

European Approaches to Japanese Language and Linguistics

edited by Giuseppe Pappalardo and Patrick Heinrich

Linguistic Factors Affecting Moraic Duration in Spontaneous Japanese

Giuseppe Pappalardo

Università Ca' Foscari Venezia, Italia

Abstract Japanese is often referred to as a mora-timed language (Ladefoged 1975): the mora has been described as the psychological prosodic unit in the spoken language, and it is the metric unit of traditional poetry (Bloch 1950). However, it is clear that morae are not strictly isochronous units (Beckman 1982). Thus, experimental studies have focused on detecting compensation effects that make average mora durations more equal through the modulation of the inherent duration of the segments involved (Han 1962; Port, Al-Ani, Maeda 1980; Homma 1981; Hoequist 1983a; 1983b; Warner, Arai 2001). Kawahara (2017) used the *Corpus of Spontaneous Japanese* to verify whether the durational compensation effect within a /CV/ mora occurs in natural speech, in addition to read speech in the lab. He observed a statistically significant compensation effect of /CV/ morae, in which vowel duration tends to vary in response to the duration of the preceding consonant. However, as the same author has pointed out, the compensation is not absolute because there are several linguistic factors that potentially affect segments' duration profiles. This study will support the idea that moraic isochrony does not occur in spontaneous Japanese by presenting empirical data on how linguistic factors can considerably affect variation in the average duration of morae.

Keywords Spoken corpora. Moraic isochrony. Durational compensation. Inherent segment duration. Vowel devoicing. Pitch accent.

Summary 1 Introduction. – 2 Methods. – 3 Results and Discussions. – 3.1 Inherent Differences in Segmental Duration. – 3.2 Vowel Devoicing. – 3.3 Pitch Accent. – 4 Conclusions.


Edizioni
Ca' Foscari**Ca' Foscari Japanese Studies 13 | Linguistics and Language Education 1**

ISSN 2724-1203

ISBN [ebook] 978-88-6969-428-8 | ISBN [print] 978-88-6969-429-5

Peer review | Open access

Submitted 2020-02-05 | Accepted 2020-05-19 | Published 2020-07-03

© 2020 |  Creative Commons Attribution 4.0 International Public License

DOI 10.30687/978-88-6969-428-8/001

15

1 Introduction¹

Pike (1945, 35) classified world languages according to two types of rhythmic/prosodic patterns: stress-timed and syllable-timed. According to this classification, stress-timed languages, like English and German, tend to have isochronous interstress intervals, while syllable-timed languages, like Italian and Spanish, tend to have equal syllable duration.² Ladefoged (1975, 224) added the mora-timed type, in which isochrony is maintained at the level of the mora, a sub-syllabic constituent that includes either onset and nucleus, or a coda. The Japanese mora can be constituted by a sequence of a consonant and a vowel (/CV/), a single vowel (/V/), a moraic nasal (/N/), the first part of a long consonant (/Q/) or the second part of a long vowel (/H/). What makes all these sequences morae is that, in theory, they have the same duration.

Bloch (1950) was one of the first scholars to claim that the mora is a unit related to timing in Japanese. He states that morae have the same duration or are perceived as having equal length:

The most striking general feature of Japanese pronunciation is its staccato rhythm. The auditory impression of any phrase is of a rapid pattering succession of more-or-less sharply defined fractions all of about the same length. In any one utterance, or indeed in any one conversation or style of discourse, the perceived relative duration of successive phrases can be adequately compared in terms of these fractions: two phrases containing twice or three times as many fractions as another is heard as lasting just twice or three times as long. (Bloch 1950, 90-1)

Since he describes morae as sequences perceived as having the same duration, his description is limited to perception. Hockett (1955, 59) clarifies this point by pointing out that the mora “is defined fundamentally in terms of duration and nothing else”. Instrumental analyses have been conducted in order to demonstrate the isochronous nature of the mora. Han (1962) claims that what gives Japanese its staccato quality is the fact that the actual length of each *onsetsu*³ is approximately the same. Her instrumental analysis based on the observation of spectrograms indicates that a long syllable, that is a syllable consisting of a /CV/ mora followed by a long vowel mora (/H/) or a geminate mora (/Q/), is approximately twice as long as a short syl-

1 Parts of this paper have been presented at the XI International Conference on Corpus Linguistics, held at the University of Valencia in Spain in 2019, May 15-17.

