Behavior Research Methods, Instruments, & Computers
2004, 36 (3), 531-547

Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper

KATSUO TAMAOKA
Hiroshima University, Hiroshima, Japan

and

SHOGO MAKIOKA
Osaka Women's University, Osaka, Japan

On the basis of the lexical corpus created by Amano and Kondo (2000), using the *Asahi* newspaper, the present study provides frequencies of occurrence for units of Japanese phonemes, morae, and syllables. Among the five vowels, /a/ (23.42%), /i/ (21.54%), /u/ (23.47%), and /o/ (20.63%) showed similar frequency rates, whereas /e/ (10.94%) was less frequent. Among the 12 consonants, /k/ (17.24%), /t/ (15.53%), and /r/ (13.11%) were used often, whereas /p/ (0.60%) and /b/ (2.43%) appeared far less frequently. Among the contracted sounds, /sj/ (36.44%) showed the highest frequency, whereas /mj/ (0.27%) rarely appeared. Among the five long vowels, /aR/ (34.4%) was used most frequently, whereas /uR/ (12.11%) was not used so often. The special sound /N/ appeared very frequently in Japanese. The syllable combination /k/+V+/N/ (19.91%) appeared most frequently among syllabic combinations with the nasal /N/. The geminate (or voiceless obstruent) /Q/, when placed before the four consonants /p/, /t/, /k/, and /s/, appeared 98.87% of the time, but the remaining 1.13% did not follow the definition. The special sounds /R/, /N/, and /Q/ seem to appear very frequently in Japanese, suggesting that they are not *special* in terms of frequency counts. The present study further calculated frequencies for the 33 newly and officially listed morae/syllables, which are used particularly for describing alphabetic loanwords. In addition, the top 20 bi-mora frequency combinations are reported. Files of frequency indexes may be downloaded from the Psychonomic Society Web archive at http://www.psychonomic.org/archive/.

The table of fundamental Japanese sounds, which are described in phonological scripts of two types of kana symbols, *hiragana* or *katakana*, consists of 101 basic kana, including the special sound of the nasal /N/ (written in hiragana as ん and in katakana as ン). However, in modern spoken and written Japanese, it is not unusual to find some morae that are not described in the traditional kana table. One of the reasons for this drastic change in Japanese sounds has been the recent, massive adoption of alphabetic loanwords, especially from English, resulting in various new sounds in modern Japanese. Taking this new trend into consideration, a list of an additional 33 kana (or morae) was officially issued for these loanwords through a cabinet announcement by the Government of Japan (1991). Since kana is a shallow script, whereby orthographic symbols and phonological units correspond to each other on a regular basis, each single kana (occasionally two kana for contracted sounds) corresponds to the phonological unit of a *mora*. Thus, by counting kana units, the present study was able to calculate the overall frequencies of occurrence for Japanese phonological units. Through the use of a lexical corpus (Amano & Kondo,[1] 2000) produced on the basis of the *Asahi* newspaper, the present study provides type and token frequencies of occurrences (hereafter, if not clearly mentioned, *frequency* refers to *frequency of occurrence* or *frequency counts*) for the units of phonemes, morae, and syllables.

The frequency counts for phoneme, mora, and syllable units are useful for estimating the phonological familiarity of a Japanese word by adding or averaging frequencies involved in the target word. For example, the Japanese writing system includes three scripts: kanji, hiragana, and katakana. Although each word is more likely to be written in one script than in another, it is possible to transcribe the target word into other scripts. By transcribing words into unfamiliar scripts, the orthographic familiarity will be severely reduced. Using this process, Besner and Hildebrandt (1987) found that loanwords normally written in katakana were named more quickly than both nonwords and kanji compound words presented in katakana. However, since the same words in different scripts are still pronounced in the same way, it is quite possible that phonological familiarity (frequency in sounds) could be independent of the orthographic familiarity (frequency in

The authors express their gratitude to Takayuki Arai and Yasushi Hino for their valuable assistance in reviewing this article. Correspondence concerning this article should be addressed to K. Tamaoka, International Student Center, Hiroshima University, 1-1, 1-chome, Kagamiyama, Higashi-Hiroshima, 739-8523 Japan (e-mail: ktamaoka@hiroshima-u.ac.jp).

print) of the target words. Likewise, Tamaoka and Murata (2001) indicated that CVCVCV-structured nonwords with the same vowel repetitions (e.g., /kohomo/) were processed more slowly and less accurately than those with no repetition (e.g., /kohami/). This tendency is called *vowel repetition effects*. However, Tamaoka and Murata did not use the phonological frequencies of phonemes and morae as a control condition in their experiments. Previous studies involved with the processing units of Japanese phonology (e.g., Cutler & Otake, 1994; Otake, Hatano, & Yoneyama, 1996; Otake, Hatano, Cutler, & Mehler, 1993; Tamaoka & Terao, 2004) and with mora timing (e.g., Han, 1992; Sugito, 1989; see Vance, 1987, and Warner & Arai, 2001, for literature reviews) also have not provided any frequencies of occurrence for the morae unit when constructing target words and nonwords.

As can be seen from these works, to develop more precise experimental conditions, frequency counts for the phonological units of phonemes, morae, and syllables will have to make a greater contribution to various studies in psychology, linguistics, psycholinguistics, and other related areas in estimating phonological familiarity.

Changes of Japanese Sounds and Spellings

In the history of the Japanese sound palette, a great alteration occurred through the adoption of various vocabularies from different languages. In the Nara period (A.D. 710–784), Japanese sounds were fundamentally constructed with a consonant and a vowel (Koizumi, 1990; Komatsu, 1981; Kubozono, 1999; Kubozono & Ota, 1998; Numoto, 1987). There was no need to distinguish between syllables and morae, because they were identical in this period. The first alteration came as a result of the influence of the Chinese language. Three special sounds—the nasal /N/, the geminate (or voiceless obstruent) /Q/, and the long vowel /R/—were created in the process of adopting Chinese words into the Japanese language. These sounds, which were regarded as independent morae, created a distinctive difference between a mora and a syllable. Sound sequences in which these sounds occur, such as /kaN/, /kaQ/, and /kaR/, are counted as two morae, but only a single syllable.

More recently, additional sounds have come from the adoption of alphabetic languages. These loanwords contained many sounds that could not be easily described in the traditional Japanese sound system. As a result, the Japanese developed ways to present the specific sounds of loanwords, using katakana. Kokuritsu Kokugo Kenkyujo [or the National Institute for Japanese Language] (1983) found 162 loanwords that could be written in two ways. Endo (1989) has provided a variety of examples, such as アクセサリー /akusesariR/ and アクセサリ/akusesari/ for *accessories*, ウェーブ /weRbu/ and ウエーブ /ueRbu/ for *wave*, ウィンドー /wiNdoR/ and ウインドウ /uiNdou/ for *window*, アマチア /amacjia/[2] and アマチュア /amactjua/ for *amateur*. The introduction of dual (or even triple) spellings into the Japanese vocabulary has been accelerated, especially in the areas of sports, computers, fashion, and the sciences. Even vocabulary entries in various Japanese dictionaries have a variety of katakana spellings for loanwords (Endo, 1989). As such, in daily Japanese publications, various katakana spellings have been used to describe the same loanwords beyond the limits of the basic 100 morae and three special sounds (see Appendix A).

Responding to this massive adoption of alphabetic loanwords, the Government of Japan (1991) officially declared a list of 33 additional morae (see Appendix B) to be used for presenting these loanwords. This administrative decision implies an official acceptance of katakana multiple spellings for alphabetic loanwords. Except for /wu/ (ヴ), the remaining 32 morae are all contracted sounds, such as /fo/ in フォークfor *fork*, /kwo/ in クォーツ for *quartz*, and /vwe/ in ヴェール for *veil*, which were created to describe more accurately the sounds of foreign words. The type and token frequencies of 100 basic morae, the special sounds of /N/, /Q/, and /R/, and 33 additional morae (a total of 136 morae) are outlined in Appendixes A and B.

Japanese Sound Structure

Japanese phonemes, morae, and syllables can be portrayed in a single hierarchical structure (Haraguchi, 1996; Kubozono, 1989, 1995, 1999, 2002; Kubozono & Ota, 1998; Terao, 2002). As is shown in Figure 1, the lowest phonological level is the phonemic level, representing consonants (C), semivowels (S), and vowels (V). The next level is the moraic level. The first mora (μ_1) is constructed from four different combinations: a single vowel (φV, such as /a/, /e/, and /o/), a consonant and a vowel (CV, such as /ka/, /se/, and /mu/), a semivowel and a vowel (SV, such as /ja/, /ju/, and /wa/), and a consonant, a semivowel, and a vowel (CSV, such as /tja/, /pjo/, and /mju/). The second mora (μ_2) is represented by four special sounds: two consonants, represented by /N/ or /Q/, and two vowels represented by /R/ and /J/. The dual vowels, such as /ae/, /oi/, and /ou/, are described by a vowel plus /J/.

The CV-structured morae are also considered to be *light syllables*; in this sense, these phonetic combinations are classified as syllables, as well as morae. The highest level in the figure is the syllable level, which involves a combination of the first and the second morae ($\mu_1 + \mu_2 = \sigma$). These combinations are termed *heavy syllables*. For example, a CVN syllable /kaN/ consists of three phonemes (/k/, /a/, and /N/), two morae (/ka/ and /N/), and one syllable /kaN/. It is also possible to create a long single syllable, such as CSVRN, as in /tjaRN/, and CSVRQ, as in /pjuRQ/ (see more details in Matsuzaki, 1994), but frequencies related to combinations of two special sounds are excluded in this study, due to the rarity of their occurrence. On the basis of these classifications, the present study calculated frequencies of phonemes, morae, and syllables.

