Teaching English plosives: A syllable-based approach

Eiichi Yubune

Abstract
This essay discusses English plosives in Japanese-English interlanguage production in different phonological contexts. We focus our attention on aspiration in word-initial plosives, vowel length preceding word-final plosives, and devoicing of sonorant in word-initial clusters involving plosives. I argue that features of plosives in the IL can be attributed to the transfer of the syllable structure of Japanese and its phonological constraints. The preference for open syllables, no occurrence of initial consonant clusters, and the relatively identical duration of each syllable in Japanese are considered to be the most influential factors of the transfer. To discuss these claims, I take into account data from experimental studies on phonetics including VOT, interlanguage syllable structure and Japanese syllables. From the present study, I suggest that an allophonic level of explicit instruction based on the nature of the syllable structure of English, which is drastically different from that of Japanese, is needed for the learners to acquire more intelligible and native-like speech.

1. Teaching plosives to Japanese learners of English: How is it done?

It is common practice to represent sounds in different languages which are phonetically similar, but not absolutely identical, with the same symbol or set of feature specifications. Sometimes this practice is simply for a convenience or ease in transcription. At other times, the sounds in question are said to be different only at the phonetic (allophonic) level, and at the phonemic level, they are thought of as the same sound (Keating, 1984: 286). In this regard, to my knowledge, the allophonic features of English plosives are paid little attention to in prevalent classroom pronunciation teaching and even in specific courses for phonetics or pronunciation practice usually done at the university level in Japan. Rather, greater attention has been paid to typical foreign consonants to Japanese such as /f, v, l, r/ and so on. In other words, since all major plosives such as /b, d, g, p, t, k/ are also found in the Japanese kana syllabary, it is largely considered wrongly, particularly by the learners, that no special pronunciation training is needed to acquire the correct pronunciation of these consonants. In addition, although the lexical stress of English words are taught with much emphasis, the primary concern is on the allocation of the primary or secondary stress of a word (because this knowledge is considered to be part of grammar), and the allophonic nature of the adjacent plosives of the lexical stress is then tend to be disregarded.

As a matter of fact, English plosives involve a number of features in their phonetic realization in different phonological contexts (e.g., aspiration, devoicing, intensity, different types of clusters, and so forth) which are, in general, not observable in the realization of the Japanese counterparts.

I argue in this essay that these differences arise largely from the difference of the syllable structure and its phonological constraints between the two languages. Then I also provide some pedagogical implications on how Japanese learners of English should acquire English plosives in different phonological contexts on the basis of the interlanguage syllable structure.
2. The realization of voice opposition of English plosives in different phonological contexts

The English plosives are classified by voice opposition into two phonemic groups: /b, d, g/ and /p, t, k/. The complete articulation of a plosive, or stop, consonant consists of following three stages: (1) the closing stage: the articulate organs move together in order to form the obstruction. (2) the compression stage: lung action compresses the air behind the closure. (3) the release stage: the articulate organs forming the obstruction part rapidly, allowing the compressed air to escape abruptly (cf. Gimson and Cruttenden, 1994: 139).

2.1 Voicing

Usually, the voiced series /b, d, g/ may have the vibration of vocal cords during the second stage shown above (i.e. during the closure), when they occur particularly in positions between voiced sounds, while voiceless series /p, t, k/ do not have such voice (ibid: 141). In the spectrograph, we can find a black shadow indicating the glottal tone for the voiced plosives whereas there is no such trace for the voiceless plosives.

In initial and final positions, however, that is following or preceding silence, they may be only partially voiced or completely voiceless (ibid: 141). They are also likely to be voiceless when they are followed by a voiceless sound as in obstacle, bedstore and bagpipe (ibid: 149-154). Furthermore, even in the intervocalic positions /b, d, g/ may sometimes be subject to devoicing (ibid: 141). It has also been claimed that even when vocal cord vibration is not present, glottographic and laryngoscopic studies (Catford, 1977: 112) show that whisper-like narrowing is present.

It is also pointed out in Brown (1990: 29, 37) that both voiced and voiceless plosives (stops) tend to be articulated very lightly even when intervocalic before an unstressed syllable and are frequently not realized by a stop. Rather, this would have to be described as a fricative articulation in general phonetic terms.

2.2 Intensity of articulation

Particularly in the initial and the final position of a word, the voice opposition of consonants may be lost, so that the energy of articulation becomes a significant factor. Thus, it may be important to define [p], for instance, as strong or fortis and [b] as weak or lenis (Gimson, 1980: 35-36). Acoustically, fortis plosives have greater intensity at their release stage than lenis plosives (O'Connor, 1973). Thus, at the phonetic level, it is more reasonable to discriminate them in such a way as they sound weak or strong.

These criteria postulated by energy opposition of plosives are manifested in speech realization with some notable characteristics. Brown (1990: 27-34) characterizes how voiced/voiceless plosives behave and what kinds of effect they give to the adjacent phonemes in different phonological contexts. She claims that it is 'aspiration' which chiefly distinguishes initial voiceless stops from their voiced counterparts; the difference between them lies in the timing of the onset of voicing (VOT) immediately following the release of the closure (ibid: 28-29). And the voicelessness normally has the quality of the following vowel (Brown, 1974: 29) On the other hand, in word final position, the voiceless stops are distinguished by being preceded by a comparatively short vowel with tight voicing
and glottalization; and a relatively long vowel with full voicing preceding a voiced stop (ibid: 32).

