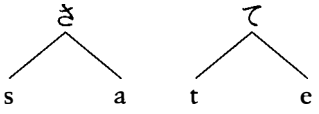
2 Phonetics

When we try to list all the sounds in a language, it is important not to confuse orthography, i.e. the writing system, with the actual sounds. For example, the sound [k] in English can be exemplified by various spellings, as in *kiss*, *sick*, *choir*, *quit*, *cow*, *Iraq*, and *unique*. (Note: The brackets "[ ]" are used when representing the pronunciation of each sound. This symbol will be further explained in chapter 3.) Even though we are looking at seven different ways of spelling here, we are nonetheless referring to the single sound [k].

Confusing spelling and actual sounds is particularly problematic in Japanese because, as figure 2.1 (p. 6) shows, the majority of the Hiragana syllabary – one of the writing systems used for Japanese – consists of the combination of two independent sounds, i.e. a consonant and a vowel.

Take さ and て for instance.

(1)



The さ stands for the consonant [s] and the vowel [a] in that order; and て stands for the consonant [t] and the vowel [e] in that sequence. Thus, the Hiragana syllabary is unable to isolate the individual sounds of Japanese and does not reflect the phonetic inventory (i.e. the list of sounds available) of the language.

Another option for listing Japanese sounds would be to consider utilizing the specific Romanization system that has been developed for Japanese in order to isolate the individual sounds. However, even this system does not necessarily reflect the actual sounds. Consider the series in (2), which is part of a commonly used Romanization system for Japanese (i.e. the "Kunrei"-style Romanization).[1]

An Introduction to Japanese Linguistics, Third Edition. Natsuko Tsujimura.

あ	い	う	え	お
a	i	u	e	o
か	き	く	け	こ
ka	ki	ku	ke	ko
さ	し	す	せ	そ
sa	si	su	se	so
た	ち	つ	て	と
ta	ti	tu	te	to
な	に	ぬ	ね	の
na	ni	nu	ne	no
は	ひ	ふ	へ	ほ
ha	hi	hu	he	ho
ま	み	む	め	も
ma	mi	mu	me	mo
や		ゆ		よ
ya		yu		yo
ら	り	る	れ	ろ
ra	ri	ru	re	ro
わ				
wa				
を				
o				
ん				
n				

Figure 2.1 Hiragana syllabary[2]

(2) a. ta
b. ti
c. tu
d. te
e. to

Under this Romanization system, the same consonant is shared by each of the five in (2a–e), and the only difference is supposed to be the vowel that accompanies it. However, the consonant that is represented by t in (2a–e) is not pronounced

in the same manner. The consonant *t* in (2a, d, e) is a sound similar to the t sound in the English words *top*, *ten*, and *tone*; the consonant in (2b) is similar to the sound represented by ch as in the English word *cheese*; and the consonant in (2c) is pronounced very much like the sequence ts as in the English word *cats*.

This illustration of the Hiragana syllabary and the Romanization system clearly shows that writing systems do not correlate with the phonetic inventory of the language, and we need a system that enables us to describe a sound as it is pronounced. This is why we focus on *spoken* language rather than *written* language when we investigate the phonetic system of a language. To avoid the confusion that we have demonstrated above, we use phonetic symbols to transcribe the sounds that exist in a language.

1 Phonetic Inventory

Before describing the sounds of Japanese, it is helpful to understand the general mechanism of how speech sounds are produced. While there exist some sounds in the world's languages that are produced as we inhale, most sounds are produced as air flows from our lungs. The stream of air, which originates in the lungs, travels through the **trachea** – commonly called the windpipe – and then reaches the **larynx**, less technically known as the Adam's apple. The larynx is an important speech organ because this is where the **vocal cords** (or vocal folds) are found: depending on the nature of the vocal cords' activities, different types of sounds are produced. For instance, place your fingers at the Adam's apple, and sound out each of [s] and [z] continuously. The presence and absence of vibration you feel at your throat is attributed to the activities of the vocal cords. The vocal cords are made of two muscular folds, and when they are pulled together leaving a slight opening, the air from the lungs tries to escape through the narrow opening of the vocal folds, which leads to their vibration. This is what you feel when [z] is pronounced. In contrast, when the vocal cords are open, creating a space referred to as the **glottis**, air flows freely without causing their vibration. This is the situation with [s]. The presence or absence of the vibration of the vocal cords, as is reflected by the difference between [z] and [s] in our example, is called **voicing**: voiced sounds like [z] are produced with vocal cord vibration while voiceless speech sounds like [s] are produced without such vibration. Figure 2.2 illustrates the contrastive states of the vocal cords.

Once the air from the lungs passes through the larynx into the mouth, a variety of sounds are produced, and they are generally divided into two groups, consonants and vowels. Consonants are sounds made with an obstruction in the mouth, while no such obstruction occurs with vowels.