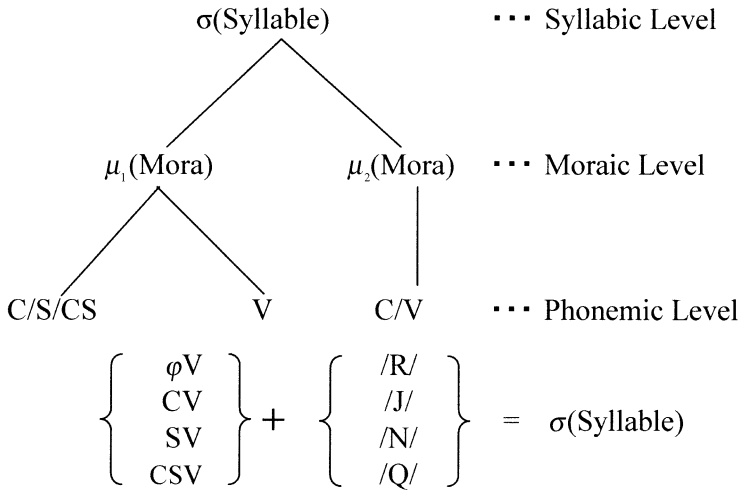

Figure 1. Phonological structure of Japanese syllables.

Kana Type and Token Frequencies

As a result of the word printed frequency index created by Amano and Kondo (2000), a very large lexical corpus of 341,771 words was established from newspapers containing 287,792,797 words, all of which were taken from the *Asahi* newspaper printed from 1985 to 1998. At present, this is the largest and the most up-to-date word corpus created from calculating the word frequency of occurrence in Japanese written texts. Thus, the present study used this database to calculate frequencies of phonemes, morae, and syllables. The programming language of MacJPerl 5.15r4J for Macintosh was used to run a calculation procedure.

When referring to frequency of occurrence, two different frequencies are used: type and token. Type frequency counts a word once. For example, according to the database (Amano & Kondo, 2000), /keitai/ has four distinguished lexical meanings: 6,457 times for 携帯 (*portable*), 3,980 for 形態 (*form*), 58 for 継体 (*succession*), and 1 for 敬体 (*polite form*). Whereas the value of the token frequency for /keitai/ is 10,496, the value of type frequency for this sound remains as 4. Likewise, a mora /ke/ is counted 4 times for type frequency, whereas token frequency yields a value of 10,496 for /ke/. A phonemic (a single V) or a moraic (*φ*V type) frequency of /i/ for the sound /keitai/ is counted 8 times (/i/ phonemes twice in each of four homophonic words) for type frequency and 20,992 for token frequency. It is interesting to compare frequency counts in print and sound. The /keitai/ for *succession*, written with two kanji (継体), has a very small value of 4 for token frequency in print, but a very large value of 10,496 for token frequency in sound. As such, this word is very unfamiliar in the written form but is extremely familiar in the phonological form.

In order to investigate the differences in frequency of occurrence, the type and token frequencies for 101 basic morae (or kana), are plotted in Figure 2. A majority of morae are within a range of 0–5,000,000 for token frequency and 0–10,000 for type frequency. The largest value of token frequency was given to イ (/i/, *φ*V type). The largest value of type frequency was given to ウ (/u/, *φ*V type). It is rather surprising that the special sound /N/ appeared to be very high in both type and token frequencies. As was explained for the alteration of Japanese sounds, the /N/ sound appeared in the process of adopting various Chinese words, so that it is often included in many Japanese words (see details in Kubozono & Ota, 1998; Matsuzaki, 1994; Numoto, 1987). Although the /N/ sound is identified as *special*, it is often seen in Japanese words. Vowels of /o/ and /a/, which are regarded as the first to be acquired among Japanese children, do not appear as often as an independent kana (mora) unit of *φ*V.

As is shown in Figure 2, Pearson's correlation was significantly high between the two frequency counts ($n = 101$; $r = .908$, $p < .01$). This result indicates that the frequency indexes of token and type for 101 morae exhibit almost no difference. A linear regression line is drawn on a plotted graph ($y = 485.15x - 140{,}755$, $R^2 = .824$). The resulting correlation was calculated on the basis of mora units presented by kana symbols, so that phonemic frequencies are not taken into consideration. In the present study, token frequency is used to calculate frequency counts of phonemes, morae, and syllables. Thus, *frequency of occurrence* refers hereafter to token frequency, unless otherwise specified.

Five Vowels (*φ*V) and (+V)

Simply speaking, vowels and consonants are distinguished by whether or not they can function as independent morae/syllables. According to this definition, there are five vowels in Japanese: /a/, /i/, /u/, /e/, and /o/. When they are independent morae/syllables, their phonological structure is described as *φ*V. Since semivowels /w/ and /y/ cannot be independent morae/syllables by themselves and share characteristics of consonants, they

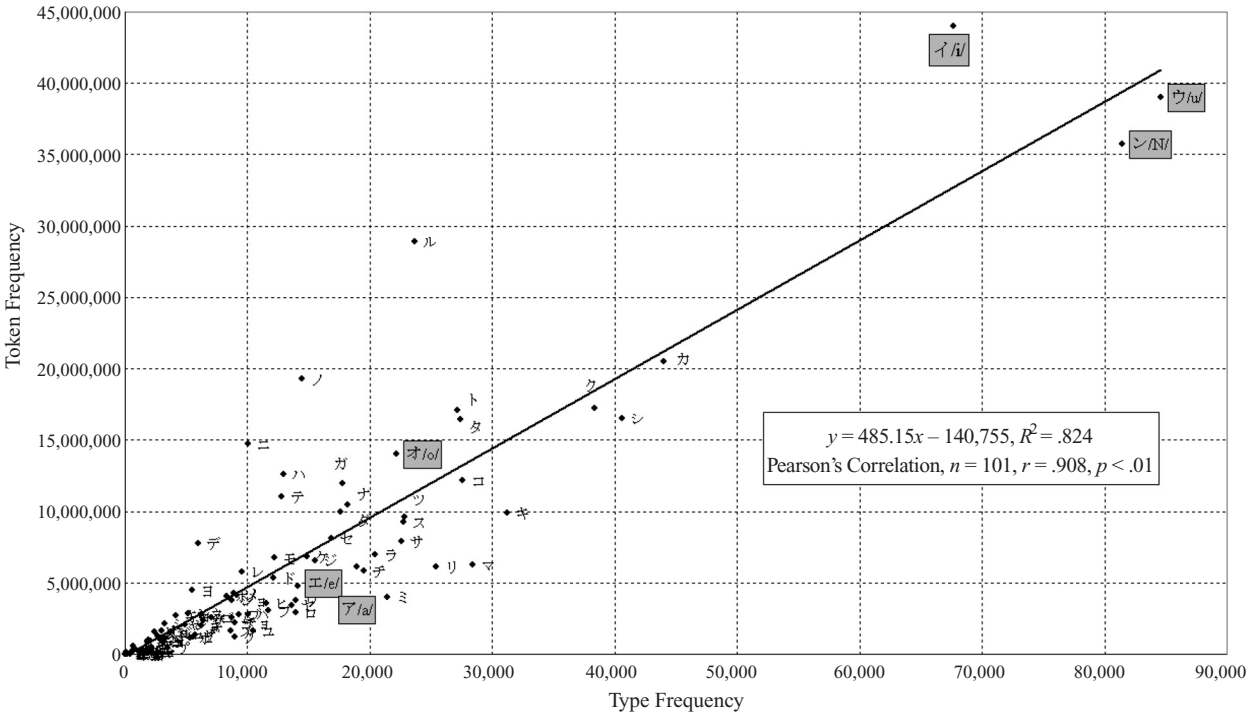

Figure 2. Plotting of mora type and token frequencies.

are categorized slightly differently than regular vowels. The special sounds /N/, /Q/, /R/, and /J/ are, however, considered as single morae.

Table 1 shows frequency counts of the five Japanese vowels as single moraic (kana) units. Among them, /i/ was the most frequently used vowel in the form of φV, having a frequency value of 43,985,426 times, or 40.72%. This may be caused by the ending of adjective inflections, which are very often a single phoneme and a mora /i/, such as /nemui/ (*sleepy* in English), /samui/ (*cold*), and /amai/ (*sweet*). The second most frequently used vowel was /u/, having a frequency value of 39,052,254 times, or 36.16%. Again, this may be caused partly by verb inflections, such as /utau/ (*to sing*), /arau/ (*to wash*), and /warau/ (*to laugh*). The least frequently used vowel was /e/ at 4.41%.

However, when the frequencies of the possible appearances of vowels (+V) are counted, all five vowels seem to have a much smaller range from 58,172,645 occurrences, or 10.94%, for /e/ to 124,790,724 times, or 23.47%, for /u/. The vowels /a/ and /o/ were less frequently seen in the φV structure, but in the +V structure their frequency increased to 23.42% for /a/ and 20.63% for /o/. Therefore, once all sound combinations were considered, /a/, /i/, /u/, and /o/ showed similar frequencies of occurrence. Yet /e/ remained the least frequently used vowel among the five.

Consonants (C) and Their Combinations With Vowels (CV)

As can be explained by the changes in Japanese sounds, the combinations of a consonant and a vowel are considered as typical Japanese morae, evident in the large total frequency of 386,480,338. There are 12 consonants in the CV structure shown in Table 2. Among them, /k/ was the most frequently used consonant, totaling 66,644,942 times,

Table 1
Frequency Counts of Vowels (φV and +V)

	φV Type		All Vowels	
Vowel	Count	Frequency (%)	Count	Frequency (%)
/a/	6,149,909	5.69	124,536,587	23.42
/i/	43,985,426	40.72	114,568,843	21.54
/u/	39,052,254	36.16	124,790,724	23.47
/e/	4,767,153	4.41	58,172,645	10.94
/o/	14,053,377	13.01	109,714,325	20.63
Total	108,008,119	100.00	531,783,124	100.00

Table 2
Frequency Counts of Consonants and Their Clusters (CV)

Voice	Articulation	Consonant	Total	%	Vowel /a/	/i/	/u/	/e/	/o/
Voiceless	Stop	/p/	2,312,976	0.60	984,460	192,847	445,557	182,551	507,561
		/t/	60,002,817	15.53	16,442,465	5,808,199	9,612,206	11,037,767	17,102,180
		/k/	66,644,942	17.24	20,530,065	9,926,212	17,211,261	6,823,149	12,154,255
	Fricative	/s/	46,153,365	11.94	7,932,086	16,518,016	9,271,734	8,127,590	4,303,939
		/h/	24,536,108	6.35	12,594,714	3,554,252	3,067,681	1,254,559	4,064,902
Voiced	Stop	/b/	9,383,168	2.43	2,749,654	1,177,824	2,758,994	1,591,698	1,104,998
		/d/	23,005,066	5.95	9,953,427	–	–	7,744,822	5,306,817
		/g/	20,094,629	5.20	11,996,376	2,009,223	1,210,586	2,328,945	2,549,499
	Fricative	/z/	12,163,220	3.15	1,384,993	6,550,231	1,644,763	1,646,454	936,779
	Nasal	/m/	23,748,918	6.14	6,289,027	3,972,466	2,577,386	4,118,162	6,791,877
		/n/	47,766,161	12.36	10,498,607	14,757,618	411,356	2,771,906	19,326,674
	Liquid	/r/	50,668,968	13.11	6,947,107	6,116,529	28,899,891	5,777,889	2,927,552
Total			386,480,338	100.00	108,302,981	70,583,417	77,111,415	53,405,492	77,077,033

