are other ways of transcribing diphthongs (Finegan, 2004, p.96).

əj/, /ɔj/).

In order to differentiate each vowel, articulatory features including tongue height, tongue frontness, tenseness, rounding, lengthening, nasalization and tone play an important role in pronunciation (Finegan, 2004).

Regarding frontness, in English, there are five front vowels (i.e., /i/, /ɪ/, /e/, /ɛ/, /æ/), two central vowels (i.e., /ə/, /ʌ/), and five back vowels (i.e., /u/, /ʊ/, /o/, /ɔ/, /ɑ/), while in Japanese there are only two front vowels (i.e., /i/, /ɛ/), one central vowel (i.e., /a/) and two back vowels (i.e., /ɯ/, /ɔ/).

Tenseness is another articulatory feature and it is characterized as tense (i.e., /i/, /ɛ/, /u/, /o/, /ɑ/) and lax (i.e., /ɪ/, /ɛ/, /ʊ/, /ɔ/, /ə/, /æ/, /ʌ/) monophthong distinctions. Celce-Murcia, Brinton, & Goodwin, (1996) clarify that "muscle tension serves to stretch the articulation of tense vowel sounds to more extreme peripheral positions in the mouth" (p.96). Compared to tense vowels, lax vowels tend to be shorter in duration and do not occur at the end of stressed syllable. The approximate positions of the tongue during their articulation include center other than front and back. Moreover, lax vowels aren't produced with more muscular tension than tense vowels. Some examples of lax vowels are: /ɪ/ (e.g., "bit", "pig"), /ɛ/ (e.g., "pet, "bet"), /ʊ/ (e.g., "put", "foot") /æ/ (e.g., "pat", "bat"), /ʌ/ (e.g., ("put", "but"), /ɔ/ (e.g., "port", "bought") and /ə/ (e.g., "about", "sofa"). Likewise some examples of tense vowels are: /i/ (e.g., "bee", "beat"), /e/ (e.g., "bait", "late"), /u/ (e.g., "pool", "boot"), /o/ (e.g., "poke", "boat") and /ɑ/ (e.g., "pot", "father") (Finegan, 2004). There is no such differentiation in Japanese vowel inventory and thus, such distinction tends to be one of the most problematic areas for Japanese learners of English (Ohata, 2004). Strange, Akahane-Yamada, Kubo, Trent, Nishi, and Jenkins (1998) revealed that AE /æ/, /ʌ/, /ɑ/ are assimilated to Japanese /a/ and AE /ɔ/ is assimilated to Japanese /ɔ/. As a result, Japanese learners of English tend to have difficulty making distinctions of "hut" VS. "hat", and "pat" VS. "pot."

In a similar fashion, Japanese listeners tend to assimilate more than one AE vowel to a Japanese category using primarily spectral cues (e.g., vowel height, vowel frontness, tenseness, rounding). When duration differences are large, Japanese listeners seem to be able to differentiate the differences (e.g., /i/, /ɪ/); however, when spectral (e.g., F3) and temporal (e.g., transition duration) differences are very small (e.g., /u/, /ʊ/), they have difficulty telling the differences (Nishi & Kewly-Port, 2007). Nishi and Kewly-Port (2007) pointed out that "the accuracy of discrimination between contrasting L2 sounds depends on the similarity of their assimilation patterns into L1 categories" (p.1496).

2.2.2. Consonants

Japanese contains mainly fifteen consonants[10] (/p/, /b/, /t/, /d/, /k/, /g/, /m/, /n/, /s/, /z/, /h/, /ɾ/, /j/, /Φ/, /Ç/) (Finegan, 2004; Ohata, 2004), whereas American English contains twenty four or twenty five consonants[11] (/b/, /p/, /m/, /w/, /f/, /v/, /θ/, /ð/, /t/, /d/, /n/, /s/, /z/, /r/, /l/, /ʃ/, /ʒ/, /tʃ/, /dʒ/, /j/, /k/, /g/, /ŋ/, /h/, /ʔ/) (Finegan, 2004). Both languages have unique distribution patterns of consonants. Such critical differences tend to make Japanese difficult to produce some sounds which do not exist or are not distinguished in Japanese, resulting in transferring

---

[10] Phoneme inventory varies depending on classification and some contains other sounds (e.g., /ɣ/, /ɰ/).

[11] Depending on classification, phonetic representations are varied as to exactly which, and how many, symbols are used (Ur, 1991, p.47).

Japanese (L1) phonological knowledge into English (Piske et al., 2001).

A particularly troublesome case is the discrimination of /r/ and /l/, which are not included in the Japanese phonetic system. Many studies have examined the pronunciation attainment of Japanese learners of English in regard to /r/ and /l/ (e.g., Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Lively, Logan & Pisoni, 1993; Aoyama, Flege, Guion, Akahane-Yamada, Yamada, 2004). It is widely notable that Japanese learners tend to substitute Japanese alveolar flap /ɾ/ for both of the phonemes because that is the similar L1 counterpart (Guion, Flege, Ahahane-Yamada, & Pruitt, 2000), resulting in pronouncing /l/ and /r/ identically (e.g., "light" VS. "right" or "late" VS. "rate"). That is, native speakers of English perceive the auditory flap /ɾ/ as English /l/ (Sekiyama & Tohkura, 1993).

In fact, acoustic properties of Japanese counterpart /ɾ/ are substantially similar to English /l/ (Hattori & Iverson, 2009). In Hattori and Iverson (2009) study, acoustic analysis[12] revealed that there were differences among English /r/, /l/ and Japanese flap /ɾ/ in terms of acoustic domains (e.g., F3, F2, F1 and transition duration (td)). Approximate frequency range and transition duration range are shown in *Table* 1.

