Mind the Gap: An Analysis of Japanese and English Phonology and Its Implications for Learners

Richard M. Woodward

1. Introduction

This paper analyzes the phonological differences between Japanese and English to pinpoint potential challenges for native Japanese speakers learning English. Initially, the paper outlines the phonetic inventory of Japanese, which is comparatively smaller than English, consisting of sixteen consonants, two unique consonant phonemes (/N/ and /Q/), and five short vowels. A subsequent section compares the syllable structure of English with the moraic structure of Japanese. Phoneme charts illustrate the predominantly open-ended nature of Japanese sound units. The paper then contrasts the role of syllables and morae in word stress, identifying English as a stress-accent language and Japanese as a pitch-accent language. It is proposed that mastering word stress is more effective within a broader discourse context, as it can be influenced by other suprasegmental features. The third section explores these suprasegmental features, revealing that despite differences in timing and stress patterns, the two languages share similar intonation contours. The final section addresses common difficulties encountered by Japanese English learners. These issues are directly linked to the phonological contrasts presented earlier, as the transfer of Japanese pitch accent and syllable-timed characteristics often results in a monotone or "robotic" speaking style. The paper concludes by detailing classroom and self-study activities designed to help Japanese students develop an ear for the stress contours and rhythm of English.

2. The Japanese Sound System

As members of distinct language families—English (Indo-European) and Japanese (Japonic)—it is expected they will differ significantly in their phonological systems. The number of sounds provides a clear initial contrast. English has twenty-four consonants (including complex consonants and semivowels), compared to just sixteen in Japanese. English has as many as twenty vowel sounds when considering both American and British English, while Japanese is limited to five basic vowels. Despite this numerical disparity,

other features of Japanese phonology contribute to a greater number of actual consonant and vowel-related sounds. These are also explored below.

Consonants

The sixteen fundamental consonant sounds of Japanese are categorized below by manner of articulation. There are six groups, one fewer than in English, as Japanese lacks lateral sounds. Within these six groups, consonants are further classified based on the voiced-voiceless distinction and their place of articulation.

Stops/plosives

Like English, Japanese has six stop consonants: the voiceless bilabial /p/, alveolar /t/, and velar /k/, with their corresponding voiced counterparts /b/, /d/, and /g/.

Fricatives

Japanese has two basic fricative sounds: the voiceless alveolar /s/ and its voiced counterpart /z/. The inclusion of the voiceless alveopalatal /ʃ/ as a distinct phoneme varies among scholars, as it is largely considered a positional variant of /s/. However, the absence of its voiced counterpart /ʒ/ in Japanese may explain the challenges native Japanese speakers face when pronouncing French words like *je* and *genre*.

Affricates

Unlike English, which has two complex affricate consonants (/tʃ/ and /dʒ/), Japanese is often considered to have only one—the voiceless alveolar /ts/—or none at all. This might seem surprising to learners, as sounds like /tʃ/, /dʒ/, /ts/, and /dz/ do appear in Japanese. However, as Shibatani (2005) explains, these are regarded as allophones (positional variants) of the phoneme /t/.

Shibatani uses a widely accepted interpretation to define the sixteen fundamental consonant sounds of Japanese, which are based on the standard Tokyo dialect. Over time, the pronunciation of this native dialect has been influenced by Sino-Japanese vocabulary, loan words, and hybrid terms. This suggests that while /s/, /z/, and /t/ were present in the original Tokyo dialect, other fricative and affricate sounds were introduced through later linguistic influences.

This shift in pronunciation is also reflected in different Japanese romanization systems. The Kunreisiki system adheres to the sixteen-consonant premise, while the modern Hepburn system, based on English pronunciation, includes specific representations for fricatives and affricates. As a result, Hepburn offers a more intuitive spelling for

English speakers. For example, the Kunreisiki spelling of *kirisutan* for the word Christian is less helpful for English speakers than the Hepburn spelling, *kirisuchan*, which more accurately represents its Japanese pronunciation.

Nasals

While English has three voiced nasal sounds, Japanese originally also had three: the bilabial /m/, alveolar /n/, and velar /ŋ/. However, the velar /ŋ/ is now typically considered an allophone of the velar stop /g/. Historically, /g/ was used at the beginning of words and /ŋ/ internally. Citing research that shows a steady decline in the use of /ŋ/, Shibatani (2005, pp. 171-2) argues against classifying /ŋ/ as a distinct phoneme, given that most speakers now use /g/ in both word-initial and internal positions.

Liquids

While English has two distinct liquid sounds, /l/ and /r/, Japanese has only one: the voiced alveolar liquid, /r/. This sound is produced as a flap, where the tongue makes a brief, single contact with the alveolar ridge or the area directly behind it.

Glides/semi-vowels/approximants

Both English and Japanese have two approximant sounds (also called semi-vowels or glides): the voiced palatal /j/ and the voiced velar /ɰ/. Similar to English, these sounds are produced with a partial blockage of airflow and a clear movement of the tongue towards and then away from the place of articulation. Although /ɰ/ is classified as a velar approximant, its articulation involves significant lip rounding, which is why it is sometimes listed with the bilabial consonants.

Special consonants

Thus far, we have discussed thirteen consonants. The remaining three include /h/, with its allophones /ɸ/ and /ç/, as well as two unique consonants not found in standard IPA charts, /N/ and /Q/.

While English classifies /h/ as a vowel due to its variable articulation, Japanese considers it a voiceless glottal consonant that can become both voiced and bilabial. This change is demonstrated in sequential voicing, a phenomenon where a consonant's place and manner of articulation change when it is part of a compound word. For instance, *hana* (flower) becomes *bana* in the compound *ikebana* (flower arranging), where /h/ changes to /b/. We could be forgiven for mistaking /b/ as an allophone of /h/. However, we have already noted that /b/ is an independent consonant.

Furthermore, /h/ has two positional variants. Before the vowel /ɯ/, it becomes the bilabial fricative /ɸ/, as in *fugu* (pufferfish) and *Fujisan* (Mt. Fuji). Before the vowel /i/, it becomes the palatal fricative /ç/, as in *hina* (doll). The Hepburn romanization system

shows the similarity between /ɸ/ and the English /f/. The Kunreisiki system, on the other hand, writes *fugu* as *hugu* and *Fuji* as *huji*.