Note—The stop /t/ with /i/ is represented by /cji/ (see Appendix A), although it is not indicated in Table 2. /z/ is pronounced /zj/ or /dj/ when it is combined with /i/, but in this study /zi/ is used to represent this mora (see Appendix A). In this table, /z/ is categorized as a fricative, although /z/ is pronounced either as a fricative or an affricate allophonically.

or 17.24% of the total CV frequency. The second most frequently used consonant was /t/, appearing 60,002,817 times, or 15.53%. The third was /r/ at 50,668,968 times, or 13.11%. The consonant /p/ showed the lowest frequency of 2,312,976 times, or 0.60%. Following this, /b/ was counted 9,383,168 times, or 2.43%. As is shown in Table 2, /ru/ was the most frequently used among CV morae (or syllables, since they are identical), with a frequency count of 28,899,891 times. This is because inflections of verbs are often /ru/. The second most frequently used mora was /ka/, being recorded 20,530,065 times. The third was /no/, the fourth was /ku/, and the fifth was /to/. It should be noted that /z/ is pronounced /zj/ or /dj/ when it is combined with the vowel /i/. However, in the present study, /zi/ is used to represent this mora (see Appendix A). Interestingly, the frequency of /zi/ is high—6,550,231 times of occurrence, far beyond the frequencies of other vowel combinations with /z/.

Frequencies Involving Semivowels (S)

In Japanese phonology, from the perspective of the kana script, φV and CV mora/syllable sounds are referred to as *direct sounds* (*tan'on* in Japanese). Since SV and CSV mora/syllable sounds are represented by two kana symbols (the second is written in a smaller size kana), such as キャ (/kja/), ショ (/sjo/), and ニュ (/nju/), these are called *contracted sounds*, or *palatalized sounds* (*yoo'on* in Japanese). The contracted sounds are constructed with the semivowel /j/[3] in their moraic structures, which appear with three vowels—/a/, /u/, and /o/. The vowels /i/ and /e/ in contracted morae do not appear in Japanese (they do not contain frequency counts) and are thus called *open spaces* (*akima*). They are indicated by – in Table 3.

Apart from this, both the direct and the contracted sounds are identical in terms of the use of a single mora and a single syllable. From the perspective of the kana

Table 3
Frequency Counts of Semivowels and Contracted Sounds (SV and CSV)

Phonemic Structure	Voice	Articulation	Phonemes	Total	%	Vowel /a/	/i/	/u/	/e/	/o/
SV	Voiced	Semivowel	/w/	3,777,201	28.37	3,777,201	–		–	
			/j/	9,536,584	71.63	3,432,278	–	1,620,066	–	4,484,240
CSV (C+/j/+V)	Voiceless	Stop+/j/	/pj/	155,382	0.65	95	–	29,907	–	125,380
			/tj/	3,401,638	14.18	199,251	–	995,900	–	2,206,487
			/kj/	3,817,809	15.92	166,750	–	929,206	–	2,721,853
		Fricative+/j/	/sj/	8,737,457	36.44	2,138,138	–	2,858,438	–	3,740,881
			/hj/	634,527	2.65	53,284	–	3,304	–	577,939
	Voiced	Stop+/j/	/bj/	179,278	0.75	511	–	16,016	–	162,751
			/gj/	1,011,811	4.22	59,424	–	15,343	–	937,044
		Fricative+/j/	/zj/	3,712,578	15.48	188,263	–	1,482,387	–	2,041,928
		Nasal+/j/	/mj/	63,922	0.27	14,646	–	15,494	–	33,782
			/nj/	460,379	1.92	3,087	–	446,886	–	10,406
		Liquid+/j/	/rj/	1,806,101	7.53	50,769	–	214,108	–	1,541,224

Note—The overall total of SV frequency is 13,313,785, whereas the overall total of CSV frequency is 23,980,882. The total frequency of /j/ is 31,711,365, whereas the total frequency of /w/ is 3,777,201.

script, the contracted sounds are combinations of 11 consonants with /j/, which are described by three small kana, ヤ (/ja/), ュ (/ju/), and ョ (/jo/), depending on the following vowels (for details, see Amanuma, Otsubo, & Mizutani, 1996). All these contracted sounds contain semivowels within a phonological structure. It should be noted that the sounds /dj/ and /zj/ are not distinguished in standard Japanese, although they could be written differently in kana. Thus, they are treated as the same sound, as represented by /zj/. This resulted in 11 consonants with the semivowel /j/ in contracted sounds shown in Table 3.

There are two semivowels, /w/ and /j/. As is indicated in Table 3, /j/ had a much higher frequency (9,536,584 times) than /w/ (3,777,201 times). Among morae of the SV structure, /jo/ had the highest frequency (4,484,240 times). Only /j/ is combined with consonants, and as is shown in Table 3, there are 11 CS combinations. The combination /sj/ appeared 8,737,457 times, or 36.44%. The CS combination /kj/ was the second most frequent, occurring 3,817,809 times. The least used CS combination was /mj/, appearing only 63,922 times, or 0.27%. The second least used CS combination was /pj/, counted 155,382 times, or 0.65%. Among the CSV moraic units, /sjo/ was the most frequently used, appearing 3,740,881 times. The mora /sju/ was the second most frequently used, appearing 2,858,438 times. The third was /kjo/ at 2,721,853 times. In contrast, /pja/ was the least frequent, being counted only 95 times.

Vowel combinations with /w/ are limited to /a/. According to Okumura (1977), /kwa/ and /gwa/ were used in old Japanese but disappeared by the end of the 16th century. However, /wi/, /we/, and /wo/, including /kwa/, /kwi/, /kwe/, /kwo/, /gwa/, /vwa/, /vwi/, /vwe/, and /vwo/, are now revived and used only for newly adopted katakana loanwords. These morae are included as a part of the aforementioned officially announced 33 sounds (see Appendix B). In contrast, /j/ appears frequently in *Wago* (words of Japanese origin) and *Kango* (words of Chinese origin). The sounds /ja/, /ju/, and /jo/ are used following various consonants, such as /p/, /s/, /h/, /b/ /g/, /z/, /m/, and /n/ in both Wago and Kango.

Thirty-Three Additional Morae Used for Alphabetic Loanwords

According to the 1991 Japanese government guidelines, 33 additional morae (or kana),[4] along with the 100 basic kana and 3 kana for three special sounds, are officially specified for describing alphabetic loanwords (Tamaoka & Miyaoka, 2003). The type and token frequencies of these 33 morae, used only for alphabetic loanwords, are reported in Appendix B. These additional morae showed a wider range of token frequencies. The most frequently used mora was /ti/ (ティ), being registered 66,313 times. In fact, /ti/ is used for presenting common loanwords, such as ティーム (/tiRmu/, for *team*), ティー (/tiR/, for *tea*), パーティー (/paRtiR/, for *party*), and ボランティア (/boranTia/, for *volunteer*). According to the basic Japanese sound system (i.e., the 101 basic morae), /cji/ for チ (allophone of the /t/-line mora sounds) was not distinguished from the consonant /t/ for /ta/ (タ), /tu/ (ツ), /te/ (テ), and /to/ (ト): These five morae are all classified as /t/-plus-vowel CV morae. However, as Ishihara (1997) has pointed out, modern Japanese can clearly distinguish ミルクティー (/mirukutiR/, for *milk tea*), from ミルクチー (/mirukucjiR/). Thus, they can easily identify the difference between /ti/ and /cji/. Furthermore, Ishihara (1997) stated that Japanese also distinguish ツィ(/ci/) from ティ(/ti/) and チ (/cji/) in both speaking and listening. For example, the name of former Russian president *Yeltsin* is described by エリツィン (/ericiN/), rather than エリチン (/ericjiN/) or エリティン (/eritiN/). A majority of native Japanese speakers can correctly perceive and properly pronounce them as different sounds.

In contrast, the morae, including the contracted sounds of /kw/ with vowels—/kwa/ (クァ), /kwi/ (クィ), /kwe/ (クェ), and /kwo/ (クォ)—are rarely seen. The mora /kwu/ (クゥ) can exist by simply adding the vowel /u/. However, instead, /ku/ (ク) is used, so /kwu/ does not appear in Japanese. Although these are new sounds, they follow the Japanese phonological structure of CSV depicted in Figure 1. The least frequently used mora was /kwe/, which appeared only twice. In the lexical corpus of Amano and Kondo (2000), the mora /kwe/ was used only for the English loanword *sequence*, presented as シークェンス (/siRkweNsu/). Similarly, /kwi/ was the second least used mora, appearing 20 times. Examples of /kwi/ include クィーン (/kwiRN/, for *queen*) and クィンテット(/kwiNteRto/, for *quintet*). The mora /kwa/ was used only 21 times: 12 for *quartet*, presented as クァルテット (/kwaruteRto/), and 9 for *esquire*, presented as エスクァイア (/esukwaia/). The mora /kwo/ has a slightly higher frequency of 85. This was used for loanwords such as *quarter*, presented as クォーター (/kwoRtaR/), *quality*, presented as クォリティー (/kworitiR/), and *quarterly,* presented as クォータリー (/kwoRtariR/). Judging from these examples, /kw/ is basically used to describe the English spelling *qu*, which also represents the sound /kw/ in English.

