

Pronunciation Difficulties of Japanese Speakers of English: Predictions Based on a Contrastive Analysis

Steven W. Carruthers

Abstract

A contrastive analysis of English and Japanese phonology can help to identify potential, even likely, challenges for Japanese speakers of English (JSE). Learners of English will encounter differences in many areas: the distribution of allophones and phonemes, the number of vowels and consonants, the variety of environments in which fricatives occur, the discrimination and production of /r/ and /l/ and other English approximates, the variety of syllable structures, and prosody. Knowledge of these phonological contrasts would aid instructors and learners in selecting critical features for focused attention.

Introduction

A contrastive analysis of English and Japanese phonology can help to identify potential, even likely, challenges for Japanese speakers of English (JSE). This proposition is founded on Lado's (1957) claim that learners "transfer the forms and meanings" (p. 2) from their first language (L1) to the second language (L2). His assertion is the basis of the Contrastive Analysis Hypothesis (CAH), which states that a comparison of two languages can be used to "predict areas that will be either easy or difficult for learners" (Gass & Selinker, 2001, p. 72). Lado (1957) emphasized that knowledge of language similarities and differences is critical for developing teaching materials, creating tests for pronunciation and vocabulary, designing research, and understanding cultures (pp. 2-8). Lado warned, however, that these predictions must be confirmed by evidence from actual learner production (p. 72). Lado recognized that some *hypothetical* problems are not realized in production, and, conversely, significant difficulties not predicted may arise. Moreover, due to variations among individuals, not all learners will encounter the same difficulties. (For this reason, and to overcome the weakness of a pure contrastive analysis, examples in this paper are based on realized errors.) Despite these caveats, Lado maintained, "The problems will nevertheless prove quite stable and predictable for each language background" (p. 72).

Although knowledge of L1 transfer is not the only factor affecting pronunciation of an L2, it is certainly an important com-

ponent in a balanced approach to more intelligible production (Avery & Ehrlich, 1992, p. xvi). Thompson (2001) observed that JSE "find the more complex [vowel and consonant] distinctions and sound combinations of English very hard to produce" (p. 297). Moreover, there are numerous phonological features on which English and Japanese differ. When pronouncing English, speakers of Japanese are likely to encounter an array of difficulties stemming from differences in sound inventory, distribution of sounds among the categories of phonemes and allophones, syllable constraints, and prosody.

Segmental Difficulties

In pronouncing English, Japanese learners face two basic *segmental* issues: (a) sounds present in English but not in Japanese and (b) differences in the distribution of phonemes and allophones. Stockwell, Martin, & Bowen (1965, as cited in Gass & Selinker, 2001) offered a hierarchy of difficulties, with the most difficult being situations in which there is "differentiation" (p. 76) (i.e., two L2 sounds correspond to one L1 sound), the next difficult being situations in which a sound is present in the L1 but not the L2, and the least difficult being situations in which L1 and L2 sounds are roughly equivalent. Accordingly, English phones not found in Japanese will be an initial hurdle in producing intelligible English pronunciation. A contrast of Japanese and English provides many examples. Kobayashi (as cited in Nakai, 2005) counted the number of sounds in Japanese as 108, 1700

fewer than in English (p. 14). Although these numbers may seem exaggerated, they do reflect the disparity in sound inventory between the two languages. Lado (1957) also claimed that learners of an L2 will have greater difficulty with an L2 phoneme absent from the L1 than an L2 sound that is similar to an L1 sound (p. 13). Thus, we would expect JSE to be challenged by the English lax vowels and the consonants /ð/, /θ/, and /v/.