Table 1. The result of acoustic analysis of English /r/, /l/ and Japanese flap /ɾ/ by Iverson (2009)

	English /r/	English /l/	Japanese flap /ɾ/
F1	250-350 Hz	300-400 Hz	300-350 Hz
F2	1100-1300 Hz	1100-1300 Hz	1500-2000 Hz
F3	1600-2000 Hz	3300-3700 Hz	2600-3100 Hz
td	50-100 ms	10-20 ms	5-20 ms

The result showed that there was no significant difference was found for F1; however, significant differences were found for F2 and F3. Precisely, significant differences were observed between English /r/ and Japanese flap /ɾ/ for F3 (Hattori & Iverson, 2009). English /r/ articulatory features are related to F3 which is "a front cavity resonance where the front cavity includes a lip constriction formed by the tapering gradient of the teeth and lips (with or without rounding) and a large volume cavity behind it that includes the sublingual space" (Espy-Wilson, Boyce, Jackson, Narayanan, & Alwan, 2000, p.354). In a perception study, American listeners were sensitive to distinguish differences between English /r/ and /l/. In contrast, Japanese adults typically had difficulty hearing F3 variation and were likely to weight the onset frequency of F2 heavily (Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert, 2003). Considering Japanese phonetic system, it is somewhat reasonable that F3 is not used as essential information to differentiate any consonantal or vocalic sounds. F2 is used to distinguish the Japanese approximant categories (i.e., /j/, /w/) and it is more familiar to Japanese.

Other crucial articulatory complex segmental elements are inter-dental and labio-dental

---

[12] American English /r/ can be characterized along multiple dimensions, such as (a) third formant (F3) which involves in the differentiation of roundedness, (b) second formant (F2) which is associated with tongue advancement, (c) first formant (F1) which is associated with tongue height, and (d) transitional duration of F1 and F3. (Saito, 2015, p.379).

F1 associated with tongue height	F2 associated with tongue advancement
/i/ & /u/ (high vowels): Low F1 /æ/ & /a/ (low vowels): High F1	/u/ & /a/ (back vowels): Low F2 /i/ & /æ/ (front vowels): High F2

consonants: /ð, θ, v/.　Japanese learners of English tend to substitute /z, s, b/ for /ð, θ, v/ sound. Given that there are no inter-dental or labio-dental fricatives in Japanese, Japanese learners tend to have difficulty pronouncing or discriminating the differences (Kavanagh, 2007).　In this way, "then," "think", "very" tend to be confusing distinctions for Japanese to pronounce properly and their pronunciation might be wrongly perceived as "zen", "sink", and "berry" by native speakers of English.

3. Pedagogical implications for phonetic teaching

3.1. Learner's factors

There are various factors affecting pronunciation learning, such as learners' ages, first language, proficiency level in the target language, exposure to the target language, amount and type of prior pronunciation instruction, attitude toward the target language and motivation to achieve intelligible speech patterns in the target language (Celce-Murcia, et al., 2010; Nation & Newton, 2009).　For designing a course all the factors above need to be taken into account.

3.2. Teachers' role

Teachers should be flexible as well as creative in order to meet the needs of diverse learners (e.g., Carey et al., 2015: Yoshida, 2006), and on top of that, teachers' knowledge base (Baker & Murphy, 2011) of pronunciation teaching is a crucial issue.　Teachers should have a good command of phonetic systems, including the knowledge of how the various organs of speech are involved in the articulation of segmental sounds, as well as the ways in which sound vary in context and the knowledge of how suprasegmental features function to express meaning within discourse (Celce-Murcia et al., 2010).　Besides, teachers must determine how much information should be offered and in what order it should be presented to learners.

For instance, if learners' goal setting is to achieve native-like speech, "an instructional focus on accent minimization or reduction should not be rejected" (Saito, Trofimovich, & Isaacs, 2016, p.15) in that it may be safe for such learners to master phonetic features other than some salient pronunciations (Saito, 2014).

3.3, Suprasegmental VS. Segmentals

Previous studies have widely discussed whether segmentals or suprasegmentals (e.g., Anderson-Hsieh, Johnson, & Koehler, 1992; McNerney & Mendelsohn 1992; Saito, 2011) or integration of both (e.g., Celce-Murcia et al., 2010; Jenkins, 2000) should be the focus of instruction.　An important similarity across languages may be that "each L1 and each individual is likely to vary in the type of segmental and suprasegmental instruction required, depending on the phonological inventory of the learners' L1 and idiosyncratic speech patterns" (Carey et al., p.A.20).

Various empirical studies have confirmed the importance of suprasegmental features in learners' overall intelligibility and perceived comprehensibility (Tanner & Landon, 2009).　McNerney and Mendelsohn (1992) placed great importance on suprasegmentals.　"Suprasegmentals are extremely important in the communication of meaning in spoken language. /.../ because individual sounds can usually be inferred from the context" (p.186).　By the same token, Levis and Grant (2003) documented that segmentals are not important in communication, but suprasegmentals are more clearly connected to functions of spoken English (p.13).　They went on to explain that suprasegmentals which affect not only words but also the entire utterances are more directly

relevant to speaking skills rather than segmentals. Some other researchers expressed a similar view (e.g., Anderson-Hsieh, Johnson, & Koehler, 1992; Munro & Derwing, 1995; Matsuura, Chiba, & Ara, 2012).