Moraic consonants

The topic of mora will be discussed later, but it is important to note that in Japanese, consonants are never isolated; they always appear with a vowel, forming a consonant-vowel (CV) structure. This is evident in words like *ha-na*, *i-ke-ba-na*, *fu-gu*, and *hi-na*. However, two consonants, /N/ and /Q/, can function as independent "syllables."

The consonant /N/ should not be confused with the uvular nasal on IPA charts, although Shibatani notes their articulatory similarities. It is distinct from /n/ because it can form an independent "syllable" within or at the end of a word, as seen in *Fu-ji-sa-n*.

The symbol /Q/ represents the lengthening of a consonant sound. In Japanese script, this is marked by its own character and is counted as a separate mora (syllable-like unit), just like /N/. This articulation can be described as a form of glottalization. Similar to the brief glottal stop in English words like *button* and *mutton*, there is a comparable action in the Japanese word *yatto* (finally), transcribed as /jaQto/. Shibatani (2005) explains this phenomenon as a lengthening of the following consonant, which is reflected in romanization by a double consonant, as in *yappari* (as expected), transcribed as /jaQpari/.

The first table on the next page lists the sixteen Japanese consonants. When two consonants are shown together, the voiceless consonant is on the left.

Japanese Vowels

Compared to its consonants, the Japanese vowel system is quite straightforward. Standard Japanese has five short vowels: /i/, /e/, /a/, /o/, /ɯ/. Unlike English, Japanese does not have tense vowels, complex vowels, or diphthongs. The two back vowels, /o/ and /ɯ/, are similar to their English counterparts in terms of lip rounding. The second table below illustrates the height and front-back characteristics of these vowels. The tongue's position for the three more central vowels—/e/, /a/, and /o/—is slightly lower than for their IPA and English counterparts.

		Bilabial	Alveolar	Palatal	Velar	Glottal
Plosive		p b	t d		k g	
Fricative			s z			h
Nasal		m	n			
Liquid				r		
Approximant		ɯ̟		j		
Moraic					N	Q

Figure 1. Consonant phonemes in standard Japanese

	Front		Center		Back
High	i				ɯ
Lower-mid		ɛ		ɔ	
low			a		

Figure 2. Vowel phonemes in standard Japanese

Long vowels

According to Kaiser (1994), "there is no systematic difference in quality between short and long vowels" in Japanese (Chapter 3, p. 14). This is because Japanese long vowel sounds are not an inherent part of the syllable, as they are in English, but are created by adding an extra vowel sound as an independent mora, resulting in a CV-V structure. For example, Kaiser (1994) contrasts the four-mora *obasan* 'aunt' with the five-mora *obaasan* 'grandmother', which is lengthened by the duplication of a vowel sound (Chapter 3, p. 15).

おばさん (four mora sound units, o-ba-sa-n) 'aunt'

おば<u>あ</u>さん (five mora sound units, o-ba-<u>a</u>-sa-n) 'grandmother'

Japanese linguist Kindaichi Haruhiko proposed using a separate phonetic symbol, /R/, to represent vowel lengthening by duplication, as described by Shibatani (2005, p. 162-3). This proposal aimed to distinguish duplication from the lengthening that occurs when adding a different vowel sound, as seen in *keizai* 'economy' (ke-i-za-i).

けいざい (four mora sound units, ke-i-za-i) 'economy'

This is written /kƐizai/, whereas *grandmother* is written /ɔbaRsaN/ by adherents of Kindaichi's proposal.

A third way to express vowel lengthening in Japanese script is the use of a long bar in katakana, the script used for foreign loanwords. This is seen in the borrowed English word *economy*, which is written with a bar representing the duplicated vowel sound.

エコノミー (five mora sound units, e-ko-no-mi-i) 'economy'

The duplicated vowel /i/ is represented by the use of the *bar* symbol.

Vowel devoicing

Kaiser (1994) identifies two types of vowel devoicing in Japanese—"obligatory and optional devoicing" (Chapter 3, p. 29)—a phenomenon not found in English. While vowels are typically voiced, the high vowels /i/ and /ɯ/ in Japanese may lose their voicing.

Obligatory Devoicing

The vowels /i/ and /ɯ/ are always devoiced when they are between voiceless consonants, as shown in the words *kishi* (/kiʃi/) 'river bank' and *kushi* (/kɯʃi/) 'skewer'. However, this devoicing does not occur when the vowel is at the beginning of a word or in a stressed syllable.

Optional Devoicing

The vowel /ɯ/ can be optionally devoiced when it appears at the end of a word in combination with /s/ to form polite verb endings, such as /sɯ/. In most cases, devoicing is so pronounced that the vowel is almost completely inaudible. Consequently, in everyday speech, the final vowels of certain phrases become virtually imperceptible.

じこしょうかいします	ji-ko-sho-u-ka-i shi-ma-sɯ	'Let me introduce myself.'
リチャードです	ri-cha-a-do de-sɯ	'I am Richard.'

This section has analyzed the modern Japanese sound system, contrasting its consonants and vowels with those of English. The sound charts reveal that Japanese has fewer phonemes than English. However, a more comprehensive understanding of the Japanese sound inventory was achieved by also considering fricatives, affricates, and the mechanisms used to create long vowels. The next section will analyze the Japanese syllable system, including how the use of glide syllables further expands the sound repertoire.

3. Japanese Syllable Structure and Pitch Accent

According to Avery & Ehrlich (2012), English has five main syllable types, including Consonant-Vowel (CV), CVC, CCVC, CCVCC, and CCCVCC (p. 53). They point out that differences in syllable structure between English and Japanese can create challenges for Japanese learners of English. One major difference is that Japanese syllables are primarily CV in structure, which means they do not have the consonant clusters common in English. This is why Japanese learners often insert vowels between consonants, a direct result of their native language's suprasegmental units. The following analysis will examine Japanese syllable structure and its relationship to word stress.

Syllables and Mora

Japanese is characterized by its open syllable structure (CV), which contrasts with the predominantly closed syllable structure of English. The CV structure of Japanese refers to its rhythmic units, or mora. Essentially, morae are units that correspond to the kana—the Japanese writing system that organizes phonemes similarly to an alphabet. Although Japanese has fewer phonemes than English (sixteen consonants and five vowels), these combine to form forty-five basic mora units. The accompanying chart shows how the kana are arranged in columns, read from right to left, beginning with individual vowels and then each consonant-vowel combination. There are no symbols for individual consonant sounds.