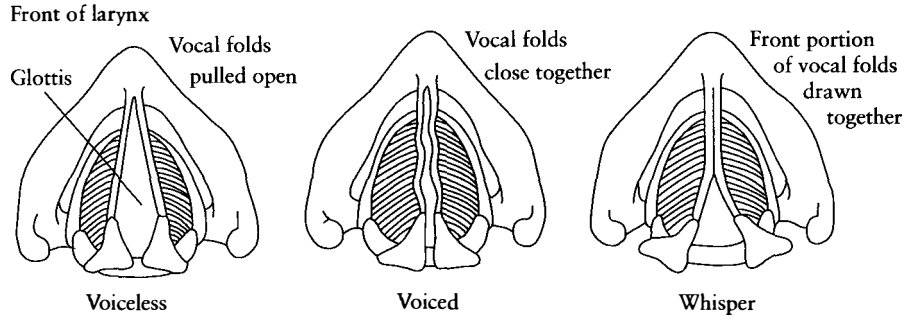


Figure 2.2 Vocal cords (O'Grady et al. 1989: 17)

1.1 Consonants

We have already discussed that voicing is important in distinguishing sounds. There are two additional factors that help us to isolate one sound from another. They are **place of articulation** and **manner of articulation**. With reference to the anatomical structure of the mouth, place of articulation indicates the place in the mouth where the sound is made. The lips and the various regions along the roof of the mouth, as is illustrated in figure 2.3, are all places of articulation. Different sounds are made at different places of articulation by placing articulators such as the tongue at these different locations.

Manner of articulation refers to how the articulators, such as the lips and the tongue, achieve contact with the places of articulation. For instance, the sound produced when an articulator has complete contact with a place of articulation and the sound produced with partial contact are different.

In addition to these three factors, it is also important to keep in mind whether the airflow from the lungs passes through the oral cavity (the mouth) or through the nasal cavity. When the velum is raised, the passage through the nasal cavity is completely blocked, and the air from the lungs goes out through the oral cavity. Sounds made in this way are called **oral** sounds. When the velum is lowered, on the other hand, the airflow from the lungs travels through the nasal cavity as well as into the oral cavity. The sounds created in this manner are referred to as **nasal** sounds.

Thus, consonants in a language are described by place of articulation, manner of articulation, voicing, and whether they are oral or nasal. A specific combination of these factors identifies a single sound to which a unique phonetic symbol is assigned.

1.1.1 Stops

The first group of sounds that we will consider is one that is characterized by its manner of articulation, namely, **stops**. Stops are further divided into two types,

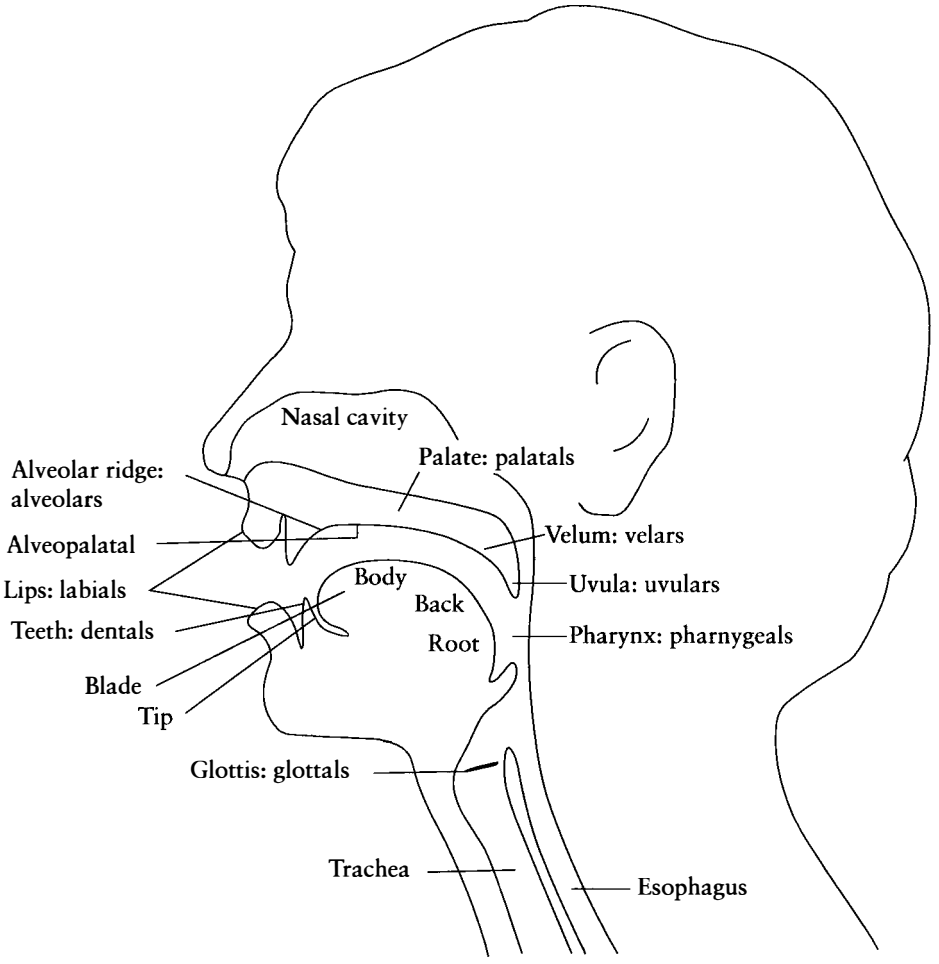


Figure 2.3 Sound-producing system (O'Grady et al. 1989: 19)