2 Pike (1945)'s idea was further bolstered by Abercrombie (1967).

3 The term *onsetsu* is used here with the meaning of mora, not syllable.

lable. She also points out that there is a compensation effect within the mora, whereby a consonant and a vowel balance each other in order to obtain equal duration with neighbouring units (Han 1962, 74). Homma (1981) confirms Han's theory on the isochrony of the Japanese mora: by demonstrating that the ratio of duration of a /CVQCV/ word (three morae) to a /CVCV/ word (two morae) is approximately 3:2, she argues that mora lengths remain roughly equal because of temporal compensation. Hoequist (1983a) reports that the durational ratio of /CVN/ and /CV/ is approximately 1.8:1: it is less than the expected 2:1 ratio, yet still higher than the ratio found in a syllable-timed language like Spanish. Port, Al-Ani, Maeda (1980) found evidence of compensation within /CV/ syllables, showing that vowels tend to get longer after inherently short consonants. Port, Dalby, O'Dell (1987) point out that the compensation effect is activated not only within /CV/ syllables but also at a higher level, in words. By investigating the duration of words with different numbers of morae, they discovered that the correlation between the duration of the word and the number of morae is stronger than that between the duration of the word and the number of syllables. However, they note that words with a geminate obstruent and a devoiced vowel uttered in fast speech are shorter than other words.

The theory of mora isochrony in Japanese has been called into question in the light of Beckman (1982)'s experimental analysis, whose results provide "no convincing evidence for the phonetic reality of the mora" (Beckman 1982, 133-4). Beckman measures segment durations in particular of /CV/ syllables with a devoiced vowel and long consonants, and examines the compensation effect within a /CV/ syllable and across mora boundaries using as stimuli 75 test words uttered by five native speakers. Even though several /CV/ morae in which a compensation effect is applied have been observed, she claims that negative correlations between adjacent segments may be unreliable evidence for compensation, since some of them can be viewed as linguistic universals (for example, vowels in most languages are longer before a voiced than a voiceless obstruent). Beckman explains the staccato rhythm of Japanese, pointed out by Bloch (1950), as an influence of the writing system. The (moraic) spelling *kana* system and traditional poetic meter both date back to Old Japanese (712-794), which had only /CV/ syllables and few initial vowels.⁴ Japanese native speakers divide words into morae not because these are real units in the spoken language, but because their intuitions are influenced by the segmentation of the *kana* writing system.

⁴ Moraic nasals, moraic second parts of a long consonant and moraic second parts of a long vowel emerged from /CV/ syllables through sound changes which took place in Early Middle Japanese.

More recent studies have focused on structural factors other than isochrony, which may contribute to the perception of the so-called *staccato rhythm* of Japanese. Ramus, Nespø, Mehler (1999)'s study identifies structural factors that provide an effective criterion to distinguish between mora-, syllable- and stress-timed languages.⁵ These factors are the proportion of the speech stream constituted by vowels and the standard deviation of consonantal intervals. While stress-timed languages, like English and Dutch, have high consonantal variability and low vocalic proportion, and syllable-timed languages, like Italian and Spanish, have relatively lower parameters, among the various languages analysed Japanese shows distinctive features, with a very high vocalic proportion and a very low variation in consonantal duration. Indeed, the distinguishing feature of Japanese is that it has few consonantal clusters and long vocalic intervals that make its rhythm different from that of other languages.

Warner, Arai (2001) undertake a thorough review of previous studies which attempted to demonstrate the isochrony of the mora and conclude that, rather than being a temporal and isochronous unit in Japanese rhythm, the mora plays a more structural role and influences duration only indirectly. They claim that the experimental studies that have hitherto sought to verify mora-timing are inconsistent and have serious methodological flaws.

The previous studies reviewed by Warner, Arai (2001) base their assumptions on experimental analyses which make use of small sets of stimuli read by speakers in the lab. In order to verify whether the compensation effect within a /CV/ mora, as pointed out by Port, Al-Ani, Maeda (1980), occurs in natural speech in addition to read speech in the lab, Kawahara (2017) uses a large corpus of spontaneous speech which includes all types of consonants. Whereas Port, Al-Ani, Maeda examine the compensation effect using morae which include only /a/ and /u/, Kawahara's study takes into account all the Japanese vowels and statistically examines the robustness of the compensation effect claimed in previous research. Kawahara's results show a statistically significant compensation effect, with vowel duration varying in response to the duration of the preceding consonant: the shorter the consonant, the longer the vowel tends to be. However, there are various factors that may have blurred the compensation effect claimed in Kawahara's analysis. First, he measured the me-

5 The contribution of structural factors to the rhythm of languages was pointed out before Ramus, Nespø, Mehler (1999) by Dauer (1982). She compares data from syllable-timed and stress-timed languages and concludes that the two most essential factors are a) the presence/absence of vowel reduction and b) the presence/absence of complex consonantal clusters. According to her, conventional rhythm categories have little to do with segmental isochrony, with the characteristic "rhythm" of a language being determined largely by the phonotactic structure.