わ	ら	や	ま	は	な	た	さ	か	あ
ɯa	ra	ja	ma	ha	na	ta	sa	ka	a
	り		み	ひ	に	ち	し	き	い
	ri		mi	hi	ni	ʧi	ʃi	ki	i
	る	ゆ	む	ふ	ぬ	つ	す	く	う
	rɯ	jɯ	mɯ	hɯ	nɯ	tsɯ	sɯ	kɯ	ɯ
	れ		め	へ	ね	て	せ	け	え
	rɛ		mɛ	hɛ	nɛ	tɛ	sɛ	kɛ	ɛ
を	ろ	よ	も	ほ	の	と	そ	こ	お
ɯɔ	rɔ	jɔ	mɔ	hɔ	nɔ	tɔ	sɔ	kɔ	ɔ

Figure 3. Forty-five basic Japanese kana, or mora.

As previously mentioned, there are also three special morae-/N/ (ん), /Q/ (っ), and /R/ (represented by a vowel kana)-that are typically not on the main charts. When the voiced and semi-voiced variants are included (indicated by the diacritics ゛and ゜), the total number of morae comes to seventy, as shown in the next chart.

Figure 4. Twenty-five voiced and semi-voiced mora in Japanese

Speed of Speech and the Perception of Syllables, Mora

Shibatani (2005) provides a clear distinction between the terms syllable and mora. He defines a mora as "a unit that can be represented by one letter of kana and functions as a rhythmic unit in the composition of Japanese poems (…). While ordinary syllables include a vowel, moras need not" (p. 158).

Kaiser (1994) demonstrates this difference using the word for "red snapper" (たい). While it is pronounced as a single syllable in typical speech, it actually consists of two rhythmic units (*ta* + *i*) which combine to form a long vowel sound. When a Japanese speaker deliberately slows down the word, for example to clarify pronunciation or spelling, they will make a clear distinction between the two units. This can be compared to how an English speaker might separate syllables to clarify spelling, as in *sna-pper*.

Shibatani further clarifies the difference between morae and syllables using examples containing the three special morae: *shinbun* (newspaper), *hakkiri* (clearly), and *ookii* (big) (p. 158-9). He explains that while *shinbun* has two syllables (*shin-bun*), it has four morae (しんぶん, *shi-N-bu-N*). Similarly, *hakkiri* has three syllables (*ha-kki-ri*) but four morae (はっきり, *ha-Q-ki-ri*), and *ookii* has two syllables (*oo-kii*) but four morae (おおきい, *o-R-ki-R*). The distinction between syllables and morae is crucial for understanding Japanese word stress, or accent.

Syllable Structure

Considering the mora as a distinct type of syllable allows us to expand the Japanese syllable structure beyond the basic CV pattern. As demonstrated by words like *ta-i* and *o-o-ki-i*, vowels can function as individual rhythmic units. Similarly, the example

shi-n-bu-n shows that certain consonants can also be counted separately.

In addition to these three structures (CV, V, and C), Kaiser (1994) identifies a fourth, CGV, which consists of a consonant, the glide or semi-vowel /j/ (G), and a full vowel. This structure, also known as palatalization, is seen in the morae や /ja/, ゆ /jɯ/, and よ /jo/. These thirty-six CGV morae bring the total number of Japanese rhythmic units to 106.

Figure 5. Thirty-six 'glide mora' in Japanese

Starting from the right-hand side, the chart above combines seven columns of unvoiced consonant-glide-vowel mora and five columns of voiced and semi-voiced variants. As with the other types of mora, these are regarded as individual rhythmic units.

Word Stress in English and Japanese

Just as word stress varies across English dialects, so too are there many variations among Japanese dialects. For simplicity, we will consider the characteristics of the standardized form, which is based on the Tokyo dialect.

Burns, Avery, and Ehrlich (Avery & Ehrlich, 2012) describe how English is a stress-accent language, with vowels in stressed syllables being "longer and louder" and unstressed vowels often reduced to a short schwa sound (/ə/) (p. 63). In contrast, Japanese is a pitch-accent language, where the morae have the same length and volume. Furthermore, while English words can have both major and minor stress, Japanese words have a maximum of one accented element.

Unlike Chinese or Vietnamese, Japanese has only two pitch levels, high (H) and low (L). Izubachi (2005) describes three pitch accent patterns in Japanese: an accent on the first mora, an accent on a subsequent mora, or no accent at all. If the pitch accent is on the first mora, that mora has a higher pitch than all the morae that follow. For example, the noun だいじん (minister) is pronounced 'DA-i-ji-n' (H-L-L-L).

If the accent is on another mora, the first mora has a low pitch, while the following morae up to the accented mora all take a higher pitch. Any morae after the accented one are lower in pitch. For instance, the noun うぐいす (nightingale) is said 'u-GU-i-su' (L-H-L-L). If there is no pitch accent, the first mora takes a low pitch, and all subsequent morae have a higher pitch. The noun せんたく(choice) is therefore pronounced 'se-N-TA-KU' (L-H-H-H).

The similarity between the initial low pitch in the last two patterns (L-H) may cause problems for non-native speakers attempting to distinguish between two-mora homophones. A classic example involves the words for chopsticks, bridge, and edge, all of which are pronounced はし, or /haʃi/. For chopsticks, the pitch accent is on the first mora ('HA-shi', H-L). However, both bridge and edge have the same L-H pattern ('ha-SHI'). Bridge and edge are only distinguishable when followed by grammatical particles, as the pitch pattern continues across word boundaries: 'ha-SHI-ga' (L-H-L) means 'the bridge', whereas 'ha-SHI-GA' (L-H-H) means 'the edge'. Fortunately, it is very rare for two-mora homophones to take on all three patterns.

General Rules

Just as there is no reliable way to predict word stress in English, there is no consistent method for determining the pitch accent of individual Japanese words. In both languages, the patterns must be memorized as part of the learning process. Burns, Avery, and Ehrlich (Avery & Ehrlich, 2012) provide four general rules for English that help distinguish stress patterns in homophones, three-syllable words, compound nouns, and words with affixes (p. 70-71). Similarly, Izubachi (2005) and Kaiser (1994) identify some general principles for Japanese.