Other researchers counter to the claim above and discussed that segmental features are placed more in comparison with suprasegmental features (Carey et al., 2015; Jenkins, 2000). Compared to the number of studies, there are more segmental studies than suprasetmental ones (Thomson and Derwing, 2014). Especially among them, there is considerable debate regarding vowel pronunciations and their effects on listeners' comprehensibility (e.g., Lee, Guion, &Harada, 2006; Munro, Flege, & Mackay, 1996; Saito & Lyster, 2012). Nishi and Kewley-Port (2007) found that vowel pronunciations somewhat affect listeners' understanding and added that only the subset of vowels may not be helpful in learning a complete set of vowels unlike the consonant training. Efficient learning of nonnative vowels requires exposure to a full set of vowel categories, both easy and difficult. Jenkins (2000) on the other hand doubted that mispronunciation in vowel quality, which is concerned with tongue and lip position, constantly affects listener's understanding. Some possible reasons may be related to the fact that vowel continuum tends to be complicated to distinguish. "The vowel continuum between /æ/, /ɛ/, /ɑ/ /ʌ/ and /ɔ/ is difficult to notice not only for students but also for teachers, especially compared to the /r/ -/l/ contrast" (Saito & Lyster, 2012, p.396). In addition, vowel inventories dramatically differ between regional dialects of English (Fox & Jacewicz, 2009).

It may be worth noting that not all segmental features have equally been tested in experimental studies. I postulate that it might be because some phonological features such as prevocalic /r/ and postvocalic /r/ are challenging to measure, in that "the realization of these individual sound features can be significantly influenced by the preceding and following phonetic contexts" (Saito, 2012, p.852).

Today, phonetic teaching has been moving from either segmental or suprasegmental debate to a more balanced view of segmental and suprasegmental features (Celce-Murcia et al., 2010) to meet learners' needs and their proficiency level (Yoshida, 2006). Saito et al. (2016) also commented that "the relative weight of instructional focus on segmentals versus suprasegmentals particularly with the view of improved comprehensibility, may vary as a function of learner ability level" (p.14). For maintaining comprehensibility, attention should be paid to suprasegmental features throughout all level of L2 oral development, minimum level of segmental accuracy and fluency are required at a beginner level, segmental precision is expected at an advanced level of language proficiency. Even speakers who reach a certain level of phonological ability produce comprehensible pronunciations, their segmental inaccuracies have been pointed out (Saito et al., 2016).

3.4. Setting priorities

Teachers who teach phonetic features in practical situations have been struggled whether all phonological features should be taught or some particular features should be mainly picked up and focused more than others.

In response to practitioners' needs, Saito (2014) conducted a study in which experienced teachers both native speakers of English (NEs) and native speakers of Japanese (NJs) whose background were similar judged tokens of English pronunciation produced by Japanese learners of English. The study indicated that experienced teachers shared consensus on relatively important pronunciation features for Japanese learning English and revealed that what phonetic features should be taught and what order they should be taught in. Priority setting by Saito (2014) is shown

in *Table* 2.　As shown in *Table* 2, the result indicated that priority of consonants is higher than that of vowels.

One may wonder that raters' variability such as teaching experiences, their familiarity to learners' L2 accents as well as their knowledge in phonetics and pronunciation teaching may possibly affect reliability of the priority-setting process.　To reduce the variability, Saito (2014) paid close attention to assure both quality and validity of the teacher judgment with special effort (i.e., raters in the study were all experienced teachers with similar teaching background in EFL classrooms in Japan).　In this regard, the findings contribute to a solution of pronunciation teachers' questions and may serve as a useful framework for planning pronunciation lessons.

3.5. History of pronunciation teaching

Pronunciation teaching has placed a special emphasis on getting the sound right mainly in isolated, controlled and contrived sentences with "listen and repeat" practice.　At the same time, there have been some doubts whether such exercises help learners improve their pronunciation for a long time (Nation & Newton, 2009) and in spontaneous conversation (see 3.8.1. FFI).

Historically, there are three general approaches to the pronunciation teaching: intuitive-imitative, analytic-linguistic and integrative approach (Celce-Murcia et. al., 2010; Chela-Flores, 2001).　A brief history of pronunciation teaching with reference to the three resources (Celce-Murcia et al., 2010; Ur, 1991; Mitchel & Myles, 2004) is summarized in Table 3.

According to Celce-Murcia et al. (2010), intuitive-imitative approach gives opportunities to students to listen and imitate the phonetic features of the target language without explicit explanation.　This type of approach became popular in 1960s and 1970s and contributed to support the form of audio-lingual method.　Analytic-linguistic approach was developed to complement the intuitive-imitative approach.　It provides learners with an explicit intervention of pronunciation and let them pay attention to the sounds and rhythms of the target language.　Various information and tools such as phonemic chart, articulatory descriptions, and explanations of the form and function of prosody are utilized in the approach.　Practical exercises such as minimal pair drills, and rhythmic chants are often adopted (p.2).　Current approach takes the idea of the integrative approach (Chela-Flores, 2001) which emphasize on "integral part of oral communication" (Morely, 1991, p.496).　It is extensively accepted that pronunciation practice should be integrated with the rest of language learning activities such as oral communicative practice, instead of dealing pronunciation as an isolated practice.　That is, practice of phonetic features within meaningful contextual-

Table 2. Phonetic features that Japanese learners need to learn and three levels of priority (Saito, 2014)

Priority order	What to learn
1	• segmentals /l, r, ð, θ, v/ • Complex syllables (e.g., CCVV VS. CCVCCC) in comparison with Katakana words.
2	• assimilation rules • segmentals /æ, ʌ, f/ • vowel quality (producing long and loud vowels) to mark lexical stress • wide range of pitch to mark sentence stress • dynamic intonation patterns
3	• diphthongs /aʊ, aʊ, oʊ, ɔɪ, eɪ/ • segmentals /p, t, k, w, n, ŋ, h/ • speed of rate —speech rate • reduction of pauses and repetitions

Table 3. A brief history of pronunciation teaching based on various methodologies (Celce-Murcia et al., 2010; Ur, 1991; Mitchel & Myles, 2004)