Hawkins, P (1984: 33-34) describes the interlocking nature of vowel length preceding word final voice opposition of consonants on a physiological (i.e. a mechanical) basis. According to him, the voiceless sounds require greater energy of articulation, because they do not have the support of voicing to amplify the sound wave. The fortis sounds thus require greater oral (i.e. supraglottal) air pressure to maintain audibility. This in turn reduces the pressure differential across the vocal cords, and the cords stop vibrating. Before a lenis sound, or finally, voicing continues for a longer period (ibid: 34). It should be concluded, from what has been said above, that "vowels are 'naturally' longer but deviate from this target (by being shortened) when they precede a fortis consonant" (ibid: 34).

Therefore, it would be reasonable and more precise to say that vowels preceding word-final fortis consonant (and more relevantly in this paper, /p, t, k/) become naturally (physiologically) shorter than those preceding lenis counterparts (/b, d, g/). In normal speech, for example, the long vowel /i:/ may be pronounced shorter than the short vowel /i/ when the former precedes /t/ and the latter /d/. According to Kalevi and Wiik’s (quoted by Gimson, 1980: 98) experiment, seat has its mean time 12.3 csecs. while hid 14.7 csecs.

Note that the same effect also operates when plosives occur medially in a word; cf. the length of /ai/ in rider/writer, although in this situation voicing throughout the compression stage is likely to be present in /b, d, g/ (Gimson and Cruttenden, 1994: 141).

3 Voice onset time of plosive consonants

Lisker and Abramson (1964) point out that the timing of onset of voicing (VOT) is the inclusive cue to distinguish /b, d, g/ from /p, t, k/ particularly in initial and medial positions. In other words, as Abramson (1977: 295) claims, temporal variations in glottal settings for phonation would differentiate most homorganic consonants said to be distinguished phonologically by such features as voicing, aspiration and intensity. Now, then, we can consider VOT as the primary cue to the phonetic descriptions of the voice opposition of English plosives when we discuss those particularly in initial position.

If the onset of vocal cords vibration starts before the burst then it is called voicing lead (-VOT), whereas if it starts after the burst it is called voicing lag (+VOT or VOT). Then the term aspiration is used to describe the auditory effect of the glottal turbulence accompanying the voicing lag (Port and Rotunno, 1979: 654 quoted by Chiba, 1994). +VOT, in particular, is divided further into two domains: the short lag (0 msec - 25 msec after the burst) and the long lag (60 msec - 100 msec after the burst).

3.1 VOT of English plosives

From the view point of VOT studies (e.g. Lisker and Abramson, 1964; Zlatin, 1974; Keating, Linker and Huffman, 1983; Flege and Eefting, 1987), English speakers generally employ the short lag for /b, d, g/ while the long lag for /p, t, k/, although some exceptions of this are reported as in Keating, Linker and Huffman (1983: 281) in which they claim that "while initial /b, d, g/ are generally voiceless unaspirated, some speakers sometimes
prevoiced”.

On the other hand, some other languages (e.g. Dutch, Spanish, Hungarian, Tamil and so on) as well as Japanese use the lead (-VOT) for /b, d, g/ while the short lag for /p, t, k/ (e.g. Lisker and Abramson, 1964; Homma, 1980). Thus the voice opposition of English initial plosives is characterized by way of the positive VOT, in other words, aspiration. Also, the value of VOT varies rather systematically according to its place of articulation; as the place of articulation goes back in the oral cavity, the value of VOT increases. In other words, there is a rule such as /b/ < /d/ < /g/, /p/ < /t/ < /k/. And this seems to be a universal generalization (e.g. Lisker and Abramson, 1964).

VOT is also said to function as a crucial perceptual cue by native speakers of English. The evidence of categorical perception is reported by Lisker and Abramson (1970), according to which, for instance, the perceptual shift from /ba/ to /pa/ occurs 25 msec after the burst. But, as a matter of course, the faster the speaker speaks, the smaller the difference between the two VOT becomes.

It is also noted that some studies dealing with context effects on VOT have determined that VOT is conditioned by such factors as the following vowel quality, stress, consonant cluster effects and nature of speech samples (i.e. sentential vs. citation material) (Lisker and Abramson, 1967; Klatt, 1975; Weismer, 1979). I will discuss some of these factors on VOT, which are considered to be most relevant to this essay, in the following sections.

3.2 VOT and the following vowel quality

It has been reported that VOT is closely related with the following vowel quality (e.g. Klatt, 1975; Port and Rotunno, 1979; Weismer, 1979). Lisker and Abramson (1967), however, reported that vowel identity had no systematic influence on VOT. But, as it is pointed out in Klatt (1975: 691) that their data sample may have been insufficient, we should follow the later studies in this essay.