dian duration of each consonant in Japanese in relation to the median duration of the following vocalic segment, making no distinctions between vowels. As the author himself points out, the distribution of vowels after particular consonants may skew the results of the analysis. For example, the mora /dV/ seems to be a good example of compensation, since (1) the duration of the consonant is one of the shortest and (2) the following vowel is rather longer compared to vowels after different consonants. This can be explained by considering the fact that the vowels of the /dV/ segment are always non-high vowels,⁶ which universally tend to be longer than the high vowels. Another example is the mora /cV/, for which a relatively long duration of the consonant and a relatively shorter duration of the vowel have been calculated. As one would expect, the phoneme /c/, which is phonetically realised in Japanese with the palato-alveolar affricate [tʃ] or the alveolar affricate [ts], tends to be realised as relatively long consonants in natural languages. However, the short duration of the following vowel may be attributed not only to the compensation effect but also to the fact that most of the vowels in /cV/ are high vowels, which are inherently shorter than non-high vowels. Second, there are several linguistic factors, not considered in his study, that may potentially affect segment duration in Japanese, like vowel devoicing (Beckman 1982) and pitch accent (Hoequist 1983a; 1983b).⁷

The aim of the current study is to expand the results of Kawahara (2017)'s study by analysing and quantifying the effect of linguistic factors, such as inherent segment duration, vowel devoicing and pitch accent, that may affect moraic duration in spontaneous Japanese and blur the potential compensation effect claimed in previous studies.

2 Methods

The empirical analysis that follows is based on the *Corpus of Spontaneous Japanese* (henceforth CSJ), which has been jointly developed by the National Institute for Japanese Language and Linguistics (NINJAL), the National Institute of Information and Communications Technology (NICT), and the Tokyo Institute of Technology (Maekawa, Kikuchi, Tsukahara 2004).⁸ This richly annotated corpus of spontaneous Japanese, which was also used by Kawahara (2017), contains more

⁶ This is true only for the conservative variety. In loanwords (*katakana*) the /dV/ segment is also attested with high vowels. However, the occurrence rate is rather low.

⁷ Contextual effects on segmental durations are also found in the results of Venditti, van Santen (1998)'s analysis of read speech data.

⁸ CSJ (The Corpus of Spontaneous Japanese) (2004). National Institute for Japanese Language and Linguistics and National Institute of Information and Communications Technology. URL https://pj.ninjal.ac.jp/corpus_center/csj/en.

than 650 hours of spoken language. The data considered in the present study is the so-called *Core*, which is a smaller database extensively annotated with a mixture of phonemic and sub-phonemic labels. After an automatic alignment, all the annotations in the *Core* were checked by human labellers, who made further corrections. The *Core* includes about 45 hours (half a million words and almost a million segmental intervals) of spontaneous Japanese by 139 speakers - 79 males and 60 females - living in the Tokyo area, whose age ranges from 20 to 69 years. All speakers in the corpus spoke so-called Standard Japanese. As for the type of speech, the CSJ-*Core* includes three speech types: monologue (academic presentation speech [APS] and simulated public speaking [SPSJ]), dialogue (interviews on the contents of APS and/or SPS talks, task-oriented dialogues and free dialogues by the same speakers as in the monologues) and reproduction speech (reading aloud of the transcribed APS or SPS by the speakers who produced the original spontaneous speech) (Maekawa 2015a). The *Core* is in the RDB format, which can be queried using the SQL language (Maekawa 2015b). The use of this large corpus is extremely efficient for this kind of research because it allows us to perform various types of analysis by setting many search parameters simultaneously.

Using the CSJ-*Core* released in 2013, the average durations of morae with different characteristics will be calculated and compared through the analysis of the natural speech produced by 139 speakers. The duration of morae can be precisely calculated since in the CSJ-*Core* the start time and the end time of segments at each level (phone, phoneme, mora, *bunsetsu*, etc.)⁹ is specified. For the purpose of this research, a new table with the following specification has been created: mora type, duration of the mora, duration of the consonant, duration of the vowel, vowel devoicing, and perceived accent. In creating the new table, the duration of the closure “<cl>” typical of plosives and affricates has been included in the duration of the consonant. Since only /V/ and /CV/ morae will be considered for this study, all the special morae with a long consonant, long vowel and moraic nasal - indicated respectively as /Q/, /H/ and /N/ in the CSJ - have been excluded from this table. Furthermore, morae preceded and followed by /Q/ and morae followed by /H/ or /N/ have been excluded from the analysis, since it is difficult to determine the mora boundaries in long consonants and in long vowels, and since vowels become longer in closed syllables (Port, Dalby, O’Dell 1987; Kawahara 2017). Previous analyses based on the CSJ confirmed the general assumption that special morae are shorter than independent morae with a ratio of 0.52:1, as they have quite a similar average duration to inde-

⁹ The smallest parts of a sentence which can be uttered separately from each other in actual speech.

pendent morae constituted by the single high vowels /i/ and /u/ (Pappalardo 2020). Furthermore, morae included in fillers, such as *eeto*, *ma*, etc., have been excluded from the analysis, since in fillers segmental duration tends to be altered for prosodic and pragmatic reasons. In sum, the morae analysed in this study are all types of /V/ and /CV/ morae, not preceded by /Q/ and not followed by /Q/, /H/ and /N/ (633653 intervals in total).