Approximate established years	Method	Classification of three types of approach and other types of approach[13]	Focus of instruction/ Teaching aids / Examples of tasks and activities
mid-late1800s	The Direct Method	intuitive-imitative others	• Listening, mimicking, and repeating a native-like speech model (teachers or recordings)
mid-1800s	Grammar-translation	others	• Communication is not a primary objective, rather grammar and text comprehension are central. • Teaching pronunciation is largely irrelevant
1940s, 1950s	Audiolingualism in the United States and of the Oral Approach in Britain	intuitive-imitative analytic-linguistic	• Pronunciation is important • Imitating and repeating after a teacher or a recording model • Explicit teaching from the beginning • Use of visual transcription system or charts that demonstrate the articulation of sounds • Minimal-pair drills
1960s	Cognitive Approach or Cognitive-Code Approaches	analytic-linguistic others	• Learning learnable items such as grammatical structures and words, rather than aiming at native-like speech patterns • Drills such as sound discrimination, pronunciation of specific elements before learners participate in real communication activities • Phonemes need to be learned before words, words before phrases and sentences, simple sentences before more complicated ones, and so forth
1960s, 1970s	Total Physical Response	intuitive-imitative integrative others	• Responding to language input with body motion • Learning pronunciation without explicit pronunciation instruction
1970s	Silent Way	intuitive-imitative others	• Focusing on the accurate sounds and structure of the target language without learning phonetic alphabet or explicit linguistic information
	Community Language Learning	intuitive-imitative analytic-linguistic integrative	• Learning pronunciation items that learners want to practice and using teachers as a resource

[13] Each approach is classified on the basis of three types of approach proposed by Celce-Murcia (1996) and a type of approach which is different from those three is labeled as "others".

1970s	Natural Approach[14]	intuitive-imitative integrative others	• Initially focusing on listening without pressure to provide learners with opportunities to internalize the target sound system • Manipulative visuals • Act out • Use of high-frequency vocabulary, • Short sentences • Yes/no questions • Either/ or questions, and other questions that require only one-word answers • Getting learners physically involved with the target language. • Learning pronunciation without explicit pronunciation instruction
mid-late 1970s	Communicative Approach	intuitive-imitative analytic-linguistic integrative	• Using language to communicate is central • Aiming at intelligible pronunciation rather than native-like • Listening and imitating • Phonetic training • Minimal-pair drills • Contextualized minimal pairs which is a developed version of minimal-pair drills • Visual aids, • Tongue twisters, • Developmental approximation drills • Practice of vowel shifts and stress shifts related by affixation for intermediate or advanced learners • Reading aloud/recitation, • Recordings of learners' production

ized contexts has been suggested.

3.5.1. Minimal pair drills

As shown in Table 3, traditionally, teachers have made use of focused production tasks such as minimal-pair drills with the combination of modeling and imitating teacher's model along with articulatory descriptions.

Minimal-pair drills include word and sentence drills. Word drills are for instance, "sing" VS. "thing," or "veil" VS. "bale," or "scythe" VS. "size" which have been represented to differentiate / s, b, z/ from /θ, v, ð/. Similarly for teaching /r/ and /l/ differences, presentation of the minimal

---

[14] The natural approach and its extensions are used with many other compatible methods and activities (total physical response and the audio-motor unit, chants, music, games, role play, storytelling, affective activities, etc.) (Ur, 1991, p.171).

set "ri", "lee", and the Japanese word "り" (ri) is offered to differentiate between /r/ and /l/ as well as the flapped Japanese /ɾ/. Sentence drills are composed of two types: syntagmatic drills (e.g., "Don't sit in that seat") and pragmatic drills (e.g., "Don't slip on the floor" VS. "Don't sleep on the floor") (Celce-Murcia et al., 2010).

In a laboratory setting, a result showed that L2 learners especially beginners benefited from intensive exposition of minimally paired word input such as "rock" or "lock" produced by a number of native English speaking people (Saito, 2015, p.380). Since some beginners who have insufficient abilities to discriminate English phonemes tend to take up to much time trying to decode the sound, focusing on specific phonemic recognition skills through minimal-pair drills may be helpful.

Yet up until now, many researchers have seen little sustained benefit in minimal- pair drills (e. g., Carey et al., 2015; Makino, 2013). Two possible reasons may be: how long positive effects of minimal-pair practice would last after intervention has not been reported and the extent to which outcome of minimal-pair drill practice can be applied in a real life setting remains uncertain.

According to a conceptualization, Transfer Appropriate Processing (TAP) theory (Lightbown, 2008), both a type of practice and a kind of learning that takes place are closely related. Text manipulation activities will be successful in controlled settings as mechanical drills involve conscious processing, still they are unlikely useful in spontaneous face-to-face communication (e.g., Ellis, 2016; Trofimovich & Gatbonton, 2006). In order to ensure effectiveness of minimal-pair drills, further longitudinal experimental studies with a variety of datasets of various populations in natural settings are suggested.

3.5.2. Articulatory descriptions / diagrams

Different researchers uphold the idea of providing explicit phonetic information such as provision of explanation on the relevant articulatory configurations or articulatory settings (e.g., illustration which involves the relationship of the English vowels to one another, the approximate positions of the tongue or lips during their articulation) (e.g., Celce-Murcia et al., 2010; Noguchi, 2014). (see Figs. 1, 2 and Table 4.) As it is necessary especially for beginners to receive perceptional aspects of new sounds receptively to establish new phonetic representations (Saito, 2015), effectiveness of using diagrams is the accepted view. Learners are encouraged to do some experiments that changing position of tongue, or lips or teeth can change sounds. In pronunciation teaching, it is suggested that teachers should offer the manner of producing sounds rather than merely asking learners to copy the teacher's model.

3.6. Current trend of pronunciation instruction

Currently, pronunciation instructions have been taking some new ideas from other fields such as drama, psychology, and speech pathology (Celce-Murcia et al., 2010) and have been trying to achieve a balance between segmental and suprasegmental features, between repetition and contextualized communicative practice.