First, VOT is conditioned by tenseness of the following vowels. In Port and Rotunno’s (1979 quoted by Chiba, 1994: 78) study, they use /i, A, u/ as tense vowels and /i, *, u/ as lax vowels to make CVC word samples which start with /p, t, k/ and all end with /n/. And they report that tense vowels make preceding VOT values longer than lax vowels and their ratio are 12% for /p/, 10% for /t/ and 12% for /k/ (ibid).

Weismer’s (1979) study also concerns VOT in a similar context. His speech samples were also constructed to yield a set of CVC words which, instead, include /i, ei, u/ and /i, *, */ as tense and lax vowels respectively (which are followed by the classification of tense/lax contrast given by Ladefoged (1975: 73-76)) and have voiced/voiceless characteristics of the final consonants. His study shows that tense vowels make preceding VOT values 11% longer than lax vowels regardless of the voicing features of the final consonants (ibid: 199-200). (However, Klatt (1975: 691) reported that the VOT for /p, t, k/ is greater by about 10 to 40 msec before sonorant consonants.)

It is clear, from these studies, that the VOT in voiceless plosives is greater before tense vowels than lax vowels. In other words, English initial /p, t, k/ followed by tense vowels have greater aspiration.

VOT may also be conditioned by the following vowel tongue height. Klatt (1975) reported a statistically significant difference in VOT values for voiceless plosives, depending on whether a high or low vowel followed the plosive. In his study, the VOT in voiceless plosives followed by a vowel was found to be greater if the syllable
nucleus was a high vowel (ibid: 691). And he concluded that the VOT is 15% longer before the high vowels /i, u/ than before /ay, */ with significance at the 0.01 level (ibid: 691). The possible explanation of longer VOT before high vowels is that "high vowels seem to influence the behavior of the larynx such that the laryngeal fundamental frequency is higher and voicing is less easy to initiate or sustain than in other vowels" (ibid: 694).

On the other hand, Weismer (1979: 202) reported that tongue height did not appear to exert a systematic influence on VOT, with the exception of values associated with /k/. And he also pointed out that the Klatt's data was biased with /kV_/ words. In this light, he supposed that ",/k/ data reflect the fact that the transglottal pressure drop necessary for the initiation of voicing is realized more slowly when /k/ is paired with a high vowel" (ibid: 202).

In conclusion, we can say that VOT for English initial plosives is greater before tense vowels and, therefore, it follows that greater aspiration, and accordingly greater force of articulation, is manifested in the context. In contrast, it seems at present that vowel height effects on VOT are relatively smaller, with the exception of values associated with /k/.

3.3 Influence of stress on VOT

Most phonetic descriptions agree that the voiceless plosives are invariably aspirated when they begin stressed syllables and unaspirated when they begin unstressed syllables that are non-initial within a word. Moreover, some descriptions add that in word-initial position the amount of aspiration in /p, t, k/ may vary considerably with degree of stress (Gimson, 1962: 146 quoted by Lisker and Abramson, 1967). Thus, since we have good reason to believe that the feature of aspiration is directly related to the timing of voice onset, it follows that we should expect differences in stress to be reflected in VOT values.

Lisker and Abramson (1967) examined the difference in VOT values of word-initial /p, t, k/ in stressed/unstressed syllables both in isolated words and sentences. (Note that their specification of stress opposition primarily follows Trager and Smith, 1951). They reported that the VOT difference between stressed and unstressed /p, t, k./ is about four times greater in isolated words than in sentences (a figure of 24 msec vs 6 msec) (ibid: 17). This is because, as they pointed out, the stressed stops are reduced more than are the unstressed in sentences, so that the stress effect on VOT is itself very much reduced in the case of stops in sentences. And the larger stress effect in isolated words is almost entirely a matter of the overlong delay of voice onset in the production of the stressed stops (ibid: 17).

Keating, et al. (1983) (also in Keating (1984)) conducted a similar experiment which, however, differs in that she used only isolated words each containing one of the six stops in English before a low vowel, and stops occurred either initially or intervocally; the following vowel had primary stress, secondary stress or was reduced. She reports that the degree of aspiration of the voiceless stops is sensitive to stress level although the VOT values of initial /t/ are the same before vowels with secondary stress and no stress (Keating, 1984: 305-306). (It is considered that the high VOT values, in her study, obtained for initial voiceless stops in unstressed syllables result from reduction and devoicing of the entire unstressed syllables. That is, the high VOT values may not be the same as aspiration of the initial consonant (ibid.).) She also shows that medial voiceless stops have long-lag VOT values
before main and secondary stress whereas, before reduced vowels, the values are much lower, falling within the
range of short rather than long lag, and that initial voiced stops usually have short-lag VOT values regardless of
stress (ibid: 307).

Thus, it is reasonable to conclude that initial and medial voiceless plosives have greater VOT values before
stressed syllables and tend to increase the values according to the greatness of stress. In other words, it has been
substantiated that greater aspiration is manifested in more stressed syllables.

3.4 VOT of English consonant clusters involving plosives

There is a wide variation in the average voice onset time for plosives in different consonant cluster
environments. Klatt measured, in his study (1975), the mean voice onset times in different clusters involving
plosive consonants (i.e. /b, d, g, p, t, k/ + /r, l, w/; the resulting clusters conform to the English phonotactics)
including the clusters that begin with /s/ (i.e. /spr, str, skr, spl, skw/). Note that all his sample words are
monosyllabic legitimate English words.