Japanese people educated since the Second World War study English for at least 3 years from Grades 7–9, and usually for 6 years up to Grade 12. As the consonant /v/ is commonly seen in English, some new morae have started being used to approximate this English sound. These have included /va/ (ヴァ), /vi/ (ヴィ), /vu/ (ヴ), /ve/ (ヴェ), and /vo/ (ヴォ) for describing the /v/ sound in loanwords. Among them, /vi/ is the most common mora, appearing 733 times as examples in such words as ヴェニス (/venisu/, for *Venice*), ソヴィエト (/sovieto/, for *Soviet*), ダイヴィング (/daiviNgu/, for *diving*), and リヴィング (/riviNgu/, for *living*). As all these examples indicate, the English spelling of *vi* is described using the contracted sound of /vi/ in Japanese. Since these words had previously been written as /bi/ (ビ), modern Japanese has begun to distinguish /v/ and /b/ when adopting

alphabetic loanwords. Yet native Japanese speakers most probably pronounce a fricative consonant /v/ like a stop consonant /b/ (Matsuzaki, 1993), although they distinguish /v/ from /b/ in the katakana script.

Another interesting aspect of the new 33 morae is the existence of the consonant /f/, which is not described in the table of 101 basic Japanese morae. Jouo (1977) has claimed that a combination of /f/[5] plus /a/ is already accepted by native Japanese speakers to describe alphabetic loanwords as examples of ファイト (/faito/, for *fight*) and ファースト (/faʀsuto/, for *first*). Koizumi (1989, 1990) and Ishihara (1995) also have suggested that /f/ is an independent consonant within the collection of Japanese sounds, although it is not included in the table of 101 fundamental Japanese morae. In fact, according to our index of frequency counts, the consonant /f/ extended to other vowels and is now found in /fa/ (ファ), /fi/ (フィ), /fe/ (フェ), and /fo/ (フォ). In comparison with other additional morae (see Appendix B), the frequencies of these morae are relatively high at 60,484 times for /fa/, 50,615 times for /fi/, 16,129 times for /fe/, and 31,399 times for /fo/. Since frequencies concerning the combination of /f/ plus vowels are high among the 33 morae, the consonant /f/ seems to have infiltrated into Japanese sounds. However, Ishihara (1995) has noted that, in general, native Japanese speakers have not yet been able to distinguish /fa/ from /ha/, /fe/ from /he/, and /fo/ from /ho/ in speaking (i.e., production).

Syllables With Dual Vowels/Diphthongs (+VJ) and Long Vowels (+VR)

As is depicted in Figure 1, the structure of CVV (both dual vowels/diphthongs of CVJ and long vowels of CVR) is regarded as two morae, CV and V (V can be either /ʃ/ or /ʀ/), but as a single syllable. The frequency of occurrence for these sounds is shown in Table 4. Taking two vowels for V_1V_2 positions that are interchangeable, there were 25 different combinations of the two vowels. When the first and the second vowels were the same, they were considered long vowels. In Table 4, these were shown in five diagonal cells. When the first and the second vowels differed, they were recorded as dual vowels, totaling 20 cases. Although the 25 two-vowel combinations added up to a total frequency of 72,657,680 times, these sounds were not unusual in terms of occurrence in Japanese words. It should be noted that a word such as *tousan*, meaning *father*, is very often pronounced as /tosaN/ with a long vowel, not two vowels. Furthermore, vowel combinations of *-ei* and *-ou* are often pronounced as long vowels. However, since the original database of Amano and Kondo (2000) does not distinguish whether or not these two consecutive vowels are long vowels, these differences are not taken into consideration in the present frequency counts.

The total frequency of the five long vowels was 3,600,096 times. Among the five long vowels, /aʀ/ showed the highest frequency of 1,236,532 times, 1.70% of the total dual/long-vowel frequency, or 34.35% of the total for the five long vowels. The long vowel of /iʀ/ was the second highest in frequency, appearing 733,413 times, or 1.01% (or 20.37% of the five vowels). Third was /eʀ/, appearing 634,079 times, or 0.87% (or 17.61% of the five vowels), and fourth was /oʀ/, appearing 560,222, or 0.77% (or 15.56% of the five vowels). The lowest was /uʀ/, which occurred just 435,850 times, or 0.60% (12.11% of the five vowels). However, it should be noted that, due to difficulties involved in the differentiation of words' moraic segmentations, the token and type frequencies of a long vowel /ʀ/ include two continuous vowels, such as in /baai/, meaning *case*, which is usually segmented into two morae of /ba/ and /ai/.

Twenty dual vowels were counted up to 69,057,584 times. Among them, the highest frequency rate was /ou/, showing 27,448,213 times, or 37.78%. The second was /ai/ at 17,710,419 times, or 24.38%. The third was /ei/ at 12,199,230 times, or 16.79%. Surprisingly, these three combinations make up 78.95% of all the 25 combinations. More specifically, the term *dual vowel*, or *diphthong*, refers to two continuous vowels (e.g., /ai/, /au/, /oi/) that change from an open vowel (i.e., /a/, /o/) to a closed vowel (i.e., /i/, /u/). The vowel /e/ is between the open and the closed vowels. Two continuous vowels, such as /ia/, /uo/, and /io/, which change from a closed

Table 4
Frequency Counts of Dual Vowels (+/ʃ/) and Long Vowels (+/ʀ/)

Following Vowels (/ʃ/ or /ʀ/)	Previous Vowels (V+)														
	/a/			/i/			/u/			/e/			/o/		
	Pair	Count	%	Pair	Count	%	Pair	Count	%	Pair	Count	%	Pair	Count	%
/a/	aʀ	1,236,532	1.70	ia	1,049,302	1.44	ua	246,548	0.34	ea	493,440	0.68	oa	78,570	0.11
/i/	ai	17,710,419	24.38	iʀ	733,413	1.01	ui	2,048,443	2.82	ei	12,199,230	16.79	oi	1,306,835	1.80
/u/	au	994,228	1.37	iu	1,223,360	1.68	uʀ	435,850	0.60	eu	12,575	0.02	ou	27,448,213	37.78
/e/	ae	1,329,310	1.83	ie	561,233	0.77	ue	695,410	0.96	eʀ	634,079	0.87	oe	369,639	0.51
/o/	ao	372,343	0.51	io	411,375	0.57	uo	255,949	0.35	eo	251,162	0.35	oʀ	560,222	0.77
Total		21,642,832	29.79		3,978,683	5.48		3,682,200	5.07		13,590,486	18.70		29,763,479	40.96

Note—The five shaded diagonal cells are long vowels. The overall total of frequencies in Table 4 is 72,657,680. Specifically, the term *dual vowel*, or *diphthong*, refers to two continuous vowels (e.g., /ai/, /au/, /oi/) that change from an open vowel (i.e., /a/, /o/) to a closed vowel (i.e., /i/, /u/). The vowel /e/ is between the open and closed vowels. Two continuous vowels, such as /ia/, /uo/, and /io/ which change from a closed vowel to an open vowel, are considered to be two independent vowels (see Kubozono, 1999).

vowel to an open vowel, are considered to be two independent vowels (see Kubozono, 1999). There are 8 dual vowels out of 20 combinations based on this definition (i.e., /ai/, /au/, /ae/, /oi/, /ou/, /oe/, /ei/, and /eu/), which were counted 61,370,449 times, or 84.48% of the 20 vowel combinations. This suggests that the majority of two continuous different vowel combinations are *open-to-closed* dual vowels. Since Table 4 provides frequency counts for all possible consecutive vowels, all the different two-vowel combinations are labeled "dual vowels" in Table 4. Thus, the difference between continuous (i.e., dual vowel or diphthong) and independent vowels should be carefully treated when using frequency counts in Table 4.

Long vowels placed after consonants (C+), semivowels (S+), and consonants plus semivowels (CS+) are considered as syllabic units, as is depicted in Figure 1. Their token frequencies are shown in Table 5. Among these combinations, /paR/ appeared most frequently, at 407,453 times. The second was /taR/ at 229,783 times. The third was /meR/ at 174,212 times, and the fourth was /riR/ at 157,583 times. The consonant /p/ was combined most frequently with five long vowels and was counted 581,256 times, whereas the second was /r/ at 507,423 times. Long vowels are also combined with the semivowels /w/ and /j/. The semivowel /w/ is compounded only with /aR/ and was counted 42,893 times. In contrast, the semivowel /j/ is combined with three long vowels—/jaR/, /juR/, and /joR/. In addition, various contrastive sounds with /j/ are also tied with these three different long vowels, shown in Table 5. Among them, the most frequently appearing sound was /njuR/ at 72,726 times. The second was /pjuR/ at 28,837 times, and the third was /zjaR/ at 21,755 times.

Long Vowels in Katakana Loanwords

A katakana symbol of ー for the long vowel /R/, as in コーヒー (/koRhiR/, meaning *coffee*), is used for alphabetic loanwords. The total frequency of katakana appearances was 28,464,753 times in the database of the *Asahi* newspaper. By counting only katakana, Table 6 provides the top 20 frequently printed katakana. The result is noteworthy in that the special sound of the long vowel /R/ was the highest, appearing 2,860,333 times, or

Table 5
Frequency Counts of Syllables With Long Vowel /R/ (+R type)

	Total		Vowel				
	Frequency	%	/aR/	/iR/	/uR/	/eR/	/oR/
With Consonants (φVR or CVR)							
φ	229,142	6.74	51,753	31,871	1,806	50,810	92,902
/p/	581,256	17.10	407,453	70,941	7,240	33,850	61,772
/t/	358,111	10.54	229,783	46,703	2,489	51,418	27,718
/k/	309,283	9.10	128,888	31,200	17,828	63,609	67,758
/s/	217,669	6.40	69,847	96,096	25,708	18,067	7,951
/h/	76,375	2.25	14,443	6,388	2,517	503	52,524
/b/	206,895	6.09	73,329	52,467	18,442	26,756	35,901
/d/	78,784	2.32	33,435	_	_	32,034	13,315
/g/	77,532	2.28	10,857	32,410	2,015	17,723	14,527
/z/	65,767	1.93	13,814	43,638	1,838	518	5,959
/m/	236,140	6.95	33,050	6,556	10,031	174,212	12,291
/n/	106,216	3.12	34,166	31,217	5,452	19,008	16,373
/r/	507,423	14.93	38,329	157,583	141,091	108,136	62,284
With Semivowels (SVR)							
/w/	42,893	1.26	42,893	–	–	–	–
/j/	87,314	2.57	5,392	–	34,678	–	47,244
With Contracted Sounds (CSVR)							
/pj/	28,897	0.85	0	–	28,837	–	60
/tj/	25,394	0.75	15,579	–	8,055	–	1,760
/kj/	8,090	0.24	0	–	7,658	–	432
/sj/	20,120	0.59	5,698	–	5,454	–	8,968
/hj/	3,127	0.09	0	–	2,824	–	303
/bj/	15,460	0.45	0	–	15,440	–	20
/gj/	272	0.01	15	–	12	–	245
/zj/	32,055	0.94	21,755	–	5,690	–	4,610
/mj/	6,387	0.19	48	–	6,339	–	0
/nj/	72,922	2.15	3	–	72,726	–	193
/rj/	5,634	0.17	8	–	5,587	–	39
Total	3,399,158	100.00	1,230,538	607,070	429,757	596,644	535,149

Note—There are more combinations of consonants with /+R/ that are not listed in Table 5. Thus, the total frequency of a single /R/ (3,607,169) in Appendix A is higher than the total of 3,399,158 shown in Table 5.