Contrasting the Sound Inventories of Vowels

Japanese uses only five vowels, compared to 12 in English, as it makes no tense and lax distinction (Avery & Ehrlich, 1992, p. 137). Tense vowels in English are those with slightly more muscle tension (/iy/, /ey/, /uw/, and /ow/) and are contrastive to their lax counterparts and other vowels in English (/I/, /ε/, /ʊ/, /ʌ/, and /ə/) (Avery & Ehrlich, p. 31). To further complicate the situation, Japanese vowels are actually “between the tense and lax vowels of English” (Avery & Ehrlich, p. 137). Kewley-Port, Akahane-Yamada, and Aikawa (1996) estimated that JSE need to acquire six or more vowels; in addition, learners actually need to redistribute their existing vowels. When speaking English, JSE pronounce /ʌ/ and /æ/ as /a/ (Thompson, 2001, p. 297) or interchange these sounds (K. Cook, personal communication, April 7, 2006); and /ə/ becomes /a/ (Thompson, p. 298). In an experiment designed to measure the intelligibility of JSE, Kewley-Port et al. found intelligibility judged by English native speakers particularly low for /ʌ/ (p. 2). They added that new vowel acquisition was more problematic for vowel sounds closely clustered in place of articulation (p. 3) (i.e., sounds concentrated in the mid-central, mid-front, and high-front regions). Of lesser concern, because of constraints on Japanese pronunciation, JSE may devoice /I/ and /ʊ/ between voiceless consonants (Thompson, p. 297-298). K. Cook (personal communication, March 15, 2006) offered the following ex-

ample: Learners may pronounce *situation* as /sɪtsuweɪʃʌn/.

Other vowel features make speech of JSE less intelligible to native speakers of English. English vowels vary in duration, sometimes depending on the environment. It is generally known that the duration of the tense high-front vowel /iy/ tends to be longer than its lax counterpart /I/; additionally, both have longer duration when preceding voiced consonants in the same syllable (Celce-Murcia, Brinton, and Goodwin, 1996, p. 64). “Failure to make these [vowel] distinctions,” as Avery and Ehrlich (1992) pointed out, “can lead to misunderstandings” (p. 96). Kewley-Port et al. (1996) suggested, however, that “since the Japanese language has a vowel length contrast...this contrast can be easily transferred to the correct production of short-long vowels in English” (p. 3).

Contrasting the Sound Inventories of Consonants

The sound inventory of English also includes consonants not found in Japanese. Well noted are difficulties with /θ/ and /ð/, often substituted by /s/ and /z/, respectively (Thompson, 2001, p. 298). Before high-front vowels, /ʒ/ or /dʒ/ may be substituted (Thompson, p. 298). Lambacher et al. noted that when listening, learners confuse /θ/ before /iy/ with other fricatives, such as /ʒ/, as /s/ does not occur before /iy/ in Japanese (p. 335). This difficulty with perception may affect production; consequently, learners may produce *thin* similarly to the English *shin*. I say “similarly” because the Japanese sound inventory actually utilizes [ɕ] before [i], which is similar to [ʒ] except “articulated with the blade and body of the tongue higher in the mouth” (Lambacher et al., 2001, p. 335). Also prevalent is the challenge of the English /v/; JSE may instead produce /b/, sounding *berry* for *very* (Avery & Ehrlich, 1992, p. 135; Thompson, p. 298).

Contrasting the Distribution of Sounds as Phonemes and Allophones

Not unexpectedly, phones are distributed differently in English and Japanese among the categories of phoneme and allophone. Learner difficulties in perception and production, however, can be predicted following Best's Perceptual Assimilation Model (PAM) (Lambacher et al., 2001, p. 335). According to PAM, during perception, sounds without an L1 equivalent are assimilated in the following three ways:

1. L2 sound is assimilated to an L1 category.
2. L2 sound is assimilated as an uncategorizable speech sound.
3. L2 sound is not assimilated to speech. (Best as cited in Lambacher et al., 2001, p. 341)

It follows that in production, a category 1 sound would be substituted with a strongly similar available L1 sound, such as the Japanese [ɕ] for the English /ʒ/. Sounds in category 2 are likely substituted with a sound with similar manner, approximate place of articulation, or acoustics; for example, JSE replace /θ/ with /s/. Category 3 sounds are absent in production. Best (as cited in Lambacher et al., 2001) shared some additional refinements: Pairs of phones in the L2 that correspond to different sounds in the L1 would be "easy to discriminate" (as cited in Lambacher et al., p. 335). An L2 sound that does not exactly match an L1 sound is more difficult (p. 335). Best added that most difficult to discriminate would be a set of L2 phones that are allophones of one L1 phone (p. 335), such as the English /f/ and /h/, which are allophones in Japanese. These fricatives, as well as English liquids /r/ and /l/ and other approximates, are problematic for JSE, both receptively and productively, and these areas will be the focus of the following subsections.