In Japanese, interrogative pronouns and adjectival nouns consistently have their accent on the first mora. Examples include いつ (when), which is pronounced I-tsu, and the adjectival nouns どんな (which), pronounced DO-n-na, あか (red), pronounced A-ka, and ながさ (length), pronounced NA-ga-sa.

Adjectives ending in い (/i/) always conclude with two vowel sounds, and the pitch accent consistently falls on the first of these two. This pattern is seen in adjectives like たのしい (fun), pronounced ta-NO-SHI-i (L-H-H-L), and かなしい (sad), pronounced ka-NA-SHI-i (L-H-H-L).

Finally, verbs exhibit consistent pitch patterns across their conjugations. All verbs in their affirmative-polite and past-polite forms place their pitch accent on the final ま

(ma) mora. For example, いきます (I go) is pronounced i-KI-MA-su, and いきました (I went) is pronounced i-KI-MA-shi-ta.

In this section, we have contrasted the closed-syllable structure of English with the open-syllable structure of Japanese, noting that Japanese syllables break down into smaller units called mora. Unlike English syllables, which usually contain both consonants and vowels, morae can consist of both or just one. English is a stress-accent language, where word stress is indicated by vowel length and volume. In contrast, Japanese is a pitch-accent language, lacking this distinction. English words can have two stress points, while Japanese words have only one or none. Burns, Avery, and Ehrlich (Avery & Ehrlich, 2012) suggest teaching English words within context rather than in isolation to ensure accurate word stress. This approach is also beneficial for Japanese, as surrounding elements in a sentence can influence the pitch patterns of individual words, which will be the focus of the next section.

4. Contrasting the Suprasegmental Aspects of Japanese with English

Avery, Ehrlich, and Jull (Avery & Ehrlich, 2012) introduce the primary suprasegmental components of English speech: sentence stress and intonation. They also detail the features of connected speech, such as contraction, linking, deletion, assimilation, and palatalization, and how these affect pronunciation.

English is a stress-timed language, where unstressed syllables are reduced in length and clarity. This allows speakers to pronounce sentences with the same number of stresses in roughly the same amount of time, regardless of the number of syllables. This is in contrast to Japanese, which is a syllable-timed language. As discussed earlier, each mora, or syllable, has the same length and volume. Consequently, the time it takes to say a sentence in Japanese is determined by its number of syllables. Given this significant difference, one might expect major disparities in sentence stress and intonation. However, as will be shown, there are also some similarities.

Prosodic Phrases and Sentence Pitch Patterns

Igarashi (2021) explains that Japanese sentences are divided into prosodic phrases that generally mirror the syntactic divisions of noun and verb phrases (p. 8). These phrases are delineated by particles, and as seen earlier, pitch patterns can extend across word boundaries to include these particles. When considering sentence-level pitch patterns,

it becomes clear that the principles of word pitch extend to encompass entire phrases. Ayryziger (2019) outlines six characteristics of these phrasal pitch patterns, which are essentially an expansion of the word-level patterns.

1. Each phrase has one accented syllable (mora). All accented mora have a high pitch.
2. The first mora of a phrase is high pitch if it is accented. Otherwise, it is low.
3. Each high pitch is followed by a *downstep*, causing subsequent mora to have a low pitch.
4. Once the pitch goes low after an accented mora, it stays low for the rest of the phrase.
5. The pattern starts anew with each new phrase. A phrase might be as short as one word, or a whole clause in complex sentences.
6. All mora before the accented mora have a high pitch, except the first, which must be low since it is not accented.

Native English speakers do not consciously apply major and minor sentence stress, just as Japanese speakers do not consciously think about these stress patterns. In connected speech, Japanese speakers automatically begin with a low pitch, rise on the first accented mora, and then descend in pitch for the rest of the phrase. Ayryziger illustrates this with a weather forecast example, where each new line signifies a new phrase and the accented morae are in bold. The transcription and translation are provided below.

きの**う**　まで　は
ＬＨＬ　ＬＬ　Ｌ
すごし　や**す**い　すずしい　ひ　が　つづきました　が
ＬＨＨ　ＨＨＬ　ＬＬＬＬ　Ｌ　Ｌ　ＬＬＬＬＬＬ　Ｌ
ま**た**
ＬＨ
まなつ　の　**あ**つさ　が　もどってきました
ＬＨＨ　Ｈ　ＨＬＬ　Ｌ　ＬＬＬＬＬＬＬＬ

ki-nɔ-ɯ ma-dɛ ɯa
(Until yesterday)

sɯ-gɔ-ʃi　ja-sɯ-i　sɯ-zɯ-ʃi-i　hi　ga　tsɯ-ʣɯ-ki-ma-ʃi-ta　ga
(the cool and comfortable weather had continued)

ma-ta
(but now)
ma-na-tsɯ nɔ a-tsɯ-sa ga mɔ-dɔ-Q-tɛ-ki-ma-ʃi-ta
(the mid-summer heat has returned)

The downstep after an accented mora and the subsequent lack of pitch change within a phrase likely contribute to the perception of Japanese as somewhat monotone. However, other features also affect the rhythm of connected speech. Threse will be explored below.

Prominence

Avery, Ehrlich, and Jull (Avery & Ehrlich, 2012) state that English has major and minor sentence stress, with major stress typically placed on the final content word of an utterance, making it longer and louder. Igarashi (2021) explains that Japanese uses a similar process, where the information focus of a sentence stands out with a higher pitch relative to other prosodic phrases (p. 10). The prosodic phrase containing the sentence's focus has a wider pitch range than it would in isolation. Conversely, the pitch ranges of other prosodic phrases are compressed, mirroring the vowel reduction seen in unstressed syllables in English. Amanuma (1989) calls this a "shift in prominence," noting how pitch range and pauses can clarify the meaning of otherwise ambiguous statements (p. 140). In the following example sentences, capitalization indicates higher pitch, and parentheses mark prosodic boundaries.

a. (NAKAMURASAN WA) (kirai da) (to iimashita)
Mr. Nakamura hate said
b. (Nakamurasan wa) (KIRAI DA) (to iimashita)
Mr. Nakamura hate said

As the simple translation shows, the lack of case-marking pronouns or subject-specific verb conjugation makes the meaning ambiguous. In example a., if Mr. Nakamura is pronounced with a higher pitch than the other phrases, the meaning becomes "I hate Mr. Nakamura." Conversely, if hate is pronounced with a higher pitch, as in example b., the meaning becomes "Mr. Nakamura said he hates (it)."