He reported that averaging across place of articulation, the mean VOT for voiced series is 18 msec before a
vowel (i.e. single consonants) and 23 msec before a sonorant consonant (i.e. /l, r, w/). The corresponding means for
voiceless plosives (not preceded by /s/) are 61 msec before a vowel and 81 msec before sonorant consonants (ibid:
689), and they range from 47 msec for /p/ to 102 msec for /tw/. In contrast, when a voiceless plosive is preceded
by the consonant /s/ in the same word, the VOT is considerably reduced; the average VOT following /s/ is 22 msec
before vowels and 29 msec before sonorant consonants (ibid: 691).

The increased time lag of voice onset in /p, t, k/-plus-sonorant clusters has been described in the phonetics
literature (e.g. Jones, 1975; O'Connor, 1973; Gimson, 1980). The sonorant, /r, l, w/, is said to be voiceless and have
the equivalent quality to fricatives in these environments. As the result, particularly voiceless plosives followed by
one of these sonorants have longer aspiration as a whole before the onset of voicing. Thus, the manifestation of
voice opposition of initial plosives is made much clearer in this context in terms of aspiration, that is, the
voicelessness of the following sonorant. To the contrary, initial clusters that start with /s/ saliently reduce aspiration
whether or not they are followed by a sonorant.

3.5 VOT of Japanese plosives

We can not find many studies on VOT of Japanese plosives. In this section, I want to discuss some major

In Homma's (1981) study, twenty-four real and nonsense test words were prepared, and the words contained
vowel /a/ and voiceless and voiced stops /p, t, k, b, d, g/ at the initial and also at the medial position to yield a /CaCa/
word with accent (i.e. Japanese pitch accent) on the first syllable (i.e. Japanese mora). Also these words were
placed in the sentence frame: Kore wa ____ desu. ('This is ____'). Half of the test words have gemination of
stops at the medial position. Note that Japanese is a mora-counting language. The length of an utterance
phonologically depends on the number of moras. For example, kan (‘a can’) and kana (‘syllabary’) are two mora
words which linguistically have the same duration. The ratio of duration of *gaka* ('painter') and *gakka* ('lesson') corresponds to the number of maras namely 2:3, because the first constituent of the geminated /kk/ is counted as one mora (Homma, 1981: 273). Thus, the former would be divided into two moras /ga-ka/ and the latter three moras /ga-k-ka/.

The following were some notable points observed in her study:

1. VOT increased as the place of closure moved toward the back of the mouth.
2. VOT was shorter before voiceless stops than voiced ones, but the difference was negligible except with velars.
3. VOT was clearly related with accent but not with gemination of stops.
4. Japanese stops have shorter VOT than English. (Homma, 1981: 280)

Among the above findings, part of (1), (2) and (3) are also found in earlier studies on VOT of English plosives (e.g. in Lisker and Abramson, 1964; 1967). However, the last evidence should be of greater significance specific to Japanese. For instance, the VOT value for the accented initial /t/, regardless of the voicing feature of following consonants, is 32 msec; the value is even smaller than the half of the values observed in earlier data on VOT of English plosives in the same context (e.g. Keating, et al, 1983). It is rather closer to the values of English /d/. This generally holds true for other voiceless plosives.

Chiba (1994) reported on VOT of initial plosives and clusters in English words spoken by three native speakers of English and by twenty five Japanese secondary students. The words are found in a conversational script so that the context differs from one word to another. His result lacks the conformity with the universal feature of VOT (i.e. VOT increases as the place of articulation moves toward the back of the mouth) because the words are influenced by the following vowel quality, voicing of the following consonants and sentencial stress. However, the mean values for the VOT for initial /p, t, k/ of the native speakers of English are still at least twice those of the Japanese learners in the same contexts (ibid: 86). Among these, he claims that Japanese learners are especially poor in producing /p/, because where, in the spectrogram, /t/ and /k/ show relatively clear vertical striations, that of /p/ is so vague or sometimes even regarded as a fricative substitute /Φ/ (ibid: 87).

Also, most of the native English speakers employ +VOT for voiced initial plosives with a few exceptions for /b/ whereas most of the Japanese learners employ -VOT for the voiced series; 27 cases out of 33 tokens for *good*, 11 out of 18 for *going* and 14 out of 16 for *bag* (ibid: 86). He points out that it seems that these negative VOT shows the influence of the manner of articulation of Japanese plosives (ibid: 86); voicing is the primary cue for the realization of voice opposition of plosives in any context.

Chiba's (1994) study also dealt with VOT for such clusters involving plosives as /tr/ and /kl/. He reported that the VOT of Japanese learners is considerably shorter than that of native English speakers. This is because the clusters are often divided into two sounds with a vowel /u/ inserted in between the consonants so that the sonorants are never devoiced as English. Furthermore, the initial plosives are often pronounced without aspiration. Consequently they often sounded as *dree* for *tree* or *gleen* for *clean* (ibid: 88).
Thus, from the above results, we can characterize the features of Japanese plosives as following:

1) Japanese voiceless plosives, in particular in initial position, have considerably less aspiration than those of native speakers of English to the extent that they may be heard as voiced counterparts; this may well be attributable to less force of articulation.