oral and nasal. Oral stop sounds – often referred to simply as stop sounds or plosives – are produced when the airflow originating from the lungs and coming through the oral cavity, or mouth, is completely blocked. What separates oral stops from nasal stops is that the former occur when the velum is raised, preventing airflow escaping through the nasal cavity. Nasal stops are articulated when a complete obstruction takes place in the oral cavity and the air goes through the nasal cavity because of the lowered velum. Nasal stops will be discussed in section 1.1.6 below.

The location of complete air blockage in the oral cavity reflects the place of articulation. When the blockage is made at the lips by placing the upper lip and lower lip together, a bilabial stop is produced. Furthermore, when the **bilabial** stop is accompanied by vocal cord vibration, a **voiced** bilabial stop is produced,

and we transcribe it as [b]. The voiced bilabial stop in Japanese is exemplified by words such as *binboo* "poor" and *kooban* "police box". On the other hand, when the vocal cords do not vibrate, a **voiceless** bilabial stop is articulated. We represent this sound as [p]. Examples of words with the voiceless bilabial stop [p] are *pan* "bread" and *denpoo* "telegram". The bilabial stops in English are found in words like *bird* and *label* for the voiced ones and *pen* and *pop* for the voiceless ones.

A stop sound made at the alveolar ridge, where the airflow would be completely blocked by the blade of the tongue at the alveolar ridge, is referred to as an **alveolar** stop. The voiced alveolar stop is represented as [d] and its voiceless counterpart as [t]. They are exemplified by *doko* "where" and *denwa* "telephone" for the former and *tensai* "genius" and *tokidoki* "sometimes" for the latter. The alveolar stops in English words like *sad/adapt* and *tiny/pet* are typically produced by the tip of the tongue, rather than the tongue blade, contacting the alveolar ridge. That is, [t] and [d] in Japanese are pronounced slightly further forward than those in English with the tongue tip almost touching the back of the upper teeth (cf. Vance 1987, 2008). Despite this difference, however, we use the same set of phonetic symbols of [d] and [t] to represent the alveolar stops in both languages.

When the air obstruction takes place by raising the tongue body to the velum or soft palate, **velar** stops are produced, and sounds such as the first consonant of words like *gakusei* "student", *gojuu* "fifty", *konbu* "seaweed", and *kinoo* "yesterday" are articulated. The first two examples contain a voiced velar stop [g], and the second a voiceless velar stop [k]. The voiced and voiceless velar stops appear in English words *agree* and *cat*, respectively.

Although the stop sounds are roughly the same in Japanese and English, a slight difference can be detected with the voiceless members [p, t, k]. When these voiceless stops occur at the beginning of a word or syllable in English, they are accompanied by **aspiration**, i.e. a puff of air. This can be felt by placing a piece of paper in front of your mouth and then trying to pronounce the English word *pin*. The initial consonant should be accompanied by a puff of air, and hence is aspirated, as is indicated by the movement of the paper. When a native speaker of Japanese pronounces the word *pan* "bread", by contrast, we notice that the paper does not move as much as in English. This is because the degree of aspiration in the Japanese voiceless stops is not as large as in English. Even these sounds, however, can be aspirated, especially when they are pronounced with great emphasis (cf. Vance 2008).

1.1.2 Fricatives

Fricatives are sounds produced when the airflow in the oral cavity is forced through a narrow opening in the vocal tract, so that air turbulence is generated, resulting in a friction noise. The friction noise sounds different depending on where the narrow opening occurs. To start with fricative sounds that are created at the front part of the mouth, there is a voiceless **bilabial** fricative sound, which

is similar to the sound that is made when one blows out a candle but without much lip protrusion. There is a narrow opening between the lips, and the air from the lungs escape through them. This sound occurs as the first sound in Japanese words *hurui* "old" and *hukai* "deep", and is represented by the symbol [ɸ]. There is no voiced bilabial fricative in Japanese; and there is neither voiced nor voiceless bilabial fricative in English.

There are a few fricative sounds that exist in English but are missing in Japanese: they are **labio-dental** and **interdental** fricatives. The first consonants of English words like *fin* and *vase* are produced when the upper teeth and the lower lip achieve contact. These sounds are referred to as labio-dental fricatives. The air passes between the upper teeth and lower lip, creating friction noise. The voiceless labio-dental fricative is [f] and its voiced counterpart is [v]. A partial blockage of the airflow can also occur when the tip of the tongue is between the upper and lower teeth. Since the sound is made between the teeth, these fricative sounds are called interdental. The voiceless and voiced interdental fricatives are [θ] and [ð], respectively, and they are exemplified by *three/truth* and *they/smooth*.