The average duration of morae has been calculated considering the entire CSJ-*Core*. However, since the speaking rate may vary from speaker to speaker, in order to observe a potential individual variation, the average durations of four speakers will be also presented: a male and a female speaker (M1 & F1) aged between 25 and 29 at the time of the recording, and a male and a female speaker (M2 & F2) aged between 40 and 44 at the time of the recording. The speech of all four speakers is a monologue.

Through the comparison of the average durations of morae with or without particular characteristics, it will be possible to determine to what extent linguistic factors such as segmental structure, vowel devoicing and pitch accent contribute to altering the potential moraic isochrony and preventing segment compensation.

3 Results and Discussions

3.1 Inherent Differences in Segmental Duration

Among the linguistic factors that cause mora duration to vary, Warner, Arai (2001) mention the segmental structure of the mora, which can be constituted by a single vowel /V/, a consonant and a vowel /CV/ (independent morae), a glottal stop /Q/, the second part of a long vowel /H/ or a moraic nasal /N/ (special morae). As one would imagine, in the absence of any compensation, a /CV/ mora would be longer than a /V/ or /H/ mora. Pappalardo (2020) used the CSJ in order to calculate the ratio of duration of special morae to independent morae in spontaneous Japanese, which is approximately 0.52:1. This result may in itself prove that mora isochrony is not maintained in spontaneous speech. However, even in independent morae, the duration may vary depending on the type of vowel or consonant. Universally, high vowels are shorter than low vowels and plosives are shorter than fricatives. To what extent does consonant or vowel type contribute to varying the average duration of a mora? Table 1 illustrates the average duration (AD) of /V/ and /CV/ morae with different structures throughout the whole corpus and among the four speakers considered.¹⁰

¹⁰ Morae with a devoiced vowel have been excluded from this analysis.

Table 1 Average durations in seconds of morae with different structures

	/i/		/a/		/ri/		/ra/		/si/		/sa/	
	Token	AD	Token	AD	Token	AD	Token	AD	Token	AD	Token	AD
CSJ-Core	38297	0.060613	16268	0.086899	10924	0.092998	11071	0.118096	7356	0.154227	6161	0.149593
M1	289	0.045605	89	0.070351	81	0.078758	69	0.085823	19	0.151656	44	0.122114
F1	354	0.073229	125	0.091951	45	0.103274	64	0.118848	59	0.15953	45	0.186982
M2	197	0.059865	85	0.091474	78	0.08447	66	0.111707	35	0.132314	87	0.17082
F2	407	0.060617	417	0.097308	119	0.110421	111	0.126727	64	0.189812	26	0.143991

Both in the general results and across the four speakers, average durations are quite homogeneous: /i/ is always shorter than /a/ and /ri/ is always shorter than /ra/. There is a slight difference between the average durations of /si/ and /sa/: in F1 and M2 /sa/ is longer than /si/ in contrast with the general results. Since /a/ is generally longer than /i/, one would expect /sa/ to be always longer than /si/. However, the onset consonant is not the same at the surface level in the two morae, as the consonant in /si/ is allophonically realised as a palato-alveolar fricative [ç] instead of the alveolar fricative [s] of /sa/. The ratio of the shortest /i/ to the longest /si/ in general results is of approximately of 0.39:1, that is: /i/ is less than half the duration of /si/, very far from any form of mora isochrony.

Table 2 Average durations in seconds of /ra/ and /sa/ morae

	/ra/				/sa/			
	Token	AD-mora	AD-consonant	AD-vowel	Token	AD-mora	AD-consonant	AD-vowel
CSJ-Core	11071	0.118096	0.0276437	0.090452	6161	0.149593	0.077963	0.071629
M1	69	0.085823	0.027413	0.058409	44	0.122114	0.063911	0.058202
F1	64	0.118848	0.020539	0.098308	45	0.186982	0.084457	0.102524
M2	66	0.111707	0.027217	0.084488	87	0.17082	0.089272	0.081547
F2	111	0.126727	0.030958	0.095767	26	0.143991	0.076183	0.067807

Table 2 illustrates the data about /ra/ and /sa/ morae, with details for the average duration of morae, consonants and vowels. The inherent difference in the segmental duration of /r/ and /s/ is homogeneous both in the general results and across the four speakers, with /r/ being always shorter than /s/. If a compensation effect is applied, the vowel /a/ should be slightly longer after /r/, but this has not been verified in all cases. While in the general results the vowel in /ra/ is longer than that in /sa/, this is not consistent across the four speakers: only in F2 can a clear compensation effect be observed. However, even in this case of compensation within a /CV/ mora, the ratio of /ra/ to /sa/ in F2 remains approximately 0.88:1.