An example of suggested types of instruction is form-focused instruction (FFI) which helps "learners in communicative or content-based instruction to learn features of the target language that they may not acquire without guidance" (Spada & Lightbown, 2008, p.181). Besides, a little more focus has been placed the use of technology for teaching pronunciation, in addition to learner-centered approach (e.g., Morley, 1991; Olson, 2014; Thornbury, 1993).

Figure 1. The NAE vowel quadrant and sagittal section of the mouth (Celce-Murcia et al., 2010)

Figure 2. Vowel quadrilateral showing the complete monophthongs found in English (Low, 2015)

Table 4. Japanese vowel chart (Finegan, 2004)

	FRONT UNROUNDED	CENTRAL UNROUNDED	BACK UNROUNDED	BACK ROUNDED
HIGH	i		ɯ	
MID	ɛ			ɔ
LOW		a		

Table 5. English and Japanese consonants chart (Finegan, 2004; Ohata, 2004)

MANNER OF ARTICULATION AND VOICING	BILABIAL	LABIO-DENTAL	INTER-DENTAL	ALVEOLAR	ALVEO-PALATAL	VELAR	PALATAL	UVULAR	GLOTTAL
STOPS voiceless voiced	**p**[15] **b**			**t** **d**		**k** **g**			ʔ
NASALS	**m**			**n**		ŋ			
FRICATIVES voiceless voiced	Φ[16]	f v	θ ð	**s** **z**	ʃ ʒ		ç		**H**
AFFRICATES voiceless voiced					tʃ dʒ				
APPROXIMANTS voiced central voiced lateral	w			r l			j		
OTHERS voiced trill voiced flap				ɾ					

---

[15] Consonants indicated in bold type are common in English and Japanese.

[16] The symbols enclosed in squares are typical in Japanese.

3.6.1. FFI

Based on a broad perspective of second language acquisition (SLA) research, integrating language focuses into meaning-oriented classrooms is hypothesized beneficial for students to establish form-meaning mapping and accuracy of L2 (e.g., Doughty & Williams, 1998; Lyster, 2007). As stated above, based on TAP theory, focusing only on pronunciation practice with repetition or minimal-pair drills in class may not be applied in natural settings (Lightbown, 1998). A beneficial type of instruction is to "draw students' attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning or communication" (Long, 1991, p.45–46).

In a similar way, concerning suprasegmental features, various researchers (e.g., Jenkins, 2004; Levis & Pickering, 2004) emphasized the importance of teaching intonation in context rather than within isolated sentences.

However, it is risky to conclude that isolated tasks are insignificant. A recent study elaborated on the use of FFI for teaching L2 segmental sounds in accordance with learners' development (Saito, 2013). According to FFI model developed by Lyster and Ranta, (Lyster 2007) there are three stages of learners' development: noticing, awareness and practice. At the noticing stage, learners need to be pushed to "attend to sound-sized units of L2 phonological information" and "notice the perceptual difference between a new sound and its L1 counterpart" (Saito, 2013, p.25). Accordingly, isolated intervention appears to play a role. At the awareness stage, as learners need to internalize the phonetic representation, communicative tasks within meaningful contexts are necessary. At the practice stage, learners need to work on to move the target phonetic representations from declarative to procedural knowledge[17] through extensive repetitive practice in communicative authentic contexts.

3.6.2. Use of computer technology

There have been various attempts to develop computer technology systems in pronunciation and the use of technology has been investigated for its potential in L2 pronunciation teaching that are not available in traditional classroom learning. Although it is still in its infancy, it has somewhat contributed to enhance learners' "discovery" (Olson, 2014) unless it only provides input and only trains receptive skills (p.5)

Computer Assisted Pronunciation Training (CAPT) offers immediate feedback through the integration of Automatic Speech Recognition which is able to indicate when sounds are not being produced in a way that the machine recognizes as intelligible (Thomson, 2011).

Initially, suprasegmental speech features, in particular intonation contours, were studied within the visual feedback paradigm (Olson, 2014). *BetterAccentTutor* was developed to teach three components of American English prosody (i.e., intonation, stress and rhythm) to non-native speakers of English (Kommissarchik & Kommissarchik, 2000). It provides learners with immediate audio-visual feedback to compare their own production and production pronounced by native-English speakers.

Today, there are various types of software (e.g., SPECO, PRAAT), which focus on the identifi-

---

[17] According to skill acquisition theory (Anderson, 1990), there are two types of knowledge: declarative and procedural which is automatized declarative knowledge. Further, Dekeyser (2007) drew an inference from the theory and claimed as follows: once learners have successfully converted declarative knowledge into procedural knowledge, access to the procedural knowledge becomes "spontaneous, effortless, fast and errorless" (p.3) and practice can serve as a bridge between the two types of knowledge.

cation and production of segmental features (Jenkins, 2004). SPECO Project which includes segmental practice is a speech technology whose primary purpose is for clinical remediation of children's speech problems (Sfakianaki, Roach, Vicsi, Csatari, Öster, Kacic, & Barczikay, 2001). PRAAT Program, which was developed for vowel and diphthong teaching, offers visual image of learner's own productions to compare to those of native-English speakers. Saito (2007) used PRAAT speech analysis software and revealed that Japanese learners of English showed improvement in accuracy of a segmental feature, /æ/. Carey et al. (2015) suggested L1 point of reference approach (L1POR) which involves "production practice including a combination of multiple sensory input: articulatory explanation, visual cues such as pictures and videos modeling articulation, real-time acoustic visual feedback for vowels and learners reflecting on stored sound memory of L2 prototypes" (Carey et al., 2015, p.A-22).

Recent dictionaries issued with CD-ROMs have also been improving and some "recent dictionaries offer learners a range of features such as the opportunity to hear words in isolation and, in some cases, in connected speech, and the possibility of recording and listening to themselves in order to compare their own pronunciation with the dictionary version" (Jenkins, 2004, p.118).