The first consonants of Japanese words like *zoo* "elephant" and *san* "three" are fricative sounds. These sounds are produced when we try to create a narrow opening by keeping the tip of the tongue very close to the alveolar ridge. They are called **alveolar** fricatives. The voiced alveolar fricative as in *zoo* "elephant" is transcribed as [z], while the voiceless counterpart, as in *san* "three", is represented by the symbol [s]. Thus, the voiceless alveolar fricative [s] and the voiceless alveolar stop [t] share the same place of articulation, alveolar, and voicing feature, voiceless; but they differ in manner of articulation. Similarly, [z] and [d] share the same place of articulation and voicing, but the former is a fricative and the latter is a stop. That is, they have different manners of articulation. Voiced and voiceless alveolar fricatives in English are exemplified by the respective first sound of the words *song* and *zero*.

When a fricative sound is made with the blade of the tongue just behind the alveolar ridge, the sound is characterized as a **palato-alveolar** or **alveo-palatal** fricative. The voiceless alveo-palatal fricative is represented as [š], and is exemplified by the first sound in Japanese words like *sinbun* "newspaper" and *sika* "deer". The same set of sounds is found in the first sounds of English words *shoe* and *shine*. The voiced alveo-palatal fricative is [ž], and is not usually found in Japanese.[3] There are not many words that have this sound in English either, but it can be exemplified by the middle consonant of *vision* and the second consonant of *measure*. For many English speakers, the alveo-palatal fricatives in English are accompanied by rounding of the lips.

When the middle of the tongue behind the blade, i.e. the tongue body, approaches the middle of the roof of the mouth, called the **hard palate** or just **palate**, creating a narrow passage, friction sounds are produced. They are **palatal** fricatives. Japanese has only the voiceless palatal fricative [ç], as in the words *hiroi* "spacious" and *hitori* "one person". This sound is similar to the sound in the German word *ich* "I" or to the first sound in the English word *huge* as pronounced by many native speakers of American English.

Finally, when the air is partially blocked at the narrow opening between the vocal cords (i.e. the glottis), it creates friction and a glottal fricative sound is produced. The **glottal** fricative is represented as [h], and can be found in the words *hanbun* "half" and *hosoi* "thin" for Japanese and *heart* and *hotel* for English. There is no vibration of the vocal cords in pronouncing [h], and hence the glottal fricative is voiceless.[4] Careful attention should be paid to the Romanization representations of the glottal fricative [h] on the one hand and bilabial and palatal fricatives, [ɸ] and [ç], on the other. The Japanese words that include these sounds are uniformly written in Romanization with *h*, as in *haru* "spring", *huro* "bath", and *hirune* "nap", but the actual pronunciation that corresponds to *h* in each of these words has a distinct sound quality that is identified by different places of articulation. The letter *h* in *haru* is glottal, *h* in *huro* is bilabial, and *h* in *hirune* is palatal. This is why we must focus on a sound as it is actually pronounced rather than as it is written when we consider the range of sounds available in a given language.

1.1.3 Affricates

When we pronounce English words such as *church* and *judge*, we notice that the first and the last consonants of each word are produced by the combination of a brief stop followed by a fricative. At first, the air is blocked at a designated place of articulation, but then it is released with a partial blockage. A sequence of a stop immediately followed by a fricative is called an **affricate**. Japanese has four affricate sounds: the alveolar [t^s] (voiceless) and [d^z] (voiced), and the alveo-palatal [č] (voiceless) and [ǰ] (voiced). The voiceless alveolar affricate is made with a short alveolar stop that is released into an alveolar fricative. The voiceless alveolar affricate [t^s] is exemplified in Japanese by the first sounds in the words *tumi* "sin" and *turi* "fishing". The voiced alveolar affricate, [d^z], does not seem to display a clear contrast with the alveolar fricative, [z], in the pronunciation of most Japanese people (cf. Vance 1987; Shibatani 1990). For example, the pronunciation of the second consonant of *mazusii* "poor" and that of the third consonant of *mikazuki* "crescent moon" varies between [d^z] and [z], with perhaps a slight preference for the voiced alveolar affricate. In pronouncing the word *tizu* "map", on the other hand, the second consonant for many speakers is invariably the alveolar fricative [z]. The alveo-palatal affricates, [č] and [ǰ], are exemplified by the first sounds in the words *tikaku* "near" and *zikan* "time", respectively. English has these two alveo-palatal affricate sounds. The voiceless one, [č], is exemplified by the consonants in *church*; and its voiced counterpart, [ǰ], is found in *judge*. Associated with the pronunciation of these English affricate sounds is a slight rounding of the lips.