2) Japanese initial voiced plosives, in general, have voicing feature during the closure (i.e. -VOT); in word initial, voicing is the primary cue for the realization of voice opposition of plosives.

3) Initial clusters involving plosives (except those with preceding /s/) have considerably less aspiration than those of native speakers of English without devoicing the following sonorant; this is partly because the plosive itself have a relatively weak intensity of articulation and partly because an intrusive vowel is inserted between the plosive and the sonorant.

4. The syllable structure of Japanese

4.1 Phonological syllables of Japanese

It has been generally agreed on that the phonological syllable of Japanese is dealt with in terms of the temporal unit called haku (Kamei, 1956; Kindaichi, 1967; Sugitou, 1989). The haku corresponds to each kana, which has separate symbols for each V, CV, and CyV syllable of the language, and each haku tends to be pronounced, in principle, almost in the same duration in normal speech of Japanese. Note that these syllables are all open syllables. Japanese haku also includes syllables such as:

1) the latter constituent of a set of two consecutive vowels
2) the latter-half constituent of a long vowel: /R/ (chou-on)
3) a syllable-final gemination of a following obstruent onset: /Q/ (soku-on)
4) a nasal that is homorganic in place of articulation with a following consonant or just a nasalized vowel of the preceding vowel before a V syllable: /N/ (hatsu-on)

Note that Japanese is not a tonal language although it has two types of tonal accents: high pitch and low pitch, which, thus, distinguish the meanings of the two homorganic words such as hashi with a low-high pitch ('bridge') and hashi with a high-low pitch ('chopsticks') respectively. It is also noted that among the above hakus, 2), 3) and 4) are called special hakus (cf. Kindaichi, 1967) and they do not occur in word initial and in isolation. Each haku in the above, then, constitutes an additional mora (haku), juxtaposed to the mora which the basic (C(y))V syllable constitutes.

In short, Japanese syllables are open syllables except /N/ and /Q/, and Japanese does not have consonant clusters except CyV, /N/ and /Q/ syllables. This follows that there is no contact assimilation between adjacent consonants in Japanese. Thus, as is often said (e.g. Murakawa, 1988; Imai, 1980; Makino, 1977; Fujii; 1986), Japanese learners have a strong tendency to add or invert parasitic vowels to English consonant clusters resulting in breaking not only the sound quality of the clusters but also the overall rhythm of English speech.

4.2 Realization of different types of syllables
If we analyze the normal speech of Japanese, then how are the phonological syllables which has been discussed above, namely haku, phonetically realized? I want to discuss it particularly in terms of duration.

English is a rhythmic stress-timed language. Rhythm tends to fall with the same amount of time between two primary stresses, regardless of the number of syllables; the more syllables a rhythmic unit has, the shorter becomes the duration of the segments (Homma, 1978; 1981). Therefore, stress and the number of syllables have great effects on vowel duration (Homma, 1981: 280). On the other hand, Japanese is a mora-counting language; each mora (haku) constitutes the rhythmic unit, and given a certain number of moras, word duration is relatively fixed (cf. Homma, 1981; Jouu, 1977; Kindaichi, 1967), although the duration of each syllable in a word is phonetically slightly different. Thus, Japanese is regarded as a language of syllable-timed rhythm (Pike, 1947), and is often said to have staccato rhythm (Bloch, 1950 in Joos, 1958: 331) or machinegun rhythm (O'Connor, 1967: 126).

By looking at spectrographs of normal Japanese speech, we can observe relatively the same length of each haku schematically.

In Sugito's (1989) spectrographic study on the duration of each syllable of words, she examined some words which include the special haku as well as words which do not include any of those special haku. The words were spoken by a Japanese speaker and a native English speaker at a normal speed. Her study reveals the following four striking features/difference of the relative duration of syllables between the tokens of English and Japanese:

1. ‘Yamamoto’ (a person's name) does not have any special haku in it. We can observe the relatively same duration of each haku in /ya-ma-mo-to/.

2. An English word ‘runner’ pronounced by the Japanese speaker, that is /ra-n-na-a/ (four moras), includes a nasal that has the same place of articulation with a following consonant (in this case, /n/ in the second mora). Thus it is also described as /ra-N-na-a/ We can notice the considerably longer /n/ in the Japanese token compared to the English one. And this long /n/ constitutes two moras, which has almost twice the length of an ordinary mora.

3. An English word ‘batter’ pronounced by the Japanese speaker, that is /ba-t-ta-a/ (four moras), includes a gemination of a following obstruent onset (in this case, /t/ in the second mora). Thus it is also described as /ba-Q-ta-a/. We can notice the considerably longer /t/ (i.e. the closure time) in the Japanese token compared to the English one. And this long /t/ constitutes two moras, which has almost twice the length of an ordinary mora.

Note that even if the /t/ has a longer closure time, it does not affect the VOT (Homma, 1981: 276). In other words, this compression stage is not for the greater force in the following release stage but merely for obtaining sufficient duration as one haku.