Fricatives. In certain environments, selecting appropriate English fricatives may challenge JSE. Lambacher et al. (2001) observed that the English /f/ and /h/ are often unidentifiable for JSE (p. 443) and difficult to produce in certain vowel contexts. In

Japanese, /h/ has three allophones: "the palatal fricative [ɕ] before /i/, the bilabial fricative [ɸ] before /u/, and [h] elsewhere" (Lambacher et al., 2001, p. 335). Thus, in production before /uw/, /h/ may be sounded as [ɸ]. I have often heard *bula* pronounced as [ɸula]. Other fricatives have allophonic variants dependent on the vowel which follows. Avery and Ehrlich (1992) stated that before high front vowels, /s/ and /t/ are pronounced /ʃ/ and /tʃ/, respectively (p. 135). JSE may produce *sit* as /ʃIt/ or [ɕIt]. Thompson (2001) warned that /d/ and /z/ before the high vowels /iy/ and /uw/ may also be affected, being sounded instead as /dʒiy/ or /dʒuw/, and /tuw/ goes to /tʰu/ (p. 298). All of the above phenomena are a result of these sounds being in complementary distribution before their respective high vowels (Shibatani, 1987, pp. 865-866).

Liquids. In addition to fricatives, English liquids are a notorious challenge for Japanese learners. The English /l/ and /r/ are often described by teachers and learners as corresponding to a single Japanese sound; however, this categorization is not entirely accurate. Avery and Ehrlich (1992) described these two sounds not as allophones of a single Japanese sound but rather as "one liquid sound which is between the English /r/ and /l/" (p. 135). Thompson (2001) characterized the Japanese /r/ as a flap-like sound, like short /d/ (p. 298). Price (as cited in Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004) agreed, describing the sound as an "apico-alveolar tap [ɾ]" (p. 234), a flap articulated with just the tip of the tongue. In this paper, I have chosen to treat these sounds as allophones, as Aoyama et al. (2001) proposed: "Despite the articulatory difference between Japanese and the English liquids /l/ and /r/, Japanese speakers seem to perceptually assimilate both English liquids to Japanese /r/" (p. 234). Thus, although they are not technically allophones of Japanese [ɾ], these sounds act as allophones because of perceived similarities in place and manner of articulation.

However described, both English /l/ and /r/ are a challenge to discriminate and produce for JSE. Moreover, the native speaker of English likely misunderstands the JSE's production of these sounds. Avery and Ehrlich (1992) stated that because the JSE produces [ɾ] for both English /l/ and /r/, the native speaker listener will hear "the opposite of what the Japanese speaker intended" (p. 135). Further confusing the situation, as Aoyama et al. (2001) observed, although both are assimilated as Japanese [ɾ], English /r/ is "more dissimilar phonetically" (p. 234), despite being regularly transliterated as *r* orthographically (i.e., romaji). In a study on the intelligibility of /l/ and /r/ production, Aoyama et al. found that JSE, particularly children, are better able to improve /l/ pronunciation (p. 245). Also, the Japanese [ɾ] is more likely to be perceived as English /l/ when substituted for /l/ (p. 245). Interestingly, Aoyama et al. stated, "An L2 sound that is similar...may enjoy an advantage in the early stages of L2 acquisition" (p. 245) because of the intelligibility of the L1 sound, but ultimately, learners may achieve greater accuracy with the more dissimilar sound. However, for the English /r/, Lado's previously mentioned assertion—sounds not in the L1 sound inventory are the most difficult to acquire—seems to apply.