1.1.4 Liquids

In comparing the first sounds of the English words *sun* and *run*, we notice that the two sounds are made at the same area of the mouth, namely, at the alveolar

ridge. In both pronunciations, the tip of the tongue approaches the alveolar ridge and furthermore the airflow from the lungs continuously escapes through the space between the tongue tip and the alveolar ridge. The major difference between the two sounds, however, is the degree to which the obstruction of the airflow is created: the passage between the tongue tip and the alveolar ridge is narrower in the voiceless alveolar fricative [s] in *sun* than in the first sound of *run*. The latter sound is produced with constriction in the vocal tract, but the air flows more freely from the mouth with a lesser degree of blockage of air or friction than in fricative sounds. The sounds that are articulated in such a manner are called **liquids**. English has two liquid sounds, [l] and [r], both alveolar, and the initial consonants of *land* and *run* exemplify each. The difference between the two is that with [l] the air channel is on the sides of the tongue while with [r] it is in the middle of the mouth. English [l] is normally pronounced with the tongue tip touching the alveolar ridge, but with the sides of the tongue lowered. The air flows freely over one or both sides of the tongue. Consequently, [l] is referred to as a lateral sound. The American English [r] sound is often made with the tongue tip curled back. Both [l] and [r] are accompanied by vibration of the vocal cords, and are voiced.

Japanese has one liquid sound, the voiced alveolar liquid, as is exemplified by the words *roku* "six" and *ringo* "apple". Although the sound in *roku* and *ringo* generally share the place and manner of articulation with English [l] and [r], the initial sound of these Japanese words is quite different from [l] or [r] in English. The liquid sound is technically called the alveolar tap, represented by the phonetic symbol [ɾ], and is produced by placing the tongue tip at the alveolar ridge followed by an immediate release of that contact. The alveolar liquid in Japanese sounds very similar to the "d" sound in English words like *tidy* and *steady* in the American English pronunciation (technically called flap): with both sounds, the tongue achieves very rapid contact at the alveolar ridge. This similarity between English and Japanese explains why native speakers of English learning Japanese are often unable to make a distinction between [ɾ] and [d] in Japanese words. For the sake of simplicity, we will use the transcription symbol [r] for the Japanese alveolar liquid sound throughout this book.[5]

1.1.5 Glides

Glides are sounds that have characteristics of vowels and consonants, and are often called semivowels. Like vowels, they are produced without (much) obstruction in the mouth. They, however, function like consonants in that they occur at the beginning of a syllable immediately before a vowel. There are two glides in Japanese, as is exemplified by the first sounds of *wakaru* "understand" and *yasui* "cheap". In the articulation of the first sound of *wakaru*, the body of the tongue approaches the velum, so the place of articulation is velar and is represented as [w]. The first sound of *yasui* is made by the tongue body approaching the hard palate; that is, it is the palatal glide and is transcribed as [y]. These two glide

sounds are also accompanied by vocal cord vibration, and are, therefore, voiced. English also has two glides, [w] and [y], as in *wish* and *yield*, respectively. The English [w], however, is accompanied by lip rounding while the Japanese [w] is not generally observed with the same type of lip movement, at least in normal speech. This difference is captured by identifying the glide as the labio-velar glide for English and the velar glide for Japanese. Although there are phonetic symbols that strictly represent the difference in the lip rounding, we will use [w] for the glide in both languages.

1.1.6 Nasals

In producing a nasal consonant, the velum is lowered and a complete obstruction occurs in the oral cavity. The airflow, however, passes freely through the nasal cavity. The sounds produced in this way are **nasal stops** or simply **nasals**. The nasals are like the oral voiced stops [b, d, g] in that they are voiced and are produced with a complete blockage in the oral cavity. There are three nasal sounds in Japanese: [m], [n], and [ŋ]. The first one, [m], is a bilabial nasal sound, and appears in words like *mikan* "orange" and *mame* "beans". The second nasal sound, [n], is alveolar, and is found in *neko* "cat" and *naka* "inside". Finally, the occurrence of the velar nasal [ŋ] within a word depends on the speaker. For example, some speakers, especially older speakers, pronounce the "g" in the word *kagaku* as [ŋ], while others pronounce it as the velar stop [g]. Even when a speaker consistently uses the velar nasal sound in her or his speech, however, it never appears in word-initial position in Japanese. Thus, the initial consonant of the word *gakkoo* "school" can only be pronounced as [g].

Some works describe the nasal sound before a pause, in words like *yon* "four" and *ken* "ticket", as being pronounced with the tongue body touching the uvula (cf. Vance 2008). The uvular nasal is represented as [ɴ]: the words mentioned above are, thus, transcribed as [yoɴ] and [keɴ] before a pause. Sometimes a nasal before a pause may make no contact with the root of the mouth. Such a nasal could be described as a nasalized glide.

The same set of nasal sounds found in Japanese occurs in English. The bilabial nasal [m] is exemplified by the first sound in the word *mother*; the alveolar nasal [n] occurs at the first and last consonants in the word *nun*; and the final consonant of the word *sing* is an example of the velar nasal [ŋ]. The velar nasal does not occur in word-initial or syllable-initial position in English.

Nasal sounds in Japanese present an illustration of a phonetic phenomenon called **coarticulation**. Coarticulation occurs when place of articulation of one sound extends to a neighboring sound and as a result the pronunciation of adjacent sounds overlaps. Consider the following examples. (The label "Nom" is the abbreviation of Nominative Case particle.)