Table 6
The 20 Most Frequently Used Katakana Symbols

Ranking	Katakana	Sound	Frequency	%
1	ー	/R/	2,860,333	10.05
2	ン	/N/	2,119,234	7.45
3	ル	/ru/	1,448,895	5.09
4	ス	/su/	1,336,665	4.70
5	ト	/to/	1,188,192	4.17
6	イ	/i/	936,791	3.29
7	ア	/a/	932,222	3.28
8	ラ	/ra/	905,031	3.18
9	リ	/ri/	783,483	2.75
10	ク	/ru/	742,272	2.61
11	ッ	/Q/	631,694	2.22
12	カ	/ka/	620,694	2.18
13	タ	/ta/	559,735	1.97
14	ド	/do/	558,191	1.96
15	ロ	/ro/	528,947	1.86
16	レ	/ri/	449,324	1.58
17	コ	/ko/	418,221	1.47
18	マ	/ma/	411,896	1.45
19	メ	/me/	411,418	1.45
20	プ	/bu/	406,179	1.43

Note—The total of katakana frequencies was 28,464,753. Percentages of katakana frequencies were calculated on the basis of the overall total of 28,464,753. Shaded rows indicate the three special sounds.

10.05% of the total katakana frequency. The second most frequently used katakana was the special sound of the nasal /N/, showing a high frequency of 2,119,234 times, or 7.45%. The special sound of the geminate /Q/ came in at the 10th position, appearing 631,694 times, or 2.22%. These three special sounds were counted 5,611,261 times, or 19.72%. This result interestingly implies that the special sounds are not *exceptional* but are, rather, inclined to be commonly used in Japanese alphabetic loanwords.

Syllables Ending With a Nasal (+N)

Phonetically speaking, the Japanese geminate /N/ has four sounds: [n], [m], [ŋ] and [ɴ].[6] Typical examples of these include the [n] in *kandai* (/kaNdai/), meaning *generous*, [m] in *danboo* (/daNboR/), meaning *heating*, [ŋ] in *bunka* (/buNka/), meaning *culture*, and [ɴ] in *kekkon* (/keQkoN/), meaning *marriage* ([ɴ] is used in the ending of words). However, these four phonetic sounds are all perceived as the single phoneme /N/ by native Japanese speakers. It is also surprising that /N/ was counted at an extremely high frequency of 35,719,268 times. When Japanese borrowed a variety of Chinese words, the special sound /N/ was heavily adopted to Japanese sounds (for details, see Tamaoka, 2003; Tamaoka, Kirsner, Yanase, Miyaoka, & Kawakami, 2002). In this sense, /N/ is no longer *special* with regard to frequency of occurrence.

As is depicted in Figure 1, the single syllables of *φ*VN, CVN, SVN, and CSVN are considered as two morae of CV+N, SV+N, or CSV+N, since /N/ is regarded as a single mora. The frequencies of +N syllables are listed in Table 7. Combinations of /k/+V+/N/ (i.e., /kaN/, /kiN/, /kuN/, /keN/, and /koN/) showed the highest frequency at 6,986,514 times, or 19.91%. Among them, /kaN/ was the most frequently repeated syllable, being recorded 3,044,363 times. The second most frequently repeated syllable combination was /s/+V+/N/ (i.e., /saN/, /siN/, /suN/, /seN/, and /soN/). This combination was counted up to 5,998,668 times, or 17.10%, although /suN/ had a very low count—only 7,513 times. As compared with morae of *φ*V and CV with /N/ combinations, morae with semivowels (i.e., SV and CSV) with /N/ are much less frequently constructed. Among these less frequent morae, the highest was /zj/+V+/N/, being recorded 275,110 times, or 0.78%. As is shown in Table 7, the six combinations /pjaN/, /pjuN/, /hjaN/, /hjuN/, /bjaN/, and /bjoN/ did not appear at all.

Syllables With a Geminate (+Q and Q+)

The geminate /Q/ is depicted by a small kana symbol of っ in hiragana and by ッ in katakana. However, a larger version of the same symbol つ in hiragana or ツ in katakana represents the mora sound /tu/, which has nothing to do with the geminate /Q/. This sound difference is much easier to recognize when words using the geminate /Q/ are depicted in the alphabetized script *roma-ji*. A few examples from the Japanese language are *hattatsu* (/haQtatu/ in phonemic symbols, meaning *development*), *kippu* (/kiQpu/, meaning *ticket*), *kakki* (/kaQki/, meaning *vitality*), and *masshiro* (/maQsiro/, meaning *snow-white*). As can be seen in *roma-ji*, any one of the consonants /p/, /t/, /k/, and /s/ may be geminates. In fact, the geminate /Q/ represents the duration, approximately the time of one mora, in which one pauses with one's mouth in the shape of the following consonant before pronouncing it (Kawakami, 1977; Komatsu, 1981). In this sense, /Q/ is also referred to as a *double consonant*.

The geminate /Q/ does not occur in the onset and the coda of words but is always found in the middle of words. Thus, phoneme and mora frequencies (a syllable with Q) before /Q/ are shown in Table 8 as +Q, and those after /Q/ are shown in Table 9 as Q+.

As is shown in Table 8, the vowel /i/ in *φ*V was most frequently positioned before /Q/. The frequency of /iQ/ syllable was 545,475 times. In fact, it is easy to find some frequently used expressions, such as /iQpaN/, meaning *general*, /iQsyo/, meaning *together*, and /iQkai/, meaning *once*. The second was the /kaQ/ syllable, being recorded 454,858 times, often seen in /kaQpatu/, meaning *active*, /kaQsoRRo/, meaning *a runway*, and /kaQkazaN/, meaning *an active volcano*. The third was the /keQ/ syllable, counted 363,962 times. In contrast, the six syllables of /pjaQ/, /pjuQ/, /pjoQ/, /bjoQ/, /mjuQ/, and /mjoQ/ did not exist in the lexical corpus. Among the +Q syllables with semivowels, /sjuQ/ was the most frequently recorded at 174,082 times. An example of this is /sjuQseki/, meaning *attendance*.

On the basis of a common definition (e.g., Kawakami, 1977), the actual sound of /Q/ is not voiced but has a single mora duration of one of the five consonants [p], [t],

Table 7
Frequency Counts of Syllables With Nasal /ɴ/ (+ɴ type)

	Total		Vowel				
	Frequency	%	/a/	/i/	/u/	/e/	/o/
With Consonants (φVN or CVN)							
φ	3,346,672	9.54	576,458	1,311,153	253,485	1,062,207	143,369
/p/	271,632	0.77	136,061	40,626	11,250	25,736	57,959
/t/	1,502,415	4.28	443,181	95,099	1,185	796,614	166,336
/k/	6,986,514	19.91	3,044,363	975,378	52,105	2,221,991	692,677
/s/	5,998,668	17.10	1,764,878	1,915,600	7,513	2,106,924	203,753
/h/	2,956,124	8.43	886,479	233,514	243,219	274,897	1,318,015
/b/	1,287,350	3.67	295,866	64,245	774,031	132,820	20,388
/d/	1,277,236	3.64	836,265	–	–	369,078	71,893
/g/	1,860,202	5.30	151,909	193,500	275,756	1,219,511	19,526
/z/	2,076,189	5.92	101,720	813,732	8,560	1,109,929	42,248
/m/	2,213,553	6.31	448,183	793,675	364	268,992	702,339
/n/	3,139,702	8.95	341,757	1,082,676	737	1,693,767	20,765
/r/	1,502,421	4.28	293,375	254,448	5,465	642,304	306,829
With Semivowels (SVN)							
/w/	91,975	0.26	91,975	–	–	–	–
/j/	92,974	0.27	5,707	–	809	–	86,458
With Contracted Sounds (CSVN)							
/pj/	203	0.00	0	–	0	–	203
/tj/	45,357	0.13	44,736	–	209	–	412
/kj/	20,125	0.06	20,089	–	2	–	34
/sj/	129,765	0.37	14,514	–	60,730	–	54,521
/hj/	244	0.00	0	–	0	–	244
/bj/	2	0.00	0	–	2	–	0
/gj/	1,471	0.00	1,423	–	45	–	3
/zj/	275,110	0.78	15,994	–	249,357	–	9,759
/mj/	6,005	0.02	4,277	–	1,722	–	6
/nj/	238	0.00	30	–	131	–	77
/rj/	201	0.00	192	–	7	–	2
Total	35,082,145	100.00	9,519,432	7,773,646	1,946,684	11,924,770	3,917,613

Note—There are more combinations with /ɴ/. Thus, the total frequency of a single /ɴ/ (35,719,268) is higher than the total of 35,082,145.

[k], [s], and [ʃ] placed after /Q/. Table 9 shows the frequencies of the four phonemes /p/, /t/, /k/, and /s/ ([ʃ] is included in /s/) placed after /Q/. These combinations are depicted by shadow in Table 9, with frequencies equaling 4,683,681 times, or 98.87% of the total combinations. /Qka/ was the highest frequency, having a rate of 624,174 times. /Qte/ was the second at 507,742 times, and /Qto/ was the third at 421,717 times.

There were 53,656 cases (1.13% of Q+ combinations) of deviations from the traditional definition of /p/, /t/, /k/, and /s/ after /Q/. Among them, /d/, /h/, /g/, and /z/ after /Q/ were counted 52,969 times. Among these consonants, /d/ was the most frequently placed after /Q/, being counted 22,872 times, or 0.48%. A phonemic combination of /Qdo/ showed the highest frequency at 22,379 times. These new combinations with /Q/ suggest that Japanese sounds have gradually diversified by adopting various alphabetic loanwords, especially from English.