Avery and Ehrlich (1992) identified a further concern: In syllable-final position following a vowel, JSE substitute a dissimilar native sound, [ɾ], for /l/ and /r/ (p. 135). This analysis, however, is unsatisfying and incomplete. Because a review of the literature had thus far not provided a clear explanation, I surveyed Miura's (1979) *English Loanwords in Japanese: A Selection*, a collection of loanwords (in dictionary form), to look for patterns. Loanwords and other English words spoken by JSE generally conform to the same phonological constraints (although there are some exceptions due to orthographic influences). This survey of 78 loanwords with /r/ or /l/ showed that in word-final unstressed syllables, JSE more likely substitute a combination of a reduced vowel sound followed by /r/ (/ə/ or other re-

duced vowel plus /r/) with a long full vowel sound (32 of 38 dictionary entries). For example, *charter* is pronounced as /tʃa:ta:/ by JSE (Miura, p. 36). In word-final stressed syllables ending in /ar/, the /r/ is dropped and the vowel lengthened (4 of 6 dictionary entries). For example, *car* is pronounced /ka:/. However, this pattern is not true for all vowel plus /r/ combinations. Following /o/ or /i/, /r/ is substituted by /a/ (2 of 6 dictionary entries). For example, *door* is pronounced /doa/ (p. 47) and *volunteer* is pronounced /boraNtia/ (p. 30). Thus, the only tendency that could be identified with such a small sample is that word-final stressed syllables ending in /ar/ are pronounced /a:/, but for others—those ending in /or/ or /ir/—the /r/ is replaced by /a/. In syllable final clusters (vowel plus /r/ clustered with another consonant), the combination of a vowel plus /r/ is generally pronounced /a:/ (9 of 12 entries). For example, the first syllable of *charter* /tʃa:ta:/ follows this pattern (p. 36). As for the English /l/, JSE retain a consonant sound when it occurs alone (not in clusters), substituting with the Japanese [ɾ] (or similar liquid) and adding an epenthesis vowel in all cases, either stressed (11 of 11 entries) or unstressed (10 of 10 entries). This is true even with occurrences of the English syllabic-*l* [ɫ] as found in words like *bustle* [hʌsɫ], which is pronounced [hus:uru] (p. 68) by JSE. The survey of 3 words containing /l/ occurring in syllable final clusters is inconclusive; it can be said that /l/ may be omitted (1 of 3) or retained and epenthesis (2 of 3), but no tendency could be identified. In sum, because learners of English are likely to omit /r/ or substitute it with a vowel sound, and /l/ is often retained but articulated as [ɾ], JSE will need special attention to and practice of these sounds.

Other approximants. Other approximants challenge the JSE. The English glides /y/ and /w/ in word-initial position before the corresponding high vowel might be omitted by JSE, although these sounds are not as much of a problem in other environments. Therefore, they may produce /ir/ for *year* or

/ʊd/ for *wood* (Avery & Ehrlich, 2002, p. 136). For *w*, lack of jaw movement is part of the problem, but the lesser degree of lip rounding of JSE also leads to low intelligibility (Thompson, 2001, p. 297). Thompson stated that lip rounding is not only critical to pronouncing /w/ but also important in pronouncing the lip-rounded English high-back vowel /uɰ/ (p. 297-298), which in Japanese is unrounded /ʊ/.

Lip rounding actually denotes two movements. Ladefoged and Maddieson (1996) described these as “vertical lip compression” (i.e., decreasing aperture) and “protrusion” (p. 295). In English and other languages these are coordinated movements, but Japanese utilizes the former, not the latter. Similarly to increasing protrusion, decreasing “lip aperture tends to lower all formant [peak] frequencies” (p. 295). An alteration in the frequency of a sound can reduce intelligibility. Ladefoged and Maddieson, citing a personal communication with Pulleyblank, noted that this same lip compression without protrusion phenomenon happens in the /h/ allophone [ɸ] that occurs before Japanese /ʊ/, an unrounded high back vowel (p. 295).