(3) a. ken made [kem made]
ticket even

b.	ken	desu	[ken̲ desu]
	ticket	is	
c.	ken	ga	[keŋ̲ ga]
	ticket	Nom	

Careful attention should be paid to the pronunciation of the nasal sound at the end of the word *ken* "ticket". Its pronunciation depends on the nature of the consonant that immediately follows it: when the bilabial nasal [m] follows the word *ken*, as in (3a), the sound at the end of *ken* is pronounced as the bilabial nasal; when the voiced alveolar stop [d] appears immediately after *ken*, as in (3b), the final sound of the word is also pronounced as the alveolar nasal; and when the voiced velar stop immediately follows *ken*, as in (3c), the velar nasal surfaces at the end of the word. In these examples, the place of articulation of the immediately following consonant, i.e. [m, d, g], extends to the preceding nasal consonant, so that the nasal sound is pronounced with the same place of articulation. That is, coarticulation accounts for the variety of the nasal sounds observed with the pronunciation of the last consonant of *ken* "ticket" in (3). Nasal coarticulation is a widely recognized phenomenon in other languages and is observed in English as well. For instance, examine how the nasal sound in the negative prefix *in-* is produced when it is added to the adjectives *possible*, *sufficient*, and *complete*, generating *impossible*, *insufficient*, and *incomplete*. We notice that the place of articulation of the nasal sound differs in each word: it is bilabial in *impossible*, alveolar in *insufficient*, and velar in *incomplete*. The variation in the place of articulation observed with the identical prefix reflects the influence of the sound immediately following *in-*: [p], [s], and [k]. Interestingly, coarticulation is sometimes mirrored in the spelling of a word, as is the case with *impossible*. Another example is found in the word *illogical*: although at a first glance, it may not be easily discernible, this word consists of *logical* and the negative prefix *in-*; and the place of articulation of the nasal sound in the prefix is influenced by the following liquid sound [l], also reflected in the spelling.

Although we have discussed four nasal sounds [m, n, ŋ, ɴ] as nasal stops that are available in Japanese, we may additionally include two more, alveo-palatal [n̆] and palatal [ɲ], in connection with the coarticulation phenomenon. Examples include *ken zya (nai)* "it's not a ticket" and *ken ya (kane)* "things like tickets and (money)": in the first example, the sound that immediately follows *ken* is the voiced alveo-palatal affricate [ǰ], and the word final nasal sound in *ken* is pronounced as the alveo-palatal nasal [n̆]; and in the second example, the palatal glide [y] appears right after *ken*, causing the nasal sound of the word to be realized as the palatal nasal [ɲ]. It should be noted that while alveo-palatal and palatal nasal sounds do exist in Japanese, they surface only as a result of coarticulation rather than as sounds that freely occur irrespective of the nature of neighboring sounds. We will discuss more of this point in chapter 3.

Tables 2.1 and 2.2 show the English and Japanese consonantal systems, respectively.

Table 2.1 Summary of English consonants.

		bilabial	*labio-dental*	*interdental*	*alveolar*	*alveo-palatal*	*palatal*	*velar*	*labio-velar*	*glottal*
Stops:	[+V]	b			d			g		
	[–V]	p			t			k		
Fricatives:	[+V]		v	ð	z	ž				
	[–V]		f	θ	s	š				h
Affricates:	[+V]					h				
	[–V]					č				
Approximants:										
liquid	[+V]				r, l					
glide	[+V]						y		w	
Nasals:	[+V]	m			n			ŋ		

[+V] = voiced
[–V] = voiceless

Table 2.2 Summary of Japanese consonants.

		bilabial	*alveolar*	*alveo-palatal*	*palatal*	*velar*	*uvular*	*glottal*
Stops:	[+V]	b	d			g		
	[−V]	p	t			k		
Fricatives:	[+V]		z	(ž)*				
	[−V]	ɸ	s	š	ç			h
Affricates:	[+V]		d^z	ǰ				
	[−V]		t^s	č				
Approximants:								
liquid	[+V]		r					
glide	[+V]				y	w		
Nasals:	[+V]	m	n	(ň)	(ɲ)	ŋ	N	

[+V] = voiced
[−V] = voiceless
*See note 3 at the end of this chapter.

1.1.7 Further Notes on Consonants

In addition to palatal (and alveo-palatal) sounds, all other consonants in Japanese can be **palatalized** (cf. Bloch 1950; Vance 2008). A palatalized sound is produced by raising the tongue body toward the hard palate when a certain consonant is pronounced. We will transcribe a palatalized consonant with a superscript [^{y}]. For instance, the bilabial stops [p, b, m] can be palatalized by keeping the same place and manner of articulation while moving the tongue body closer to the hard palate, yielding [p^{y}, b^{y}, m^{y}]. Examples of palatalized consonants in Japanese include *sanbyaku* [samb^{y}aku] "three hundred", *ryokan* [r^{y}okan] "inn", *myaku* [m^{y}aku] "pulse", and *kyaku* [k^{y}aku] "guest". Alveo-palatal and palatal consonants are produced in the area of the mouth that involves the hard palate; that is, they are in a sense already palatalized. For this reason, Japanese sounds represented by the symbols [ž, š, ǰ, č, ç, y] are not further transcribed with the palatalization symbol of [^{y}]. Additionally, palatalization of alveolar fricatives [z, s] and alveolar affricates [d^{z}, t^{s}] results in alveo-palatal fricatives and affricates, respectively, i.e. [ž, š] and [ǰ, č]. This relationship suggests that the characterization and representation of sounds by way of place and manner of articulation have the advantage of capturing what is common among the sounds and how one set of sounds can relate to another set of sounds in a language. The palatalization phenomenon will be taken up again in the discussion of mimetics in chapter 3.