Index of Bi-Mora Frequency Counts

An index for how often two moraic units are combined among Japanese words is called a *bi-mora frequency of occurrence*. Since all Japanese words can be presented in kana, bi-mora frequencies are calculated on the basis of two kana combinations. For example, the combination of the morae /hi/ (ヒ) and /to/ (ト) appears as a word unit 229,773 times, whereas the same combination appears 615,869 times as a part of words, including its combination as a single word /hito/, meaning *human*. The index of bi-mora frequencies of occurrence regarding 4,624 two-kana (mora) combinations with a total of 298,339,034 times are provided as an Excel file in the order of high to low, including 224 bi-mora combinations of 0 frequency counts.

The top 20 bi-mora frequency counts are reported in Table 10. There were 10 bi-mora combinations of CV plus V in the top 20. Furthermore, 5 bi-mora combinations were CV plus the special sound /ɴ/ among the top 20. Only 3 bi-mora combinations were CV plus CV in the top 20 (i.e., /su/ + /ru/, /re/ + /ru/, and /ko/ + /ku/). In the frequency ranking, the combination of /ka/ and /i/ was the highest, being counted 4,269,940 times. Second was /te/ and /i/ at 3,851,691 times. Third was /ko/ and /u/, occurring 3,669,655 times; fourth was /se/ and /i/, being counted 3,661,544 times. These top 4 bi-mora combinations are all CV plus V combinations. The combination of /ka/ plus /ɴ/ was fifth, appearing 3,044,363 times.

Table 8
Frequency Counts of Syllables With Geminate /Q/ (+Q)

	Total		Vowel				
	Frequency	%	/a/	/i/	/u/	/e/	/o/
With Consonants (φVQ or CVQ)							
φ	693,336	13.79	67,060	545,475	52,435	19,642	8,724
/p/	38,604	0.77	8,378	16,216	745	8,473	4,792
/t/	316,090	6.29	98,351	14,586	8,153	81,718	113,282
/k/	1,088,220	21.65	454,858	50,820	9,056	363,962	209,524
/s/	415,492	8.26	90,730	102,525	9,479	185,670	27,088
/h/	531,613	10.57	344,826	48,097	50,784	3,255	84,651
/b/	168,535	3.35	42,753	16,511	75,964	23,594	9,713
/d/	104,491	2.08	90,999	–	–	3,453	10,039
/g/	206,361	4.10	181,921	5,717	3,128	12,728	2,867
/z/	330,426	6.57	24,293	272,866	10,640	19,338	3,289
/m/	240,152	4.78	61,064	54,495	1,802	14,237	108,554
/n/	237,301	4.72	21,264	162,392	34	48,631	4,980
/r/	201,010	4.00	38,113	78,517	2,250	35,694	46,436
With Semivowels (SVQ)							
/w/	8,899	0.18	8,899	–	–	–	–
/j/	131,420	2.61	36,591	–	7,159	–	87,670
With Contracted Sounds (CSVQ)							
/pj/	0	0.00	0	–	0	–	0
/tj/	41,906	0.83	8,745	–	3	–	33,158
/kj/	12,754	0.25	12,324	–	40	–	390
/sj/	210,130	4.18	15,613	–	174,082	–	20,435
/hj/	12,700	0.25	11,673	–	50	–	977
/bj/	390	0.01	116	–	274	–	0
/gj/	4,108	0.08	3,839	–	92	–	177
/zj/	27,113	0.54	5,462	–	21,289	–	362
/mj/	8	0.00	8	–	0	–	0
/nj/	119	0.00	86	–	13	–	20
/rj/	6,256	0.12	118	–	4,832	–	1,306
Total	5,027,434	100.00	1,628,084	1,368,217	432,304	820,395	778,434

The bi-mora frequencies are also arranged in 100 × 102 bi-mora matrix providing bi-mora frequency counts not only for φV and CV combinations, but also for the contracted sounds CSV and SV and two special sounds /N/ and /Q/. A 68 × 68 matrix excluding the contracted and special sounds is also provided, since the simple moraic combinations of φV and CV are often preferred for constructing stimuli in psychological experiments in phonology. Although it is well known that word frequency affects phonological processing (e.g., Fushimi, Ijuin, Patterson, & Tatsumi, 1999; Hino & Lupker, 1998; Taft, 1979, 1991), it has not yet been clearly indicated whether or not bi-mora frequency of occurrence affects Japanese phonological processing independently from word frequency. Thus, the bi-mora frequencies of occurrence are also provided by the Web-accessible database.

Web Sites of the Frequency Indexes of Phonemes, Morae, Syllables, and Bi-Morae for Downloading

Four separate zip files containing Excel files may be downloaded from the Psychonomic Society Web archive, http://www.psychonomic.org/archive/. The database is stored as (1) bi-mora_frequency_index.zip (58.15K), providing the bi-mora frequency index, (2) bi-mora_matrix_100.zip (24.77K), providing the 100 × 102 bi-mora matrix Excel table, (3) bi-mora_matrix_68.zip (17.74K), providing the 68 × 68 bi-mora matrix Excel table, and (4) frequency_index.zip (11.78K), containing the frequency indexes of phonemes, morae, and syllables. The data above may also be accessed from the homepage of Katsuo Tamaoka on the Web site of Hiroshima University, Japan, http://home.hiroshima-u.ac.jp/ktamaoka/ (Excel, zip, and PDF files).

Summary

In order to provide phonological familiarity, the present study has reported frequency counts for units of Japanese phonemes, morae, and syllables, using the word printed frequency index constructed from 1985 to 1998 editions of the *Asahi* newspaper (originally prepared by Amano & Kondo, 2000).

Frequency counts provide a clear picture of trends in Japanese phonology. The type and token frequency of 101 moraic units represented by kana indicated a high correlation (r = .908), suggesting no difference between type and token frequency counts of kana units. Among the five vowels appearing as a single unit (φV), /i/ was the most frequent, whereas /e/ was the least frequent. Once all possible combinations with the five vowels

Table 9
Frequency Counts of Geminate /Q/ and Mora Combinations (Q+)

	Total		Vowel				
	Frequency	%	/a/	/i/	/u/	/e/	/o/
With Consonants (Q+φV or Q+CV)							
φ	0	0.00	0	0	0	0	0
/p/	549,181	11.59	216,237	24,649	95,274	26,377	186,644
/t/	1,320,303	27.87	275,356	93,811	21,677	507,742	421,717
/k/	1,524,225	32.17	624,174	150,228	208,466	135,416	405,941
/s/	724,177	15.29	120,915	237,532	59,722	206,956	99,052
/h/	12,538	0.26	1,365	83	10,461	26	603
/b/	500	0.01	91	0	398	1	10
/d/	22,872	0.48	492	–	–	1	22,379
/g/	10,190	0.22	142	38	9,988	0	22
/z/	7,369	0.16	1	5,514	1,854	0	0
/m/	0	0.00	0	0	0	0	0
/n/	0	0.00	0	0	0	0	0
/r/	37	0.00	31	1	0	0	5
With Semivowels (Q+SV)							
/w/	0	0.00	0	–	–	–	–
/j/	0	0.00	0	–	0	–	0
With Contracted Sounds (Q+CSV)							
/pj/	120,286	2.54	34	–	0	–	120,252
/tj/	81,362	1.72	27,878	–	26,742	–	26,742
/kj/	143,512	3.03	4,150	–	27,303	–	112,059
/sj/	220,635	4.66	31,402	–	75,847	–	113,386
/hj/	0	0.00	0	–	0	–	0
/bj/	0	0.00	0	–	0	–	0
/gj/	3	0.00	0	–	3	–	0
/zj/	147	0.00	62	–	0	–	85
/mj/	0	0.00	0	–	0	–	0
/nj/	0	0.00	0	–	0	–	0
/rj/	0	0.00	0	–	0	–	0
Total	4,737,337	100.00	1,302,330	511,856	537,735	876,519	1,508,897

Note—Shaded rows indicate syllables following the traditional Q+ structures.

were included in frequency counts, this trend changed: /u/ became the most and /e/ remained the least frequent. Among the 12 consonants, /k/ showed the highest counts, whereas /p/ showed the lowest. The present study has also reported contracted sounds. The /sj/ with vowels appeared most frequently. Among the CSV moraic units, /sjo/ appeared most frequently, whereas /pja/ was the least, being counted only 95 times. The Japanese government provided guidelines of katakana usages for alphabetic loanwords, by officially recognizing an additional 33 moraic expressions. Among these, /ti/ appeared most frequently. The consonant /f/ for /fa/, /fi/, /fe/, and /fo/ is already accepted in Japanese sounds, appearing at a relatively high frequency of 60,484 times.

The present study has provided frequency counts for special sounds. Among the five long vowels, /aR/ showed the highest frequency, whereas /uR/ was the lowest frequency. It was also interesting to find that the long vowel /R/ (represented by a horizontal line ー in katakana) appeared the most frequently in katakana-presented loanwords. The study also showed that the special sound /N/ appeared very often, being counted 35,719,268 times. Among various syllabic combinations including /N/, the combination of /k/+V+/N/ showed the highest frequency counts. At 7,858,931 times, the special sound /Q/ was also found to appear very frequently. The frequency of /iQ/ was the highest among the combinations with /Q/. As such, frequency counts tell us that the three well-known special sounds are actually not *special* in terms of occurrence in a Japanese newspaper.

Table 10
The Top 20 Bi-Mora Frequency Counts

No.	Kana	Phonemes	Frequency	%
1	カイ	/ka/+/i/	4,269,940	1.43
2	テイ	/te/+/i/	3,851,691	1.29
3	コウ	/ko/+/u/	3,669,655	1.23
4	セイ	/se/+/i/	3,661,544	1.23
5	カン	/ka/+/N/	3,044,363	1.02
6	イル	/i/+/ru/	3,016,280	1.01
7	スル	/su/+/ru/	2,935,841	0.98
8	トウ	/to/+/u/	2,708,538	0.91
9	ナイ	/na/+/i/	2,653,620	0.89
10	レル	/re/+/ru/	2,643,258	0.89
11	タイ	/ta/+/i/	2,632,623	0.88
12	ウシ	/u/+/si/	2,282,422	0.77
13	ケン	/ke/+/N/	2,221,991	0.74
14	ダイ	/da/+/i/	2,120,498	0.71
15	セン	/se/+/N/	2,106,924	0.71
16	ヨウ	/jo/+/u/	1,999,208	0.67
17	ドウ	/do/+/u/	1,930,217	0.65
18	シン	/si/+/N/	1,915,600	0.64
19	コク	/ko/+/ku/	1,851,379	0.62
20	サン	/sa/+/N/	1,764,878	0.59

Note—The total of bi-mora token frequencies is 298,339,034.