Some advantages that can be gained from such computer technology may be that it would address learners' individual problems or offer specific skills that learners want to develop. It also provides learners with immediate individualized feedback regardless of teachers' knowledge base and help learners notice their errors. Moreover, it makes it possible to offer private and stress-free practice (Stracke, 2012), which in turn might result in reducing learners' foreign language anxiety (Neri, Cucchiarini, Strik, & Boves, 2002).

However, some researchers have raised concerns over the use of speech analysis programs in that "computer assisted pronunciation training (CAPT) applications available today merely replicate classroom instruction but on a computer monitor (e.g., minimal pair practice, descriptions and diagrams of articulatory gestures, etc.)" (Thomson, 2011, p.747). To what degree CAPT provides learners with meaningful practice that can be applied to real conversation is uncertain. In a similar vein, whether learners have opportunities to engage in social interaction with others through CAPT remains an unsettled question. What is crucial to keep in mind is that interaction plays a significant role in L2 learning process based on a notion of SLA. In addition, error diagnosis by CAPT is only possible with a limited degree of detail (Neri et al., 2002). "Even if pedagogically desirable, detailed diagnosis is simply not feasible because the performance levels attained are too poor" (p.9). Jenkins (2004) also expressed a concern that some software is still not able to perform as well as human listeners listening to non-native speech.

4. Future directions

For further investigation, to what degree learners need to learn for acquiring compressible pronunciation of both segmental and suprasegmental features is a debatable issue. Indeed, it is difficult for teachers to draw a line, "what amounts to minimal gains in productive control of this feature" (Carey et al., 2015, p.A-19). Similarly, despite the importance of incorporating pronunciation practice into communicative contexts, how to accomplish the integration is left unclear: "teachers have received little clear direction about how to accomplish this integration" (Levis & Grant, 2003, p.13).

In terms of pronunciation research, whether the degree of effectiveness varies between kinds of phonetic features and populations remains an open question. Appropriateness of types of tasks

exploited in studies needs further investigation (Thomson and Derwing, 2014). For instance, reading a passage aloud has been often used to assess pronunciation abilities, it is questionable whether such task is appropriate to measure learners' phonetic abilities in spontaneous speech.

In order to obtain further corroboration for effective pronunciation instruction, continuous collaborative efforts between teachers and researchers are indispensible.

5. Conclusion

There are significant phonetic differences between English and Japanese. In pronunciation teaching, teachers need to determine what phonetic features should teach in what order, in accordance with learners' needs, goals and proficiency level. With respect to focused phonetic features, integration of both segmentals and suprasegmentals is considered to be ideal, instead of either segmentals or suprasegmentals. In terms of priorities in pronunciation teaching, consonants placed higher than vowels (Saito, 2014).

Through the years, it is well known that pronunciation teaching has relied heavily on decontextualized controlled minimal-pair drill practice with a combination of listen and repeat technique. These days, the teaching paradigm has been shifting toward more interactive with FFI to attain comprehensible phonetic features that can be applied in spontaneous natural settings. However, appropriate instruction should be determined based on learners' proficiency level or learners' development stage. Given that beginners who need to establish new phonetic representations, phonemic distinction exercises with a series of articulatory descriptions or diagrams seem beneficial at the beginning of their learning (Saito, 2015). As subsequent practice, it is suggested to focus on phonetic features with more integrated instruction in meaningful contextualized contexts.

Learners are encouraged to pay attention to the discrepancies between their own production and comprehensible pronunciation through various exposures, such as computer technology and interaction with others. In virtue of significant advances in technology, opportunities to incorporate computer-based instruction have become available (i.e., CAPT). While such CAPT has come in a variety of forms, some concerns such as opportunity of interaction with others have been raised by various researchers. Based on the notion of SLA, interaction with various people should not be downplayed. Carey et al. (2015) stated that "the ability to perceive second language speech is shaped over time on the basis of exposure and social interaction" (p.A21). Interactional theory also supports that conversational interaction including negotiation of meaning helps learners to make connections between form and meaning and hence such interaction facilitates second language (L2) acquisition (Long, 1996).

From now on, effective pronunciation instructions need to be corroborated by further longitudinal studies and indications from SLA research and teaching in class should be taken into account. With respect to students, it is considered to be significant to increase their awareness of the importance of pronunciation through various activities (Nation & Newton, 2009).

References

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure, *Language Learning 4 (2)*, 529–555.

Anderson, J. R. (1990). *Cognitive psychology and its implications*. New York: W. H. Freeman.

Aoyama, K., Flege, J., Guion, S., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimi-

larity and L2 speech learning: The case of Japanese /r/ and English /l /and /r/. *Journal of Phonetics, 23*, 233-250.

Baker, A. A. & Murphy, J. (2011). Knowledge base of pronunciation teaching: Staking out the territory. *TESL Canada Journal, 28 (2)*, 29-50.

Bongaerts, T., Van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition, 19* (4), 447-465.

Bradlow, A., Akahna-Yamada, R, Pisoni, D, and Tokhura, Y. (1999). Training Japanese Listeners to identify English /r/ and /l/: Long term retention of learning perception and production. *Perception and Psychophysics 61 (5)*, 977-985.

Cai, Y. W., & Lee, H. Y. H. (2015). A Contrastive Analysis of Mandarin Chinese and Thai: Suggestions for Second Language Pronunciation. *Theory and Practice in Language Studies, 5* (4), 719-728.

Carey, M. D., Sweeting, A., & Mannell, R. (2015). An L1 point of reference approach to pronunciation modification: Learner-centred alternatives to 'listen and repeat'. *Journal of Academic Language and Learning, 9* (1), A18-A30.

Celce-Murcia, M., Brinton, D. & Goodwin, J. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. New York: Cambridge University press.