Also both (2) and (3) include the special haku: the latter-half constituent of a long vowel (in this case, /a/ in the last mora of the two words). Thus they are alternatively described as /ra-N-na-R/ and /ba-Q-ta-R/ respectively. Here again, we can notice the considerably longer /a/ (shown as /a a/) in the Japanese token compared to the English /E/. And this long /a/ constitutes two moras, which has almost the twice length of an ordinary mora. This shows that Japanese does not have long vowels; rather they are the sequence of two homorganic vowels to yield two hakus.

4. An English word ‘strike’ (one syllable word) pronounced by the Japanese speaker is /su-to-ra-i-ku/ (five
moras, that is five syllables). The striking difference observed here is that English /ai/ is a diphthong whereas it is represented as merely the sequence of /a/ and /i/ in the Japanese token. In other words, we can see the perfect glide of the vowel formants from /a/ to /i/ in the English token, and thus it cannot be divided further, while the clear border can be observed between /a/ and /i/ in the Japanese token. It follows that Japanese does not have diphthongs; rather, they are the sequence of two different vowels to yield two haku.

4.3 Duration of haku vs. word

Sugito (1989) also reported that the mean value of the duration of words increases as the number of the mora increases. In her study, a Japanese speaker read 556 frequent Japanese words, which are classified into groups of the same number of mora. She measured the mean value of the duration for each group of words. The result is that in Japanese the duration needed to pronounce a word is in proportion to the number of the mora in it.

In conclusion, the Japanese phonological syllables are realized in the unit of haku and each haku tends to have the same duration in the normal speech of Japanese regardless of its type.

5. General issues on the syllable structure of interlanguage phonology

Many views of the nature of interlanguage consider the learner's phonological representations as constituting a system intermediate between the native language and the target language. Among these, some researchers (e.g. Tarone, 1976, 1980; Oller, 1974; Greenberg, 1983; Sato, 1984) argue that the phenomenon of transfer extends beyond the level of individual phonemes to include syllable structure. Tarone (1980) hypothesized that there would be three separate influences on the shape of the interlanguage syllable structure. They are:

1. language transfer
2. reactivated first-language acquisition processes
3. universal processes of various kinds

The language transfer hypothesis would suggest that the learner would simply use the syllable structure from the first language (the constraints on sound sequences of their mother tongue) in his/her attempt to communicate meaningfully in the target language (Tarone 1980: 141; Brown, 1974: 42). Thus, if the hypothetical first language contains only syllables consisting of a consonant-vowel (CV) type, such as Japanese, it would be predicted that the learner would tend to transform the target language syllables into CV types.

The second hypothesis would suggest that the second language learner would tend to do what the first language learner does with syllable structure. That is, difficult syllables would be simplified by the second language learner in the same way that they are by the first language learner (ibid: 141). However, Oller (1974 quoted by Tarone, 1980) argues against this view; rather he has suggested that the processes which shape the interlanguage phonology are quite different from those which shape phonology in first language acquisition. He maintains that in first language acquisition of phonology, it is most characteristic for learners under 36 months of age to simplify by reducing or deleting difficult sounds, as cluster reduction (e.g. blue > bue), final consonant deletion (e.g. big > bi)
and weak syllable deletion (e.g. banana > nana) (ibid: 142). However, in the data reported in Oller (1974), second language learners appeared to use a very different strategy to pronounce difficult sounds:

a) instead of cluster reduction, second language learners use *vowel epenthesis* much more frequently (e.g. tree > tEree)
b) instead of final consonant deletion, second language learners more commonly favor *vowel epenthesis* (e.g. big > bigu)
c) *weak syllable deletion is very uncommon* among second language learners (ibid: 142)

Thus, it seems that epenthesis of this kind is a common strategy in second-language learners' acquisition of phonology rather than a strategy toward reduction or deletion. And Tarone (1980) argues that this vowel epenthesis strategy may also be explained in terms of a possible universal preference for the open (CV) syllable.

The third hypothesis, then, is that this simple open syllable may be a universal articulatory and perceptual unit (Tarone, 1972; 1980). In other words, in learning another language, it would seem to be possible that any leaner, regardless of first language background, might tend to break difficult sound combinations into simple CV patterns, using the kind of epenthesis (Tarone, 1980: 142).

In an attempt to bring empirical evidence to bear on this issue, Tarone (1980) conducted a pilot study of interlanguage syllable structure with data from six adult learners of English, two each from three L1 backgrounds: Cantonese, Brazilian Portuguese and Korean. Through the inspection of the findings from this study, she determined that:

1) the syllable structure of the interlanguage is often markedly different from that of the target language
2) both epenthesis and consonant deletion seem to be used as strategies for syllable modification
3) the dominant process influencing the syllable structure of the interlanguage phonology appears to be language transfer
4) a preference for the open (CV) syllable seems to operate as a process independent of a language transfer in influencing the syllable structure of the interlanguage phonology (Tarone, 1980: 148).

However, Sato (1984) argues against some of the claims from the above findings by Tarone. She points out that, while the tendency to simplify toward CV structure is apparent in those few tokens not attributable to transfer, the bulk of the data showed L1 transfer to be the main influence on the interlanguage production of Tarone's subjects (Sato, 1984: 45).