REFERENCES

Amano, N., & Kondo, K. (2000). *Nihongo-no goi tokusei* [Lexical properties of Japanese]. Tokyo: Sanseido.

Amanuma, Y., Otsubo, K., & Mizutani, O. (1996). *Nihongo onseigaku* [Japanese phonetics]. Tokyo: Kuroshio Shuppan.

Besner, D., & Hildebrandt, N. (1987). Orthographic and phonological codes in the oral reading of Japanese kana. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 335-343.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory & Language*, **33**, 824-844.

Endo, O. (1989). Gairaigo-no hyooki [A script for alphabetic loanwords]. In Y. Takebe (Ed.), *Kooza nihongo to nihongo kyooiku, dai-8-kan: Nihongo no moji hyooki (jo)* [Japanese language and Japanese language education series, Vol. 8: Script and orthography of the Japanese Language (part 1)] (pp. 213-246). Tokyo: Meiji Shoin.

Fushimi, T., Ijuin, M., Patterson, K., & Tatsumi, I. F. (1999). Consistency, frequency, and lexicality effects in naming Japanese kanji. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 382-407.

Government of Japan (1991). *Gairaigo-no hyooki* [Describing foreign words]. Cabinet Announcement No. 2, June 28, 1991.

Han, M. S. (1992). The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica*, **49**, 102-127.

Haraguchi, S. (1996). Syllable, mora and accent. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp. 45-75). New York: Mouton de Gruyter.

Hattori, S. (1955). On'inron (1) [Phonology (1)]. *Kokugo-gaku* [A study of the Japanese language], **22**, 279-301.

Hattori, S. (1956). On'inron (3) [Phonology (3)]. *Kokugo-gaku* [A study of the Japanese language], **26**, 302-322.

Hattori, S. (1958). Nihongo-no On'in [Japanese Phonology]. *Sekai dai hyakka jiten*, **22** [World Encyclopedia 22], 360-364.

Hino, Y., & Lupker, S. L. (1998). Effects of word frequency for Japanese kana and kanji word naming and lexical decision: Can the dual-route model save the lexical-selection account? *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 1431-1453.

Ishihara, J. (1995). Nihongo-ni okeru masatsuon, haretsuon, hasatsuon-no bunpu pataan [A distributed pattern of Japanese fricatives, plosives and affricates]. *Nagoya Working Paper in Linguistics*, **11**, 133-157.

Ishihara, J. (1997). Nihongo on'in taikei-ni okeru yoo'on no ityizuke-ni kansuru koosatu [A study of the locus of contracted sounds in Japanese phonology]. *Nagoya Working Paper in Linguistics*, **13**, 71-90.

Jouo, H. (1977). Gendai nihongo-no on'in [Phonology of modern Japanese]. In S. Oono & T. Shibata (Eds.), *Iwanami Kooza Nihongo: Vol. 5. On'in* [Iwanami Lecture Series on Japanese: Vol. 5. Phonology] (pp. 107-145). Tokyo: Iwanami Shoten.

Kawakami, S. (1977). *Nihongo onsei gaisetu* [An introduction to Japanese phonetics]. Tokyo: Baihusha.

Koizumi, T. (1989). Onsei to on'in [Phonetics and phonology]. In M. Sugihuji (Ed.), *Kooza nihongo to nihongo kyooiku: Vol. 2. Nihongo-no onsei on'in* [Japanese language and Japanese education series: Vol. 2. Japanese phonetics and phonology] (pp. 1-20). Tokyo: Meiji Shoin.

Koizumi, T. (1990). Watasi-no gojuu'onzu kan [Ideas regarding the fifty-sound table]. *Nihongogaku* [Study on the Japanese Language], **9**, 4-9.

Kokuritsu Kokugo Kenkyujo [the National Institute for Japanese Language] (1983). *Kokuritsu kokugo kenkyuujo houkoku 75: Gendai hyooki no yure* [Report 75 of the National Japanese Language Research Institute: Tremble of modern script]. Tokyo: Author.

Komatsu, H. (1981). *Nihongo-no on'in* [Japanese phonology]. Tokyo: Chuo Koronsha.

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language & Speech*, **32**, 249-278.

Kubozono, H. (1995). Perceptual evidence for the mora in Japanese. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV* (pp. 141-156). Cambridge: Cambridge University Press.

Kubozono, H. (1999). *Nihongo-no onsei* [Japanese sounds]. Tokyo: Iwanami Shoten.

Kubozono, H. (2002). *Shingo-wa kooshite tsukurareru* [How new words are made]. Tokyo: Iwanami Shoten.

Kubozono, H., & Ota, S. (1998). *On'in koozoo to akusento* [Phonological structure and accents]. Tokyo: Kenkyusha Shuppan.

Matsuzaki, H. (1993). Gairaigo to gendai nihongo on'in taikei [Loanwords and the modern Japanese phonological system]. *Tsukuba Daigaku Kokugo Kokubungaku Kai* [Society of Tsukuba University Japanese Language and Literature], **18**, 22-30.

Matsuzaki, H. (1994). Wago, kango, gairaigo-no gokei to tokusyu haku-no onhairetu joo-no seiyaku [Frequencies of the Japanese special sounds in the lexical categories of *wago*, *kango* and *gairaigo*]. *Journal of the Department of Japanese, Tohoku University*, **4**, 75-86.

Numoto, K. (1987). *Nihon kanzion-no kenkyuu* [A study of Japanese kanji sounds]. Tokyo: Tokyodo Shuppan.

Okumura, M. (1977). On'in no hensen (2) [Transition of the Japanese phonology (2)]. In S. Ohno & T. Shibata (Eds.), *Nihongo 5: On'in* [Japanese 5: Phonology] (pp. 221-252). Tokyo: Iwanami Shoten.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory & Language*, **32**, 358-378.

Otake, T., Hatano, G., & Yoneyama, K. (1996). Speech segmentation by Japanese listeners. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp. 183-201). Berlin: Mouton de Gruyter.

Sugito, M. (1989). Haku ka onsetsu ka [Mora or syllable?]. In M. Sugito (Ed.), *Nihongo no onsei on'in* [Japanese phonetics and phonology] (pp. 154-177). Tokyo: Meiji Shoin.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, **7**, 263-272.

Taft, M. (1991). *Reading and the mental lexicon*. Hillsdale, NJ: Erlbaum.

Tamaoka, K. (2003). Where do statistically-derived indicators and human strategies meet when identifying on- and kun-readings of Japanese kanji? *Cognitive Studies*, **10**, 1-28.

Tamaoka, K., Kirsner, K., Yanase, Y., Miyaoka, Y., & Kawakami, M. (2002). A Web-accessible database of characteristics of the 1,945 basic Japanese kanji. *Behavior Research Methods, Instruments, & Computers*, **34**, 260-275.

Tamaoka, K., & Miyaoka, Y. (2003). The cognitive processing of Japanese loanwords in katakana. *Japanese Psychological Research*, **45**, 69-79.

Tamaoka, K., & Murata, T. (2001). OCP effects on Japanese phonological processing. *Phonological Studies*, **4**, 119-126.

Tamaoka, K., & Terao, Y. (2004). Mora or syllable: Which unit do Japanese use in naming visually-presented stimuli? *Applied Psycholinguistics*, **25**, 1-27.

Terao, Y. (2002). *Iiayamari wa dooshite okoru ka?* [How do speech errors occur?]. Tokyo: Iwanami Shoten.

Vance, T. J. (1987). *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.

Warner, N., & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica*, **58**, 1-25.

NOTES

1. In this article, including references, an alphabetic description of Japanese names follows the commonly used Hepburn style. Since the Hepburn style does not distinguish between long and short vowels (e.g., the proper name of *Kondo* is pronounced /kondor/ with a long vowel at the end), this article uses the spelling of *Kondo*, not *Kondoo*. However, to represent precise sounds, Japanese titles of research papers that include long vowels are shown by repeating the same vowels twice, such as *oo*.

2. The phonemic symbol /j/ is used for semivowels in contracted sounds, such as /kja/ (キャ), /pja/ (ピャ), and /bjo/ (ビョ). Instead of /j/, some phonologists use /y/ for the same sound. The capital J or /J/

refers to the second vowel in dual-vowel combinations. For example, since /kai/ contains dual vowels of /ai/, it is described as CVJ (consonant+vowel+vowel). As such, a small letter /j/ and a capital J refer to different phonological elements.

3. The phonemic symbol /j/ usually refers to contracted moraic sounds. However, the expression /j/ is occasionally used for noncontracted sounds. Although Hattori (1955) used /ci/ for the same sound, Matsuzaki (1993) used /j/ for /cji/ (チ). In this article, we adapted the system of phonemic expressions provided by Matsuzaki.

4. There is no clear outline to express all Japanese sounds perfectly using phonemic symbols. In order to avoid overlaps of phonemic symbols as much as possible, we set a standard to use /cji/ for チ (allophone of the /t/-line mora sounds), /ti/ for ティ, and /ci/ for ツィ. Hattori (1955) described the /t/-line morae as /ta/, /ci/, /cu/, /te/, and /to/. However, he did not discuss about a newer mora, ティ. After this mora became popular, Matsuzaki (1993) adapted a different way, using the phonemic symbols /ci/ for ツィ, /ti/ for ティ, and /cji/ for チ. The phonemic expressions in the present study are basically adapted from Matsuzaki (1993). The phonemic expression /c/ is also used for /ca/ for ツァ, /ce/ for ツェ, and /co/ for ツォ (see Appendix B). The actual debate on this issue is much more complicated (see the detailed discussions in Hattori, 1955, 1956, 1958; Ishihara, 1997; Matsuzaki, 1993, 1994). Since this issue is beyond the scope and purpose of this article, this question is only briefly mentioned.