3.2 Vowel Devoicing

Vowel devoicing is a salient phenomenon of the Japanese language, which involves the complete disappearance of the sonority of close vowels (/i/ and /u/) when they occur between voiceless consonants or between a voiceless consonant and a pause (Fujimoto 2015). Although from a phonetic point of view the /CV/ mora in which the vowel loses its sonority is a segment that comprises only a consonantal sound, the status of the mora is maintained, since native speakers still “hear” the vowel. Beckman (1982) tests the mora hypothesis and tries to verify whether the duration of a mora with a devoiced vowel becomes shorter by comparing the length of /CV/ morae with voiced and devoiced vowels. She uses 54 pairs of morae uttered by five informants and concludes that in only four pairs (7%) the mora with a devoiced vowel is longer than that with a voiced vowel. Furthermore, by applying the less strict version of the mora hypothesis, Beckman compares only the duration of consonants, in order to verify whether the consonant within a mora with a devoiced vowel is longer than a consonant within a mora with a voiced vowel, that is whether there is a compensatory lengthening of the former. As a result, consonants which precede a devoiced vowel are neither consistently nor significantly longer than consonants which precede a voiced vowel. By using the CSJ-*Core*, the current study aims to verify whether Beckman’s assumptions are also true for spontaneous speech and examines the ratio of duration of a /CV/ mora with a devoiced vowel to that of a /CV/ mora with a voiced vowel.

Figure 1 illustrates average duration in seconds of the morae /ki/, /ku/, /si/, /su/, /ti/ and /tu/ with a voiced vowel, on the left, and with a devoiced vowel on the right. As is clear, morae with a voiced vowel are comparatively longer than ones with a devoiced vowel. Moreover, this is due not only to the lesser length of the devoiced vowel but also to the duration of the consonant, which is always shorter than its counterpart in morae with a voiced vowel. In order to verify the reliability of general results based on the speech of 139 speakers, the average durations of morae with a voiced and a devoiced vowel have been observed in the four speakers selected (Figure 2). The results are consistent: morae with a devoiced vowel are generally shorter, with a minor duration of the consonant in most cases.¹¹ No compensation effect can be observed in the segment duration within the /CV/ morae analysed. If we take a close look at the morae /ki/, /si/ and /ti/, we can notice that the length of the vowel does not consistently vary

11 The only exception is the mora /su/ in M2 which is longer when the vowel is devoiced. This is probably due to a particular speaking style of the speaker, who presumably tends to pronounce the auxiliary *-masu* by lengthening the final devoiced vowel.

Linguistic Factors Affecting Moraic Duration in Spontaneous Japanese

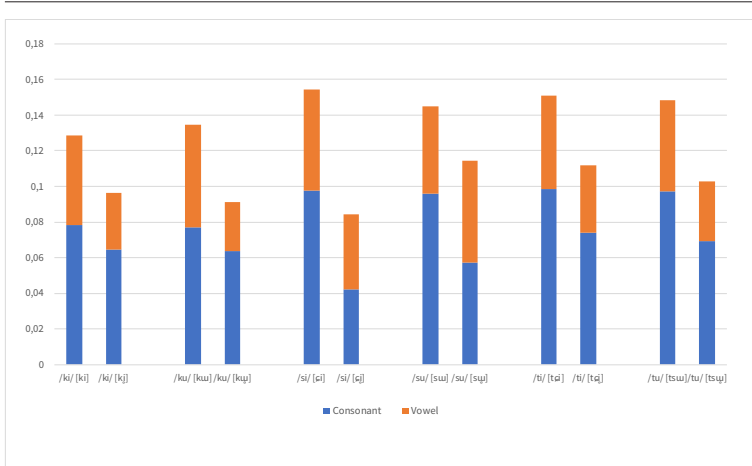


Figure 1 Average duration in seconds of /CV/ morae with voiced and devoiced vowels in CSJ-Sore

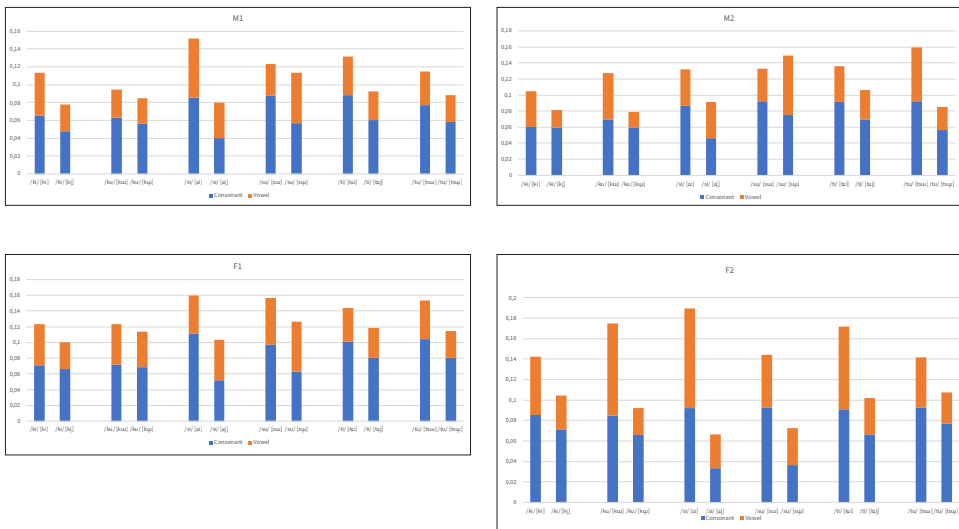


Figure 2 Average duration in seconds of /CV/ morae with voiced and devoiced vowels in M1, M2, F1 and F2