Japanese is one of many languages in which a single consonant and its lengthened counterpart contrast, as in [s] vs. [ss]. Long consonants, called **geminates**, are reported to exist in Arabic, Finnish, Italian, and Norwegian, among many more languages. Geminates are observed in English as well. Examples include *white tape* and *black kite* as pronounced in rapid speech, where the [t] in *white tape* and the [k] in *black kite* are each pronounced as one elongated stop consonant. While geminates in English are typically restricted to word boundaries, excluding a possibility of appearing within a single word, Japanese geminates can occur internal to a word. Japanese examples of words with geminates are contrasted with words with single consonants in (4).

(4)	a.	sakka	"author"	vs.	saka	"hill"
	b.	katta	"won"	vs.	kata	"shoulder"
	c.	assari	"simply"	vs.	asari	"clam"

A similar contrast between single consonants and geminates is found in other languages including Italian and Luganda (cf. Ladefoged 1982). For example, Italian *papa* "pope" and *pappa* "porridge, baby food" are minimally different in the length of the second consonant.

In Japanese, geminates normally occur with voiceless consonants,[6] although voiced geminate consonants can be seen particularly in loanwords like *beddo* "bed" and *baggu* "bag". Furthermore, when a word appears in isolation, i.e.

neither preceded nor followed by another word or sound, we do not see a geminate consonant at the beginning of the word; nor does the word end with a geminate, because Japanese consonants generally require vowels to follow them. That is, a geminate consonant is always preceded and followed by a vowel in Japanese. Languages with geminates vary in this respect, some allowing them in word-initial and word-final positions. For example, Cairene Arabic has words like *sitt* "lady, grandmother" and *hamm* "a worry, concern", which end in geminates; and in Moroccan Arabic, words like *bba* "father" and *ttrma* "be thrown" start with geminates.

The contrast between a single consonant and its geminate counterpart is shown by the difference in duration that is manifested in a wave form and a spectrogram. Figure 2.4 represents the contrastive sounds of (4a), *sakka* vs. *saka*, by way of wave forms (the top image) and spectrograms (the bottom image): the horizontal axis indicates the duration in milliseconds of each sound contained in the word.

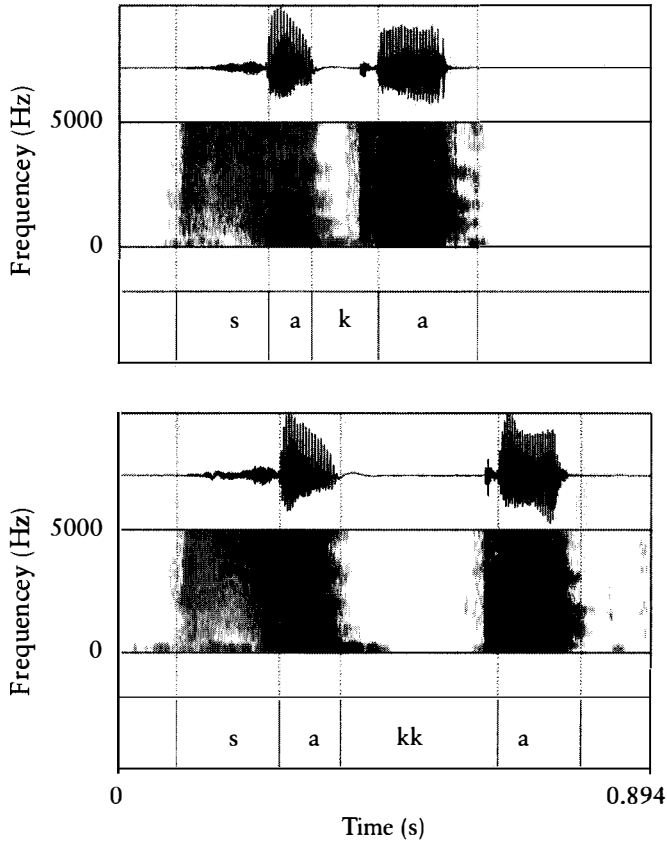


Figure 2.4 Contrast between single (*saka* "hill") and long (*sakka* "author") consonants