Celce-Murcia, M., Brinton, D., Goodwin, J. M. (2010). *Teaching Pronunciation Hardback with Audio CDs (2): A Course Book and Reference Guide*. New York: Cambridge University Press.

Chela-Flores, B. (2001). Pronunciation and language learning: An integrative approach. *IRAL-International Review of Applied Linguistics in Language Teaching, 39* (2), 85-101.

Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam: John Benjamins.

Cutler, A., Oahan, D., & Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review, *Language and Speech, 40 (2)*, 141-201.

DeKeyser, R. (1997). Beyond explicit rule learning. *Studies in Second Language Acquisition, 19*, 195-221.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39* (3), 379-397.

Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom and second language acquisition* (pp.197-261). Cambridge: Cambridge University Press.

Ellis, R. (2016). Grammar teaching as consciousness raising. In Hinkel, E. (Ed.), *Teaching English grammar to speaker of other languages*. (pp.128-150). New York: Routledge.

English skills of Japanese students fail to meet government targets. (2016, February 2). *The Japan Times*. Retrieved August 5, 2016 from http://www.japantimes.co.jp/news/2016/02/02/national/english-skills-japanese-students-fail-meet-government-targets/#.V6NOdaMcRjo

Espy-Wilson, C., Boyce, S., Jackson, M., Narayanan, S., & Alwan, A. (2000). Acoustic modeling of American English /ɹ/. *Journal of the Acoustical Society of America, 108*, 343-356.

Finegan, E. (2004). *Language: Its structure and use* (5th ed.). Boston: Thomson Wadsworth.

Flege, J. E., Bohn, O-S., & Jang, S. (1997). The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25 (2)*, 169-186.

Fox, R. A., and Jacewicz, E. (2009). "Cross-dialectal variation in formant dynamics of American English vowels," *The Journal of the Acoustical Society of America, 126 (5)*, 2603-2618.

Graham, C. (2013). Revisiting f0 Range Production in Japanese-English Simultaneous Bilinguals. Annual Report of UC Berkeley Phonology Lab, 110-125.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America, 107* (5), 2711-2724.

Guiora, A. Z., Beit-Hallami, B., Brannon, R. C., Dull, C. Y. & Scovel, T. (1972). The effects of experimentally induced changes in ego states on pronunciation ability in second language: An exploratory study. *Comprehensive Psychiatry, 13 (5)*, 421-428.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for

second-language acquisition. *Psychological Science, 14*, 31-38.

Hattori, K., & Iverson, P. (2009). English /ɹ/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America, 125*, 469-479.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /ɹ/-/l/ to Japanese adults. *Journal of the Acoustical Society of America, 118*, 3267-3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87* (1), B47-B57.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics, 24*, 109-125.

Kavanagh, B. (2007). The phonemes of Japanese and English: A contrastive analysis study. *Aomori University of Health and Welfare, 8 (2)*, 283-292.

Kirkpatrick, A. (2007). *World Englishes: Implications for international communication and English language teaching*. Cambridge: Cambridge University Press.

Kommissarchik, J., & Komissarchik, E. (2000). Better Accent Tutor-Analysis and visualization of speech prosody. *Proceedings of InSTILL 2000*, 86-89.

Kondo, M. (2009). Is acquisition of L2 phonemes difficult? Production of English stress by Japanese speakers. In *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009)* (pp.105-112). Somerville, MA: Cascadilla Proc. Project.

Lee, B., Guion, S., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals, *Studies in Second Language acquisition, 28*, 487-513.

Leaver, B. L., Ehrman, M., & Shekhtman, B. (2005). *Achieving success in second language acquisition*. Cambridge University Press.

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*, 367-377.

Levis, J. M., & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal, 12 (2)*, 13-19.

Levis, J. M., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System, 32*, 505-524.

Lightbown, P. M. (1998). The importance of timing in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp.177-196). Cambridge, UK: Cambridge University Press.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /I/: Il. The role of phonetic environment and talker variability in learning new perceptual categories, *Journal of Acoustic Society of America, 94*, 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994).

Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the acoustical society of America, 96* (4), 2076-2087.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp.39-52). Amsterdam: John Benjamins.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition*. (pp.413-468). San Diego, CA: Academic Press.

Low, E. L. (2015). *Pronunciation for English as an international language: From research to practice*. London: Routledge.

Lowie, W. (2013). L2 phonology. In Peter Robinson, *The Routledge encyclopedia of second language acquisition (pp.347-353)*. London: Routledge.

Lyster, R. (2007). *Learning and teaching languages through content: a counterbalanced approach*. Amsterdam/Philadelphia: John Benjamins.

Mackey, A., Gass, S. M., McDonough, K. (2000). How do learners perceive interactional feedback?, *Studies in Second Language Acquisition. 22 (4)*, 471-497.

Makino, M (2013). Report on the effective teaching and learning of English pronunciation. *Kansai University forum for foreign language education, 12*, 121-134.

Matsuda, A. (2002). International understanding through teaching Word Englishes. *World Englishes, 21 (3)*, 436-440.

Matsuura, H., Chiba, R., & Ara, S. (2012). International Intelligibility of Learner English. 商学論集 , *81 (2)*, 63-74.

McNerney, M., & Mendelsohn, D. (1992). Suprasegmentals in the pronunciation class: Setting priorities. In Peter Avery & Susan Ehrlich (Eds.), *Teaching American English Pronunciation*. (pp.185-196). Oxford: Oxford University Press.

Meng, H., Tseng, C. Y., Kondo, M., Harrison, A. M., & Visceglia, T. (2009). Studying L2 suprasegmental features in Asian Englishes: a position paper. *Proceedings of Interspeech 2009*, (Sep. 6-10), 1715-1718, Brighton, U.K.