I would compare information focus in English and Japanese as follows: English

sentence focus is a horizontal phenomenon, where stressed syllables are lengthened and unstressed ones are shortened. In contrast, Japanese focus is a vertical phenomenon, signified by a wider pitch range for the information focus, while other sentence elements have a reduced pitch range.

Intonation

The basic intonation pattern in Japanese is similar to English, with a rising-falling contour across an utterance. In Japanese, this same pattern is also present within individual prosodic phrases, with the lowest pitch at the very end of the complete utterance. Fujii (2012) describes other fundamental Japanese intonation patterns, comparing them to English. As in English, a simple statement can be turned into a question by using rising intonation at the end of the utterance. This allows speakers to express doubt on a subject using intonation alone, similar to how "This is a book" becomes "This is a book?". While rising intonation is used for Yes/No questions in both languages, a key difference is that Japanese can change meaning solely through intonation, without any alteration to its syntax.

これは本ですか？ (kore wa hon desu ka?) Is this a book?
これは本ですか？ (kore wa hon desu ka?) This is a book, is it?

Falling intonation can convey disbelief or disappointment in Japanese. In English, similar emotions would most likely be expressed with rising intonation.

Intonation for commands, exclamations, and WH questions is similar in English and Japanese, with both languages using falling intonation for these utterances. However, like the previous example, a change in intonation can add specific emotion or nuance to WH questions. For example, a final rising intonation expresses increased curiosity, as seen in the second case below.

なにをかったって？　Nani wo katta te?　(What did you buy?)
なにをかったって？　Nani wo katta te?　(What did you say you bought?)

Finally, Fujii (2012) identifies a subtle difference in alternative questions between English and Japanese. While English uses a rising tone for each alternative before a final falling intonation, the Japanese pattern changes depending on whether or not a verb is

used.

どちら、りんご、みかん？　Dochira, ringo, mikan?
(Which do you want) Apple or orange?

どちらをたべるの、りんご、みかん？　Dochira wo taberu no, ringo, mikan?
(Which do you want to eat, an apple or an orange?)

In the first case, intonation rises on both *ringo* (apple) and *mikan* (orange). In the second, intonation rises on *no* (question particle) and then falls on both *ringo* and mikan.

This section has briefly introduced the main characteristics of connected Japanese speech. Despite some similarities in intonation patterns, English and Japanese differ fundamentally in their accent systems. These differences, along with those in the syllabic and phonetic systems discussed previously, can create challenges for Japanese learners of English. The next section will address these difficulties and propose solutions.

5. Phonological Problems Experienced by Japanese Learners of English

This section examines problems caused by segmental differences between English and Japanese. I will suggest activities to prevent or correct pronunciation errors and then discuss difficulties with stress, rhythm, and intonation in English, with examples of practice activities.

Consonants

Section one contrasted the 24 consonant sounds of English with the 16 in Japanese. Since English uses sounds that don't exist in Japanese, L1 Japanese speakers may have initial difficulty pronouncing these new sounds.

/f/ and /v/

Although both English and Japanese have six stop consonants (e.g., /p/ and /b/), Japanese lacks the labiodental fricatives /f/ and /v/. While the Japanese sound /ɸ/ (a positional variant of /h/) assists L1 Japanese speakers in pronouncing /f/ at the start of words, the sound is often underarticulated or replaced by /h/ in the middle of words (e.g., *cohee* for *coffee*). Similarly, the voiced labiodental fricative /v/ is commonly substituted

with /b/ in words like *very*, *vest*, and *volunteer*. This error also occurs in mid-word positions, as in *favorite*, *lovely*, and *avoid*. To correct these errors, exaggerating the placement of the teeth on the lower lip and pausing as the articulators meet can highlight the correct articulation of /f/ and /v/. Rapidly alternating between these voiced and voiceless consonants without moving the articulators can also reinforce the connection between the two sounds.

/s/ and /t/

In section two, we saw that Japanese phonemes are arranged into consonant-vowel (CV) pairs called mora. This characteristic causes L1 transference errors, particularly with the consonants /s/ and /t/ when they are followed by the vowels /ɪ/ and /iy/. As Figure 3 shows, the 'sa-SHI-su-se-so' and 'ta-CHI-tsu-te-to' sequences in Japanese account for the tendency of Japanese learners to replace /sɪ/ with /ʃ/ (e.g., *sip*, *missing*, *city*) and /tɪ/ with /ʧɪ/ (e.g., *tip*, *meeting*, *romantic*).

Before using minimal pairs, it's helpful to have students isolate and pronounce /sɪ/ and /ʃ/, or /tɪ/ and /ʧɪ/ in rapid succession. This allows them to physically feel the difference between the sounds.

/θ/ and /ð/

Japanese lacks interdental consonant sounds. Because of this, Japanese L1 speakers often substitute /θ/ with the 'sa-shi-su-se-so' sequence and /ð/ with the voiced 'za-ji-zu-ze-zo' sequence (see Figure 4). This can cause errors such as *think* and *thank* becoming *shink* and *sank*, and *this* and *that* becoming *jis* and *zat*.

To help students with these sounds, exaggerate the pronunciation by isolating them and doing rapid, alternating repetitions of the voiced and voiceless consonants. This creates a physical and sensory memory of the sounds, something they cannot experience in their native language. As Underhill (Macmillan Education ELT, 2011) states, "(...) we are helping people to make contact with the muscles that make the sounds, (...) to find their own muscular criteria...to use muscular memory," (Macmillan Education ELT, 2011, 18:36).

/r/ and /l/

The distinction between /r/ and /l/ is arguably the most common consonant problem for Japanese L1 speakers. As we saw in section one, Japanese has only one liquid consonant, the alveolar /r/, which is pronounced as a brief flap. Because this sound is phonetically in between the English /r/ and /l/, Japanese speakers often confuse the two, saying, for example, *liver* instead of *river* or *flee* instead of *free*.

When teaching the /l/ sound, emphasize that the tongue tip must touch the alveolar ridge. Students can even press their tongue firmly against the ridge to help them become

familiar with the sensation. For the /r/ sound, instruct students to begin with their tongue in a neutral position, as they would for the Japanese vowel /ɯ/. They should then pronounce a short vowel sound, drop their jaw, widen their lips, and move their tongue toward the front of the mouth, stopping just before the alveolar ridge. Using minimal pairs like *light* and *right* can reinforce that the tongue touches the alveolar ridge for /l/ but not for /r/.