in order to compensate the different duration of the consonants. We would expect a longer [i] after the plosive [k], which is comparatively shorter than the fricative [ç] and the affricate [tç], but the compensatory effect cannot be observed, since the vowel [i] in /ki/ is the shortest (table 3). For this reason, the data obtained by observing only morae with the voiced vowel [i] are not consistent with Kawahara (2017)'s findings.

Table 2 Average duration in seconds of morae with a voiced [i] in CSJ-Core.

	Token	AD-mora	AD-consonant	AD-vowel
/ki/ [ki]	8612	0.128771	0.078493	0.050278
/si/ [çi]	7356	0.154227	0.097849	0.056378
/ti/ [tçi]	5234	0.151043	0.09832	0.052723

Table 4 reports detailed data on the duration of morae with or without vowel devoicing, together with the ratio of duration between the two. The ratio of the morae with a devoiced vowel to those with a voiced vowel goes from 0.55:1 of the mora /si/ to 0.79:1 of the mora /su/. The data hitherto presented give further support to Beckman (1982)'s assumptions and confirm that vowel devoicing is a linguistic factor which considerably affects moraic duration in Japanese.

Table 4 Average duration in seconds of morae with a devoiced and a voiced vowel

	Devoiced vowel		Voiced vowel		Ratio
	Token	AD	Token	AD	
/ki/	3613	0.096419	8612	0.128771	0.75:1
/ku/	5760	0.091139	11773	0.134624	0.68:1
/si/	17578	0.084271	7356	0.154227	0.55:1
/su/	16018	0.114273	8152	0.145024	0.79:1
/ti/	1775	0.111703	5234	0.151043	0.74:1
/tu/	3675	0.102919	6717	0.148526	0.69:1

3.3 Pitch Accent

In languages like Italian, the typical accentual system is dynamic and consists in emphasising the accented syllables by increasing loudness and by lengthening the vowel in open syllables. In the word *casa* ['ka:sa] 'house', the first accented syllable is longer not for segmental reasons, but because the dynamic accent affects its duration prosodically (Bertinetto 1980, 1981). Japanese has a pitch accent, whereby the acoustic correlate is the fundamental frequency determined by the rate at which the vocal cords open and close in voicing (Vance

1987; Beckman 1986). The accented syllable is marked by a drop from a relatively high pitch to a relatively low one. In this kind of accentual system, the length of the vowel in accented syllables is generally not subject to variation. Hoequist (1983a; 1983b) examines the effect on segment duration of dynamic accent in Spanish and English, and of pitch accent in Japanese, using as stimuli test words read in frame sentences by a few informants (the Japanese native speakers are five). In particular, Hoequist (1983a) reports that high pitch morae show a small, consistent and statistically significant duration increase compared to low pitch morae (the ratio calculated is 1.08:1), a lengthening which probably does not play any role in the perception of the accent, as is instead the case with the perception of duration in languages with a dynamic accent. In this study, we have tried to verify Hoequist's findings using the CSJ-Core, addressing the question of whether the minute lengthening in high pitch morae occurs consistently in spontaneous speech in addition to read speech in the lab. The morae examined are those with the consonant /k/, a voiceless velar plosive whose duration is comparatively short, followed by all five vowels (devoiced vowels have been excluded from this analysis).

Table 5 Average durations in seconds of lexically accented and unaccented

	/ka/				Ratio
	Accented		Unaccented		
	Token	AD	Token	AD	
CSJ-Core	3623	0.140924	21322	0.143815	0.98:1
M1	23	0.119004	94	0.116125	1.02:1
M2	15	0.116643	67	0.119546	0.98:1
F1	22	0.147482	147	0.140155	1.05:1
F2	31	0.149667	288	0.151091	0.99:1

	/ki/				Ratio
	Accented		Unaccented		
	Token	AD	Token	AD	
CSJ-Core	1404	0.126558	7208	0.129201	0.97:1
M1	12	0.092054	44	0.119369	0.77:1
M2	3	0.105978	40	0.104832	1.01:1
F1	11	0.130833	85	0.122065	1.07:1
F2	18	0.123505	65	0.147331	0.83:1

	/ku/				Ratio
	Accented		Unaccented		
	Token	AD	Token	AD	
CSJ-Core	1362	0.129209	10411	0.135332	0.95:1
M1	2	0.105833	49	0.093862	1.12:1