Sato (1984) examined whether this preference for open syllables is not universal but more dominantly influenced by transfer. It was hypothesized in her study that, because of transfer, Vietnamese-English IL would show a preference for closed rather than open syllables in the modification of English syllable-final consonant clusters. According to her, Vietnamese prefers the closed syllable (a total proportion of closed syllables of 81%),
and the language allows consonant clusters only in syllable-initial position (Sato, 1984: 47). The analysis revealed that modification of syllables containing consonant clusters more frequently yielded closed rather than open syllables. Specifically, it was shown that cluster reduction by one segment was favored over other processes: cluster deletion, vowel epenthesis, and feature change (ibid: 55). These results, therefore, stand against the hypothesis of a universal preference for the open syllable. Rather, the syllable modification strategy in IL phonology tends to be influenced primarily by transfer.

It was also reported in her study that the Vietnamese subjects hardly used the vowel epenthesis strategy in their IL production (ibid: 55). She argues that the virtual absence of epenthesis in the Vietnamese-English IL data suggests that this phonological process is strongly influenced by constraints on syllable structure in the first language (ibid: 55). This finding is also contrary to the preference of open syllables and the vowel epenthesis strategy claimed by Tarone.

Furthermore, in the same study, L1 transfer also was clearly demonstrated in terms of the effect of syllable position on IL cluster production. The subjects had much more difficulty with syllable-final than syllable initial clusters (ibid: 55) because Vietnamese does not allow syllable-final clusters. In other words, the tendency toward more radical restructuring of coda than onset clusters can be explained in terms of the L1 transfer.

It should be concluded, from what has been shown above, that the syllable structure of the interlanguage is often markedly different from that of the target language, and it seems that L1 transfer most strongly affects the IL syllable structure.

English, in fact, has its strict phonotactic constraints on consonant clusters within a syllable and they may often be simplified, in actual speech, following certain rules in relation to reduced syllables and co-articulation (cf. Brown, 1974; 1990) which, thus, may sometimes be different from those of the interlanguage phonology.

6. Further discussion

As we have discussed so far, the syllable structure of Japanese is very different from that of English. And Japanese learners may often use the Japanese syllable structure in their IL production because presumably L1 transfer most strongly affects the IL syllable structure.

To my knowledge, we can find, unfortunately, very few systematic studies on the IL syllable structure of Japanese (e.g. Greenberg, 1983). However, what was mentioned above enables us to predict, to some extent, that Japanese learners also use the Japanese syllable structure in their IL production. And the kind of syllabic inaccuracies may lead to nonnative-sounding utterances by breaking not only the rhythmic pattern of English but also the quality of the subordinating phonological constituents, or segments (cf. Pennington and Richards, 1986; Jensen, 1993).

Tokieda (1941) also points out that the fixed form of the rhythm of a language has a fundamental control over the characteristics of the sound representation of that language. Thus, according to him, because the features of those segmentals such as vowel and consonants are subject to their superordinate framework, this should be the
starting point of any phonetic and phonological study of a given language (ibid. 155-161). And the unit by which the rhythm is realized is the syllable (and the pause) (Brown: personal communication).

In this last section, based on our discussion so far, I want to discuss and reconsider teaching English plosives in different phonological contexts to Japanese learners in relation to the transfer of syllable structure of Japanese.

6.1 Teaching aspiration in initial /p, t, k/

As we have seen, Japanese speakers tend to pronounce a certain (C)V syllable in relatively the same time span as one single haku in the normal speech of Japanese. Thus, in Japanese, it is saliently restricted to pronounce a certain consonant very long or strong, while this is often done in stressed syllables of English words, because this would violate the Japanese syllable-timed rhythm of speech (Hashimoto, 1980; Murakawa, 1988). In particular, it is often said (e.g. Murakawa, 1988; Hashimoto, 1980; Imai, 1980; Makino, 1977; Fujii; 1986) and has been well substantiated in the VOT studies in earlier sections, that Japanese learners tend to pronounce English voiceless plosives relatively short in terms of aspiration and, thus, in a rather weak manner even though it is in the initial position of a stressed syllable. Thus, for example, pin may be heard as bin (Gimson and Cruttenden, 1994: 142).

Thus, I argue that it should be emphasized for Japanese learners of English to abandon their duration-based syllables and pronounce /p, t, k/ with sufficient aspiration in word initial or other stressed syllables. In order to obtain greater aspiration, the learners need to make the compression stage firm enough to encourage the release to be strong and long. Note, again, that it is reported in Chiba's (1994: 87) VOT study that Japanese learners are particularly poor in pronouncing initial /p/ and it is often realized as the fricative substitute /ɸ/. He points out that the striation of initial /p/ which shows the burst in the spectrograph is very vague compared to /t, k/ (ibid: 87). This evidence clearly shows that Japanese plosives are much weaker than those of English in initial position.

Japanese learners, then, need to learn that /p, t, k/ are fortis consonants and they involve acoustically greater force of articulation and, thus, greater aspiration in initial position, and that consequently this makes the duration of the whole syllable in question considerably longer. Otherwise they may be heard as their voiced counterparts: /b, d, g/.