5. The phoneme /f/ refers to an unvoiced labio-dental fricative or an unvoiced bilabial fricative.

6. The brackets, [ ], refer to phonetic symbols, whereas backslashes, / /, refer to phonemic symbols.

APPENDIX A
Frequency Counts of the 100 Basic Morae and Three Special Sounds

	Phonemic Combination	Mora	Cons/Semi	Vowel	Katakana	Token Frequency	Type Frequency
			Vowels /a/, /e/, /i/, /o/, and /u/				
1	φV	a	φ	a	ア	6,149,909	18,898
2	φV	i	φ	i	イ	43,985,426	67,643
3	φV	u	φ	u	ウ	39,052,254	84,624
4	φV	e	φ	e	エ	4,767,153	14,103
5	φV	o	φ	o	オ	14,053,377	22,195
			/p/, /t/, and /k/ (Voiceless, Stop)				
6	CV	pa	p	a	パ	984,460	2,937
7	CV	pi	p	i	ピ	192,847	1,343
8	CV	pu	p	u	プ	445,557	3,480
9	CV	pe	p	e	ペ	182,551	1,165
10	CV	po	p	o	ポ	507,561	2,183
11	CV	ta	t	a	タ	16,442,465	27,384
12	CV	cji	cj	i	チ	5,808,199	19,482
13	CV	tu	t	u	ツ	9,612,206	22,853
14	CV	te	t	e	テ	11,037,767	12,822
15	CV	to	t	o	ト	17,102,180	27,131
16	CV	ka	k	a	カ	20,530,065	43,982
17	CV	ki	k	i	キ	9,926,212	31,183
18	CV	ku	k	u	ク	17,211,261	38,359
19	CV	ke	k	e	ケ	6,823,149	14,873
20	CV	ko	k	o	コ	12,154,255	27,551
			/s/ and /h/ (Voiceless, Fricative)				
21	CV	sa	s	a	サ	7,932,086	22,624
22	CV	si	s	i	シ	16,518,016	40,585
23	CV	su	s	u	ス	9,271,734	22,763
24	CV	se	s	e	セ	8,127,590	16,836
25	CV	so	s	o	ソ	4,303,939	8,857
26	CV	ha	h	a	ハ	12,594,714	12,919
27	CV	hi	h	i	ヒ	3,554,252	11,507
28	CV	hu	h	u	フ	3,067,681	11,707
29	CV	he	h	e	ヘ	1,254,559	2,633
30	CV	ho	h	o	ホ	4,064,902	8,336
			/b/ and /g/ (Voiced, Stop)				
31	CV	ba	b	a	バ	2,749,654	10,011
32	CV	bi	b	i	ビ	1,177,824	5,569
33	CV	bu	b	u	ブ	2,758,994	9,263
34	CV	be	b	e	ベ	1,591,698	3,769
35	CV	bo	b	o	ボ	1,104,998	5,317
36	CV	da	d	a	ダ	9,953,427	17,622
37	CV	de	d	e	デ	7,744,822	5,965
38	CV	do	d	o	ド	5,306,817	12,158
39	CV	ga	g	a	ガ	11,996,376	17,737
40	CV	gi	g	i	ギ	2,009,223	6,249
41	CV	gu	g	u	グ	1,210,586	8,951
42	CV	ge	g	e	ゲ	2,328,945	6,422
43	CV	go	g	o	ゴ	2,549,499	7,088
			/z/ (Voiced, Fricative)				
44	CV	za	z	a	ザ	1,384,993	5,736
45	CV	zi	z	i	ジ	6,550,231	15,529
46	CV	zu	z	u	ズ	1,644,763	8,670
47	CV	ze	z	e	ゼ	1,646,454	3,024
48	CV	zo	z	o	ゾ	936,779	3,125
			/m/ and /n/ (Voiced, Nasal)				
49	CV	ma	m	a	マ	6,289,027	28,364
50	CV	mi	m	i	ミ	3,972,466	21,459
51	CV	mu	m	u	ム	2,577,386	8,694
52	CV	me	m	e	メ	4,118,162	9,174
53	CV	mo	m	o	モ	6,791,877	12,180
54	CV	na	n	a	ナ	10,498,607	18,207
55	CV	ni	n	i	ニ	14,757,618	10,038
56	CV	nu	n	u	ヌ	411,356	2,034

APPENDIX A (Continued)

	Phonemic Combination	Mora	Cons/Semi	Vowel	Katakana	Token Frequency	Type Frequency
57	CV	ne	n	e	ネ	2,771,906	6,260
58	CV	no	n	o	ノ	19,326,674	14,413
				/r/ (Voiced, Liquid)			
59	CV	ra	r	a	ラ	6,947,107	20,442
60	CV	ri	r	i	リ	6,116,529	25,431
61	CV	ru	r	u	ル	28,899,891	23,656
62	CV	re	r	e	レ	5,777,889	9,566
63	CV	ro	r	o	ロ	2,927,552	13,925
				/w/ and /j/ (Voiced, Semivowel)			
64	SV	ja	j	a	ヤ	3,432,278	13,622
65	SV	ju	j	u	ユ	1,620,066	10,491
66	SV	jo	j	o	ヨ	4,484,240	5,464
67	SV	wa	w	a	ワ	3,777,201	13,922
				/pj/, /tj/, and /kj/ (Voiceless, Stop)			
68	CSV	pja	pj	a	ピャ	95	9
69	CSV	pju	pj	u	ピュ	29,907	147
70	CSV	pjo	pj	o	ピョ	125,380	82
71	CSV	tja	tj	a	チャ	199,251	1,216
72	CSV	tju	tj	u	チュ	995,900	1,898
73	CSV	tjo	tj	o	チョ	2,206,487	9,007
74	CSV	kja	kj	a	キャ	166,750	551
75	CSV	kju	kj	u	キュ	929,206	1,814
76	CSV	kjo	kj	o	キョ	2,721,853	4,139
				/sj/ and /hj/ (Voiceless, Fricative)			
77	CSV	sja	sj	a	シャ	2,138,138	3,208
78	CSV	sju	sj	u	シュ	2,858,438	5,175
79	CSV	sjo	sj	o	ショ	3,740,881	8,688
80	CSV	hja	hj	a	ヒャ	53,284	170
81	CSV	hju	hj	u	ヒュ	3,304	69
82	CSV	hjo	hj	o	ヒョ	577,939	700
				/bj/ and /gj/ (Voiced, Stop)			
83	CSV	bja	bj	a	ビャ	511	45
84	CSV	bju	bj	u	ビュ	16,016	83
85	CSV	bjo	bj	o	ビョ	162,751	504
86	CSV	gja	gj	a	ギャ	59,424	207
87	CSV	gju	gj	u	ギュ	15,343	117
88	CSV	gjo	gj	o	ギョ	937,044	2,084
				/zj/ (Voiced, Fricative)			
89	CSV	zja	zj	a	ジャ	188,263	1,166
90	CSV	zju	zj	u	ジュ	1,482,387	3,623
91	CSV	zjo	zj	o	ジョ	2,041,928	4,929
				/mj/ and /nj/ (Voiced, Nasal)			
92	CSV	mja	mj	a	ミャ	14,646	113
93	CSV	mju	mj	u	ミュ	15,494	201
94	CSV	mjo	mj	o	ミョ	33,782	400
95	CSV	nja	nj	a	ニャ	3,087	69
96	CSV	nju	nj	u	ニュ	446,886	738
97	CSV	njo	nj	o	ニョ	10,406	175
				/rj/ (Voiced, Liquid)			
98	CSV	rja	rj	a	リャ	50,769	183
99	CSV	rju	rj	u	リュ	214,108	1,122
100	CSV	rjo	rj	o	リョ	1,541,224	2,442
				/N/ (Voiced, Nasal)			
101	N	N	–	–	ン	35,719,268	81,463
				/R/ (Long Vowel Including Two Same Vowels Simply Continued)			
102	R	R	–	–	ー	3,607,169	23,598
				/Q/ (Geminate)			
103	Q	Q	–	–	ッ	7,858,931	14,435

APPENDIX B
Frequency Counts of 33 Additional Morae (or Kana) Used for Describing Alphabetic Loanwords

	Phonemic Combination	Mora	Cons/Semi	Vowel	Katakana	Token Frequency	Type Frequency
102	CSV	sje	sj	e	シェ	14,924	149
103	CSV	cje	cj	e	チェ	37,633	269
104	CV	ca	c	a	ツァ	2,465	43
105	CV	ce	c	e	ツェ	4,802	38
106	CV	co	c	o	ツォ	675	14
107	CV	ti	t	i	ティ	66,313	1,205
108	CV	fa	f	a	ファ	60,484	681
109	CV	fi	f	i	フィ	50,615	474
110	CV	fe	f	e	フェ	16,129	346
111	CV	fo	f	o	フォ	31,399	408
112	CSV	zje	zj	e	ジェ	30,738	255
113	CV	di	d	i	ディ	65,994	947
114	CSV	dju	dj	u	デュ	8,344	65
115	SV	je	j	e	イェ	20	7
116	SV	wi	w	i	ウィ	12,605	170
117	SV	we	w	e	ウェ	25,909	226
118	SV	wo	w	o	ウォ	7,747	109
119	CSV	kwa	kw	a	クァ	21	2
120	CSV	kwi	kw	i	クィ	20	6
121	CSV	kwe	kw	e	クェ	2	1
122	CSV	kwo	kw	o	クォ	85	10
123	CV	ci	c	i	ツィ	13,067	15
124	CV	tu	t	u	トゥ	1,854	78
125	CSV	gwa	gw	a	グァ	25	5
126	CV	du	d	u	ドゥ	1,542	55
127	CSV	vwa	vw	a	ヴァ	641	80
128	CSV	vwi	vw	i	ヴィ	733	81
129	CV	vu	v	u	ヴ	283	29
130	CSV	vwe	vw	e	ヴェ	311	38
131	CSV	vwo	vw	o	ヴォ	66	15
132	CSV	tju	tj	u	テュ	72	8
133	CSV	fju	fj	u	フュ	206	7
134	CSV	vju	vj	u	ヴュ	34	4

Note—The 33 additional kana are listed on the basis of the Cabinet Announcement of the Government of Japan (1991). The frequencies of the 33 kana were counted using words described in katakana. The phonemic expressions of these kana were basically adapted from Matsuzaki (1993), with a few exceptions, such as /v/ for ヴァ, ヴィ, ヴ, ヴェ, and ヴォ.

(Manuscript received July 23, 2003;
revision accepted for publication June 4, 2004.)