M2	5	0.110676	19	0.132189	0.84:1
F1	6	0.141124	38	0.120068	1.17:1
F2	7	0.156062	145	0.175859	0.89:1

	/ke/				Ratio
	Accented		Unaccented		
	Token	AD	Token	AD	
CSJ-Core	1062	0.123772	4832	0.119918	1.03:1
M1	1	0.064218	5	0.105576	0.6:1
M2	2	0.113772	16	0.103805	1.1:1
F1	5	0.1257	39	0.139827	0.9:1
F2	7	0.145579	64	0.128336	1.13:1

	/ko/				Ratio
	Accented		Unaccented		
	Token	AD	Token	AD	
CSJ-Core	2181	0.11956	13010	0.12112	0.99:1
M1	14	0.094263	192	0.100752	0.94:1
M2	9	0.154827	62	0.102442	1.51:1
F1	24	0.121734	73	0.113275	1.07:1
F2	30	0.120857	143	0.124741	0.97:1

Table 5 illustrates the average durations in seconds of the morae /ka/, /ki/, /ku/, /ke/ and /ko/ divided into lexically accented and unaccented. Looking at the ratios of morae from the entire CSJ-Core, the effect of pitch accent on the moraic duration seems to be inconsistent and almost inexistent. Hoequist (1983a) reports that high pitch morae are slightly and consistently longer than those with low pitch, but from the data obtained from the CSJ, cases in which the duration decreases are prevalent. Across the four speakers selected, the variation in duration between accented and unaccented morae is very irregular and presumably does not depend on the effect of the accent. In the general results, the ratio of duration of unaccented morae to accented morae varies from 0.95:1 to 1.03 and this is not consistent with Hoequist's findings. Therefore, it is possible to conclude that pitch accent cannot be included among the linguistic factors that affect moraic and rhythmic duration in Japanese.

4 Conclusions

In the present study we have analysed the effect of three linguistic factors that can potentially affect the duration of morae, using a large-scale corpus of spontaneous Japanese speech. We have reached the following conclusions:

- Even though Japanese is often referred to as a mora-timed language, the data obtained in this study confirm the claims of previous research (Beckman 1982 among others) that moraic isochrony is not a characteristic of spontaneous Japanese. The existence of mora in the perception of native speakers is probably due to factors other than a segmental equal duration in the spoken language.
- The potential compensation effect, claimed in previous studies, which is activated in order to adjust the duration of morae and make their duration within a word or a sentence more homogeneous, is hampered by linguistic factors which determine a variation in segmental duration.
- The varying number of elements in morae (/V/, /CV/, /Q/, /H/, /N/) is one of the reasons why, in the absence of any compensation effect, a /V/ mora - for example - will tend to be shorter than a /CV/ mora. Furthermore, as the results of the present analysis have demonstrated, the intrinsic and articulatory characteristics of each segment can determine a considerable and consistent difference between two different types of /CV/ morae. Even though a strong compensation effect is activated, this may not be sufficient to compensate for the considerable difference in duration of different type of morae and to guarantee a sort of mora isochrony.
- Vowel devoicing has proven to be one of the factors that most affect moraic duration. The difference in duration between a mora with a devoiced vowel and a mora with a voiced vowel is consistent, the former being shorter than the latter in almost all cases. No compensation effect between consonant and vowel has been observed within /CV/ morae with a voiced vowel analysed in this study.
- The data obtained on lexically accented and unaccented morae with the consonant /k/ reveal that pitch accent is not a linguistic factor that causes moraic duration to vary. The differences in the duration are small and inconsistent. This is not consistent with Hoequist (1983a)'s findings.
- Since linguistic factors, such as inherent segment duration and vowel devoicing, can considerably blur the compensation principle, they should be considered in all studies which attempt to analyse the compensatory mechanisms that can potentially be activated in spoken Japanese in order to make morae duration *more* equal.
- In addition to those considered in the present study, there are other linguistic factors, proposed in earlier analyses based on read speech (Venditti, van Santen 1998, Ueyama 1999 among others), that can affect segmental durations, such as the length of a sentence and the position of the mora. As for positional ef-

fects, it has been observed that there is phrase-initial shortening and/or phrase-final lengthening especially with the presence of sentence-final particles, which lengthen durations considerably. Further developments of the research will include the above-mentioned factors in order to verify their effect on mora duration in spontaneous speech.