6.2 Teaching voice opposition in final plosives

Brown (1990: 31) points out that the word final distinction between voiceless and voiced stops (as well as fricatives) is not usually stressed in manuals of English pronunciation and it is consequently often unknown to foreign teachers. In fact, this is certainly the case in Japan so that the allophonic level of descriptions of this kind should be an important implication for teaching plosives to Japanese learners of English.

As we have discussed earlier, the voice opposition of English plosives is phonetically lost in word-final position and the length of the preceding vowel is, then, the primary cue to the voicing.

However, this may make a clear contrast to Japanese. In other words, we can hypothesize the following two tendencies. First, in Japanese-English IL, final plosives tend to be pronounced relatively clearly with or without voicing according to its voicing feature as well as with an epenthetic vowel, resulting in increasing the number of
syllable. Second, since the duration of each syllable tends to be identical in Japanese, the contrast of the vowel length before word-final lenis/fortis opposition may not be manifested in Japanese-English IL. In other words, for example, the English words rope /roup/ and robe /roub/ (the vowel in the former should be phonetically much shorter than the latter because it is followed by the fortis consonant) are interpreted by Japanese learners as three-mora words: /ro-u-pu/ and /ro-u-bu/. And the two words have almost exactly the same duration because Japanese words which have the same number of moras tend to be pronounced in the same duration (Homma, 1981: 279; Sugito, 1989: 165). The first hypothesis has been substantiated in the Sugito’s (1989) spectrographic study in the earlier section, although we lack sufficient data. Also, the second hypothesis, still, would be highly probable in terms of transfer. Homma (1973, 1981) and Maeda (1979) reported that vowel duration in Japanese is more influenced by preceding consonants than the following one. Furthermore, the extent of this effect is drastically small compared to English. Thus we can say that, in Japanese, there will be no significant contrast in vowel duration before the voice opposition of following consonants.

Thus, I maintain that it should be emphasized for Japanese learners of English to abandon their duration-based and dominant (C)V syllables when they pronounce word-final plosives. Furthermore, Japanese learners need to be taught that the duration of English syllables varies greatly depending on their phonological contexts as well as on stress, and that the word-final opposition of plosives are distinguished not only or even primarily by a difference of the final consonant, but rather by quantitative (i.e. durational) contrasts extending over the greater part of the word (Gimson and Cruttenden, 1994: 141).

6.3 Teaching clusters involving plosives

As we have seen in the VOT section, initial clusters involving plosives (except those with preceding /s/) pronounced by Japanese learners have considerably less aspiration than those of native speakers of English without devoicing the following sonorant. And we have concluded earlier that this is partly because the plosive itself has a relatively weak intensity of articulation and partly because an intrusive vowel /u/ is inserted between the plosive and the sonorant.

I claim that this tendency is attributable to the transfer of the Japanese syllable structure to the English-Japanese interlanguage phonology. To begin with, Japanese does not have such consonant clusters in syllable-initial position (as well as syllable-final position). And the vowel epenthesis strategy may be a result of the transfer of the preference for open syllables.

I also maintain that the loss of such initial voice oppositions as in pray/bray, plot/blot and plead/bleed would be due primarily to the vowel epenthesis between these clusters. As we have seen, sonorants after voiceless plosives in such clusters must be devoiced, as if thus signalling aspiration. But vowel epenthesis between the segments inhibits the devoicing of the sonorants. As is also claimed in Gimson and Cruttenden (1994: 279, 286), allophones of these devoiced sonorants are important carrier of meaning in an opposition such as plot/blot. They point out, from the view point of minimal general intelligibility (cf. ibid: 283-287), that if voicing rather than aspiration is used as the distinguishing feature of /p, k/ vs /b, g/, the devoicing of /l/ after /p, k/ is less likely to occur; "if this type
of pronunciation is adopted, the listener's decoding of the message may have to rely very heavily on the general context" (ibid: 286).

Thus, it should be emphasized again for Japanese learners of English to abandon their preferred vowel epenthesis strategy in order to encourage the devoicing of the following sonorant and obtain greater aspiration in pronouncing clusters especially involving voiceless plosives (except those with preceding /s/).

7. Conclusion

As we have seen, Japanese-English IL syllable structure is remarkably different from that of English because of transfer. And the realization of English plosives in different contexts in the IL can be attributed to such characteristics of Japanese syllables as 1) the preference for open syllables, 2) no occurrence of initial consonant clusters, and, most importantly, 3) the relatively identical duration of each syllable. In other words, we have seen less aspiration in voiceless plosives in word-initial position and no devoicing of following sonorant in initial clusters involving voiceless plosives in the IL. Also it is highly predictable, in terms of transfer, that there may be no significant contrast of vowel length preceding fortis/lenis plosives in word-final position in the Japanese-English IL.

In conclusion, I suggest that Japanese learners should be made aware of the nature of English syllable structure, which is drastically different from that of Japanese, in order to acquire the allophonic level of pronunciation of English plosives in different phonological contexts.

Bibliography


Taishuukan.