Suprasegmental Difficulties

Contrasting Syllable Construction or Constraints

Regarding the *suprasegmental* level, constraints of Japanese syllable construction affect the pronunciation of JSE. English permits more syllable types than Japanese, which makes English pronunciation rife with syllable-related challenges. Japanese has only open syllables, represented as (C)V, and syllabic-*n*. Syllabic-*n*, identified here as /N/, is “a nasal sound similar to (but not identical with) English /ŋ/ as in ‘sing’” (Avery & Ehrlich, 2002, p. 136). Meanwhile, English permits V, CV, CVC, CCVC, CCVCC, and others (Avery & Ehrlich, p. 53). To deal with consonant clusters, that is, “two or more consecutive consonants or vowels in a speech segment” (Pei, 1966, p. 41), JSE unconsciously use epenthesis (Avery & Ehrlich, p. 59, 136). Celce-Murcia

et al. (1996) defined epenthesis as “the insertion of a vowel or consonant segment within an existing string of segments” (p. 164). Pei, narrowing this definition, noted that the insertion has “no etymological justification, but whose usual purpose it is to ease a difficult transition between two other adjacent sounds” (p. 81). For example, because Japanese syllable constraints do not permit consonant clusters, JSE may pronounce *strike* as [sʊtoraiɰ], inserting vowels between the consonants. Additionally, to deal with a closed syllable at the end of a word, JSE may “add a vowel...to make the word conform to the Japanese pattern” (Avery & Ehrlich, p. 54). The addition of a vowel at the end of a word is also known as vowel *paragoge* (Pei, p. 193). The specific vowel inserted in either epenthesis or paragoge follows a phonological pattern. Lovins (1975) stated that in most environments, /ʊ/ is the usual epenthesized vowel (p. 99). However, after syllable-final /t/ or /d/, /o/ is added (p. 98); following palatals—/ʃ/, /tʃ/, /ʒ/, and /dʒ/—/i/ is inserted (p. 99). The latter deviations are due to allophonic variants (in Japanese) dependent on the vowel that follows (see the discussion of *Fricatives* earlier in this paper). The rule for vowel selection follows the constraints of allophonic variation in Japanese “in order to keep the pronunciation of the consonant in question as close to that of the source language [English]” (K. Cook, personal communication, October 18, 2006). Thus, for *bed*, *match*, and *bus*, the JSE may produce /bɛd:o/, /matʃi/, and /bʌs:ʊ/. These features cannot be ignored. Unfortunately, learners often use epenthesis strategies without knowing it, making monitoring and correction difficult (Avery & Ehrlich, p. 60). Nakai (2005) also warned, “Vowel epenthesis then interferes with English rhythm and intonation, and ultimately with intelligibility” (p. 16). That is, difficulties at the segmental and suprasegmental level affect the learners’ English prosody.

Contrasting Prosody

Differences in English and Japanese prosodic systems contribute to difficulties in

pronunciation and intelligibility (Thomson, 2001, p. 299). Prosodic features include “suprasegmental phonemic phenomena” (Pei, 1966, p. 223), such as stress and pitch. Particularly in English, prosodic features “convey additional, concomitant information” (Pei, p. 223); that is, English prosody carries meaning.

Stress and pitch. First, Japanese and English differ in the manner of stress. In English, stressed syllables are “longer, louder, and higher in pitch” (Celce-Murcia et al., 1996, p. 131). Pei (1966) defined the latter specifically as a “highness or lowness of tone” (p. 208). In contrast, Japanese indicates stress primarily through pitch (Avery & Ehrlich, 2002, p. 137), not duration. This type of stress is termed *pitch accent*, giving prominence “to a syllable or word by a raised pitch, or a change of pitch” (Pei, p. 211). On a syllabic level, Japanese uses pitch mainly to contrast phonetically homophonic words, as seen in Example (1) and Example (2).

(1) aME ‘candy’

(2) Ame ‘rain’

English, however, employs stress to show contrast and to identify the focus of the sentence (Celce-Murcia et al., 1996, pp. 178–179). Because of these differences, “Japanese speakers may have difficulty both producing and perceiving the characteristic stress patterns of English” (Avery & Ehrlich, 2002, p. 137).