Clusters and final consonants

In section two, we noted that Japanese lacks consonant clusters due to the consonant-vowel (CV) structure of its mora. This trait leads to errors when pronouncing both consonant clusters and final consonant sounds in English, as Japanese L1 speakers tend to insert vowel sounds after each consonant. As Miyamoto (2019) gives, a simple example is "This is a jungle" becoming "Jisu izu a junguru" (paragraph 7). Clusters that contain /l/ or /r/ can be particularly troublesome.

While Avery & Ehrlich (2012) suggest inserting a short schwa vowel and then speeding up its pronunciation until the vowel disappears (p. 103), it may be more effective for students to slow down and enjoy the pronunciation of each part. This same technique can also help prevent adding an unwanted vowel sound at the ends of words. Encouraging students to consciously overcome the habits acquired while speaking their native language is key to eliminating the extra vowel sounds that define a Japanese accent.

Semi-vowels

Incomplete mora sequences in the 'ya' and 'wa' columns of Japanese also cause phonological transference errors. As seen in Figure 3, there are no mora for /jɪ/, /jɛ/, /ɰɪ/, /ɰɯ/, or /ɰɛ/. When these phonemes are needed, such as for loan words, simple vowel mora are used. For example, *Woodward* is written as うっどわーど (u-Q-do-wa-a-do) in hiragana script. This accounts for the tendency of Japanese L1 English learners to omit the half-vowels /w/ and /j/. For instance, *would* becomes *ood* and *year* becomes *ear*. Compounded by the consonant-vowel (CV) mora structure, *Woodward* becomes *oodoaado*. As Avery & Ehrlich (2012) suggest, reminding students of easier examples they can pronounce correctly, such as *we* and yes, may help them correct these errors (p. 136).

Vowels

In section one, we established that Japanese has just five vowel sounds compared to the twelve in English. In Japanese, long vowels are created by simply adding another vowel phoneme, and there are no diphthongs, complex, or tense vowels. Additionally, Japanese vowel sounds are never shortened in connected speech. These characteristics lead to the inability of many Japanese speakers to distinguish between various English vowel sounds.

Short vowels /ɛ/, /æ/, /ʌ/ and /a/

Japanese has just one low vowel, /a/, while English has four: /ɛ/, /æ/, /ʌ/, and /a/. This difference can make it difficult for Japanese L1 speakers to distinguish between them. To help with this, Avery & Ehrlich (2012) suggest practicing the two front vowels and the two back vowels in quick succession so students can feel the difference in how they use their mouths (pp. 99-100). The jaw-drop between /ɛ/ and /æ/ and the mouth-widening between /ʌ/ and /a/ can be exaggerated to make them more memorable. Teachers can also use minimal pairs and sentences with repeated vowel sounds, such as, "The hot dog got lost on a log in the fog."

Schwa, /ə/

Japanese does not have a neutral vowel like the English schwa /ə/. This means many students struggle to reduce full vowel sounds to schwa in connected speech. Therefore, it is crucial to teach schwa as a distinct vowel sound before working on word and sentence stress. Underhill (Macmillan Education ELT, 2011) describes how to produce the schwa by doing nothing: "The thing is to do nothing (...) the jaw is relaxed, the tongue is relaxed and the voice box is working," (Macmillan Education ELT, 2011, 16:15). Once students can make the sound in isolation, they can practice it in the context of common words. Examples in British English include the final phoneme of *mother*, *father*, *brother*, and *sister*, as well as the definite article *the*.

Tense vowels and diphthongs

As already mentioned, Japanese has just five short vowel sounds, which are duplicated to form long vowels. It lacks diphthongs, complex, and tense vowels, making these sounds difficult for Japanese students to produce in English.

Avery & Ehrlich (2012) recommend teaching tense vowels in isolation, exaggerating the pronunciation and using hand gestures, as if stretching an elastic band, to reflect the sound's length. Students should be encouraged to lengthen these vowels in one-syllable words (e.g., *beat*, *bait*, *boot*, *boat*) and then in longer words and sentences. Teachers

should draw attention to the mouth stretching for /ɪy/ and /ey/, and the lip rounding for /uw/ and /ow/.

In contrast, when practicing lax vowels, students should be instructed to keep their mouths relaxed, similar to how they produce the schwa. Lips do not need to be spread wide for /ɪ/ or /ɛ/, nor tightly rounded for /ʌ/. Once these vowels have been practiced, teachers can use minimal pairs to contrast long and short vowels.

6. Strategies for Helping Japanese Students Overcome Pronunciation Problems

Whether teaching pronunciation remedially or as part of a dedicated series of classes, it is important to keep activities meaningful, varied, and communicative. While practicing individual sounds is essential, it is equally important to practice them in context, at the sentence level.

Warm-ups

Authors like Archibald (in Avery & Ehrlich, 2012), Hadar (Accent's Way English with Hadar, 2020), and Mutonono (2016) advocate using warm-up exercises to prepare students' articulators for speaking English. These exercises, which include massaging the jaw and cheeks, stretching the tongue, and performing simple voice exercises, allow students to focus on articulator usage in English and consciously compare it with their native language. Just as athletes warm up to make their bodies more supple and reduce injury, students can make their mouths more pliable for successful sound production. This is particularly important for Japanese L1 learners, as the greater number of sounds in English requires a more extensive use of the mouth and articulators.

Modeling and listening activities

When teaching specific problem sounds, teachers should begin by modeling and demonstrating the pronunciation. This involves clearly showing the differences in the place of articulation for each sound. Overemphasizing pronunciation and overacting the articulation can make these distinctions more memorable for students.

Burns (in Avery & Ehrlich, 2012) recommends that teachers first guide students through a series of listening activities before encouraging them to produce the target sounds. Students are taught to recognize and discriminate between sounds before they are

asked to produce them. The activities that follow can be used for listening, controlled, and free practice while the teacher monitors performance.