Bibliography

- Abercrombie, David (1967). *Elements of General Phonetics*. Chicago: Aldine Atherton.
- Beckman, Mary (1982). "Segment Duration and the 'Mora' in Japanese". *Phonetica*, 39, 113-35.
- Beckman, Mary E. (1986). *Stress and Non-stress Accent*. Dordrecht: Foris Publications. Netherlands Phonetics Archives 7.
- Bertinetto, Pier Marco (1980). "The Perception of Stress by Italian Speakers". *Journal of Phonetics*, 8, 385-95.
- Bertinetto, Pier Marco (1981). *Strutture prosodiche dell'Italiano: Accento, quantità, sillaba, giuntura, fondamenti metrici*. Firenze: Academia della Crusca.
- Bloch, Bernard (1950). "Studies in colloquial Japanese IV: Phonemics". *Language*, 26, 86-125.
- Dauer, Rebecca M. (1983). "Stress-timing and Syllable-timing Reanalyzed". *Journal of Phonetics*, 11, 51-62.
- Fujimoto, Masako (2015). "Vowel Devoicing". Kubozono, Haruo (ed.), *Handbook of Japanese Phonetics and Phonology*. Berlin: De Gruyter Mouton, 167-214.
- Han, Mieko (1962). "The Feature of Duration in Japanese". *Onsei no kenkyuu*, 10, 65-80.
- Hockett, Charles F. (1955). *A Manual of Phonology*. *Indiana University Publications in Anthropology and Linguistics, Memoir 11*. Baltimore: Waverly Press. URL https://www.worldcat.org/title/manual-of-phonology/oclc/490126267&referer=brief_results.
- Hoequist, Charles Jr. (1983a). "Durational Correlates of Linguistic Rhythm Categories". *Phonetica*, 40, 19-31.
- Hoequist, Charles Jr. (1983b). "Syllable Duration in Stress-, Syllable- and Mora-timed Languages". *Phonetica*, 40, 19-31.
- Homma, Yayoi, (1981). "Durational Relationship Between Japanese Stops and Vowels". *Journal of Phonetics*, 9, 273-281.
- Kawahara, Shigeto (2017). "Durational Compensation Within a cv Mora in Spontaneous Japanese: Evidence Form the Corpus of Spontaneous Japanese". *The Journal of the Acoustical Society of America*, 142. <https://asa.scitation.org/doi/10.1121/1.4994674>.
- Ladefoged, Peter (1975). *A Course in Phonetics*. Los Angeles: University of California.
- Maekawa Kikuo; Kikuchi Hideaki; Tsukahara Wataru (2004). "Corpus of Spontaneous Japanese: Design, Annotation and XLM Representation". *Proceedings of the International Symposium on Large-scale Knowledge Resources (LKR 2004)*. Tōkyō: Tōkyō Institute of Technology, 19-24.

- Maekawa, Kikuo (2015a). "Corpus-based Phonetics". Kubozono, Haruo (ed.), *Handbook of Japanese Phonetics and Phonology*. Berlin: De Gruyter Mouton, 651-80.
- Maekawa Kikuo 前川喜久雄 (2015b). "Rirëshonarū dētābēsu" リレーショナルデータベース (Relational Database). *Nihongogaku*, 35(1), 82-7.
- Otake, Takashi (2015). "Mora and Mora-Timing". Haruo, Kubozono (ed.), *The Handbook of Japanese Phonetics and Phonology*. Berlin: De Gruyter Mouton.
- Pappalardo, Giuseppe (2020). "Il mito dell'isocronia moraic in giapponese: un'analisi quantitativa basata su corpora orali". *Kervan*, 24(1), 193-209.
- Pike, Kenneth L. (1945). *The Intonation of American English*. Ann Arbor: Michigan University Press.
- Port, Robert F.; Al-Ani, Salman; Maeda, Shosaku (1980). "Temporal Compensation and Universal Phonetics". *Phonetica*, 37, 235-52.
- Port, Robert F.; Dalby, Jonathan; O'Dell, Michael (1987). "Evidence of Mora-timing in Japanese". *Journal of Acoustic society of America*, 102(6), 3735-41.
- Ramus, Franck; Nespoulet, Marina; Mehler, Jacques (1999). "Correlates of Linguistic Rhythm in the Speech Signal". *Cognition*, 73, 265-92.
- Sagisaka Yoshinori 匂坂芳典; Tohkura Yoh'ichi 東倉洋一 (1984). "Kisoku ni yoru onsei gōsei no tame no on'in jikanchō seigyō" 規則による音声合成のための音韻時間長制御 (Phoneme Duration Control for Speech Synthesis by Rule). *Denshi tsūshin gakkai ronbunshi* 電子通信学会論文誌 (The Transactions of the Institute of Electronics, Information and Communication Engineers A), 67(7), 629-36.
- Vance, Timothy J. (1987). *An Introduction to Japanese Phonology*. Albany: State University of New York Press.
- Venditti, Jennifer J.; van Santen, Jan P.H. (1998). "Modeling Segmental Durations for Japanese Text-to-speech Synthesis". *Proceedings of the 3rd ESCA Workshop on Speech Synthesis*. Jenolav Caves.
- Ueyama, Motoko (1999). "An Experimental Study of Vowel Duration in Phrase-final Contexts in Japanese". *UCLA Working Papers in Phonetics*, 97, 174-82.
- Warner, Natasha; Aarai, Takayuki (2001). "Japanese Mora-timing: A Review". *Phonetica*, 58, 1-25.