Syllable-timed versus stress-timed. A second significant prosodic difference is that Japanese is a syllable-timed language. That is, the length of an utterance is determined by the number of syllables (Celce-Murcia et al., 1996, p. 152), each syllable having about the same duration. English, however, is a stress-timed language; the number of stressed syllables determines length (p. 152). JSE tend to pronounce all vowels fully. Accordingly, as stated by Avery and Ehrlich (2002), “Japanese speakers’ pronunciation of English words and sentences may lack the vowel reduction necessary for English rhythm” (p. 137). For example, Japanese speakers do not have reduced vowels such

as /ə/ (Avery & Ehrlich, p. 137), and although Japanese permits contractions, it does not use /ə/ in reduced speech. Thus, reduced speech forms of English words, such as *was*, *can*, and *have*, are difficult to perceive and pronounce (Thomson, 2001, p. 299; Celce-Murcia et al., p. 230).

Effect on Perception and Production

Differences in stress and pitch may encumber accurate perception and execution of English intonation patterns. Intonation is “the pitch, tone quality and melody of speech particularly when used to make a syntactical or emotional distinction” (Pei, 1966, p. 131). Thompson (2001) noted that Japanese relies on other features to make distinctions:

Many of the attitudinal colours painted by English intonation patterns find expression in Japanese by adverbials and particles. And Japanese does not share the English use of intonation to highlight information structure (for instance, to distinguish information which speaker and listener share from information which is new to the listener). (p. 299)

That is to say, where English uses sentence stress, Japanese would use a function word to highlight the focus of the utterance for the interlocutor.

Conclusion

Although it is not possible to cover all pronunciation difficulties for JSE, knowledge of these phonological contrasts aids the instructor and learner in selecting critical features for focused attention. The goal of a contrastive analysis is, according to Lado (1957), to identify “the hurdles that have to be surmounted in the teaching” (p. 3). When supported by actual learner production, as recommended by Lado (p. 72), it is beneficial to the development of teaching materials. The knowledge gained is invaluable in improving learners’ discrimination of the sounds of English and the intelligibility of their speech.

References

- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32, 233-250. Retrieved February 25, 2006, from <http://webpages.acs.ttu.edu/kaoyama/aoyama16.pdf>
- Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford, UK: Oxford University Press.
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge, UK: Cambridge University Press.
- Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course* (2nd ed.). Mahwah, NJ: Laurence Erlbaum.
- Kewley-Port, D., Akahane-Yamada, R., & Aikawa, K. (1996, October). *Intelligibility and acoustic correlates of Japanese accented English vowels*. Paper presented at the Fourth International Conference of Spoken Language Processing, Philadelphia, PA. Retrieved February 25, 2006, from <http://www.asel.udel.edu/icslp/cdrom/vol1/202/a202.pdf>
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK: Blackwell.
- Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.
- Lambacher, S., Martens, W., Nelson, B., & Berman, J. (2001). Identification of English voiceless fricatives by Japanese listeners: The influence of vowel context on sensitivity and response bias. *Acoustical Science and Technology*, 22(5), 334-343. Retrieved February 25, 2006, from http://www.u-aizu.ac.jp/~steeve/fricpaper_withfigs.pdf
- Lovins, J. B. (1975). The segmental phonology of western loanwords in Japanese. *Loanwords and the phonological structure of Japanese* (pp. 47-119). Bloomington, IN: Indiana University Linguistics Club.
- Miura, A. (1979). *English loanwords in Japanese: A selection*. Rutland, VT: Charles E. Tuttle.
- Nakai, F. (2005, November). The role of pronunciation in communication. *The Language Teacher*, 29(11), 13-19.
- Pei, M. (1966). *Glossary of linguistic terminology*. New York: Columbia University Press.
- Shibatani, M. (1987). Japanese. In B. Comrie (Ed.), *The world's major languages* (pp. 855-880). New York: Oxford University Press.
- Thompson, I. (2001). Japanese speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (2nd ed., pp. 296-309). Cambridge, UK: Cambridge University Press.