Minimal pairs

Burns describes many uses for minimal pairs, which are words distinguished by a single phoneme, such as *cut* and *cat*. First, students listen and identify which of two or more sounds the teacher makes (e.g., *ban* or *van*, *vet* or *bet*). Words can be arranged in lists or a pyramid design to add a game-like aspect to the activity. Students then draw lines between the words they hear and circle the number under the final word.

cut　　cat

sap　　sup　　sap

tuck　　tack　　tuck　　tack

lamp　　lump　　lamp　　lump　　lamp

1　　2　　3　　4　　5

Next, minimal pairs can be used for discrimination or isolation activities. In these exercises, students must identify whether consecutive sounds are the same or different, or pinpoint the "odd one out" from a series of sounds.

Discrimination	Isolation
“with” “whiz” (X)	1　2　3　4
“sing”, “thing” (X)	“see”, “see”, “she”, “see”
“think” “think” (O)	“tip” “chip” “chip” “chip”

In discrimination exercises, students listen to a pair of words and mark whether they are the same (O) or different (X). In isolation exercises, they listen to a series of words and identify the one that is different.

Phoneme recognition can be practiced using sorting activities. The teacher reads a list of words, and students group them according to the sounds they contain. For example, Japanese students might be asked to distinguish between /s/, /ʃ/, /θ/, /t/, and /ʧ/ as the teacher reads out words like *ship*, *chick*, *sink*, *tick*, *think*, and *sip*.

Pair work and communicative activities

After students have practiced recognition and discrimination of individual sounds, words, or sentences, they can move to production practice. Following group repetition, students can do the activities introduced above in pairs. This keeps both students actively involved as they try to understand their partner or make themselves understood. For example, at the sentence level, students must identify which of two sentences their partner is reading.

A	B
She has a red mouth.	She has a red mouse.
Please wash the sheets.	Please watch the seats.

Burns warns that minimal pairs should not be overused to avoid monotony. It is important to vary activity types and quickly progress from teacher-centered to student-student participation. Integrating pronunciation practice into communicative activities is also key to keeping students motivated.

For example, information gap activities can be used. In a pair-based activity, one student arranges words or pictures on a bingo-style grid and describes them to their partner. The partner must then place the items in the same order on their own grid. Similar activities can be devised that require students to speak with others around the class to determine the location of all the items on their grid.

Peer support

Jull (in Avery & Ehrlich, 2012) stresses the importance of self-correction and self-monitoring by students (pp. 216-8). Teachers can encourage this through peer correction and support. For instance, for a pronunciation assignment, students can record themselves reading a text with target sounds and send the recording to a classmate for honest critique. Another method is to have students record a short self-introduction (including their name, age, birthday, family, hobbies, and favorite foods) for their peers to critique. Peers can then suggest areas for the student to work on.

7. Strategies for Helping Japanese Overcome Problems with Suprasegmentals

Many of the activities mentioned above can also be used to practice aspects of

connected speech that may be problematic for Japanese L1 speakers. The following sections will look at the areas of stress, rhythm, and intonation, and suggest additional activities to help improve these suprasegmental elements of their speech.

Stress

In section two, Japanese was described as a pitch-accent language, while English is a stress-accent language. It can be difficult for Japanese learners to hear and produce word stress correctly. Teachers should show stress visually when introducing new words, using gestures and symbols to compare the length and loudness of stressed and unstressed vowels.

Similar to their use in the previous section, discrimination and identification activities can be used to practice stress recognition. For example, students can listen to word pairs and mark whether the stress lies on the same or a different syllable. Alternatively, they can be asked to locate the stress or group words by stress pattern as the teacher reads out polysyllabic words. Burns (in Avery & Ehrlich, 2012) describes an activity where students identify the syllable containing a schwa (p. 201), which helps them compare the length and clarity of vowels in stressed syllables with the reduced sounds of unstressed syllables.

In section three, we saw how prominence in Japanese sentences is achieved through pitch variation and range. Japanese students may have difficulty with English stress patterns, which place emphasis on content words (nouns, verbs, adjectives, and adverbs). It is important to teach the difference between content words and function words to help students locate stress. Students can work in pairs to predict the stress patterns of sentences and then perform their predictions. Burns also suggests an activity that requires students to locate the reduced unstressed vowels in sentences (Avery & Ehrlich, 2012, p. 203). McNerney & Mendelsohn (in Avery & Ehrlich, 2012) highlight the importance of stressing the information focus of a sentence (p. 190). This suggests an activity where students are given a sample dialogue using WH and Yes/No questions to elicit specific details on a topic. Students can then create their own versions of the dialogue to practice shifting stress onto the focus of each sentence.

What did you do at the weekend?	I went to Tokyo.
Did you drive there?	I went there by train.
Who did you go with?	I went there alone.

What did you do there? I went shopping in Shinjuku.

Rhythm

In section three, we described Japanese as having a syllable-timed rhythm, where each mora has the same length. Mutonono (2016) infers that this underlies the native speaker complaint that Japanese people "pronounce every single word" (Mutonono, 2016, para. 5). The activities introduced previously can be used to practice word stress and also to show Japanese students where the major and minor stresses are in sentences. In particular, teachers need to encourage Japanese students to reduce unstressed vowels to a schwa and use linking techniques.

Avery & Ehrlich (2012) suggest using simple lists, such as telephone numbers, to help students become familiar with the stress-timed rhythms of English (p. 194). Shadowing—a technique where students repeat what they hear almost simultaneously, mimicking both pronunciation and rhythm—is another way to train the ear to become accustomed to the major, minor, and unstressed parts of connected speech. The regular rhythm of nursery rhymes and limericks can also be exploited to make students more sensitive to stress patterns.

I wanted to tie up my shoe
You wanted to tie yours up too
We got caught up in a knot
And untie it could not
So together forever, we two!

Once a rhythm pattern has been introduced and practiced, perhaps by tapping out the main stresses, more advanced students can be asked to create their own original versions. An easier alternative is the expanding sentences activity suggested by McNerney & Mendelsohn (Avery & Ehrlich, 2012, p. 189). Students are shown the major and minor stresses in a series of sentences where the stress pattern remains the same even if the number of syllables increases. In pairs, students can then create and introduce their own versions.

I want to play a game.
Do you want to play a game?

What game do you want to play?
I want to play a board game.

Intonation

In section three, we noted that while the basic intonation patterns of Japanese are similar to English, the pitch pattern may not vary as much. This, along with the equal stress assigned to each mora, can make speech sound monotone. It's important to show Japanese students how to use connected speech characteristics such as contractions, linking, and palatalization to make their intonation sound less stilted.