These images distinctively illustrate that a geminate consonant in Japanese is at least twice as long as its single consonant counterpart. The ratio between a single consonant and a geminate actually varies depending on the language: it has been reported that the ratio is roughly 1:3 in Japanese and 1:1.85 in Italian (cf. Homma 1981; Han 1992; Ham 2001; Hirata and Whiton 2005; Idemaru and Guion 2008). Interestingly, the contrast between a single consonant and a geminate consonant is further manifested in neighboring sounds. More specifically, compare the duration of the vowel before the geminate consonant and that before the single consonant in figure 2.4; similarly, compare the vowel after the geminate consonant and that after the single consonant. We see that the vowel is longer before the geminate and shorter after it. This relationship between geminate consonants and surrounding vowels has been noted in various studies on Japanese (cf. Han 1994; Idemaru and Guion 2008) and also in other languages such as Arabic and Turkish (cf. Pickett et al. 1999). Languages vary in this respect, however, as examples including Finnish, Hausa, Icelandic, and Norwegian show that the vowels preceded by geminates are shorter than those preceded by single consonants (cf. Maddieson 1985).

It was noted earlier that voiced geminate consonants occur in Japanese, although voiceless geminates are more prevalent. A number of words in which voiced geminate consonants occur are loanwords, many of which are often pronounced with voiceless geminates by Japanese native speakers (cf. Lovins 1975; Kawahara 2005, 2011b; Hirayama 2008). Some examples are given in (5).

(5) a. beddo ~ betto "bed"
b. baggu ~ bakku "bag"
c. doggu ~ dokku "dog"
d. guddo ~ gutto "good"
e. gebberusu ~ gepperusu "Göbbels" (proper name)

(Kawahara 2005)

The variation in voicing, however, is not available to just any voiced geminates: loanwords in (6), for example, contrast with those in (5) in that voiceless geminates in (6) cannot substitute voiced geminates as alternative pronunciation (cf. Nishimura 2003; Kawahara 2011b) ("*" indicates that the form is not accepted by native speakers).

(6) a. webbu ~ *weppu "web"
b. reddo ~ *retto "red"
c. eggu ~ *ekku "egg"

(Kawahara 2011b)

This contrast has been accounted for by making reference to other consonants that appear in a word: if a word contains a voiced stop, fricative, or affricate, a voiced geminate can be substituted by its voiceless counterpart; otherwise, a

voiced geminate remains the only option. For instance, all the loanwords in (5) contain voiced stops at the beginning – [b] in (5a) and (5b), [d] in (5c), and [g] in (5d) and (5e) – and the voiced geminates [dd], [gg], and [bb] can be pronounced instead with [tt], [kk], and [pp], respectively. In contrast, the words in (6) do not contain voiced stops, fricatives, or affricates: in (6a) [w] is voiced but a glide (i.e. not a stop, fricative, or affricate); [r] in (6b) is also voiced but liquid; and in (6c) there are no other consonants but the voiced geminate itself. In these cases, the alternative pronunciation with a voiceless geminate is not available. The alternation between voiced and voiceless geminates in (5) – and lack thereof in (6) – suggests the important role that the nature of neighboring sounds can play. This should be reminiscent of the coarticulation phenomenon we have discussed earlier in relation to nasal sounds. In the case of voiced geminates, however, neighboring sounds influence them in such a way that repetition of the identical sound characteristics – namely, voicing – is avoided in the alternative pronunciation, i.e. the opposite effect that coarticulation exhibits. What is shared by the two sets of phenomena – nasal coarticulation and voiced geminates – is that generalizations to explain the phenomena are logically made by invoking voicing and the manner of articulation by which sounds are identified. That is, the place and manner of articulation as well as voicing constitute helpful measures not only to represent each sound available in a language but also to describe and account for common patterns among what otherwise might seem to be a random choice of sounds.

Research on voiced geminates has increasingly advanced in recent years. In light of the fact that the voiced–voiceless alternation is possible for words like those in (5), Kawahara (2005), for one, investigates how native speakers actually pronounce voiced geminates and also how they perceive them. Results from his experimental work suggest that voiceless geminates are a more natural choice for native speakers both in production and in perception. Focusing on various acoustic differences between voiced and voiceless geminate contrasts in made-up words like *kobbo* and *koppo*, Kawahara reports that in native speakers' pronunciation of voiced geminates (e.g. *kobbo*), the acoustic properties that characterize voicing were weakened, so that voiced geminates were pronounced more in line with their voiceless counterparts (e.g. *koppo*). Furthermore, when native speakers were asked to identify words with voiced and voiceless geminates, they were able to recognize words with voiceless geminates (e.g. [pp], [tt], and [kk]) correctly at a rate of more than 95 percent, while the accuracy rate dropped to 28.7 percent for words with voiced geminates ([bb], [dd], and [gg]); the native speakers incorrectly identified voiced geminates as voiceless geminates. Additionally, of the three voiced geminates, [bb], [dd], and [gg], native speakers were able to recognize [dd] (46.6 percent accuracy) with more ease than the other two (17.9 percent accuracy rate for [bb] and 22.3 percent for [gg]). As we noted earlier, voiceless geminates are more common in Japanese, but Kawahara's study indicates that even in words with voiced geminates, native speakers show a predisposition toward their voiceless counterparts in both production and perception.