MEXT (n.d.) Improvement of academic abilities (Courses of Study), lower secondary school. Retrieved August 10, 2016, from: http://www.mext.go.jp/component/english/__icsFiles/afieldfile/2011/03/17/1303755_013.pdf

Mitchell, R., and Myles, R. (1998). Second language learning theories. (2nd ed.). London: Hodder Arnold.

Morley, J. (1991). The Pronunciation Component in Teaching English to Speakers of Other Languages, *TESOL Quarterly, 25* (3), 481-520.

Munro, M., & Derwing, T. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 45*, 73-97.

Munro, M. & Derwing, T. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49 (1)*, 285-310.

Munro, M. J., Flege, J. E., & MacKay, I. R. A. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics, 17*, 313-334.

Nation, I. S. P. & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. New York: Routledge.

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer assisted language learning, 15* (5), 441-467.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research, 50*, 1496-1509.

Noguchi, J. (2014). Contrastive analysis between Japanese and American English sound systems: From an articulatory setting perspective. *The Journal of Kanda University of International Studies, 26*, 293-309.

Ohata, K. (2004). Phonological differences between Japanese and English: Several potentially problematic areas of pronunciation for Japanese ESL/EFL learners. *Asian EFL Journal, 6 (4)*, 1-19.

Olson, D. J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. *Language Learning & Technology, 18* (3), 173-192.

Piske, T., Mackay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*, 191-215.

Reischauer, E. O. (1961, November 17). *English language and Japan's role in the world*. Speech presented at Annual Conference of the National English Teachers Organization, Tokyo.

Saito, K. (2007). The influence of explicit phonetic instruction on pronunciation in EFL settings: The case of English vowels and Japanese learners of English. *Linguistics Journal, 3* (3), 16-40.

Saito, K. (2011). Identifying problematic segmental features to acquire comprehensible pronunciation in EFL settings: The case of Japanese learners of English. *RELC Journal, 42*, 363-378.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly, 46*, 842-854.

Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development. *Studies in Second Language Acquisition, 35* (1), 1-29.

Saito, K. (2014). Experienced teachers' perspectives on priorities for improved intelligible pronunciation: The case of Japanese learners of English. *International Journal of Applied Linguistics, 24 (2)*, 250-277.

Saito, K. (2015). Communicative focus on L2 phonetic form: Teaching Japanese learners to perceive and

produce English /r/ without explicit instruction. *Applied Psycholinguistics, 36*, 377–409.

Saito, K. & Lyster, R. (2012). Investigating the pedagogical potential of recasts for L2 vowel acquisition. *TESOL Quarterly 46 (2)*, 385–396.

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics, 37*, 217–240.

Seikel, J. A., King, D. W., & Drumright, D. G. (2016). *Anatomy & physiology for speech, language, and hearing* (5th ed.). New York: Delmar Cengage Learning.

Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics, 21*, 427–444.

Sfakianaki, A., Roach, P., Vicsi, K., Csatari, F., Öster, A. M., & Kacic, K. Z. (2001). SPECO: Computer-Based Phonetic Training for Children. In *Proceedings of the Phonetics Teaching & Learning Conference, London* (pp.43–46).

Smith, B. (2012). Pronunciation patterns of Japanese learners and their implications for teaching. *Polyglossia, 23*, 199–206.

Spada, N. & P. Lightbown (2008). Form-focused instruction: isolated or integrated? *TESOL Quarterly, 42 (2)*, 181–207.

Stracke, C. M. (2012). Learning Innovations and Learning Quality: Relations, Interdependences, and Future. *The Future of Learning Innovations and Learning Quality. How do they fit together*, 13–25.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., & Jenkins, J. J. (1998) Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics. 26*, 311–344

Tanner, M. W., & Landon, M. M. (2009). The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning and Technology, 13*, 51–65.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal, 28 (3)*, 744–765.

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics 36*, 1–20.

Thornbury, S. (1993). Having a good jaw: Voice-setting phonology. *ELT Journal, 47 (2)*, 126–131.

Trofimovich, P, & Gatbonton, E. (2006). Repetition and Focus on Form in processing L2 Spanish words: Implications for pronunciation instruction. *Modern Language Journal, 90 (4)*, 519–535.

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016) Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition, 19*, 122–140.

Tsujimura, N. (2014). *An introduction to Japanese linguistics* (3rd ed.). Malden, MA: Wiley Blackwell.

Ur, P. (1991). *A course in language teaching*. Cambridge: Cambridge University Press.

Yoshida, H. (2006). Analyzing pronunciation learning of EFL students using a holistic and analytic approach. *Language Education & Technology*, (43), 127–138.

APPENDIX A:

A picture of Edwin O. Reischauer, a Harvard University professor emeritus who was the former American ambassador to Japan under the Kennedy administration
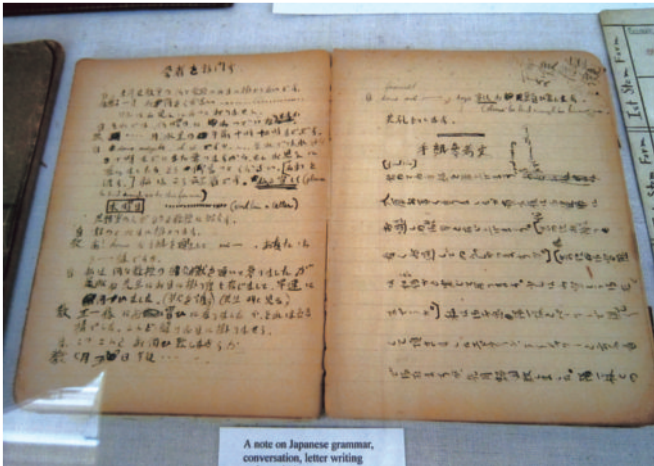
ライシャワーのハーバード大学の日本語文法に関するノート
（歴史学者であると同時に言語学者であった）。

エドウィン O. ライシャワー
（ハーバード大学名誉教授・元米国駐日大使）