Mirroring the approach for pronunciation exercises, intonation practice should start with listening activities. Burns (in Avery & Ehrlich, 2012) introduces exercises that focus on the four English intonation patterns (p. 205). For example, a teacher can read sentences as either declarative statements or Yes/No questions, and students must respond correctly by either acknowledging the statement or answering the question.

"The lesson has finished." (Falling intonation) "I know."
"The lesson has finished?" (Rising intonation) "Yes, it has."

Interactive practice can also be used for non-final rising intonation in lists. The teacher sets a situation and reads a list of items, pausing after each one. Students are asked to respond appropriately, depending on whether the list has finished or not.

"I went to the store to get ingredients for the birthday cake. I bought flour," "Yes."
"Eggs, and milk," "Uhuh."
"Sugar," "Yep."
"Salt and cocoa." "Six ingredients?"

Falling intonation on the final item of a list signals its end. When students hear the last item, they respond by confirming how many were in the list.

The intonation patterns of tag questions can be practiced in a similar way. If the teacher reads a tag question with falling intonation, students should respond to it as a statement by acknowledging it in a matter-of-fact way. If the tag question has rising intonation, they should respond as they would to a Yes/No question.

"You're from Japan, aren't you?" "Yep."
"You're not from Shizuoka, are you? "Yes, we are!"

Once students can recognize intonation contours, the teacher can conduct pronunciation drills and then use the same types of activities for pair work. Students practice producing the correct intonation, and their partners provide immediate feedback.

Naiman (in Avery & Ehrlich, 2012) suggests using questionnaires and game-like activities to practice contractions and palatalization to improve students' intonation (pp. 168-9). Activities like "Who am I" quizzes with staged content or surveys used to gather real information provide good opportunities for practicing these characteristics of connected speech.

Where*dja* go to high school?
What subject *dja* like best?
Who w*zja* favorite teacher?

Finally, linking activities are another way to wean Japanese L1 speakers away from their tendency to pronounce each word individually. Avery & Ehrlich (2012) suggest sentences that can be used for both recognition and production practice, first with the teacher and then in pairs (p. 108).

Si**tat** the bac**kof** the room.
Pa**ssou**t the books.
Put the boo**kon** to**pof** the shelf.

Naiman (in Avery & Ehrlich, 2012) describes how such practice can be extended to the creation of dialogues by pairs of students (pp. 168-9). After practicing model dialogues that show how stop consonants link with following vowels, students can create and perform their own examples. Depending on the students' level, the teacher can give suggestions regarding the types of words to include.

Yesterday I wen**tto** Tokyo.
What di**dja** do?
I bough**ta** new computer.
Di**dja** ge**titin** Akihabara? …

This final section has concentrated on the common pronunciation problems experienced by Japanese L1 English learners, the causes of which lie in the phonological differences between the two languages, as presented in the first three sections. Based on suggestions from various authors, this paper has provided a general approach to teaching pronunciation and connected speech in the classroom, progressing from listening to group and then pair work. These activities can be adapted to suit the age, level, and needs of individual students. The importance of using a variety of recognition and production activities has been stressed.

It is also recommended that instructors encourage students to self-monitor and self-study to improve their pronunciation and suprasegmental skills. In addition to the recording activity described above, the 'Benjamin Franklin' activity presented by Rachel's English (2014) is an excellent example of a self-directed study method. Through such activities, Japanese students can use both peer and native speaker models for recognition and imitation to overcome the influence of their native language.

References

Accent's Way English with Hadar. (2020, July 1). *15 minute morning pronunciation practice for English learners*. [Video File]. Retrieved from https://youtu.be/l69yZ5xabbo.

Amanuma, Y., Ōtsubo, K., & Mizutani, O. (1989). *Nihongo onseigaku*. Tōkyō: Kuroshio shuppan.

Avery, P., & Ehrlich, S. L. (2012). *Teaching American English pronunciation*. Oxford: Oxford University Press.

Ayryziger. (2019, December 4). *Standard Japanese Phrasal Pitch Accent System* [Video File]. Retrieved from https://youtu.be/dqLq_ENOj3E.

Fujii, A. (2012, March 27). *Some notes on Japanese and English intonation*. Retrieved February 4, 2021, from https://shark.lib.kagawa-u.ac.jp.

Hudson, J. (2013, November 25). *Japanese speakers' English pronunciation errors*. Retrieved February 22, 2021, from https://pronunciationstudio.com/japanese-speakers-english-pronunciation-errors/.

Fukumori, T. (2010). *Kiso karano nihongo onseigaku*. Tōkyō: Tōkyōdō Shuppan.

Igarashi, Yosuke. (2021). *Intonation in Japanese dialects*. Retrieved February 4, 2021, from https://www.researchgate.net/publication/332249767.

Izubachi. (2005). *Pitch accent in the Japanese language*. Retrieved January 25, 2021, from https://everything2.com/title/Pitch+accent+in+the+Japanese+language.

Japanese. (n.d.). Retrieved December 31, 2020, from https://sites.google.com/pdx.edu/multicsd/languages/Japanese.

Kaiser, S. (1994). *Japanese Language II* (Vol. 3, Japanese Correspondence Course for JET Participants 1994-1995). Tokyo: PRINTX.

MacMillan Education ELT. (2011, March 5). *Introduction to teaching pronunciation workshop - Adrian Underhill (complete)*. [Video File]. Retrieved from https://youtu.be/1kAPHyHd7Lo.

Miyagawa, S. (2012). *The Oxford handbook of Japanese linguistics*. Oxford: Oxford University Press.

Miyamoto, A. (2019, October 8). *'Japanglish': why do Japanese have difficulty speaking English?* Retrieved February 22, 2021, from https://japan-forward.com/japanglish-why-do-japanese-have-difficulty-speaking-english/.

Mutonono, Elena. (2016, July 20). *Pronunciation warm-up: useless and useful activities*. Retrieved February 19, 2021, from https://www.lessonplansdigger.com/2016/07/20/pronunciation-warm-up-useless-and-useful-activities-guest-post-by-elena-mutonono/.

Rachel's English. (2014, March 14). *Esl pronunciation exercise: free time — American English*. [Video File]. Retrieved from https://youtu.be/NaJIb2y6MQA.

Shibatani, M. (2005). *The languages of Japan*. New York: Cambridge University Press.

Woodward, R. (2021, March 3). *A Contrast of Japanese and English Phonology*. Horizons University.